I. Troch, F. Breitenecker, eds. ISBN 978-3-901608-35-3

OBSERVERS FOR THE ESTIMATION OF FUNCTIONAL GROUPS OF THE ANAEROBIC DIGESTION

J. Harmand[1,4], A. Rapaport[2,4], D. Dochain[3], J. J. Godon[1]
[1]LBE-INRA, Narbonne, France; [2]UMR-ASB, Montpellier, France, [3]CESAME, UCL, Louvain-la-Neuve, Belgium, [4]INRIA MERE Research Team, Montpellier, France

Corresponding Author: J. Harmand[1,4], LBE-INRA, Avenue des étangs, 11100 Narbonne, France; harmand@supagro.inra.fr

Abstract. In this paper, we show how specific tools of automatic control, and in particular nonlinear observer theory can be exploited to estimate the most important different functional groups of the anaerobic digestion independently of its kinetics. In particular, we show how such a strategy can be used to investigate the important "assignation problem" (assign a function to a species) within the framework of microbial ecology of the anaerobic digestion.

1 Introduction

When considering a complex microbial ecosystem (here, "complex" means that we consider ecosystems involving several interacting species, each of them possibly realizing different functions), one objective is to identify which species is responsible for what function of the ecosystem. In other word, one important ecological question is to assign to a given species "one of the functions realized by the ecosystem". This "assignation problem" is related to the question of "who does what?" in this ecosystem. Obviously, this step is fundamental and is a mandatory prerequisite to the introduction into bioprocess models of a number of species higher than the number of functional groups. To do so, recent molecular probes are available to monitor functional groups, cf. for instance the principle of micro-arrays [1]. However, these techniques are usually quite specific: "you only see what you look for" or in other words "you only see what you already know". In addition, such methods are usually costly and time-consuming. An alternative method proposed in this paper consists in using modeling tools together with molecular fingerprinting techniques. Molecular techniques offer a new way of monitoring microbial ecosystems. For instance, the SSCP is based on the discrimination of the "16S DNAr" molecules using the natural conformation polymorphism of these molecules after PCR. In a recent work, it has been shown, under appropriate conditions, that the relative abundances of the most abundant species or "groups of species" could be monitored with this technique, [2]. In the following, we call FU (for Functional Units) a species or group of species that has been identified as being majoritory within the ecosystem with the help of a molecular technique as the SSCP.

Assuming a functional model is available (typically a mass-balance type model describing both biomasses and substrates/products dynamics, [3]), several "direct" or "indirect" procedures have been proposed to identify the function of each of the detected species or species groups (assuming that each group or species realizes only one function) in complex reactional biological systems cf. [4]. In this last paper, it was noticed that the approach proposed could not be used if the diversity of the ecosystem under interest is too high. Consider now that instead of monitoring "all species", we restrict our attention to important (majoritory) functional species. In this case, the approach can be used to assign specific functions to these important groups.

In this paper, we propose a modified procedure than that proposed in [4] for solving the assignation problem for functional groups of the anaerobic digestion process. Although the diversity of this system is known to be quite high, (cf. the pioneering work reported in [5] and the recent update by [6]), molecular techniques can be used to get insight about the assignation problem with respect to the major functional groups of the anaerobic digestion process such as hydrolytic or acidogenic bacteria.

2 The simplified model and the optimization approach

Consider the following simplified model of the anaerobic digestion. For the sake of simplicity, only three functions are considered here: the hydrolysis (which transforms S_1 into S_2), the acidification (a biological process realized by a bacteria consortium X_B which transforms organic macromolecules S_2 into Volatile Fatty Acids S_3) and the methanization which transforms S_3 into biogas.

The general form of the functional mass-balance model of this system in the chemostat is given by:

$$\begin{cases} \dfrac{dX_A}{dt} = (\mu_A(S_1) - D)X_A \\ \dfrac{dX_B}{dt} = (\mu_B(S_2) - D)X_B \\ \dfrac{dX_C}{dt} = (\mu_C(S_3) - D)X_C \\ \dfrac{dS_1}{dt} = -k_1\mu_A(S_1)X_A + (S_{in} - S_1)D \\ \dfrac{dS_2}{dt} = k_2\mu_A(S_1)X_A - k_3\mu_B(S_2)X_B - S_2D \\ \dfrac{dS_3}{dt} = k_3\mu_B(S_2)X_B - k_4\mu_B(S_3)X_C - S_3D \end{cases} \quad (1)$$

where S_{in} is the input substrate concentration, S_1, S_2 and S_3 are the non hydrolysed organic macromolecules, organic matter (COD) and VFA, respectively, D is the dilution rate (ratio of the input flow rate over the volume), X_A, X_B and X_C are the concentrations in hydrolytic, acidogenic and methanogenic bacteria, k_i are yield coefficients while μ_A, μ_B and μ_C are their growth rate.

In [4], the SSCP was proposed to monitor species in low biodiversity ecosystems as for instance nitrification. When considering the Anaerobic digestion process, it is to be noticed that the SSCP does not allow to monitor methanogens. Thus, we propose here a modified version of the approach proposed in [4] in order to identify – within all the detected major bacteria species – which ones belong to the hydrolytic or acidogenic functional groups only.

The problem is thus the following: given the measurement of $S_1(t)$, $S_2(t)$ and $S_3(t)$, how estimating $X_A(t)$, $X_B(t)$ and $X_C(t)$ in a first step, and how these estimations can be related to the measurements realized with a molecular technique in a second step?

Indeed, in addition to the measurements of the substrate concentrations, it is assumed that we have the measures of the "major individual species concentrations" X_i (i=1…N) (called FUs in the introduction) and the total biomass $X_T=X_A+X_B+X_C$. The assignation problem is to determine which FUs X_i are part of X_A and which ones are part of X_B (remember that we do not monitor X_C by SSCP).

To better understand our presentation, we recall here the different methods we suggested in the introduction as "direct" and "indirect" approaches on the specific example of the nitrification process.

The direct approach assumes the model (1) is available. Since X_C is not monitored, we consider the restricted model:

$$\begin{cases} \dfrac{dS_1}{dt} = -k_1\mu_A(S_1)X_A + (S_{in} - S_1)D \\ \dfrac{dS_2}{dt} = k_2\mu_A(S_1)X_A - k_3\mu_B(S_2)X_B - S_2D \end{cases} \quad (3)$$

Involving only the two first bioreactions that we simulate using successively the 2^N combinations of the FUs concentrations X_i (which are time series) where $\lambda_i \in \{0,1\}$ as inputs:

$$\begin{cases} X_A(t) = \sum_{i=1}^{N} \lambda_i X_i(t) \\ X_B(t) = \sum_{i=1}^{N} (1-\lambda_i) X_i(t) \end{cases} \quad (4)$$

The optimal solution is the combination of the N parameters λ_i for which the criterion:

$$J = \min_{\lambda_i|_{\lambda_i \in \{0,1\}}} \sum_{i=1}^{N_S} \sum_{j=1}^{M} \left(S_i^{measures}(j) - S_i^{predictions}(j) \right)^2 \qquad (5)$$

with N_S=2 and M is the number of samples over the period of time considered, is minimized. As underlined in the introduction, notice that in such an approach the restricted model derived from the model (1) - and in particular μ_A and μ_B - are assumed to be known!

In the indirect approach, we proceed in two steps. First, the trajectories X_A and X_B are generated in minimizing:

$$J = \min_{X_A, X_B} \sum_{i=1}^{N_S} \sum_{j=1}^{M} \left(S_i^{measures}(j) - S_i^{predictions}(j) \right)^2 \qquad (6)$$

with N_S=2 using either observers or control algorithms. Then, in a second step, the static mixed integer optimization problem was proposed to be solved:

$$\begin{cases} J = \min_{\lambda_i|_{\lambda_i \in \{0,1\}}} \sum_{k=1}^{M} \left[\left(X_A(t_k) - \hat{X}_A(t_k) \right)^2 + \left(X_B(t_k) - \hat{X}_B(t_k) \right)^2 \right] \\ X_A(t) = \sum_{i=1}^{N} \lambda_i X_i(t) \\ X_B(t) = \sum_{i=1}^{N} (1 - \lambda_i) X_i(t) \end{cases} \qquad (7)$$

where the notations $\hat{X}_A$ and $\hat{X}_B$ stand for the estimated trajectories of X_A and X_B whatever the method used to generate them.

3 The approach using nonlinear observers

In this section, we recall the approach proposed in [4] for a nitrification process based on the use of nonlinear observers. Consider the mass balance model (1) with only two reactions in cascade. Taking into account the fact that the total biomass X_T is supposed to be measured, we enounce the following results (for simplicity, the results were derived with $1/Y_A=k_1=k_2$ and $1/Y_B=k_3=k_4$ but can be extended to the actual more general model: hereafter, we call this modified restricted model (1bis)):

Proposition 1: The dynamical system:

$$\begin{cases} \dfrac{d\hat{Z}_1}{dt} = -D\left(\hat{Z}_1 - S_{in}\right) + G_1\left(\hat{X}_A + \hat{X}_B - X_T\right) \\ \hat{X}_A = Y_A\left(\hat{Z}_1 - S_1\right) \\ \dfrac{d\hat{Z}_2}{dt} = -D\left(\hat{Z}_2 - S_{in}\right) + G_2\left(\hat{X}_A + \hat{X}_B - X_T\right) \\ \hat{X}_B = Y_B\left(\hat{Z}_2 - S_1 - S_2\right) \end{cases} \qquad (8)$$

is a partially tunable observer for the system (1bis).

Notice that it is exactly the asymptotic observer if G_1=G_2=0.

Proof of the proposition 1: Consider the error vector:

$$\begin{cases} e_{X_A} = X_A - \hat{X}_A = Y_A\left(Z_1 - \hat{Z}_1\right) = Y_A e_{Z_1} \\ e_{X_B} = Y_B e_{Z_2} \end{cases}$$

where $Z_1 = \dfrac{X_A}{Y_A} + S_1$ and $Z_2 = \dfrac{X_B}{Y_B} + S_1 + S_2$. Its dynamics is given by:

$$\dot{e}_X = \begin{pmatrix} -D + Y_A G_1 & Y_A G_1 \\ Y_B G_2 & -D + Y_B G_2 \end{pmatrix} e_X .$$

In the case where D is constant, it is then straightforward to verify that the eigenvalues of the observer are given by:

$$\begin{cases} \lambda_1 + \lambda_2 = -2D + Y_A G_1 + Y_B G_2 \\ \lambda_1 \lambda_2 = D\left(D - Y_A G_1 - Y_B G_2\right) \end{cases} \rightarrow \begin{cases} \lambda_1 = -D \\ \lambda_2 = -D + Y_A G_1 + Y_B G_2 \end{cases} .$$

If the gains G_1 and G_2 are chosen such that $\lambda_2<0$, one has $\lim_{t\to\infty} \hat{X}_A = X_A$ and $\lim_{t\to\infty} \hat{X}_B = X_B$ and thus the system (8) is an observer for the system (1bis).■

The observer dynamics are completely independent of the bioreactions kinetics (such observers are called asymptotic observers, cf. [3]) μ_A and μ_B which can be very complicated functions of any state and/or exogen variables. The "price to pay" to build such an observer with such few knowledge is that the observer is not completely tunable (only one eigenvalue can be placed arbitrarily). Now, it may happen that part of the process kinetics is known: for instance μ_A or μ_B. In this case, one may synthesize a new observer in taking advantage of the fact that the total biomass is measured and in coupling a bounded error observer (cf. [7]) and the partially tunable observer presented hereabove. For instance, assuming μ_A is known, we get the following result:

Proposition 2: The dynamical system:

$$\begin{cases} \dfrac{d\hat{Z}_1}{dt} = -D\left(\hat{Z}_1 - \theta_1 S_{in}\right) + \left(1-\theta_1\right)\dfrac{\mu_A(S_1)}{Y_A}\hat{X}_A \\ \hat{X}_A = Y_A\left(\hat{Z}_1 - \theta_1 S_1\right) \\ \dfrac{d\hat{Z}_2}{dt} = -D\left(\hat{Z}_2 - S_{in}\right) + G_2\left(\hat{X}_A + \hat{X}_B - X\right) \\ \hat{X}_B = Y_B\left(\hat{Z}_2 - S_1 - S_2\right) \end{cases} \quad (9)$$

where the auxiliary variable Z_1 is now given by $Z_1 = \dfrac{X_A}{Y_A} + \theta_1 S_1$ and thus $\hat{Z}_1 = \dfrac{\hat{X}_A}{Y_A} + \theta_1 S_1$ while Z_2 and $\hat{Z}_2$ are unchanged, is a completely tunable observer for the system (1bis).

Again, notice that it is exactly the asymptotic observer if $\theta_1 = 1$ and G_2=0.

Sketch of the proof of the proposition 2: The dynamics of the error matrix is given by:

$$F = \begin{pmatrix} -D + \left(1-\theta_1\right)\mu_A\left(S(t)\right) & 0 \\ Y_B G_2 & -D + Y_B G_2 \end{pmatrix}$$

Notice F has only one time-varying term. A systematic approach to show that the error matrix is stable is the following. First, assume $\mu_A\left(S(t)\right) \geq \underline{\mu_A} > 0 \ \forall t$. We have $v(t) = \left(\theta_1 - 1\right)\left(\mu_A(t) - \underline{\mu_A}\right) \geq 0$ if $\theta_1 \geq 1$ and with $v(t) \in \left[0, \left(\theta_1 - 1\right)\left(\overline{\mu_A} - \underline{\mu_A}\right)\right]$. Let the constant matrix:

$$\underline{F} = \begin{pmatrix} -D + \left(1-\theta_1\right)\underline{\mu_A} & 0 \\ Y_B G_2 & -D + Y_B G_2 \end{pmatrix}.$$

Then, we can place the eigenvalues of $\underline{F}$ using the change of variable which diagonalizes it. For any positive pair $\lambda_1 \neq \lambda_2$, one can find θ_1 and G_2 such that $-D + \left(1-\theta_1\right)\underline{\mu_A} = -\lambda_1$ and $-D + G_2 Y_B = -\lambda_2$ with $\lambda_i > 0$, $i = 1,2$ and $\theta_1 > 1$. Posing $\xi_1 = e_{X_A}$ and $\xi_2 = -G_2 e_{X_A} + \left(\lambda_2 - \lambda_1\right) e_{X_B}$, one gets:

$$\begin{cases} \dot{\xi}_1 = -(\lambda_1 + v(t))\xi_1 \\ \dot{\xi}_2 = G_2 v(t)\xi_1 - \lambda_2 \xi_2 \end{cases}.$$

Introducing the candidate Lyapunov $V = \frac{1}{2}\xi_1^2 + \frac{1}{2}\xi_2^2$, it is easy to show that $\dot{V}$ is decreasing if $\lambda_1 + v(t) - \frac{(G_2 v(t))^2}{4\lambda_2} > 0, \forall t$. In particular, this holds if G_2 is chosen such that $\lambda_2 > \frac{G_2{}^2 \bar{v}}{4}$ where $\bar{v} \geq \max_{t \geq 0} v(t)$. ■

4 A modified approach

In [4], it was assumed that all bacteria could be monitored by molecular techniques. In fact, archae are not monitored by SSCP: thus the approach previously proposed cannot be used without modifications. We propose here to extend the procedure proposed in [4] for 2 bioreactional systems to the two important groups of the anaerobic digestion (hydrolytic and acidogenic functional groups) in i) building an observer for X_C based on the model (1) (the simplest being an asymptotic observer) and ii) apply exactly the method proposed in section 3 with $X_T = X_A + X_B + X_C - \hat{X}_C$.

5 Simulation example

The proposed approach is now illustrated in simulation. For simplicity, the units are adimensional. It was assumed in the following that 50 hydrolytic bacteria groups transform non hydrolyzed COD into simpler molecules and 40 acidogenic groups into VFA through a complex unknown kinetics network. Initial conditions were chosen such that only three hydrolytic and three acidogenic groups are assumed to represent more than 98 % of the respective functional groups and can be monitored through SSCP (cf. figure 1). The process is a chemostat operating with variable inputs (both dilution rate and input COD concentration are time-varying signals). First, using a classical asymptotic observer, the theoretical trajectory of the methanogenic bacteria was obtained and substracted from the total biomass. Then theoretical trajectories of hydrolytic and acidogenic bacteria were obtained. Some simulation results are shown in Figures 1 and 2.

Combining the estimations of the total hydrolytic and acidogenic groups, it is possible to find, within all possible combinations of individuals FUs, the one which best approximates the estimated trajectories.

The proposed approach allows one to give one solution to the assignation problem. Among the limitations of the approach, one can notice that all FUs are assumed to be functional.

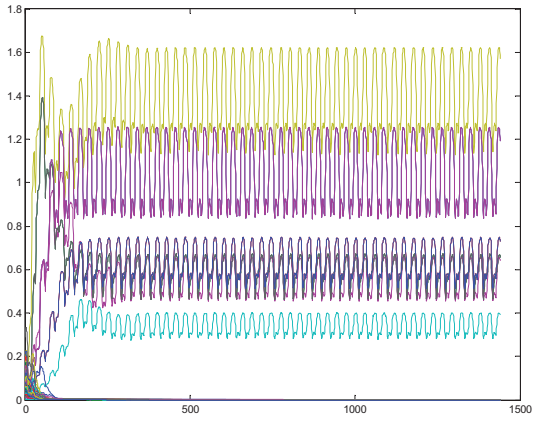

Figure 1 *: Time series of the different FUs*

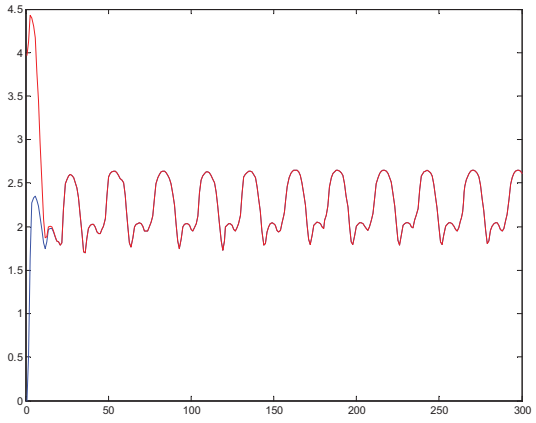

Figure 2 *: Total Acidogenic bacteria and its estimation using a nonlinear observer*

6 Conclusions

In the present paper, a procedure to solve the important assignation problem in the microbial ecology of anaerobic digestion processes was presented and evaluated using simulated data. The three-step procedure inspired from the one first proposed in [4] i) uses an observer to generate – independently of the kinetics of the process – the trajectory of the functional methanogenic biomass, ii) uses a coupled observer to generate the theoretical

trajectories of the hydrolytic and acidogenic biomasses and iii) looks for the combination of individual species concentrations detected via the use of molecular fingerprint techniques that best approximates the experimental data. For solving the step ii), the second observer (8) should be preferred if nothing is known about the kinetics. If only one kinetics is known, we can completely tune the observer (9) in coupling a robust observer (in the sense it does not depend on the unknown kinetics) with a hybrid one (cf. [7]). Finally, if the two kinetics are known, it should be noticed that there is no reason of coupling the two observers: each observer is tunable independently of the other. It seems here particularly important to notice that it is sometimes possible to build different observers from the set of measurements S_i. Indeed, the design of the proposed observers is based on the introduction of auxiliary variable Z which allows the observer dynamics to be independent on the kinetics. Depending on the quality of the different available measurements, it may happen that one change of variable gives better practical results than another one. It means that it is possible to reconstruct the same variables – here X_A, X_B and X_C – from different data: how using this redundancy obviously poses new interesting research questions. Another interesting perspective is to extend the results to cases where there are uncertainties on the input substrate concentration or in the yield coefficients.

7 References

[1] H. Sanguin, A. Herrera, C. Oger-Desfeux, A. Dechesne, P. Simonet, E. Navarro, T. M. Vogel, Y. Moenne-Loccoz, X. Nesme and G. L. Grundmann, "Development and validation of a prototype 16S rRNA-based taxonomic microarray for Alphaproteobacteria", *Environmental Microbiology,* vol. 8, no. 2, pp. 289-307, 2006.

[2] P. Loisel, J. Harmand, O. Zemb, E. Latrille, C. Lobry, J. P; Delgenes and J. J. Godon, "DGE and SSCP molecular fingerprintings revisited by simulation and used as a tool to measure microbial diversity", *Environmental Microbiology*, vol. 8, no. 4, pp. 720-731, 2006.

[3] G. Bastin and D. Dochain, " On-Line Estimation and Adaptive Control of Bioreactors ", Elsevir, 1990.

[4] Dumont, M., A. Rapaport, J. Harmand, B. Benyahia and J-J. Godon (2008) "Observers for Microbial Ecology - How Including Molecular Data into Bioprocess Modeling?", 16th Mediterranean Conference on Control and Automation, pp. 1381-1386, June 25-27, Congress Centre, Ajaccio, Corsica, France.

[5] J. .J Godon, E. Zumstein, P. Dabert, F. Habouzit, R. Moletta, "Microbial 16S rDNA diversity in an anaerobic digester", Water Science and Technology, vol. 36, no. 6-7, pp. 49-55, 1997.

[6] T. Narihiro, Y. Sekiguchi, "Microbial communities in anaerobic digestion processes for waste and wastewater treatment: a microbiological update", Current Opinion in Biotechnology, vol. 18, no. 3, pp. 273-278, 2007.

[7] V. Lemesle and J. L. Gouzé, "Hybrid bounded error observers for uncertain bioreactor models", Bioprocess Biosystems Engineering, vol. 27, pp. 311-318, 2004.