On-line estimation in a distributed parameter bioreactor: Application to the Gluconic Acid production

Míriam R. García[1] *, Carlos Vilas[2], Eva Balsa-Canto[2], Velislava N. Lyubenova[3], Maya N. Ignatova[3], Antonio A. Alonso[2]

[1] Hamilton Institute, NUIM, Ireland
[2] Process Engineering Group, IIM-CSIC, C/ Eduardo Cabello 6 36208, Vigo, Spain
[3] Institute of Control and System Research, BAS, P.O. Box 79 1113, Sofia, Bulgaria
* Correspondence concerning this article should be addressed to M.R. García miriamr@iim.csic.es

Abstract

This work presents a methodology which exploits the underlying biochemical structure of bioprocesses to estimate concentrations in aerobic fermenters from oxygen measurements. Although a number of estimators have been proposed over the years in the literature, the methodology proposed in this work is able to operate in transient conditions while does not require the knowledge of the growth kinetics. In addition, it can be also applied to fermenters where the spatial distribution of the concentrations is relevant. In this case, we propose a systematic approach to optimally locate the sensors based on the use of reduced order models. This method allows the reconstruction of the oxygen concentrations from a limited number of sensors. Finally, the methodology proposed will be illustrated on a horizontal tubular reactor for the production of gluconic acid by free-growth of *Aspergillus niger*.

Key words: On-line indirect estimation, Aerobic Fermentation, Tubular (Bio)Reactors, Optimal sensor location.

1. Introduction

Control of bioreactors has been hampered by a number of obstacles essentially associated with the lack of reliable sensors capable of providing real time measurements of the relevant variables of the process. Additionally, whereas in other fields it is possible to use reliable mathematical descriptions of the processes to design software sensors (observers), this is usually not the case in biotechnology applications where mathematical representations are not well known, specially regarding the reaction rates involved.

To overcome these limitations, a theoretical framework which takes advantage of the biochemical reactions of biomass growth and product formation was proposed in [28]. The same basic approach was used in [6] for on-line estimation in stirred tank bioreactors. In that work, the authors exploit the underlying reaction structure and transfer mechanisms to systematically design and implement identification schemes for variables and parameters. Other examples for particular types of bioreactors can be seen in the works [12, 11].

However, a general description of the methodology together with the precise conditions to apply it in aerobic fermentation still deserves attention. In particular extensions of the methodology to cope with spatial distribution of species concentrations constitute one of those open problems.

As discussed in [24] three decades ago, the use of distributed parameter reactors in biotechnology opened new opportunities which gained an increasing interest over the years [32]. The mathematical description of such systems relies on the microscopic conservation laws for mass and energy which result into a nonlinear set of partial differential equations (PDE). In most cases the analytical solution is unknown and classical numerical methods for distributed parameter systems (like finite differences or finite element methods) lead to large sets of ordinary differential equations which are computationally involved. This makes the approach unsuitable for real time tasks like on-line estimation, optimization or control [7]. In addition, the estimations of the non-measurable variables should be combined with measurements of the remaining variables covering the whole spatial domain. This requires having access to a large number of on-line sensors which may be too expensive or physically impossible to be implemented in the desired process [12, 30, 1].

In this work, an on-line identification scheme is developed for aerobic fermentation with unknown reaction rates, provided that oxygen can be measured on-line. This scheme consists of the combination of two different estimation methodologies which make use of mass invariance principles [11] and field reconstruction techniques [2]. The first one will allow the estimation of some of the relevant variables of the process without the need of expressions for the reaction rates. This approach is based on a given state transformation that leads to a description of the system independent of the reaction rates. In the case of lumped bioreactors (where the spatial distribution is not relevant) the methodology will only require modeling the transfer mechanisms, and measuring or estimating the relationships among species (yield coefficients). For distributed parameter bioreactors, on-line information of the oxygen concentration would be needed at any location of the reactor. The second methodology is the one employed to that purpose. In this respect, the approach proposed in [2] and [14] will be applied to recover the time evolution and spatial distribution of the state variables from a limited number of sensors optimally located inside the reactor.

The identification scheme presented will be illustrated and validated on a case study where the spatial distribution cannot be ignored. The example used is related to the production of gluconic acid (GA) from *Aspergillus niger*. For versatility reasons production is usually carried out in well stirred batch reactors supplied with oxygen [21, 19, 18, 35, 23]. However, some authors [25, 3] pointed out that continuous production with free-growth cells may be more advantageous than the batch one as it reduces the tendency of *Aspergillus niger* to cause clogging under high glucose concentration conditions. The bioreactor used in this work as a case study consist of a continuous tubular reactor for the production of GA from free-growth cells with oxygen supplied at any point of the fermenter. In order to avoid consumption of GA and clogging, the inlet glucose concentration has to be manipulated or controlled to maintain low, but not zero, concentration inside the reactor what calls for on-line estimation techniques.

The paper is structured as follows: in Section 2 the general formulation for both lumped and distributed parameter fermenters is presented. The description of the proposed scheme is performed in Section 3. Finally, the methodology is illustrated on a tubular reactor for GA production in Section 4. Main conclusions are summarized in Section 5.

2. General mathematical model representation

The model equations which describe the behavior of the relevant process variables (concentration of biomass, substrates, etc) inside (bio)reactors are usually obtained from mass balances. When such balances are applied to distributed parameter systems, the set of equations takes the form:

$$\frac{\partial s_i}{\partial t} = T(s_i) + R_{s_i}(s_1, s_2, ..., s_j, ...s_{n_s}), \qquad s_i(0, \vec{\xi}) = s_i^0(\vec{\xi}), \quad \forall \xi \in \Omega \tag{1a}$$

where s_i $(i = 1, .., n_s)$ stands for the concentration of each of the n_s species. The dynamic behavior is characterized by the corresponding reaction rate R_{s_i} and transfer mechanism $G(s_i)$ (input and output). s_i^0 denotes the initial bioreactor state and $\vec{\xi}$ represents the vector of spatial coordinates. The transfer term usually collects the diffusion and convection contributions. The solution of the partial differential equations will be also constrained by the presence of boundary conditions of the form:

$$\vec{n}\vec{\nabla}s_i(t, \Gamma) + qs_i(t, \Gamma) = g(t), \quad \forall t > 0 \tag{1b}$$

with Γ representing the boundary of the spatial domain and q and g being, respectively, the parameter and function describing the different types of boundary conditions. In this way, the so called Neumann boundary conditions can be obtained by choosing $q = 0$. Alternatively, Dirichlet boundary conditions can be attained by selecting q large enough so that $\overrightarrow{n}\vec{\nabla}s_i$ can be disregarded as compared to qs_i. The resulting variable at the boundary then becomes of the form g/q (see [14] for details).

The process satisfies the following assumptions [19]:

A1 Product formation is oxidative and the fermentation is aerobic.

A2 The relationships among species are constant and known or can be estimated by experiments.

A3 The transfer terms (i.e. their mathematical expression and parameters) are known for all the species $T(s_i), \quad \forall i = 1, ..., n_s$.

In addition for distributed parameter systems we have:

A4 A number of on-line oxygen sensors is available and can be located at any point on the spatial domain Ω.

Under assumptions A1-A2 the chemical mechanism for aerobic fermentation can be written as:

$$Y_{O_2}O_2 + \sum_{i=1}^{n_r} Y_{r_i} r_i \xrightarrow{s_2} \sum_{j=1}^{n_p} Y_{p_j} p_j + Y_X X,$$

where $\{r_i\}_i^{n_r}$ and $\{p_j\}_j^{n_p}$ are the set of reactants and products respectively, with their yield coefficients represented by Y. As usual, oxygen is denoted by O_2 and biomass by X.

The general formulation of the mass balance and known yield coefficients (assumption A2) can be exploited to build up the following transformations which do not depend on the production terms:

$$z_i = r_i - \frac{Y_{r_i}}{Y_{O_2}} O_2; \qquad z_j = p_j + \frac{Y_{p_j}}{Y_{O_2}} O_2; \qquad z_X = X + \frac{Y_X}{Y_{O_2}} O_2, \tag{2}$$

with $i = 1, ..., n_r$ and $j = 1, ..., n_p$. It should be noted that the only requirement to obtain the transformation is that $Y_{O_2} \neq 0$, i.e., the formation of the product has to be oxidative and the fermentation has to be aerobic (assumption A1).

The transient behavior of the new entities z_i, z_j and z_X (not produced neither destroyed by the fermentation) takes the form:

$$\frac{\partial z_i}{\partial t} = T(r_i) - \frac{Y_{r_i}}{Y_{O_2}} T(O_2); \quad \frac{\partial z_j}{\partial t} = T(p_j) + \frac{Y_{p_j}}{Y_{O_2}} T(O_2); \quad \frac{\partial z_X}{\partial t} = T(X) - \frac{Y_X}{Y_{O_2}} T(O_2) \tag{3}$$

The solution of the $n_r + n_p + 1$ PDEs requires the transport information (Assumption A3) and measurements of oxygen at any point inside the reactor. Such measurements are usually limited due to the reduced number of sensors available. In order to circumvent such drawback, the approach presented in [2] and [14] will be used to obtain the oxygen spatial distribution inside the reactor from a limited number of, optimally located, oxygen sensors. The complete proposed estimation scheme will be presented in the next section.

3. State estimation of process variables

In order to produce reliable estimates of the process variables (oxygen, biomass, reactants and products) use will be made of the model described in section 2, Eqn (3). This model is constituted by a set of partial differential equations which in order to be solved calls for discretization methods. Possibly the most popular numerical methods for this class of systems are those based on spatial discretizations, such as finite differences (FD) [26] or finite elements (FEM) [22] methods. Both methods approximate the spatial dependency by making use of locally defined basis functions. As a result, the original system is transformed into a -usually large scale- set of ordinary differential equations (ODEs). As pointed out in [8] and [10], these approaches present a number of disadvantages which can prevent their use in on-line estimation schemes namely, they are usually computationally involved, and some essential theoretical properties may be lost by the discretization method or the degree of refinement.

In our case, an additional disadvantage arises, which relates to the need of using as many sensors as discretization points [34, 2], a number usually too large in realistic implementations. In particular efficient oxygen estimation schemes need be developed to enable the reduction of the number of sensors [30]. These issues may be properly addressed in the framework of reduced order models (ROM), i.e. those based on global defined basis functions [4]. The interested reader is referred to Appendix A and references therein for an outline of the fundamentals.

The first step in the estimation scheme is to recover the whole spatial information of the oxygen field from a few sensor measurements (*field reconstruction*). To that purpose, a static observer will be developed as follows: let us define the operator $P_w \in \mathbb{R}^{w \times n}$ as that which projects any vector $\mathbf{u} \in \mathbb{R}^{n \times 1}$ in w of its n coordinates, which correspond with the number of sensor measurements, so that $\mathbf{u}_w = P_w \mathbf{u}$. Now, let us use Eqn (24) -Appendix A- to expand the estimation of the field $\hat{\mathbf{U}}$ so that:

$$\hat{\mathbf{U}} = \Phi \hat{\mathbf{C}} \quad \Longrightarrow \quad \hat{\mathbf{U}}_w = P_w \hat{\mathbf{U}} = P_w \Phi \hat{\mathbf{C}}. \tag{4}$$

The estimation of the field is computed by finding the set $\hat{\mathbf{C}}$ which minimizes the distance between measurements $\mathbf{U}_w$ and estimates $\hat{\mathbf{U}}_w$. Formally this can be stated as a least-squares optimization problem [2, 14] of the form:

$$\min_{\hat{\mathbf{C}}} \left(\mathbf{U}_w - \hat{\mathbf{U}}_w\right)^T \left(\mathbf{U}_w - \hat{\mathbf{U}}_w\right) \quad \Longleftrightarrow \quad \min_{\hat{\mathbf{C}}} \left(\mathbf{U}_w - P_w\Phi\hat{\mathbf{C}}\right)^T \left(\mathbf{U}_w - P_w\Phi\hat{\mathbf{C}}\right) \tag{5}$$

Note that for the problem to have a solution the number of sensors must be not smaller than the number of states to be estimated. In this work, we make use of the so-called *Proper Orthogonal Decomposition (POD)* method -see Appendix A for details- to compute the basis set Φ because of its optimality properties in the sense that it requires the lowest number of state equations for describing the behavior of the system with a given degree of accuracy [17]. However, it should be remarked that the main drawback of the POD technique is that the basis functions are obtained from experimental or direct numerical simulation (DNS) data. Thus, its predictive capabilities are highly dependent on the range of conditions in which the measurements were obtained. In this work, a large set of measurements covering different operating conditions will be employed to avoid such drawback.

On a second step, the oxygen estimates computed above is employed, together with Eqns (2) and (3), to obtain the estimation of the rest of the variables (r_i, p_j and X). In order to reduce the computational load, a dynamic version of the ROM, as described in Appendix A, will be employed here where the dynamic evolution of the modes will be obtained by projecting relations (3) on a given basis set Φ so as to obtain an expression for the dynamic evolution of the modes (see Eqn (21) in Appendix A). Once the dynamic evolution of entities z_i, z_j and z_X is computed, the real process variables r_i, p_j and X are recovered by relations (2). Since in this case the number of equations is not a critical issue, the *Laplacian Spectral Decomposition (LSD)* technique will be employed to compute the basis functions. The main advantage of this method regarding the POD is that it is valid for the whole range of operating conditions. Furthermore, as it is pointed out in [13], it is one of the most efficient spectral methods. We will refer to this step as *dynamic estimation*.

A schematic view of the estimation scheme is presented in Figure 1. Oxygen measurements taken at different positions (represented by continuous arrows) in the reactor are employed to reconstruct the oxygen spatial distribution (field reconstruction block). This information (dashed arrows) is then fed into the dynamic observer so as to obtain the missing information concerning the remaining variables (reactants, products and biomass).

Finally, it must be remarked that although the POD and LSD approaches seem to be the most suitable for each case other reduction techniques (spectral or pseudo- spectral) could be employed in the field reconstruction and dynamic observer blocks. In this regard several alternatives such as Chevichev or Legendre polynomials, available in the literature [15, 29, 13, 27, 16], could be attempted.

4. Case Study: Gluconic Acid Production

The proposed methodology will be illustrated with an experiment related to the production of gluconic acid in a spatially distributed bioreactor. On the first stage of the bioreactions taking place, the glucose (G) is consumed by the microorganisms (X) to produce Gluconic Acid (GA) by means of an oxidation of the sugar aldehyde group to a carboxyl group, being the glucoholactone (GOT) the intermediate compound obtained as a result of the action of Glucose Oxidase (GOD). Microorganisms also produce catalasa (CAT) that breaks the hydrogen peroxide (H_2O_2). This suggests the following reaction mechanism [21, 19, 20]:

$$Y_{G/X}G + Y_{O_2/X}O_2 \xrightarrow{X} X \tag{6a}$$

$$G + O_2 \xrightarrow{GOD} GOT + H_2O_2 \tag{6b}$$

$$GOT + H_2O \longrightarrow GA \tag{6c}$$

$$H_2O_2 \xrightarrow{CAT} H_2O + \frac{1}{2}O_2 \tag{6d}$$

where $Y_{G/X}$ (grams of G/grams of X) and $Y_{O_2/X}$ (grams of O_2/grams of X) are the yield coefficients of glucose and oxygen, respectively.

Once the glucose has been exhausted, the second stage starts and the microorganisms begin to consume GA with linear kinetics as an alternative carbon source. Note that this stage is not considered in this work because the operation conditions are such that Glucose is never totally consumed.

Preliminary experiments [19, 20] show that the peroxide is quickly consumed to produce water and oxygen so reaction (6d) is not limiting. This leads to the following simplified mechanism representative of the behavior of the system:

$$Y_{G/X}G + Y_{O_2/X}O_2 \xrightarrow{X} \quad X \tag{7a}$$

$$Y_{G/GA}G + Y_{O_2/GA}O_2 \xrightarrow{GOD,CAT} GA \tag{7b}$$

where $Y_{G/GA}$ and $Y_{O_2/GA}$ are, respectively, the grams of glucose and oxygen consumed by gram of GA.

Following [35, 23], the Monod law is employed to describe the reaction rates with glucose and oxygen as limiting substrates. In this way, the biomass growth can be mathematically represented as:

$$R_X = \mu X \qquad \text{with} \qquad \mu = \mu_{max}\frac{G}{K_G + G}\frac{O_2}{K_{O_2} + O_2} \tag{8}$$

Experimental data show that the production of GA is growth-associated and depends on the biomass formation rate in a linear way [19, 18, 35]:

$$R_{GA} = \alpha R_X$$

where α is the Luedeking-Piret equation parameter for growth-associated product formation. Therefore, the chemical reaction now reads:

$$Y_G G + Y_{O_2}O_2 \xrightarrow{R_X} \quad X + Y_{GA}GA \tag{9}$$

with

$$Y_{GA} = \alpha = Y_{GA/X}; \quad Y_G = Y_{G/X} + Y_{G/GA}Y_{GA/X}; \quad Y_{O_2} = Y_{O_2/X} + Y_{O_2/GA}Y_{GA/X}$$

The mathematical structure of the reaction rate can be finally written as:

$$R_X = \mu_{max}\frac{G}{K_G + G}\frac{O_2}{K_{O_2} + O_2}X \tag{10a}$$

$$R_{GA} = Y_{GA}R_X \tag{10b}$$

$$R_G = Y_G R_X \tag{10c}$$

$$R_{O_2} = Y_{O_2}R_X \tag{10d}$$

On the other hand, the oxygen transfer rate (OTR) is assumed to follow the Henry's law as:

$$OTR = k_{La}(O_2^* - O_2) \tag{11}$$

where k_{La} is the oxygen transfer rate per volume unit and $O_2^* = 0.0084 gl^{-1}$ is the saturation of dissolved oxygen.

It should be noted that biomass reaction rate parameters and yield coefficients appearing in Eqns (10) and the oxygen transfer rate k_{La} in Eqn (11) are unknown and should be, therefore, estimated by fitting the model to experimental data. In this work the available data from [19] were used to that purpose. The estimated values are presented in Table 1 and the best fit is shown in Figure 2, illustrating the good reproducibility of the model.

The combination of transfer Eqn (11) and reaction rates (10), leads to the following mathematical description of the tubular reactor:

$$\frac{\partial X}{\partial t} = D_X\frac{\partial^2 X}{\partial \xi^2} - v\frac{\partial X}{\partial \xi} + R_X \tag{12a}$$

$$\frac{\partial GA}{\partial t} = D_{GA}\frac{\partial^2 GA}{\partial \xi^2} - v\frac{\partial GA}{\partial \xi} + R_{GA} \tag{12b}$$

$$\frac{\partial G}{\partial t} = D_G\frac{\partial^2 G}{\partial \xi^2} - v\frac{\partial G}{\partial \xi} + R_G \tag{12c}$$

$$\frac{\partial O_2}{\partial t} = D_{O_2}\frac{\partial^2 O_2}{\partial \xi^2} - v\frac{\partial O_2}{\partial \xi} + k_{La}(O_2^* - O_2) + R_{O_2} \tag{12d}$$

where D_X, D_{GA}, D_G and D_{O_2} represent the diffusion coefficients and v is the fluid velocity. Finally, the description is completed with boundary and initial conditions of the form:

$$D_{s_i}\frac{\partial s_i}{\partial \xi} = -v(s_i^{in} - s_i) \quad \xi = 0, \tag{12e}$$

$$\frac{\partial s_i}{\partial \xi} = 0 \quad \xi = L, \tag{12f}$$

$$\text{and } s_i = s_i^0 \quad t = 0 \tag{12g}$$

with s_i $(i = 1, ..., 4)$ representing the process variables involved (X, GA, G, O_2), s_i^{in} the inlet concentration, s_i^0 the initial condition and L the longitudinal length of the reactor.Boundary conditions (12e)-(12f) hold in most reactors.

In order to illustrate the dynamics of the process, simulation experiments were performed by disturbing the system from a given steady state with inlet glucose concentration profiles as depicted in Figure 3.

The time evolution and spatial distribution along the reactor of the process variables are depicted in Figure 4. Results were obtained by means of the FEM with design and model parameters as well as operation conditions summarized in Tables 1 and 2. Several numerical tests led to the conclusion that a discretization of $n = 61$ nodes is enough to accurately solve the system.

4.1. On-line indirect estimation of biomass, glucose and GA from oxygen measurements

The observation scheme for continuous GA production is designed as discussed in Section 2 to produce estimates of biomass, glucose and GA from a limited number of oxygen measurements.

Let us first illustrate the theoretical performance of the estimator developed in Section 2 without any approximation and with measurements of oxygen along the whole reactor. The glucose inlet is perturbed as illustrated in Figure 3 and the oxygen is measured along the whole reactor. The dynamic estimation is designed as in equations (3), so that:

$$\frac{\partial Z_G}{\partial t} = D_G\frac{\partial^2 Z_G}{\partial \xi^2} + (D_G - D_{O_2})\frac{Y_G}{Y_{O_2}}\frac{\partial^2 O_2}{\partial \xi^2} - v\frac{\partial Z_G}{\partial \xi} - \frac{Y_G}{Y_{O_2}}k_{La}(O_2^* - O_2) \tag{13a}$$

$$\frac{\partial Z_X}{\partial t} = D_X\frac{\partial^2 Z_X}{\partial \xi^2} + (D_X - D_{O_2})\frac{Y_X}{Y_{O_2}}\frac{\partial^2 O_2}{\partial \xi^2} - v\frac{\partial Z_X}{\partial \xi} - \frac{Y_X}{Y_{O_2}}k_{La}(O_2^* - O_2) \tag{13b}$$

$$\frac{\partial Z_{GA}}{\partial t} = D_{GA}\frac{\partial^2 Z_{GA}}{\partial \xi^2} + (D_{GA} - D_{O_2})\frac{Y_{GA}}{Y_{O_2}}\frac{\partial^2 O_2}{\partial \xi^2} - v\frac{\partial Z_{GA}}{\partial \xi} - \frac{Y_{GA}}{Y_{O_2}}k_{La}(O_2^* - O_2) \tag{13c}$$

Note that these PDE equations do not depend on the reaction rates. In order to solve the previous equations, initial conditions are required. Since no measurements are available for G, X, GA an initial guess will be employed. In this example, we will consider zero initial concentration for G, X and GA, which implies a relative error of 100%. The observer will drive the error to zero. The concentration of the chemical compounds is recovered by using relations (2):

$$G = Z_G + \frac{Y_G}{Y_{O_2}}O_2, \quad X = Z_X + \frac{Y_X}{Y_{O_2}}O_2, \quad GA = Z_{GA} + \frac{Y_{GA}}{Y_{O_2}}O_2, \tag{14}$$

For illustrative purposes the dynamic observer is implemented by using the FEM description of the process. It must be highlighted that this requires as many oxygen sensors as discretization points (in this example $n = 61$). Figure 5 presents the glucose, gluconic acid and biomass concentration errors as a function of space and time. As it can be seen there the error between measurements and estimates decays exponentially fast.

In order to reduce the number of sensors and the computational cost, the reduced observers described in section 3 are now implemented. To begin with, the POD model is obtained from snapshots of the system behavior under a sufficiently large number of operating conditions. In this case four basis functions are enough to accurately reconstruct the oxygen distribution. Figure 7 presents the accuracy of the reconstruction using four optimally located sensors, as shown in Figure 6, and measuring every half an hour (linear behavior between successive sampling times is considered).

Next step is devoted to the dynamic observation of the remaining process variables (G, GA, X). For this purpose a LSD based model is employed. The ROM consists of 3, 4, and 8 ODEs for the glucose,

GA and biomass, respectively. Note that the reduced order model consists of 15 ODEs instead of the 183 (61*3) resulting from the FEM. Therefore, the computational time required drastically decreases when using the reduced version as compared with he FEM observer.

Using the ROM with 15 ODEs the time interval between two consecutive estimations of glucose, biomass and gluconic acid is around 5 minutes. The computations reported here were performed using a PC/Pentium 4 (3 GHz, 2 Gb RAM) platform running Windows Xp with Matlab 6.5 (The Mathworks Inc.). It should be noted (see Figure 4) that the dynamics of the process are quite slow as compared with the time between estimations. Therefore, the estimation scheme is fast enough to capture the behavior of the system.

Finally, the spatial distribution and the time evolution of the estimation error for G, GA and X obtained by the proposed scheme is depicted in Figure 8. The differences between the estimations and the process variables converge at a reasonable speed to an error below 2 % (see Figure 9).

5. Conclusions

In this work an on-line estimation methodology for a distributed parameter bioreactor has been developed. Such methodology incorporates two different approaches: a field reconstruction, employed to recover the oxygen spatial distribution from a reduced number of optimally located sensors and a dynamic observer to estimate the remaining process variables. The main advantage of this methodology is that information regarding the reaction rates is not required. It must be noted that in many biological processes such information is not available.

The approach has been successfully applied to an example related to the production of GA from free-growth cells of *Aspergillus niger* in a tubular fermenter. This methodology can be extended to other aerobic fermentation processes where on-line sensors for measuring the oxygen are available.

6. Acknowledgments

This paper includes results of the joint research project that is supported by the bilateral agreements between Bulgarian Academy of Sciences and Process Engineering Group IIM-CSIC. The authors acknowledge financial support received from the Spanish Government (MCyT Projects PPQ2001-3643 and DPI2004-07444-C04-03) and Xunta de Galicia (PGIDIT02-PXIC40209PN).

A. Reduced Order Models for PDE systems

Reduced order models (ROMs) present several advantages which encouraged many authors to apply them in different fields such as food and chemical engineering [5, 8], biology [31] or fluid dynamic problems [27, 16, 33] among others. For the sake of completeness this appendix contains the fundamental elements for the ROM synthesis.

Let us consider the following general PDE system:

$$\frac{\partial u(t,\xi)}{\partial t} = T(u) + f(u) \tag{15}$$

$u(t,\xi)$ in Eqn (15) represents the state variable, $T(u)$ is a linear operator which, as explained in section 2, collects the diffusion and convection contributions[1] and $f(u)$ is a given nonlinear function representing for instance the reaction rates. As it is the case in the family of weighted residuals methods, the distributed variable u can be expanded as:

$$\tilde{u}(t,\xi) = \sum_{i=1}^{\infty} c_i^u(t)\varphi_i^u(\xi) \tag{16}$$

The set $\{c_i^u(t)\}_{i=1}^{\infty}$ collects time dependent functions known as *modes* and $\{\varphi_i^u(\xi)\}_{i=1}^{\infty}$ is an infinite orthonormal set containing the spatial dependency of the solution. Contrary to the classical numerical methods for PDE systems mentioned in section 3, the basis functions φ_i^u in reduced order techniques are globally defined on the spatial domain. This fact together with the dissipative nature of this class

[1] For example, a standard constant diffusion-convection transport operator would read as $T(u) = D\Delta u - \vec{v}\vec{\nabla}u$

of systems [2] make possible to approximate the solution at arbitrary accuracy by a truncated series of the form:

$$\tilde{u}(t,\xi) \cong \sum_{i=1}^{m_u} c_i^u(t)\varphi_i^u(\xi) \tag{17}$$

where m_u is usually low. Global basis functions $\varphi_i^u(\xi)$ are computed by solving the following eigenvalue problem:

$$\int_\Omega R(\xi,\xi')\varphi_i^u(\xi')\mathrm{d}\xi' = \lambda_i^u\varphi_i^u(\xi) \tag{18}$$

λ_i^u in Eqn (18) is the eigenvalue associated to the i-eigenfunction φ_i^u. Depending on the nature of the kernel R, different sets of basis functions arise [2]:

- *Laplacian Spectral Decomposition (LSD).* In this case the kernel is constructed as the Green's function associated to the Laplacian operator. In practice this leads to the following eigenvalue problem [9]:

$$\Delta\varphi_i^u = \lambda_i\varphi_i^u \tag{19}$$

- *Proper Orthogonal Decomposition (POD).* Here $R(\xi,\xi')$ corresponds with the two-point correlation kernel [16], of the form:

$$R(\xi,\xi') = \frac{1}{\ell}\sum_{i=1}^{\ell} u(t_i,\xi)u(t_i,\xi') \tag{20}$$

  where each $u(t_i,\xi)$ describes the value of the field at time instant t_i (snapshot) and ℓ is the total number of snapshots.

In the dynamic version, the time evolution of the mode set is computed by projecting Eqn. (15) over each basis function so that:

$$\int_\Omega \phi_i(\xi)\left(\frac{\partial u(t,\xi)}{\partial t} - T(u) - f(u)\right)d\xi = 0 \tag{21}$$

Extending the projection over the most relevant basis functions, we end up with the following set of ODEs:

$$\frac{d\mathbf{c}^u}{dt} = \boldsymbol{A}\mathbf{c}^u + \boldsymbol{F}; \qquad \mathbf{c}^u = [c_1^u, c_2^u, ...c_{m_u}^u]^T \tag{22}$$

where matrix $\boldsymbol{A}$ and vector field $\boldsymbol{F}$ result from the projections of the transfer and the reaction terms[2], respectively. The field is then reconstructed by combining the solution of Eqn (22) with Eqn (17).

For practical reasons it is more convenient to work with a discrete version of the field. To that purpose we denote by $\phi_i^u \in \mathbb{R}^{n\times 1}$ the vector whose elements are the values of $\varphi_i^u(\xi)$ at n spatial discretization points and rewrite Eqn (17) as follows:

$$\mathbf{u}(t) \cong \Phi\mathbf{c}^u(t) \tag{23}$$

where $\Phi \in \mathbb{R}^{n\times m_u}$ is a matrix of the form $\Phi = [\phi_1, ..., \phi_{m_u}]$ and $\mathbf{u}(t) \in \mathbb{R}^{n\times 1}$ contains the values of the field at the n discretization points.

Finally for r snapshots Eqn (23) can be presented in a more compact way as:

$$\mathbf{U} \cong \Phi\mathbf{C}. \tag{24}$$

where matrix $\mathbf{C}^u \in \mathbb{R}^{m_u\times r}$ is the matrix that contains the mode set $\mathbf{c}^u$ at different r time points and $\mathbf{U}$ collects the snapshots. Expression (24) is the one employed to construct the static observer described in section 3.

[2]Mathematically, this corresponds with $\boldsymbol{A} = \int_\Omega \phi_i^u T(u)d\xi$ and $\boldsymbol{F} = \int_\Omega \phi_i^u f d\xi$

References

[1] Alonso, A. A., Banga, J. R., Martín, R. I. P., 1993. A new strategy for the control of the cooling stage of the sterilization process in steam retorts. Transac. of the IChemE 71 C, 197–205.

[2] Alonso, A. A., Frouzakis, C. E., Kevrekidis, I. G., 2004. Optimal sensor placement for state reconstruction of distributed process systems. AIChE J 50 (7), 1438–1452.

[3] Anastassiadis, S., Aivasidis, A., Wandrey, C., 2003. Continuous gluconic acid production by isolated yeast-like mould strains of *Aureobasidium pullulans*. Appl Microbiol Biotechnol 61, 110–117.

[4] Balsa-Canto, E., Alonso, A. A., Banga, J. R., 2002. A novel, efficient and reliable method for thermal process design and optimization. Part I: Theory. J Food Eng 52 (3), 227–234.

[5] Balsa-Canto, E., Alonso, A. A., Banga, J. R., 2002. A novel, efficient and reliable method for thermal process design and optimization. Part II: Applications. J Food Eng 52 (3), 227–234.

[6] Bastin, G., Dochain, D., 1990. On-line estimation and adaptive control bioreactors. Elsevier, Amsterdam.

[7] Christofides, P. D., 2001. Control of nonlinear distributed process systems: Recent developments and challenges. AIChE J 47 (33), 514 – 518.

[8] Christofides, P. D., 2001. Nonlinear and Robust Control of PDE Systems: Methods and Applications to Transport-Reaction Processes. Birkhäuser, Boston.

[9] Courant, R., Hilbert, D., 1937. Methods of Mathematical Physics. Wiley, New York, USA.

[10] Delattre, C., Dochain, D., Winkin, J., 2004. Observability analysis of nonlinear tubular (bio)reactor models: a case study. J Process Control 14, 661–669.

[11] Dochain, D., 2000. State observers for tubular reactors with unknown kinetics. J Process Control 10, 259–268.

[12] Dochain, D., Tali-Maamar, N., Babary, J., 1996. On modelling, monitoring and control of fixed bed bioreactors. Comput Chem Eng 21, 1255–1266.

[13] Fletcher, C., 1984. Computational Galërkin Methods. Springer-Verlag, New York.

[14] García, M. R., Vilas, C., Banga, J. R., Alonso, A. A., 2007. Optimal field reconstruction of distributed process systems from partial measurements. Ind Eng Chem Res 46 (2), 530–539.

[15] Gottlieb, D., Orszag, S. A., 1977. Numerical analysis of spectral methods: Theory and applications. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

[16] Holmes, P., Lumley, J. L., Berkooz, G., 1996. Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge University Press.

[17] Holmes, P. J., Lumley, J. L., Berkooz, G., Mattingly, J. C., Wittenberg, R. W., 1997. Low-dimensional models of coherent structures in turbulence. Physics Reports 287 (4), 337–384.

[18] Liu, J. Z., Weng, L. P., Zhang, Q. L., Xu, H., Ji, L. N., 2003. A mathematical model for gluconic acid fermentation by *Aspergillus niger*. Biochem Eng J 14, 137–141.

[19] Mirón, J., González, M., Pastrana, L., Murado, M., 2002. Diauxic production of glucose oxidase by *Aspergillus Niger* in submerged culture. a dynamic model. EnzymeandMicrobialTechnology 31, 615–620.

[20] Mirón, J., González, P., Vázquez, J. A., Pastrana, L., Murado, M. A., 2004. A mathematical model for glucose oxidase kinetics, including inhibitory deactivant and diffusional effects, and their interactions. Enzyme Microb Tech 34, 513–522.

[21] Rai, V. R., Constantinides, A., 1973. Mathematical modeling and optimization of the gluconic acid fermentation. AIChE symp. Ser. 69, 114–122.

[22] Reddy, J. N., 1993. An Introduction to the Finite Element Method, 2nd Edition. McGraw-Hill, United States.

[23] Rinas, U., El-Enshasy, H., Emmler, M., Hille, A., Hempel, D. C., Horn, H., 2005. Model-based prediction of substrate conversion and protein synthesis and excretion in recombinant *Aspergillus niger* biopellets. Chem Eng Sci 60, 2729–2739.

[24] Russell, T. W. F., Dunn, I. J., 1974. The tubular loop batch fermentor: Basic concepts. Biotechnol Bioeng 36, 1261–1272.

[25] Sankpal, N. V., Cheema, J. J. S., Tambe, S. S., Kulkarni, B. D., 2001. An artificial intelligence tool for bioprocess monitoring: application to continuous production of gluconic acid for immobilized *Aspergillus niger*. Biotechnol Lett 23, 911–916.

[26] Schiesser, W. E., 1994. Computational Mathematics in Engineering and Applied Science: ODEs, DAEs and PDEs. CRC Press, Inc.

[27] Sirovich, L., 1987. Turbulence and the dynamics of coherent structures. Part I: Coherent structures. Quaterly of Appl. Math. 45 (3), 561–571.

[28] Stephanopoulos, G., San, K., 1984. Studies on on-line bioreactor identification. I. theory. Biotechnol Bioeng XXVI, 1176–1188.

[29] Trefethen, L. N., 2000. Spectral Methods in Matlab. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

[30] Vande Wouwer, A., Point, N., Porteman, S., Remy, M., 2000. An approach to the selection of optimal sensor locations in distributed parameter systems. J Process Control 10, 291–300.

[31] Vilas, C., García, M. R., Banga, J. R., Alonso, A. A., 2006. Stabilization of inhomogeneous patterns in a diffusion-reaction system under structural and parametric uncertainties. J Theor Biol 241 (2), 295–306.

[32] Šantek, B., Ivančić, M., Horvat, P., Novak, S., Marié, V., 2006. Horizontal tubular bioreactors in biotechnology. Chem Biochem Eng Q 20 (4), 389–399.

[33] Willcox, K., 2006. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. Computers & fluids 35 (2), 208–226.

[34] Zerrik, E., Boutoulout, A., Bourray, H., 2001. Boundary strategic actuators. Sens. Actuator A-Phys. 94 (3), 197–203.

[35] Znad, H., Blažej, M., Báleš, V., Markoš, J., 2003. A kinetic model for gluconic acid production by *Aspergillus niger*. Chemical Papers - Chemicke Zvesti 58 (1), 23–28.

List of Tables

1 Process parameters estimated using the real data presented in [19]. 13
2 Design parameters of the horizontal tubular fermenter employed in the simulation experiments. 13

List of Figures

1 Schematic representation of the proposed estimation scheme. Continuous arrows represent oxygen measurements while dashed arrows indicate the field reconstructed in the whole spatial domain. The field reconstruction block recovers the spatio-temporal distribution of oxygen in the reactor from partial measurements. These estimate is injected in the dynamic observer block to dynamically reconstruct the remaining process variables. 14
2 Calibration results for the relevant process variables. 14
3 Perturbations introduced in the Glucose inlet for testing the estimation scheme. 14
4 Spatial distribution and time evolution of the relevant process variables under the Glucose inlet perturbations of Figure 3. 15
5 Spatial distribution and time evolution of the estimation errors for G, GA and X using the dynamic observer based on the FEM. The oxygen was continuously measured in all the FEM discretization points ($n = 61$). The estimation performance for G is expressed in terms of absolute errors whereas relative errors are considered in the case of GA and X. 15
6 Optimal location in the reactor of the four oxygen sensors. 15
7 Spatial distribution and time evolution of the estimation error for the oxygen using the field reconstruction scheme with the POD technique. Four optimally located sensors (Figure 6) were employed. 16
8 Spatial distribution and time evolution of the estimation errors for G, GA and X using the dynamic observer based on the LSD. The oxygen was measured every half an hour using four optimally located (see Figure 6) sensors. The field reconstruction scheme was employed to recover the oxygen spatial distribution (Figure 7). 16
9 Final time spatial distribution of the estimation errors for G, GA and X using the dynamic observer based on the LSD and the field reconstruction based on the POD. 16

B. Tables

Design parameter	**Symbol Value & Units**
Oxygen transfer rate per volume unit	$k_{La} = 300h^{-1}$
Biomass reaction rate parameters	**Yield coefficients**
$\mu_{max} = 0.2242\ h^{-1}$	$Y_{GA} = 44.8887$ (GA g)/(X g)
$k_G = 9.9222\ gl^{-1}$	$Y_{O_2} = -2.5598$ (O_2 g)/(X g)
$k_{O_2} = 0.00137\ gl^{-1}$	$Y_G = -51.0365$ (G g)/(X g)

Table 1: Process parameters estimated using the real data presented in [19].

Design parameters	**Symbol Value & Units**
Oxygen transfer rate per volume unit	$k_{La} = 300h^{-1}$
Glucose and Gluconic Acid mass dispersion coefficient	$D_{GA} = D_G = 0.01m^2h^{-1}$
Biomass Mass dispersion coefficient	$D_X = 0.10m^2h^{-1}$
Oxygen Mass dispersion coefficient	$D_{O_2} = 0.04m^2h^{-1}$
Flow velocity	$v = 0.01mh^{-1}$
Input stream	**Symbol Value & Units**
Glucose stream	$G^{in} = 5 - 30gl^{-1}$
Dissolved oxygen stream	$O2^{in} = 0.0084gl^{-1}$
Biomass stream	$X^{in} = 0.01gl^{-1}$
Gluconic Acid stream	$GA^{in} = 0gl^{-1}$

Table 2: Design parameters of the horizontal tubular fermenter employed in the simulation experiments.

C. Figures

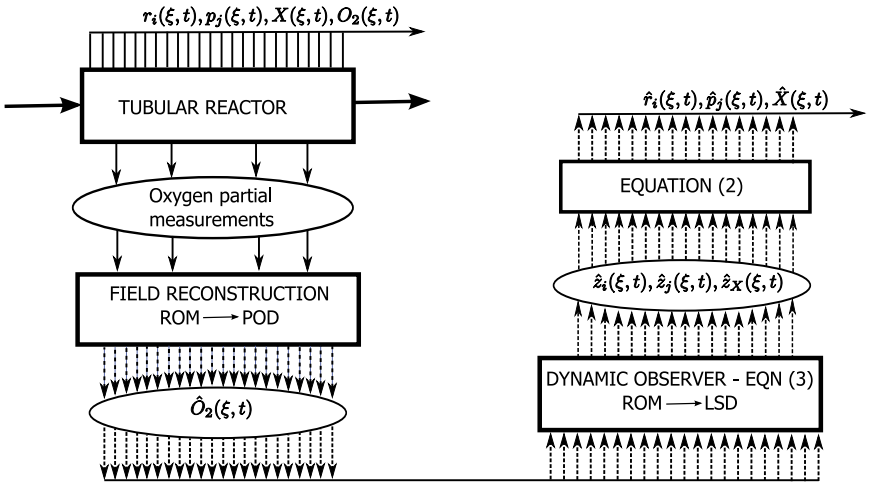


Figure 1: Schematic representation of the proposed estimation scheme. Continuous arrows represent oxygen measurements while dashed arrows indicate the field reconstructed in the whole spatial domain. The field reconstruction block recovers the spatio-temporal distribution of oxygen in the reactor from partial measurements. These estimate is injected in the dynamic observer block to dynamically reconstruct the remaining process variables.

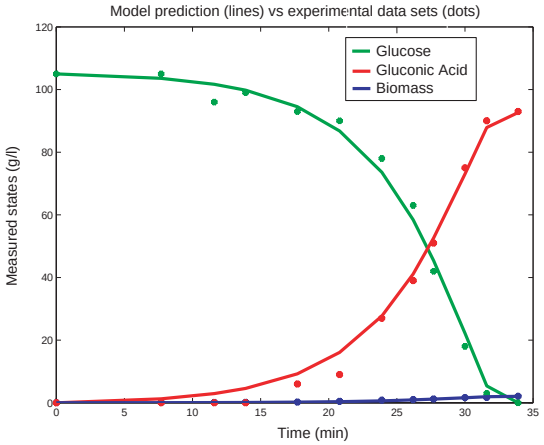


Figure 2: Calibration results for the relevant process variables.

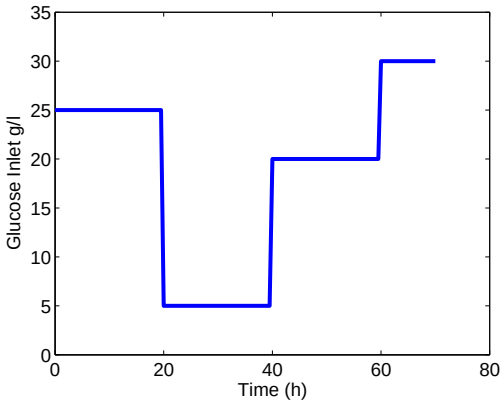


Figure 3: Perturbations introduced in the Glucose inlet for testing the estimation scheme.

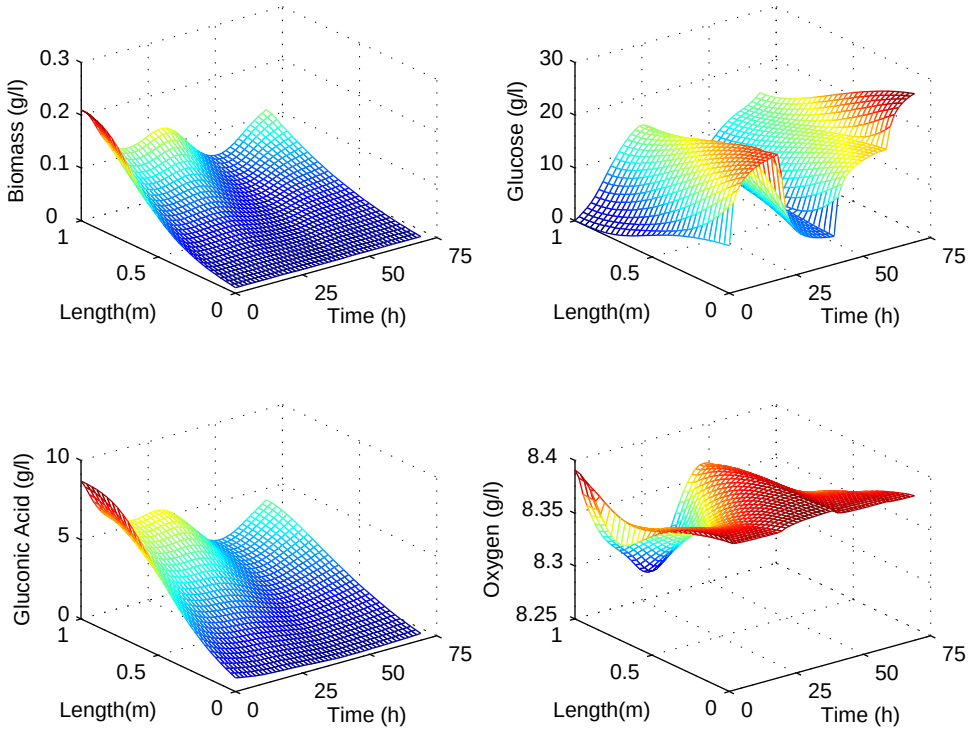
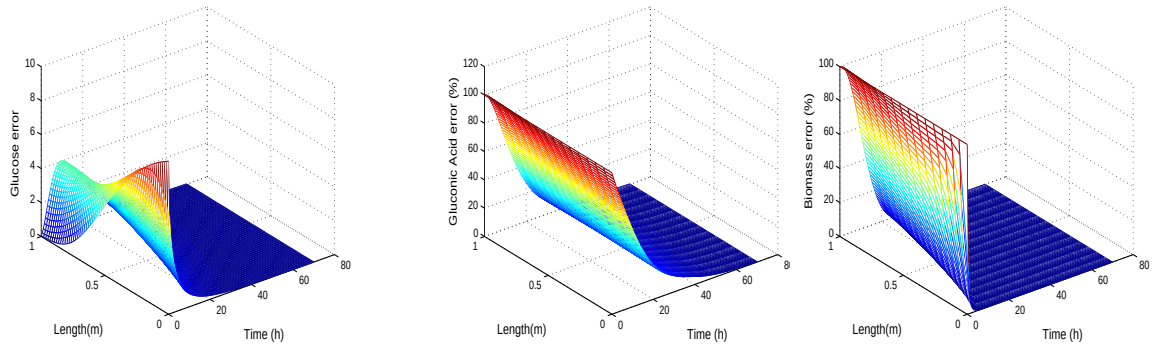


Figure 4: Spatial distribution and time evolution of the relevant process variables under the Glucose inlet perturbations of Figure 3.



(a) G absolute error (b) GA percent error (c) X percent error

Figure 5: Spatial distribution and time evolution of the estimation errors for G, GA and X using the dynamic observer based on the FEM. The oxygen was continuously measured in all the FEM discretization points ($n = 61$). The estimation performance for G is expressed in terms of absolute errors whereas relative errors are considered in the case of GA and X.

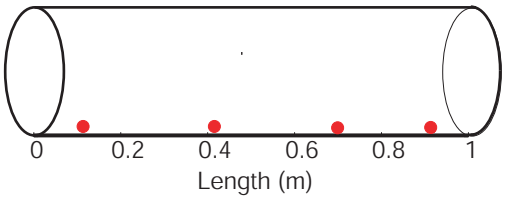


Figure 6: Optimal location in the reactor of the four oxygen sensors.

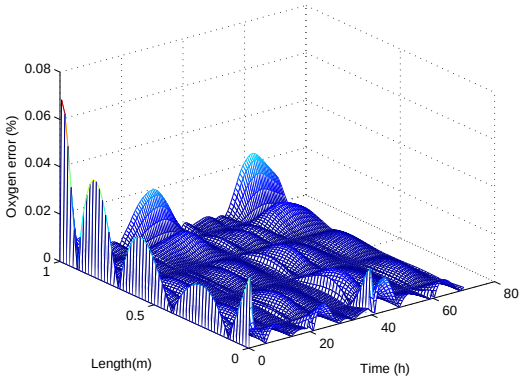
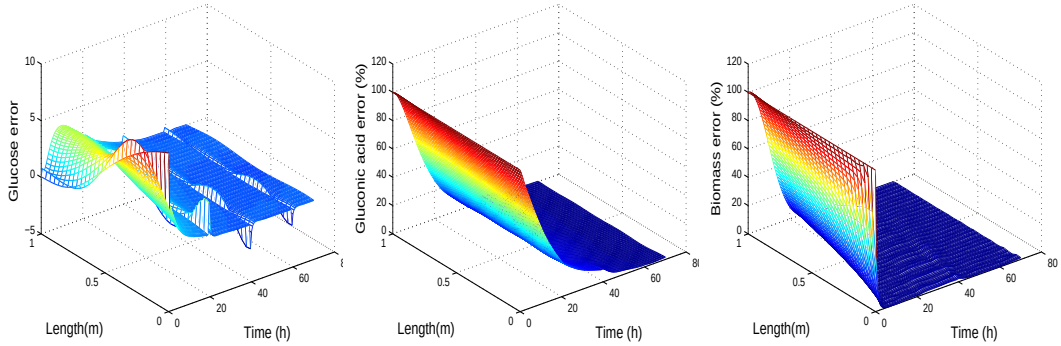
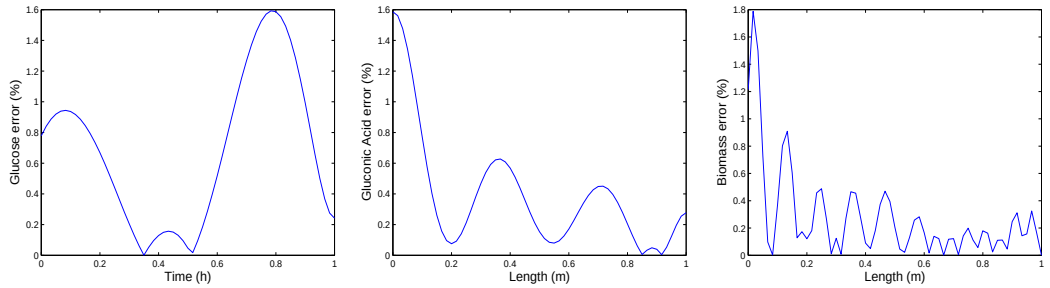


Figure 7: Spatial distribution and time evolution of the estimation error for the oxygen using the field reconstruction scheme with the POD technique. Four optimally located sensors (Figure 6) were employed.



(a) G absolute error (b) GA percent error (c) X percent error

Figure 8: Spatial distribution and time evolution of the estimation errors for G, GA and X using the dynamic observer based on the LSD. The oxygen was measured every half an hour using four optimally located (see Figure 6) sensors. The field reconstruction scheme was employed to recover the oxygen spatial distribution (Figure 7).



(a) G percent error (b) GA percent error (c) X percent error

Figure 9: Final time spatial distribution of the estimation errors for G, GA and X using the dynamic observer based on the LSD and the field reconstruction based on the POD.