united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR1327/20

Summer School on Mathematical Control Theory

(3 - 28 September 2001)

Mass balance modelling of bioprocesses

Olivier Bernard
Projet COMORE
INRIA
B.P. 93
06902 Sophia-Antipolis Cedex
France

These are preliminary lecture notes, intended only for distribution to participants

Mass balance modelling of bioprocesses

Olivier Bernard
COMORE
INRIA
BP93,06 902 Sophia-Antipolis cedex
France

September 20, 2001

1 Introduction

System modelling in general is difficult and requires time to properly understand the system and identify a model. This exercise is complicated when the system integrates living organisms. On the contrary to domains like physics where laws that are known since centuries (Ohm law, ideal gas relationship, fundamental principle in mechanics, thermodynamic principle, ...) can apply, most of the biological models rely on empirical laws. These laws result from *a priori* ideas on the working of the system (metabolism, trophic relationships, etc.) or, in some rare cases, have been estimated from some experiments. Since it is not possible to use laws that are admitted by everybody and that have been extensively validated and used, it is primordial to characterise the reliability of the laws used during the model development. This implies that the reliability of the used relationships must be classed hierarchically during the model development. In this chapter, we will see how to organise the

knowledge in the model in order to distinguish a reliable part issued from the mass balance and a more speculative part which will represent the bacterial kinetics.

The model quality and the model structure must above all be determined with respect to the model objectives. Indeed, a model can be developed for very different purposes that must be clearly identified. Will the model be used in order to:

- Reproduce an observed behaviour
- Explain an observed behaviour
- Predict the system evolution
- Understand some of the system mechanisms
- Estimate non measured variables
- Estimate process parameters
- Act on a system to regulate and impose the values for its variables
- Detect anomalies in the process working
- ...

Depending on the modelling objectives and resources, a formalism must be chosen. If the spatial heterogeneity is important and must be taken into account in the model, a parameter distributed model must be written (using e.g. partial differential equations). If the modelling aims at the improvement of a metabolite production during transient phases, the system dynamics must be represented in the model.

Moreover, besides its objectives, the model must also be in adequation with the available data. Indeed a complex model involving a large number of

parameters will also require a large amount of data to identify its parameters and to validate the model.

Finally, if we remember that most of the laws used in biology are speculative, the key step in the modelling of bioprocesses is the model validation. This step is often neglected, despite its determinant role to guaranty the model quality. In particular it is crucial to demonstrate that the model reaches properly the goals for which it was developed.

2 Principle of a bioreactor

2.1 The use of microorganisms

The fermentation principle consists in exploiting metabolic reactions that take place in the cell of a micro-organism (bacteria, yeast, phytoplankton, etc.). In order to activate the micro-organisms interesting metabolic pathways, some specific environmental conditions must be applied (temperature, pH, nutrient concentration). The microorganisms generally need nutrients to growth and precursors or activators in order to produce specific molecules. The simplest required reaction is the growth process itself in order to recover the biomass of microorganisms.

In these metabolic reactions, we can distinguish the following biochemical components:

- the substrates S_i, which are necessary for the goal of the fermentation (growth of the microorganisms and/or precursor for the metabolite to be produced). The substrate associated with growth must contain all the elements necessary to sustain growth (*i.e.* N, C, K, P, Fe, ...). In general, these elements are added in excess so that they are never limiting during the cultivation. Only the main nutrients (carbon, nitrogen or phosphorus source) are monitored along the cultivation.

- microbial biomasses (denoted X_i). The microorganisms can be of various type and species (bacteria, phytoplankton, fungi, yeast, etc) ;
- the products of the biochemical reactions,(denoted P_i). These products can be in the agro-industrial field (cheese, beer, wine, ...), chemistry (enzymes, colourings...), pharmaceutical industry (antibiotics, hormones, vitamins...) or for energy production (ethanol, biogas...)...
- catalysts: they can neither be produced nor consumed during the reaction, but they are necessary.

Depending on the objectives of the fermentation, specific microorganisms will be grown in order to enhance:

- production of biomass itself. It is for example the case for the production of backer yeast.
- production of a metabolite. The goal is to enhance the cellular synthesis of a particular compounds (ethanol, penicillin, ...).
- substrate uptake. In this case, the substrate degradation itself is the objective. This is more specially used to remove pollutants from a liquid medium. Most of the biological depollution processes are among this category.
- phenomenological studies. In this particular case the fermentation aims a better knowledge of the microorganism. The application can be to better understand how the microorganisms grow in the natural field.

2.2 The main types of bioreactors

There are a great deal of different bioreactors. Depending on the type of microorganisms that are grown, they will need a support to settle or can be

free in the liquid. They can resist to more or less intense shearing constraints which will implicate a specific steering system. These two main requirements will determine the type of bioreactor. Two classes can be identified (Bailey and Ollis, 1986):

- stirred tank reactors (CSTR) in which the medium is homogeneous and each element of volume will represent the concentrations in the whole fermenter
- the bioreactors with non homogenous concentration along space. In particular the bioreactor for microorganisms using a support to growth (called a "bed") are in this category.

When the medium is homogeneous it can be described by ordinary differential equations. When a strong spatial distribution must be taken into account a model based on partial differential equations are more appropriate. In this lecture we will present only the CSTR modelled with ODE.

2.3 Working of a bioreactor

2.3.1 Presentation

Figure 1 presents a simplified conceptual scheme explaining the principle of a bioreactor. It is mainly a culture vessel of volume V where the microorganisms grow. A pipe feeds the vessel with an influent medium (with flow rate Q_{in}) and another one withdraws the culture medium with a flow rate Q_{out}.

Depending on the way the fermenter is fed and withdrawn, 3 basic working modes can be identified (figure 2).

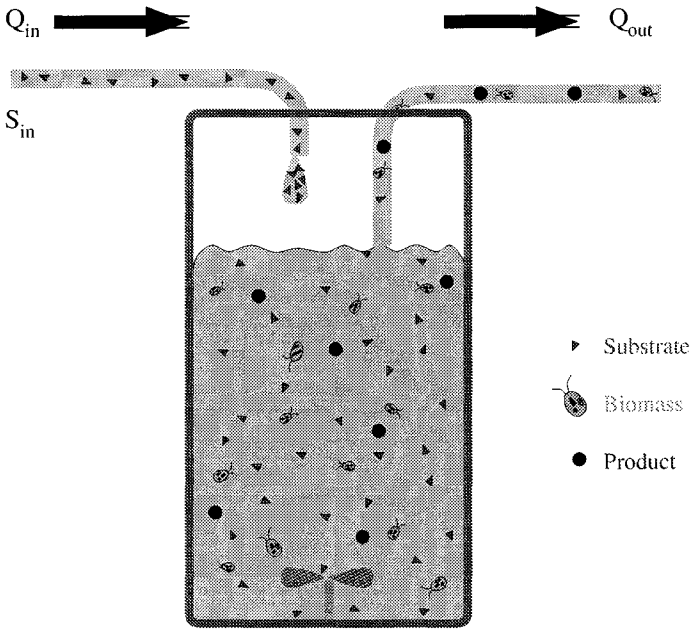

Figure 1: Principle of a bioreactor

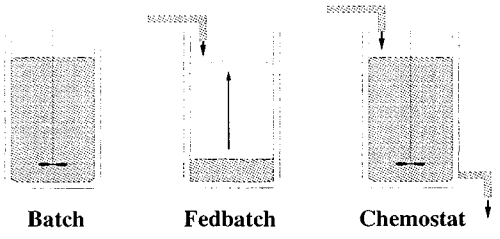

Figure 2: The various working modes of the bioreactors

2.3.2 Batch mode

The system is in batch during the fermentation, and has a constant volume, since no feeding or withdrawal are performed during the fermentation. An inoculum of micro-organisms is introduced at the initial time with all the nutrients and substrates. The biomass or the final product are recovered at the end of the fermentation. The advantage of this approach is that it avoids the contaminations with other bacteria that can come in an open system. The drawback is the limited means of action to act on the fermentation (pH, temperature, aeration...). Therefore the batch mode is often the less optimal from the automatic control point of view to optimise a cost criterion. Nevertheless, this is the most used mode in the industry.

2.3.3 Fedbatch mode

As for the batch mode the duration of a fedbatch is finite. But here the fermenter is fed and starts from a volume V_0 to reach a volume V_f at the end of the fermentation. This mode allows a better control of the growth and biotransformation process along the fermentation. The fedbatch processes are often in closed loop. This operating mode is particularly used when the product to be recovered necessitates to empty the bioreactor like e.g. for intracellular components.

2.3.4 The continuous mode (chemostat)

This is the most popular mode in the field of wastewater treatment. The volume of the bioreactor is constant since the influent flow rate is equal to the effluent flow rate. This the mode that provides the richest dynamics, and therefore which present the more latitude to optimise the process. It is also often used in laboratories to study the physiology of a microorganism. The advantage is also that it allows important productions in small size reactors.

2.3.5 The Sequencing Batch Reactors (SBR)

It is a combination of the various working mode. The idea is to recover the biomass before emptying the bioreactor. For this, the agitation is stopped to let the biomass settle. The different steps used for wastewater treatment are presented on Figure 3.

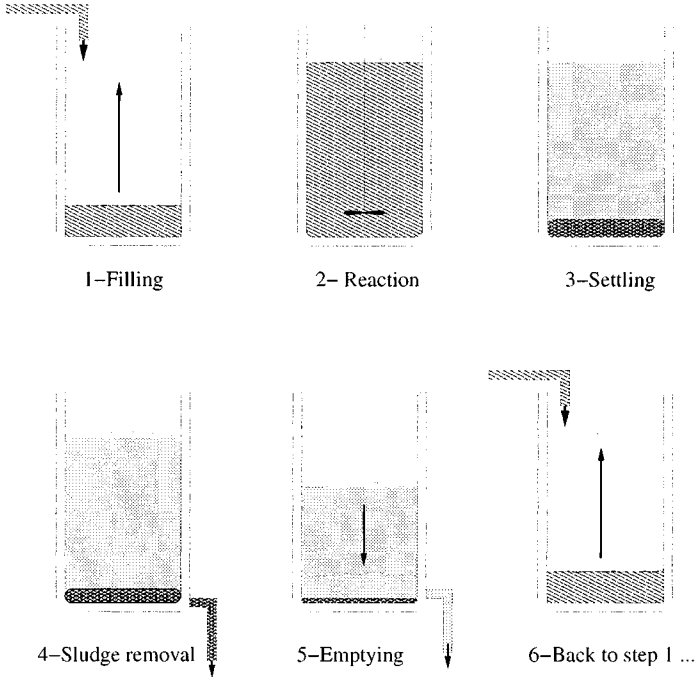

Figure 3: SBR (sequencing batch reactors): representation of the different steps

In the same way, the SFBR (sequencing fedbatch reactor) is a SBR with a stage of filling that follows a fedbatch mode.

3 The mass balance modelling

3.1 Introduction

The modelling of biological systems is delicate because it is not based on validated laws, like in other fields (mechanics, electronics, etc). The evolution of microorganisms is very complex and does not follow any clear law. Nevertheless, this system has to respect some rules, like all the physical systems. For example, the mass conservation, the electro neutrality of the solutions, etc. We will see in this section how to take these aspects into account in the model design. As a result, this mass balance approach will a guaranty a certain robustness in the model.

3.2 Reaction scheme

The reaction scheme of a biochemical process is a macroscopic description of the set of biological and chemical reactions which represent the main mass transfer within the fermenter. A formalism close to this used in chemistry is adopted (Bastin and Dochain, 1990). A set of substrates S_i are transformed into products P_i following 3 possibilities:

- The reaction is a pure chemical reaction, and no biomass is involved. The reaction is then a classical chemical reaction:

$$S_1 + S_2 + ... + S_p \longrightarrow P_1 + ... + P_q$$

- The reaction is catalyzed by a biomass X. The biomass acts only as a catalyser and the reaction is not associated with the growth of the microorganisms:

$$S_1 + S_2 + ... + S_p \xrightarrow{X} P_1 + ... + P_q$$

- The reaction is associated with growth of the microorganisms. Therefore the biomass is also a product of the reaction.

$$S_1 + S_2 + ... + S_p \xrightarrow{X} P_1 + ... + P_q + X$$

The reaction scheme is a concise way to summarise at the macroscopic level a set of reactions that are assumed to determine the process dynamics. The reaction scheme is therefore based on the assumptions related to the available phenomenological knowledge of the process.

In general only the main components of a reaction are represented. Indeed, it would be very difficult to present a real reaction for the growth of a micro-organism since a great deal of components are necessary (Fe, Pb, F, ...).

In the sequel, we will detail the reaction scheme by adding the yield coefficients associated with the consumption (k_i) or the production (k'_i) of each coefficient. Moreover, we will also indicate the rate of the reaction φ:

$$k_1 S_1 + k_2 S_2 + ... + k_p S_p \xrightarrow{\varphi} k'_1 P_1 + ... + k'_q P_q + X$$

The consumption rate of S_i is thus $k_i\varphi$, the production rate P_i is thus $k'_i\varphi$. By convention φ corresponds to the production rate of the biomass.

In the sequel we will assume that the reaction scheme is composed of a set of k biological or chemical reactions. We will considered n variables (chemical concentrations, biomass,...).

3.3 Choice of the reactions and of the variables

The choice of the number of reactions to be taken into account and the choice of the state variables is capital for the modelling purpose. It will be guided by the available knowledge on the reaction scheme on the basis of the available data set. Often the complexity of the model is too high with respect to the

amount of data that are available to test and validate the model. It must be chosen with parsimony, keeping in mind the objectives of the model.

The choice of the reactions and of the variables will mainly determine the model structure, it must be considered with care. We will see in section 6 how to validate this reaction scheme.

We briefly present in Appendix A a procedure to determine the number of reactions that must be taken into account with respect to the available data.

In the sequel, we will assume that the reaction scheme:

- represents the main mass and flow repartition between the set of reactions that intervene in the process,
- is a set of reactions whose yield coefficients are constant.

3.4 Example 1

We will consider here the example of anaerobic digestion. This process is used to remove a polluting substrate (S_1) from wastewater thanks to anaerobic bacteria. In fact, this is a very complex process which involves several different bacterial populations (Mosey, 1983). If the modelling objective is to control this intricate ecosystem in order to improve the pollution removal, then we need a rather simple model. This is why, to limit the model complexity, we consider only two main bacterial populations. We assume therefore that the dynamics can be described by two main steps:

- An acidogenesis step (with a rate $r_1(.)$) in which the substrate S_1 is degraded by acidogenic bacteria (X_1) and is transformed into volatile fatty acids (VFA) (S_2) and CO_2:

$$k_1 S_1 \xrightarrow{r_1(.)} X_1 + k_2 S_2 + k_4 CO_2 \tag{1}$$

- A methanogenesis step (with a rate $r_2(.)$), where the volatile fatty acids are degraded into CH_4 and CO_2 by methanogenic bacteria (X_2).

$$k_3\, S_2 \xrightarrow{r_2(.)} X_2\,, + k_5\, CO_2 + \; k_6\, CH_4 \tag{2}$$

The constants k_1, k_2, k_4, respectively represent the stoichiometric coefficients associated with substrate S_1 consumption, production of VFA and CO_2 during acidogenesis. k_3, k_5 and k_6 respectively represent stoechiometric coefficients associated with VFA consumption and with CO_2 and CH4 production during methanogenesis.

It is worth noting that in some sense this reaction scheme has no biological reality since biomasses X_1 and X_2 represent a set of different species. In the same way for substrates S_1 and S_2 which gathers a set of heterogeneous compounds. A lot of models can be found in the literature for this process (Hill and Barth, 1977; Mosey, 1983; Moletta et al., 1986). Generally, the description of the processes within the bioreactor are much more detailed (Costello et al., 1991; Batstone et al., 1997) but it leads to models difficult to use for control purpose.

4 The mass balance models

4.1 Introduction

We will consider an continuously stirred tank reactor that guarantees a perfect mixing. We will see that independently of the working mode (batch, fedbatch, continuous), the dynamical behaviour of the biological or chemical compounds in the reactors can be directly deduced from the reaction scheme.

We will show on a very simple example how the dynamical model can be established.

4.2 Example 2

We will consider here the very simple example of the growth of a micro-organism X on a substrate S with rate $r(.)$:

$$k\,S \xrightarrow{r(.)} X$$

The yield coefficient associated with substrate consumption is denoted k.

We assume that the influent flow rate is Q_{in} and that the effluent flow rate is Q_{out}. We denote by x and s the total amount of biomass and substrate in the volume V of the bioreactor.

Let us consider the evolution of $V(t)$, $x(t)$ and $s(t)$ between two very close time instants t and $t + dt$.

The evolution of the total liquid volume V is rather simple:

$$V(t + dt) = V(t) + Q_{in}dt - Q_{out}dt$$

For the biomass, we have to take into account the new biomass produced between t and $t+dt$. The production term in the whole volume V is $r(.)Vdt$, and thus:

$$x(t + dt) = x(t) + r(.)Vdt - Q_{out}dt\frac{x}{V}$$

Note that, in order to compute the biomass lost in the effluent (in the volume $Q_{out}dt$) we assume that the concentration in the small volume is the same as in the whole bioreactor (*i.e.* $\frac{x}{V}$). At this point the hypothesis of homogeneity in the reactor is crucial.

In the same way, for the substrate, we must also consider the quantity of substrate (with concentration S_{in}) arriving between the two time instant:

$$s(t + dt) = s(t) + Q_{in}S_{in} - kr(.)Vdt - Q_{out}dt\frac{s}{V}$$

For a very small dt, we can then derive the following equations:

$$\begin{cases} \frac{dx}{dt} = r(.)V - Q_{out}\frac{x}{V} & (3) \\ \frac{ds}{dt} = -kr(.)V + Q_{in}S_{in} - Q_{out}\frac{s}{V} & (4) \\ \frac{dV}{dt} = Q_{in} - Q_{out} & (5) \end{cases}$$

Now, let us rewrite this model in term of concentration *i.e.* using the variables $X = \frac{x}{V}$ and $S = \frac{s}{V}$). It is straightforward to see that we get the following model:

$$\begin{array}{l} \frac{dX}{dt} = r(.)V - DX \\ \frac{dS}{dt} = -kr(.) + D(S_{in} - S) \\ \frac{dV}{dt} = Q_{in} - Q_{out} \end{array} \quad (6)$$

where $D = \frac{Q_{in}}{V}$ corresponds to the dilution rate.

Model (6) simplifies for the various working modes:

- **Batch**. In this case we have $Q_{in} = Q_{out} = 0$. The volume is then constant.
- **Fed batch**. Here $Q_{out} = 0$; $\frac{dV}{dt} = Q_{in}$, V is increasing.
- **Continuous mode**. The volume V is constant since $Q_{in} = Q_{out}$.

For sake of simplicity, in the sequel we will not describe the fed batch case and we will concentrate on the batch or continuous mode. This simplifies the equation since we do not need the equation which forecasts the volume evolution.

4.3 Matrix representation

The reaction scheme leads to the following mass balance model which describes **equivalently** the mass flows within the bioreactor (Bastin and Dochain,

1990):

$$\dot{\xi} = Kr(.) + D(\xi_{in} - \xi) - Q(\xi) \tag{7}$$

Where ξ is the state vector containing all the process compounds and biomasses, ξ_{in} is the vector of the influent concentrations, $r(.)$ is a vector of reaction rates. The matrix K contains the stoechiometric coefficients (yields). $Q(\xi)$, represents the gaseous terms of exchange between the liquid and the gas phase. The dilution rate, D, is the ratio between the influent flow rate Q_{in} and the reactor volume V.

Remark 1 *In the case of the fed batch process, the state vector must also contain the volume V of the reactor. The last equation will describe the volume evolution (cf. equation (5)).*

4.3.1 Example 2 (continued)

Let us consider model (6) working in continuous mode (V is constant, $D = \frac{Q_{in}}{V}$).The model can be rewritten as follows:

$$\begin{pmatrix} X \\ S \end{pmatrix} = \begin{pmatrix} 1 \\ -k \end{pmatrix} (r(.)) + D \left(\begin{pmatrix} 0 \\ S_{in} \end{pmatrix} - \begin{pmatrix} X \\ S \end{pmatrix} \right)$$

It corresponds exactly to the general model, (7) with:

$$\xi = \begin{pmatrix} X \\ S \end{pmatrix}, \quad K = \begin{pmatrix} 1 \\ -k \end{pmatrix}, \quad \xi_{in} = \begin{pmatrix} 0 \\ S_{in} \end{pmatrix}, \quad Q(\xi) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

4.3.2 Example 1 (continued)

Now let us come back to the anaerobic digestion example (see section 3.4). We will assume that the methane solubility is very low and therefore that it directly goes into the gas phase. The carbon dioxide is stored in the liquid phase where he enters in the inorganic carbon compartment (C).

The mass balance model is then the following:

$$\frac{dX_1}{dt} = r_1(.) - DX_1 \tag{8}$$

$$\frac{dX_2}{dt} = r_2(.) - DX_2 \tag{9}$$

$$\frac{dS_1}{dt} = D(S_{1in} - S_1) - k_1 r_1(.) \tag{10}$$

$$\frac{dS_2}{dt} = D(S_{2in} - S_2) + k_2 r_1(.) - k_3 r_2(.) \tag{11}$$

$$\frac{dC}{dt} = D(C_{in} - C) - q_C(\xi) + k_4 r_1(.) + k_5 r_2(.) \tag{12}$$

where S_{1in}, S_{2in} and C_{in} are respectively the influent concentrations of substrate, VFA and dissolved inorganic carbon. The term $q_C(\xi)$ represents the inorganic carbon flow rate (of CO_2) from the liquid phase to the gaseous phase.

4.4 The gaseous flows

We have to take into account the compounds which have a gaseous phase for the mass balance. Indeed, the gaseous species can escape the bioreactor after going from the liquid to the gaseous phase (they can also enter into the bioreactor).

We use for this Henry's law which describes the molar flow rate of a compound C from its liquid phase to its gaseous phase:

$$q_c = K_L a(C - C^\star) \tag{13}$$

Remark 2 *If $q_c < 0$, it means that the gaseous flow will take place from the gaseous phase to the liquid phase.*

The transfer coefficient $K_L a$ (1/T) highly depends on the operating conditions and especially from stirring, and the exchange area between the liquid

and the gaseous phases (size of the bubbles)(Merchuk, 1977; Bailey and Ollis, 1986). The modelling of this parameter with respect to the operating conditions can be very delicate.

The quantity C^* is the saturation concentration of dissolved C. This quantity is related to the partial pressure of gaseous C (P_C) thanks to Henry's constant:

$$C^* = K_H P_C \tag{14}$$

Henry's constant can also vary with respect to the compounds in the culture medium or the temperature.

Moreover, when several gaseous species are simultaneously in the gaseous phase, they must follow the ideal gas law. This will give a relationship of constant ratio between molar flow rates and partial pressures. For m gaseous species $C_1 \ldots C_m$:

$$\frac{P_{c1}}{q_{c1}} = \frac{P_{c2}}{q_{c2}} = \ldots = \frac{P_{cm}}{q_{cm}} \tag{15}$$

4.5 Electro neutrality and affinity constants

The electro neutrality of the solutions is a second rule that the biological systems must respect: the anions concentrations weighted by the number of electrical charges must equal the concentration of cations with the same weighting.

The chemical reactions are often well known and an affinity constant is generally associated. This constant is generally related to the protons concentration H^+, and therefore to pH.

4.6 Example 1 (continued)

4.6.1 Gaseous flows

The methane flow rate is directly related to methanogenesis:

$$q_M = k_6\, r_2(.) \tag{16}$$

The gaseous CO_2 flow rate follows Henry's law:

$$q_C(\xi) = K_L a(CO_2 - K_H P_C) \tag{17}$$

where P_C is the CO_2 partial pressure.

4.6.2 Affinity constants

In the anaerobic digestion example, we will use the electro neutrality and the chemical affinity constants:

In the usual operating range of pH for these processes ($6 \leq pH \leq 8$) we assume that the VFA are under their ionised form. The dissolved CO_2 is in equilibrium with bicarbonate:

$$CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$$

The affinity constant of this reaction is then

$$K_b = \frac{HCO_3^-\ H^+}{CO_2} \tag{18}$$

4.6.3 Electro neutrality of the solution

The cations (Z), are mainly ions which are not affected by biochemical reactions (Na^+,...). Therefore, their dynamics will simply follow, without modification the cation concentration Z_{in} in the influent, so that:

$$\frac{dZ}{dt} = D(Z_{in} - Z) \tag{19}$$

The anions are mainly represented by the VFA and the bicarbonate. Electro neutrality ensures then that:

$$Z = S_2 + HCO_3^- \tag{20}$$

4.6.4 Conclusion

If we add equation (19), the model can finally be rewritten under the matrix form (7), with :

$$\xi = \begin{bmatrix} X_1 \\ X_2 \\ Z \\ S_1 \\ S_2 \\ C \end{bmatrix}, \; r(.) = \begin{bmatrix} r_1(.) \\ r_2(.) \end{bmatrix}, \; K = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ -k_1 & 0 \\ k_2 & -k_3 \\ k_4 & k_5 \end{bmatrix} \tag{21}$$

$$\xi_{in} = \begin{bmatrix} 0 \\ 0 \\ Z_{in} \\ S_{1in} \\ S_{2in} \\ C_{in} \end{bmatrix}, \; Q = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ q_C(\xi) \end{bmatrix}, \tag{22}$$

An elimination of variables HCO_3^-, CO_2, and P_C using equations (17),(15) (18) and (20), leads to the following expression for $P_C(\xi)$ (cf (Bernard et al., pear)):

$$P_C(\xi) = \frac{\phi - \sqrt{\phi^2 - 4K_H P_T (C + S_2 - Z)}}{2K_H} \tag{23}$$

setting: $\phi = C + S_2 - Z + K_H P_T + \frac{k_6}{k_L a} r_2(.)$, we finally get

$$q_C(\xi) = k_L a(C + S_2 - Z - K_H P_C(\xi)) \tag{24}$$

4.7 Conclusion

At this stage, we end up with a model based on the following physical and chemical principles:

- Mass balance
- Ionic balance
- Affinity constants
- Ideal gas law
- Henry's law

The more important hypothesis (with respect to model reliability) is the mass balance hypothesis deduced from the reaction scheme. This hypothesis will therefore require to be validated in the sequel of the modelling approach.

The mass balance model can be used in this form for monitoring or control purpose. Indeed, using the approaches developed in the framework of systems with unknown inputs (Kudva et al., 1980; Hou and Mller, 1991; Darouach, 1994) approach, the unknown reaction rates can be removed thanks to adequate state transformations (Bastin and Dochain, 1990).

Nevertheless, if the initial objective consists in simulating the system, then the reaction rates $r_i(.)$ must be written with respect to the state variables and to the system inputs (environmental variables). This step is much more delicate and a lot of hypotheses difficult to verify are requested.

5 Modelling of the kinetics

5.1 Introduction

For some specific purposes (optimal control, simulation, predictions, etc) it is necessary to have an analytical expression relating the reaction rates to

the state variables of the system. We have nevertheless to keep in mind that these expressions are most of the time approximate relationships issued from empirical considerations. Therefore we leave the background of physical modelling presented previously.

In this section we will see how to hierarchies the assumed hypotheses in order to obtain a two reliability level description of the kinetics.

5.2 The mathematical constraints

5.2.1 Positivity of the variables

A priori, some physical constraints that the model must respect are known: The variables must remain positive and they must be bounded if the amount of matter entering in the bioreactor is bounded. These physical constraints will impose constraints on the structure of the $r_i(.)$. Some quantities (percentage, ratios, etc) must remain between known bounds. To guaranty that the model respects this property, it should verify the following property:

Property 1 (H1) *For each state variable $\xi_i \in [L_{i\,min}, L_{i\,max}]$, the field $\dot{\xi}_i$ on the boundaries must be directed in the admissible space. In other words, the following conditions must be satisfied:*

$$\xi = L_{i\,min} \Rightarrow \dot{\xi}_i \geq 0$$

$$\xi = L_{i\,max} \Rightarrow \dot{\xi}_i \leq 0$$

Particular case: We must have $\xi_i = 0 \Rightarrow \dot{\xi}_i \geq 0.$ in order that variable ξ_i remains positive

5.2.2 Variables that are necessary for the reaction

The second important constraint which must be satisfied by the biochemical kinetics is related to the reaction scheme. A reaction can not take place if

one of the reactant necessary for the reaction is missing. This justifies the following property:

Property 2 *If ξ_j is a reactant of reaction i, then ξ_j can be factorized in r_i:*

$$r_i(\xi, u) = \xi_j \, \nu_{ij}(\xi, u)$$

We verify then easily that $\xi_j = 0 \Rightarrow r_i(\xi, u) = 0$

In the same way, for the reactions associated to a biomass X, we have the same property. Therefore a growth reaction can be rewritten

$$r_i(\xi, u) = \mu_i(\xi, u)X$$

The term $\mu_i(\xi, u)$ is called the growth rate.

5.2.3 Example 1 (continued)

Let us consider the anaerobic digestion model given by equations (8) to (11) and let us apply the state positivity principle:

$$X_1 = 0 \Rightarrow r_1(.) \geq 0 \tag{25}$$

$$X_2 = 0 \Rightarrow r_2(.) \geq 0 \tag{26}$$

$$S_1 = 0 \Rightarrow D(S_{1in} - S_1) - k_1 r_1(.) \geq 0 \tag{27}$$

$$S_2 = 0 \Rightarrow D(S_{2in} - S_2) + k_2 r_1(.) - k_3 r_2(.) \geq 0 \tag{28}$$

Equations (25) and (26) are not very informative. In order that (27) and (28) are respected whatever the experimental conditions, it requires:

$$r_1(.) = S_1 \phi_1(.) \text{ and } r_2(.) = S_2 \phi_2(.)$$

Moreover, biomasses X_1 and X_2 are necessary, respectively for reactions 1 and 2, and thus:

$$r_1(.) = \mu_1(.)X_1 \text{ and } r_2(.) = \mu_2(.)X_2$$

Finally, we must have:

$$r_1(.) = S_1 X_1 \nu_1(.) \tag{29}$$

$$r_2(.) = S_2 X_1 \nu_2(.) \tag{30}$$

5.2.4 Phenomenological knowledge

We will exploit the available phenomenological knowledge (even if it is often speculative) in order to propose an expression for the reaction kinetics.

First, the laboratory experiments allows one to determine the variables which act on the reaction rates. We have seen that the reactant and sometimes the biomass must be found among these variables.

Then, we must know whether the reaction is activated or inhibited by these variables. It often happens that a variable is activating and that she becomes inhibiting at high concentrations (toxicity effect).

Now, there remains to propose an analytical expression which will take into account the mathematical constraints so as the phenomenological knowledge on the process. For this, the modelling choices rely on one hand on experimental observations (when they exist!) and on the other hand on the available models in the literature. In all the cases, the parsimony principle will be privileged to guaranty that the models can be identified and validated.

The following paragraph details the list of models that are often found in the literature to describe some typical reactions. These examples are indicative and a very large number of different models can be found in the literature, in particular to describe the growth rate (Bastin and Dochain, 1990; Bailey and Ollis, 1986).

5.3 The growth rate

5.3.1 The Monod model

The most commonly used model is the Monod (Monod, 1942) model which uses the kinetics identified by Michaëlis-Menten for enzymatic kinetics :

$$\mu = \mu_{max} \frac{S}{K_s + S} \tag{31}$$

μ_{max} is the maximal growth rate and K_s the half saturation constant.

This simple model summarises the two main phases of the growth of a microorganism:

- Unlimited growth, for high values of substrate ($S >> K_S$).The growth rate is then constant, equal the maximal growth rate μ_{max}
- The limited growth, for small values of substrate. In this case the growth rate is approximately proportional to the substrate.

Note that the similitude between enzymatic reaction and growth of a microorganism are often used to justify the analytical expression of a reaction rate (Segel, 1984; Edelstein, 1988).

5.3.2 Haldane model

The Haldane model, initially proposed for an enzymatic reaction can be used to represent a substrate inhibiting the growth at high values (Andrews, 1968):

$$\mu = \mu_{max} \frac{S}{K_s + S + \frac{S^2}{K_i}} \tag{32}$$

where K_i is an inhibition constant. This model predicts that the growth rate is inversely proportional to the growth rate at high concentrations.

5.3.3 multiple limitations

When two substrates S_1 and S_2 are simultaneously limiting the growth, a usual way of modelling the reaction rates is to take the product of two Michaelis-Menten kinetics:

$$\mu = \mu_{max} \left(\frac{S_1}{K_{S_1} + S_1} \right) \left(\frac{S_2}{K_{S_2} + S_2} \right) \tag{33}$$

where K_{S_1} and K_{S_2} are the half saturation constants associated respectively to substrates S_1 and S_2.

If one of the substrate (say S_1) is at high concentration, the growth rate is then equivalent to a Monod model with respect to the other substrate (i.e. S_2).

5.4 Kinetics representation using neural networks

We expose briefly here an alternative method to represent the kinetics using a neural network. The global model will then be composed of a mass balance model based on O.D.E, and of a neural network for the reaction rates. In this sense it is an hybrid model. No *a priori* hypotheses are performed on the kinetics, except that we take into account some constraints to guaranty that the system trajectory keep an acceptable meaning. The kinetics represented by the neural network are then directly identified along the training step. Nevertheless, the variables which influence the kinetics must be determined. These variables will constitute the input of the neural network.

A schematic view of the network is presented on Figure 4 for a single hidden layer. The expression of the output of the network with respect to the inputs is as follows:

$$\mu(S_1, \ldots, S_m) = \sum_{k=1}^{n_h} \omega_k \, \phi(\sum_{i=1}^{m} v_{ki} S_i) \tag{34}$$

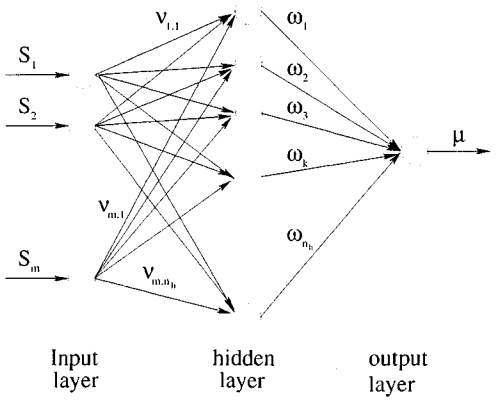

Figure 4: Scheme of a neural network including a single hidden layer.

where n_h represents the number of neurons in the hidden layer. The ω_k and les v_{ki} are respectively the weights of the input and outputs layers. Function ϕ is the activating function of the neuron. It is generally chosen among a set of functions (sigmodes, hyperbolic tangent , gaussian, etc).

The choice of the type of network and of the number of neurons is a rather classical choice and we invite the reader to refer to (Hertz et al., 1991) for more details.

Once the structure of the network has been chosen, the next step is the training phase consisting in identifying the networks weights. This operation is a bit specific for hybrid systems and we refer to (Chen et al., 2000; Karama et al., 2001) more explanations.

6 Model validation

6.1 Introduction

The last modelling step is certainly the most important, but it is also the most often neglected one. It is all the more important since we have seen that it was necessary to assume a great amount of speculative hypotheses. Before using a model, it is important to validate it properly. This stage follows generally the identification step which is not described here.

The general objective for the validation is to verify that the model fits the objectives that have been fixed. More precisely, we will see how to test **separately** the various hypotheses that have been assumed during the model development:

- the reaction scheme
- the qualitative model predictions
- the model as a whole (reaction scheme+kinetics+parameters)

It is important to note that the validation phase must be performed from a data set which was not used to establish or to identify the model. Moreover the new experiments that must be used to test the model validity must significatively differ from the previously used data set (otherwise it is a test of the experimental reproducibility rather than a test of the model validity). If these conditions are not respected, the model can not pretend to be validated

6.2 Validation of the reaction scheme

6.2.1 Mathematical principle

The proposed procedure relies on an important property, which is a consequence of the mass conservation within the bioreactor. As a result this

approach will allow us to check if the obtained mass balance is consistent with the data.

Property 3 *We assume that the $n \times k$ matrix K has more rows than columns ($n > k$). This means that there are more variables than reactions. In this conditions, we have at least $n - k$ independent vectors $v_i \in \mathbf{R}^n$ such that:*

$$v_i^t K = 0_{1\times k}$$

By convention, we normalise the first component of the vector v_i in order to have $v_{i\,1} = 1$

Consequence : let us consider the real variable $w_i = v_i^t \xi$, this variable satisfies the following equation:

$$\frac{dw_i}{dt} = D(w_{i\,in} - w_i) - v_i^t Q(\xi) \tag{35}$$

avec $w_{i\,in} = v_i^t \xi_{in}$. Let us integrate (35) between two time instants t_1 and t_2. We rewrite this equation in order to let the components v_{ij} of vector v_i appear. It leads to:

$$\sum_{j=2}^{n} v_{ij} \phi_{\xi_j}(t_1, t_2) = \phi_{\xi_1}(t_1, t_2) \tag{36}$$

where

$$\phi_{\xi_j}(t_1, t_2) = \xi_j(t_2) - \xi_j(t_1) - \int_{t_1}^{t_2} D(\tau)(\xi_{j\,in}(\tau) - \xi_j(\tau)) - Q_j(\xi(\tau))d\tau$$

The terms $\phi_{\xi_j}(t_1, t_2)$ can be estimated from the experimental measurements of ξ_j along time. An approximation of the integral can be computed *e.g.* using a trapeze formulae. Moreover if the sampling frequency is not sufficient, the data will probably require to be interpolated. We recommend for this task to use spline functions which will at the same time smooth and interpolate the data.

The relationship (36) is a linear relation linking the v_{ij} to the terms $\phi_{\xi_j}(t_1, t_2)$. Since the $\phi_{\xi_j}(t_1, t_2)$ can be computed between various time instants t_1 and t_2, (36) is a linear regression whose validity can be experimentally tested.

Important remark: In fact, relationship (36) is a linear regression which will provide us with an estimate of the v_{ij}. These terms are related with the coefficients of the yield matrix K, and will in general allow to estimate the value of these coefficients.

6.2.2 Example 4

Let us consider here the simple example of the growth of the filamentous fungi *Pycnoporus cinnabarinus* (X) on two substrates, glucose (carbon (C) source) and ammonium (nitrogen (N) source). We assume therefore that the reaction scheme is composed by a single reaction:

$$N + C \longrightarrow X$$

The stoechiometric matrix K associated to this reaction is the following ($\xi = (X\ N\ C)^t$):

$$K = (1\ -k_1\ -k_2)^t, \text{ and } \xi_{in} = (0\ N_{in}\ C_{in})^t \tag{37}$$

Let us consider the two following vectors orthogonal to the columns of K:

$$v_1 = (1\ \frac{1}{k_1}\ 0)^t \text{ and } v_2 = (1\ 0\ \frac{1}{k_2})^t$$

We can then define the following quantities:

$$\phi_X(t_1, t_2) = X(t_2) - X(t_1) + \int_{t_1}^{t_2} D(\tau)X(\tau)$$

$$\phi_N(t_1, t_2) = N(t_2) - N(t_1) - \int_{t_1}^{t_2} D(\tau)(N_{in}(\tau) - N(\tau))d\tau$$

$$\phi_C(t_1, t_2) = C(t_2) - C(t_1) - \int_{t_1}^{t_2} D(\tau)(C_{in}(\tau) - C(\tau))d\tau$$

which will allow us to rewrite the following regressions associated with v_1 and v_2:

$$\phi_X(t_1, t_2) = \frac{1}{k_1}\phi_N(t_1, t_2) \tag{38}$$

$$\phi_X(t_1, t_2) = \frac{1}{k_2}\phi_C(t_1, t_2) \tag{39}$$

It is now easy to verify if the relationships (38) and (39) are significative from a statistical point of view.

Figure (5) presents a validation example on the basis of a series of experiment. The obtained regression is highly significative. This means that relations (38) and (39) are valid. As a consequence, the rows of matrix K, which are orthogonal to v_1 and v_2 are necessarily of the type $K = (1 \ -\alpha_1 \ -\alpha_2)^t$. Therefore the reaction scheme is valid, and subsequently the mass balance model as well.

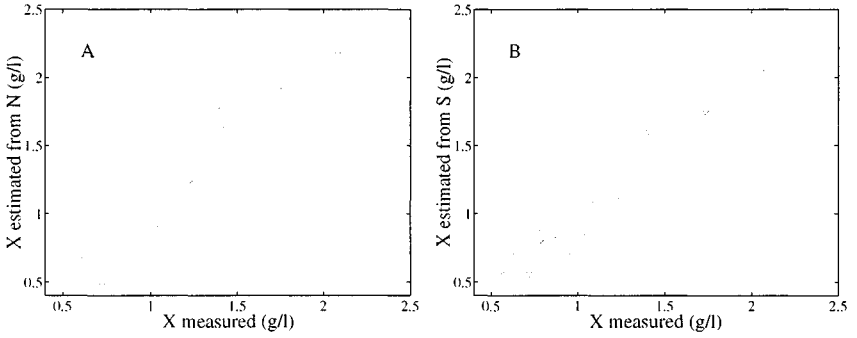

Figure 5: Validation of the linear relationship relating ϕ_X and ϕ_N (A); ϕ_X and ϕ_C (B)

Note that these techniques lead also to the estimate of the yield coefficients k_1 and k_2.

6.3 Qualitative model validation

For the third stage, we assume that the reaction scheme, and therefore the mass balance model has been validated. We will then consider a simulation model consisting of the mass balance model plus the mathematical expression of the kinetics.

The first think to do is to test whether the qualitative properties of the model respect the experimental observations.

The first qualitative behaviour that we expect the model to reproduce is the asymptotic behaviour obtained for constant inputs. Will the model predict an equilibrium, or a more complex behaviour (limit cycle, chaos,...) in agreement with experiments ?

How do these properties evolve when the inputs vary ? For example, the model will predict that an equilibrium in a bioreactor is globally stable for values of the dilution rate lower than a bound, and that for higher values the equilibrium becomes unstable. Does it correspond to the experimental observations ?

More precise qualitative property on the type of transient allowed by the model can also be compared with experimental data. For some specific systems, these transients can be rather precisely determined from a structure analysis (Jeffries, 1986; Sacks, 1990; Bernard and Gouzé, 1995; Gouzé, 1998).

Another qualitative criterion that can be discussed is the response of the system at steady state to a change in an input. Assume for example that an increase of input u_i (which is then kept constant) leads to a decrease in the steady state value of ξ_j: is it verified from an experimental point of view ?

6.3.1 Example

For example, Hansen and Hubbell (1980) study the competition between two bacterial species in a chemostat. The reaction scheme is composed of two

growth reactions:

$$k_1\, S \longrightarrow X_1$$

$$k_2\, S \longrightarrow X_2$$

The growth rate associated to these reactions is assumed to be of Monod type, *i.e.*:

$$\mu_i(S) = \mu_{max\,i} \frac{S}{S + K_{s\,i}}$$

where $\mu_{max\,i}$ and $K_{s\,i}$ are the maximum growth rate and the half saturation constant associated with substrate S for species i.

Hansen and Hubbell showed that the winner of the competition predicted by the model depends on the dilution rate. More precisely, the winner is the species with the smaller ratio $J_i = \frac{K_{s\,i}}{\mu_{max\,i} - D}$. The comparison of the 2 ratios J_1 and J_2 leads to the study of the quantity $r = \frac{\mu_{max\,1} - \mu_{max\,2}}{\mu_{max\,2} - D}$ with respect to the threshold value $\frac{K_{s\,1}}{K_{s\,2}} - 1$. If we assume that we are in the case where $D < \mu_{max\,1} < \mu_{max\,2}$, then species 2 wins for a dilution rate lower than $D_0 = \frac{\mu_{max\,2} K_{s\,1} - \mu_{max\,1} K_{s\,2}}{K_{s\,1} - K_{s\,2}}$, whereas for higher values, it is species 1 (see figure 6). These qualitative properties are verified experimentally (see Figure 7).

6.4 Global model validation

This is the classical way of validating a model: the simulation results are quantitatively compared to experimental data. The most popular criterion is the least square criterion which is computed as follows for a data set of N measurements:

$$J = \sum_{i}^{N} |\hat{\xi}(t_i) - \xi(t_i)|^2$$

where $\hat{\xi}(ti)$ is the simulated value of the state ξ at the sampling instant t_i. The criterion can be improved by weighting each component of the state ξ_j by a coefficient which takes into account the mean value of ξ_j and the measurement accuracy for this variable.

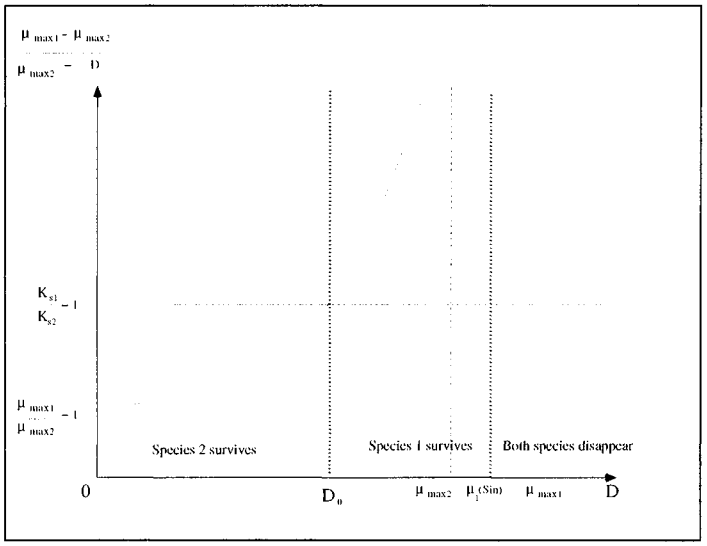

Figure 6: Competition in a chemostat with respect to the dilution rate (discussion of the quantity $\frac{\mu_{max\,1}-\mu_{max\,2}}{\mu_{max\,2}-D}$ with respect to $\frac{k_{s\,1}}{k_{s\,2}}-1$). We consider here the case where $D < \mu_{max\,2} < \mu_{max\,1}$.

This criterion should be minimum. In theory, the residuals (*i.e.* $\hat{\xi}-\xi$) must be studied from a statistical point of view. In the ideal case, it should have properties comparable to those of the measurement noise: it should at least be zero on average, and more precisely one can expect a gaussian distribution (Walter and Pronzato, 1994).

In this approach, the model is considered as a whole. If the residual analysis is not good, in the case where the previous validation steps (reaction scheme and qualitative criteria) have not been performed properly it would be impossible to know the cause of the problem. This criterion does not give any clue on the structural validity of the model (underlying reaction scheme, qualitative properties), on the validity of the type of reaction rate modelling used or on the correctness of parameter values.

If the two first validation steps have been successfully fulfilled, the problem is probably due to a an erratic parameter estimation.

In practice, in the framework of biotechnological systems, as it is difficult to validate *stricto sensu* these models, one will be satisfied with a good visual adequation between simulations and data. This subjective criterion can be reinforced by an analysis of the correlation between predictions and measurements.

7 Mass balance models properties

7.1 Boundness and positivity of the variables

We have seen in paragraph 5.2.1 that the models must be designed in order to meet constraints like the positivity of the state variables.

We will see here that the models based on mass balances are of the type BIBS (bounded input bounded state). To show this property, we use the following hypotheses which are verified for the mass balance based systems:

Hypothesis 1 (H2) *There exists a vector v^+ whose components are strictly positive, such that:*

$$v^+ K = 0_{1\times k}$$

Consequence: Let us consider the scalar quantity $w^+ = v^+\xi$. It verifies the following equation: (35):

$$\frac{dw^+}{dt} = D(w_{in}^+ - w^+) - v^+ Q(\xi) \tag{40}$$

We have to assume an hypothesis for $Q(\xi)$, which is verified in most of the cases:

Hypothesis 2 (H3) *There exists a positive real a and a real b, such that $Q(\xi)$ can be compared to a linear expression as follows:*

$$v^+Q(\xi) \geq av^+\xi + b$$

This hypothesis is verified if $v^+Q(\xi) = 0$, or if $Q(\xi)$ is described by Henry's law (see section 4.4).

Property 4 *If hypotheses (H1), (H2) and (H3) are verified, then the system is BIBS.*

Proof: The dynamics of w^+ can be bounded as follows:

$$\frac{dw^+}{dt} \leq (D+a)(\frac{Dw_{in}^+ - b}{D+a} - w^+) \tag{41}$$

if we apply property 1, we can deduce: $w^+ \leq \max(w^+(0), \frac{Dw_{in}^+ - b}{D+a})$.

In other words, $\sum w_i^+\xi_i$ is bounded. Since $w_i^+ > 0$, the state variables ξ_i are bounded.

7.2 Equilibrium point and local behaviour

7.2.1 Introduction

In this section we briefly recall the principles of the studies of the model properties. We invite the reader to consult (Khalil, 1996) for more details.

Generally, the bioreactor models are **non linear** (*e.g.* they often have multiple steady state), and they are of high dimension (large number of state variables). They often have a large number of parameters, which often intervene in nonlinear functions (nonlinearity with respect to the parameters).

Nevertheless, for dimensions greater than 3, it becomes very difficult to characterise the behaviour of a dynamical system. We will however show that the mass balance based model have structural properties than make easier the system understanding.

In this paragraph, we consider a general dynamical system:

$$\frac{d\xi}{dt} = f(\xi, u) \tag{42}$$

We keep in mind that $f(\xi, u) = Kr(\xi) + D(\xi_{in} - \xi) - Q(\xi)$. We will consider here the case where $u = (D, \xi_{in})$ is constant.

7.2.2 Equilibrium points and local stability

The equilibrium points are obtained for $\frac{d\xi}{dt} = 0$ when the inputs are maintained constant.

The non linear systems generically differ from linear systems since they can have multiple equilibrium points.

The first step in the model analysis consists in testing if these equilibrium points are locally stable. We consider the jacobian matrix of the linearised:

$$J(\xi) = \frac{Df}{D\xi}(\xi)$$

The equilibrium ξ_0 is locally stable if and only if all the eigenvalues of $J(\xi_0)$ have a negative real part. If there exists an eigenvalue with positive real part, the equilibrium is unstable. We can not conclude on the system stability if none eigenvalues have a positive real part but one (at least) eigenvalue has a zero real part.

7.2.3 Global behaviour

The dynamics of a nonlinear system can be very complicated, and complex behaviours like limit cycles, chaos, etc can appear in addition to the equilibria. It is therefore important to test whether a unique locally stable equilibrium is globally stable. In other words if for any initial conditions the trajectories will converge toward this equilibrium.

The standard method to prove that an equilibrium is globally stable relies on the Lyapunov (Khalil, 1996) approach. However it is often difficult to find

a Lyapunov function for a biological system. One can refer to (Li, 1998) for constructive methods to find Lyapunov functions in a large class of growth models.

7.2.4 Asymptotic behaviour

We have seen in paragraph 6.2.1 that in the general case where $n > k$, there exists $n - k$ vectors v_i in the kernel of K^T. These vectors allow to compute the quantities $w_i = v_i^t \xi$ whose dynamics satisfies equation (35).

Moreover, there are often q vectors v_i^0 among the v_i which verify:

$$v_i^{0\,t} Q(\xi) = 0 \tag{43}$$

The dynamics of the associated w_i^0 is then very simple:

$$\frac{dw_i^0}{dt} = D(w_{i\,in}^0 - w_i^0) \tag{44}$$

In the conditions that we consider (*i.e.* constant D and ξ_{in}), the solutions of (44) asymptotically converge towards $w_{i\,in}^0$. This means that the solutions of system (42) will converge towards the hyperplane $v_i^{0\,t} \xi = 0$.

The state of the system will then asymptotically converge toward the vectorial subspace of dimension $n - q$, which is orthogonal to the q vectors v_i^0. This allows to simplify the study of the n dimensional system (42) into a $n - q$ dimensional system.

7.2.5 Example 4 (continued)

Let us consider the model of fungal growth (equation 37). We will moreover assume that the kinetics has been represented by a Monod law with respect to the 2 substrates C and N:

$$r(\xi) = \mu_{max} \frac{C}{K_C + C} \frac{N}{K_N + N} X \tag{45}$$

The two vectors v_1 and v_2 identified in paragraph (6.2.2) verify straithforwardly equation (43).

Therefore when $t \to +\infty$, $X + \frac{N}{k_1} \to \frac{N_{in}}{k_1}$ and $X + \frac{C}{k_2} \to \frac{C_{in}}{k_2}$.

The study of the 3 dimensional system is then simplified into the study of the following system in dimension 1:

$$\frac{dX}{dt} = \mu_{max} \frac{C_{in} - k_2 X}{K_C + C_{in} - k_2 X} \frac{N_{in} - k_1 X}{K_N + N_{in} - k_1 X} X - DX \qquad (46)$$

One will verify that this system has three real equilibrium points (one of them being the trivial equilibrium $X = 0$). These equilibria, in increasing order, are respectively locally stable, unstable and locally unstable. With respect to the parameters values, the equilibria will be positive (and therefore admissible) or not. For the parametric domains where there exists a single positive equilibrium, this equilibrium is globally stable.

8 Conclusion

We have presented a constructive and systematic method to develop bioprocess models in 4 steps. Let us recall that the modelling of a bioprocess must be performed in the framework of a clearly identified objective. The modelling must correspond to the quality and the quantity of the available information so that the model can be correctly validated and identified.

The first modelling steps consists in gathering the physical and chemical principles that can apply to the system and to assume a reaction scheme in order to obtain the mass balance model.

In a second step, one must take benefit of the constraints that the model must verify and use the empirical relationships to find an analytical expression for the reaction kinetics.

The third step consists in identify the model parameters by separating those who are related to the mass balances (yield coefficients), those who

are related with the used physical principles (affinity constants, transfer constants, etc) and those who intervene in the reaction rates.

Finally, the ultimate modelling step must not be neglected: namely the model validation. During this last step the model quality must be tested using the more objective as possible criteria. The validity of the model must be assessed along its ability to properly represent the mass balance, to reproduce correctly the qualitative features of the data, and to fit quantitatively the data. The important point is that the data which must be used for model validation must not have been already used in the model construction phase. During the validation step, not only the quality of the model will be assessed, but also its validity domains: the working domains (in terms of state variable and inputs) where the model is satisfactory.

To conclude, we insist on the fact that the modelling step can be very long and expensive, but the quality of a model is a necessary conditions to ensure that a controller or an observer based on it model will properly work.

References

Andrews, J. (1968). A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrate. *Biotechnol. & Bioeng.*, 10:707–723.

Bailey, J. E. and Ollis, D. F. (1986). *Biochemical engineering fundamentals.* McGraw-Hill.

Bastin, G. and Dochain, D. (1990). *On-line estimation and adaptive control of bioreactors.* Elsevier, Amsterdam.

Batstone, D., Keller, J., Newell, B., and Newland, M. (1997). Model development and full scale validation for anaerobic treatment of protein and fat based wastewater. *Water Science and Technology*, 36:423–431.

Bernard, O. and Gouzé, J.-L. (1995). Transient behavior of biological loop models, with application to the Droop model. *Mathematical Biosciences*, 127(1):19–43.

Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A., and Steyer, J.-P. (to appear). Dynamical model development and parameter identification for an anaerobic wastewater treatment process. *Biotech.Bioeng.*

Chen, L., Bernard, O., Bastin, G., and Angelov, P. (2000). Hybrid modelling of biotechnological processes using neural networks. *Contr. Eng. Prcatice*, 8:821–827.

Costello, D., Greenfield, P., and Lee, P. (1991). Dynamic modelling of a single-stage high-rate anaerobic reactor - I. Model derivation. *Water Research*, 25:847–858.

Darouach, M. (1994). On the novel approach to the design of the unknown input observers. *IEEE Trans. Autom. Cont.*, 39(3):698–699.

Edelstein, L. (1988). *Mathematical Models in Biology.* Random House, New York.

Gouzé, J.-L. (1998). Positive and negative circuits in dynamical systems. *Journal Biol. Syst.*, 6(1):11–15.

Hansen, S. R. and Hubbell, S. P. (1980). Single-nutrient microbial competition. *Science*, 207(28):1491–1493.

Hertz, J., Krogh, A., and Palmer, R. (1991). *Introduction to the Theory of Neural Computation.* Addison-Wesley.

Hill, D. and Barth, C. (1977). A dynamic model for simulation of animal waste digestion. *Journal of the Water Pollution Control Association*, 10:2129–2143.

Horn, R. and Johson, C. (1992). *Matrix analysis.* Cambridge Univesity Press.

Hou, M. and Mller, P. (1991). Design of observers for linear systems with unknown inputs. *IEEE Trans. Autom. Contr.*, AC-37(6):871–875.

Jeffries, C. (1986). Qualitative stability of certain nonlinear systems. *Linear Algebra and its Applications*, 75:133–144.

Karama, A., Bernard, O., Genovesi, A., Dochain, D., Benhammou, A., and Steyer, J.-P. (2001). Hybrid modelling of anaerobic wastewater treatment processes. *Wat. Sci. Technol.*, 43(1):43–50.

Khalil, H. (1996). *Nonlinear Systems.* Macmillan Publishing Company.

Kudva, P., Viswanadham, N., and Ramakrishna, A. (1980). Observers for linear systems with unknown inputs. *IEEE Trans. Autom. Contr.*, AC-25(1):113–115.

Li, B. (1998). Global asymptotic behavior of the chemostat: General response functions and different removal rates. *SIAM Journal*, 59.

Merchuk, J. (1977). Further considerations on the enhancement factor for oxygen absorption into fermentation broth. *Biotechnol. & Bioeng.*, 19:1885–1889.

Moletta, R., Verrier, D., and Albagnac, G. (1986). Dynamic modelling of anaerobic digestion. *Wat.Res.*, 20:427–434.

Monod, J. (1942). *Recherches sur la croissance des cultures bactriennes.* Hermes, Paris, France.

Mosey, F. (1983). Mathematical modelling of the anaerobic digestion process: regulatory mechanisms for the formation of short-chain volatile acids from glucose. *Water Science and Technology*, 15:209–232.

Sacks, E. (1990). A dynamic systems perspective on qualitative simulation. *Artif. Intell.*, 42:349–362.

Segel, L. A. (1984). *Modeling Dynamic Phenomena in Molecular and Cellular Biology.* Cambridge University Press, Cambridge.

Walter, E. and Pronzato, L. (1994). *Identification de modèles paramétriques.* Masson.

Appendix A. Theoretical determination of the dimension of K

Let us integrate equation (7) between 2 time instants t and $t+T$:

$$\xi(t+T) - \xi(t) - \int_t^{t+T} D(\xi_{in}(\tau) - \xi(\tau))) + Q(\xi(\tau))d\tau = K \int_t^{t+T} r(\xi(\tau))d\tau\,, \tag{47}$$

Let us denote:

$$v(t) = \xi(t+T) - \xi(t) - \int_t^{t+T} D(\xi_{in}(\tau) - \xi(\tau))) + Q(\xi(\tau))d\tau$$

and

$$w(t) = \int_t^{t+T} r(\xi(\tau))d\tau$$

Equation (47) can then be rephrased:

$$v(t) = K\, w(t) \tag{48}$$

The vector $v(t)$ can be estimated along time on the basis of the available measurements. The integral value can be estimated *e.g.* with a trapeze approximation.

To avoid conditioning problem and to give the same weighting to all the state variables, we normalise the data vectors $u(t_i)$ as follows:

$$\tilde{v}(t_i) = \frac{v(t_i) - e(v)}{\sqrt{N}\sigma(v)}$$

where $e(v)$ is the average value of $v(t_i)$, and $\sigma(v)$ their standard deviation.

Now the question of the dimension of 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 are looking for a full rank matrix K. Otherwise, it would mean that the same dynamical behaviour could be obtained with a matrix K of lower dimension.

Determining the dimension of the $v(t)$ space is a classical problem in statistical analysis. It corresponds to the principal component analysis that determines the dimension of the vectorial space spanned by the vectors k_i, rows of K. To reach this objective, we consider matrix U obtained from a set of N recording of $v(t)$:

$$V = (\tilde{v}(t_1), \ldots, \tilde{v}(t_N))$$

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 of full rank. This means first that there are more measurements than reactions. It means also that the reaction are independent (none of the reaction rates can be written as a linear combination of the other ones).

Property 5 *For a matrix K of rank k, if W has full rank, then the $N \times N$ matrix $M = V^T V = W^T K^T K W$ has rank k. Since it it is a positive symmetric matrix, it can be written, by:*

$$M = P^t \Sigma P$$

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

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

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

Proof: it is direct application of the singular decomposition theorem (Horn and Johson, 1992). Since rank (M) = rank $(\Sigma) = k$, it provides the result.

Now from a theoretical point of view it is possible to determine the number of reactions in the reaction scheme: it corresponds to the rank of K or, in other words, to the number of non zero singular values of $V^T V$.

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 $M = V^T V$.

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 $v(t)$. 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).

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

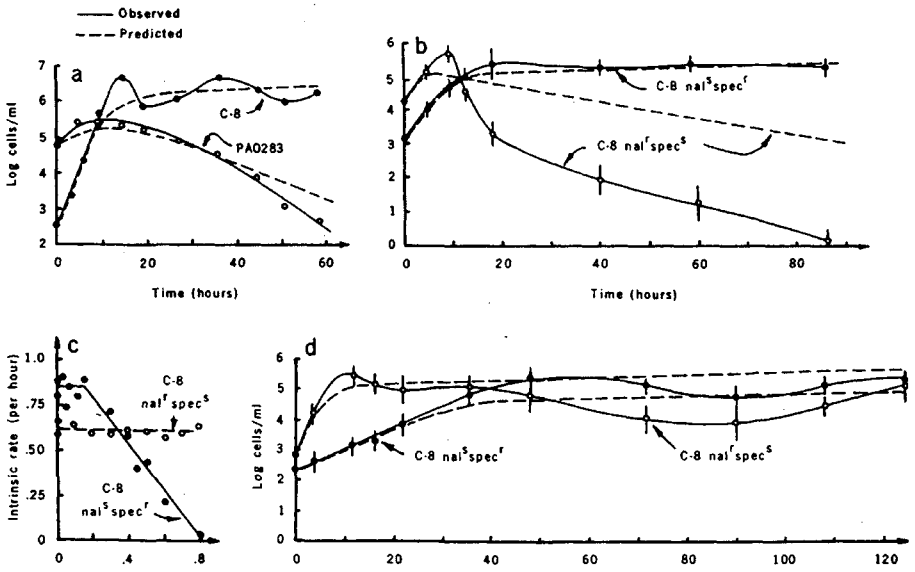

Figure 7: Experimental validation of the qualitative model behaviour. Quantitative model predictions are represented as well. The qualitative model predictions are verified for: a) Two species (*Escherichia coli*, strain C-8 and *Pseudomonas aeruginosa*, strain PA0283 which differ from their half-saturation constants. b) Two strains of *Escherichia coli* which differ from their maximal growth rates. d) Coexistence obtained with 2 strains of *Escherichia coli* which have the same parameter J_i. Figure c) represents the effect of nalidixic acid on the maximal growth rate for the 2 considered strains C-8. (from (Hansen and Hubbell, 1980)).