ON-LINE ESTIMATION OF BIOMASS IN AN *E. COLI* FED-BATCH FERMENTATION

Ana C. A. Veloso[1,2], I. Rocha[1] and E. C. Ferreira[1*]
[1]Centro de Engenharia Biológica - Universidade do Minho
[2]Escola Superior Agrária - Instituto Politécnico de Bragança

Abstract. In this work, an Extended Kalman Observer is applied to the on-line determination of biomass concentration in a high-cell density fed-batch fermentation of *Escherichia coli*.
Although the importance of this fermentation process for the biopharmaceutical industry is widely recognized, there are still several difficulties associated with the design of monitoring and control algorithms that could improve the performance of the process by decreasing the production costs and increasing the yield.
In this process, biomass concentration has an important role for model predictive control, estimation of specific growth rates, prevention of acetate accumulation and optimization of the production of recombinant proteins (regarding both productivity and moment of induction). However, nowadays it is still determined using off-line laboratory analysis, making it of limited use for control purposes.
For the development of the Extended Kalman Observer, a dynamical mathematical model of the process was used, which includes balance equations for the main state variables (biomass, glucose, acetate, dissolved oxygen and carbon dioxide concentrations) together with a complex kinetic model describing the 3 main metabolic pathways of *Escherichia coli*.
The observer applied in this work requires the on-line measurement of a subset of state variables (dissolved oxygen and carbon dioxide concentrations) together with broth weight and gaseous mass transfer rates.
State-of-the-art sensors were used for measuring dissolved oxygen and carbon dioxide concentrations and gaseous transfer rates were determined on-line using commercial gas analysers. The calculations were performed on-line in a developed LabVIEW data acquisition and control system.
The extended Kalman observer exhibited a good convergence to the real values of biomass concentration, with a very low quadratic difference between experimental and estimated data. Also, the sampling frequency for the measured variables is compatible with the existing experimental data.

Key-words: Extended Kalman Observer, Biomass Estimation, Fed-batch Fermentation.

1. Introduction

Nowadays, the ability to accurately and automatically control bioprocesses at their optimal state is of enormous importance to many industries since it can contribute for decreasing the production costs and increase the yield, keeping the quality of the metabolic products. However, the main difficulties in the design of monitoring and control systems for biological processes lie in the lack of cheap and reliable sensors capable of providing direct and on-line measurements of the biological state variables, together with the significant model uncertainty and the non-linear and time-varying nature of the system. In fact, in many practical applications, only some of the state variables involved and critical for efficient control are available for on-line measurement. For example, the dissolved oxygen concentration and gaseous flow rates are available for on-line measurement while the biomass, products and substrates concentrations are often available via off-line analysis.

* To whom all correspondence should be addressed.
Address: Centro de Engenharia Biológica - Universidade do Minho, 4710-057 Braga - Portugal
E-mail: ecferreira@deb.uminho.pt

State observers, also called software sensors (Dochain, 2003), represent an interesting alternative and have received in recent years an increased attention since they allow on-line monitoring of state variables that are not measurable in real time (Assis and Filho, 2000; Valdés *et al.*, 2003; Bernard and Gouzé, 2004, Bogaerts and Wouwer, 2004) using a model in conjunction with a limited set of state variable measurements.

In the literature, two classes of state observers are usually found. The first class includes the classical observers, such as the Luenberger, the Kalman, and the non-linear observers, which are based on the perfect knowledge of both model structure and parameters. On the other hand, the uncertainty in the model parameters can generate a large bias in the estimation of unmeasured state(s). The asymptotic observers (Bastin and Docahin, 1990), which constitute the second class of observers, do not require the knowledge of the process kinetics. Nevertheless, a potential problem concerning these observers is the dependence of the estimation convergence rate on the operating conditions (Dochain, 2003).

However, in spite of the well-developed theory behind some state observers, there are not many documented examples where those algorithms are applied to complex bioprocesses, described by dynamical models containing several balance equations and with complex kinetics.

In this work, the high-cell density fed-batch fermentation of *Escherichia coli* is studied in terms of applicability of state observers for the estimation of biomass concentration. The importance of this process for the biopharmaceutical industry is widely recognized, as *E. coli* represents the organism of choice for the production of many recombinant proteins. However, several state variables are not easily measured on-line during this process, posing additional difficulties for the implementation of control algorithms. As an example, in spite of its important role for model predictive control, estimation of specific growth rates, prevention of acetate accumulation and optimization of the production of recombinant proteins (regarding both productivity and moment of induction), biomass concentration is nowadays very difficult to measure on-line for this fermentation process.

To carry out the on-line estimation of biomass concentration, the dissolved oxygen and carbon dioxide concentrations were measured with state-of-the-art sensors and gaseous transfer rates were determined on-line using a commercial gas analyser. This on-line information was used by the software sensors for the estimation of the remaining variables included in the mathematical model, which can be regarded as one step towards the complete characterization of the process. Simultaneously, the self-developed modular supervisory system facilitates the integration of different measurements, the on-line estimation of variables and the application of those measurements in control algorithms.

This work is organized as follows: in the next section the dynamical model of *E. coli* fed-batch fermentation is described. The Extended Kalman Observer is presented in section 3. In section 4 material and methods are described. In section 5 the main results are discussed. Finally in section 6 conclusions are presented.

2. Process Modelling

The dynamics of a reaction network in a stirred tank bioreactor can be described by the following mass balance equations written in matrix form as (Bastin and Dochain, 1990):

$$\frac{d\xi}{dt} = Kr(\xi,t) - D\xi + F - Q \tag{1}$$

in which ξ is a vector representing the *n* state components concentrations (ξ ∈ $\Re^n$, r is the growth rate vector corresponding to *m* reactions (r ∈ $\Re^m$), K is the matrix of yield coefficients (K ∈ $\Re^{n\times m}$), F is the vector of feed rates and Q is the vector of gaseous outflow rates (F, Q ∈ $\Re^n$), D is the dilution rate (being D^{-1} the residence time).

As previously presented (Rocha and Ferreira, 2004), during the aerobic growth of *E. coli* with glucose as the only added substrate, the microorganism can follow three main metabolic pathways: oxidative growth on glucose, fermentative growth on glucose, and oxidative growth on acetate, the corresponding dynamical model for fed-batch fermentation being represented as follows:

$$\frac{d}{dt}\begin{bmatrix} X \\ S \\ A \\ O \\ C \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -k_1 & -k_2 & 0 \\ 0 & k_3 & -k_4 \\ -k_5 & -k_6 & -k_7 \\ k_8 & k_9 & k_{10} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} X - D \begin{bmatrix} X \\ S \\ A \\ O \\ C \end{bmatrix} + \begin{bmatrix} 0 \\ \left(\frac{F_{in}}{W}\right) S_{in} \\ 0 \\ OTR \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ CTR \end{bmatrix} \tag{2}$$

where X, S, A, O, and C represent biomass, glucose, acetate, dissolved oxygen, and dissolved carbon dioxide concentrations, respectively; μ_1, μ_2, and μ_3 are the specific growth rates; k_i are the yield (stoichiometric) coefficients; F_{in} and S_{in} are the substrate feed rate and the influent glucose concentration, respectively; W is the culture medium weight. CTR is the carbon dioxide transfer rate from liquid to gas phase, and OTR is the oxygen transfer rate from gas to liquid phase.

The variation of the culture medium weight with the time is given by:

$$\frac{dW}{dt} = F \tag{3}$$

where F includes weight variations due to the substrate feed rate, the amount of culture removed or added during sampling, base and acid additions, evaporation and mass taken from the reactor due to gas exchanges, that can not be considered negligible in small-scale high-cell density reactors.

However, the three metabolic pathways represented in the mathematical model do not occur simultaneously in the cell, originating four partial models corresponding to different metabolic regimens:

- simultaneous oxidative and fermentative growth on glucose (μ_1,μ_2,>0)
- oxidative growth on glucose (μ_1>0)
- oxidative growth on acetate and glucose simultaneously (μ_1, μ_3, >0)
- oxidative growth on acetate (μ_3>0)

3. Derivation of Observer

The Kalman Observer is the optimal state estimator for a linear process if the system model and certain properties of the measures are available. When high nonlinearities are included in the mathematical model of the process, the extended version of the Kalman observer should be used (Biagiola and Figueroa, 2004).

In order to obtain the Extended Kalman Observer (EKO), the following assumptions are made: (i) a full knowledge of the model is available: the structure of the reaction kinetics r(ξ,t) is completely known; also the numerical values of all the coefficients involved in the model (yield and kinetic coefficients) are given; and (ii) D, F and Q are known on-line, together with a *q* subset of state variables.

This vector of state variables measured is denoted ξ_1 and is related to the state of the system as follows:

$$\xi_1 = L\xi \tag{4}$$

where the *q*×*n* matrix L is an elementary matrix which selects the measured components of ξ. On the other hand, the vector of unmeasured states is denoted ξ_2, so that (ξ_1,ξ_2) constitutes a partition of ξ.

A general class of state observers for nonlinear systems of the form of Eq. (1) is as follows:

$$\frac{d\hat{\xi}}{dt} = Kr\left(\hat{\xi},t\right) - D\hat{\xi} + F - Q + \Omega\left(\hat{\xi},t\right)\left[\xi_1 - \hat{\xi}_1\right] \tag{5}$$

where $\hat{\boldsymbol{\xi}}$ denotes the on-line estimate of ξ, and $\boldsymbol{\Omega}\left(\hat{\boldsymbol{\xi}},\mathbf{t}\right)$ is an *n*×*q* gain matrix depending on $\hat{\boldsymbol{\xi}}$. The state observer design problem is then reduced to that of a reasonable choice of the gain matrix $\boldsymbol{\Omega}\left(\hat{\boldsymbol{\xi}},\mathbf{t}\right)$. To solve this problem, the observation error is introduced at this point, $e = \xi - \hat{\xi}$, and its dynamics deduced (Bastin and Dochain, 1990). Considering a linearized tangent approximation of the dynamical model of the observation error around *e=0* will give:

$$\frac{de}{dt} = \left[M\left(\hat{\xi}\right) - \Omega\left(\hat{\xi}\right)L\right]e \tag{6}$$

with:

$$M\left(\hat{\xi}\right) \equiv K\left[\frac{\partial r\left(\xi,t\right)}{\partial \xi}\right]_{\xi=\hat{\xi}} - DI_N \tag{7}$$

where I_N is the *n*×*n* identity matrix.

Considering that the model of Eq. (1) is exponentially observable, the design of the EKO is then reduced to the quadratic optimisation problem of finding the matrix $\mathbf{\Omega}\left(\hat{\xi},\mathbf{t}\right)$ that minimises the mean square observation error taking into account the constraint of the linear tangent error model (Eqs. 6 and 7). The solution of this optimisation problem is given by:

$$\Omega\left(\hat{\xi},t\right)=R\left(\hat{\xi},t\right)L^T \tag{8}$$

where the $n\times n$ square symmetric matrix $\mathbf{R}\left(\hat{\xi},\mathbf{t}\right)$ is generated by the Riccati equation:

$$\frac{dR}{dt}=-RL^TLR+RM^T\left(\hat{\xi},t\right)+M\left(\hat{\xi},t\right)R \tag{9}$$

For the fed-batch *E. coli* fermentation considered in this work, the exponential observability condition (Bastin and Dochain, 1990) was studied for 9 different combinations of measured and estimated variables for checking the applicability of the EKO for this particular process and it can be concluded that the EKO can be applied to *E. coli* fed-batch fermentation in a limited number of situations. However, it is clear that, with state-of-the-art sensors for measuring dissolved oxygen and carbon dioxide, it is possible to estimate on-line biomass and other state variables, if the cells do not exhibit only the oxidative growth on acetate or the oxidative growth on glucose regimens.

Taking the example of measuring on-line the state variables O and C, the following state partition is chosen: $\xi_1^T=\left[O\quad C\right]$ and $\xi_2^T=\left[X\quad S\quad A\right]$. The matrix L is as follows:

$$L=\begin{bmatrix}0&0&0&1&0\\0&0&0&0&1\end{bmatrix} \tag{10}$$

The matrix $M\left(\hat{\xi}\right)=M\left(\hat{X},\hat{S},\hat{A},\hat{O},\hat{C}\right)$ is given by Eq. (7), where:

$$\left[\frac{\partial r(\xi,t)}{\partial\xi}\right]_{\xi=\hat{\xi}}=\begin{bmatrix}\frac{\partial\left(\mu_1\hat{X}\right)}{\partial\hat{X}}&\frac{\partial\left(\mu_1\hat{X}\right)}{\partial\hat{S}}&\frac{\partial\left(\mu_1\hat{X}\right)}{\partial\hat{A}}&\frac{\partial\left(\mu_1\hat{X}\right)}{\partial\hat{O}}&\frac{\partial\left(\mu_1\hat{X}\right)}{\partial\hat{C}}\\\frac{\partial\left(\mu_2\hat{X}\right)}{\partial\hat{X}}&\frac{\partial\left(\mu_2\hat{X}\right)}{\partial\hat{S}}&\frac{\partial\left(\mu_2\hat{X}\right)}{\partial\hat{A}}&\frac{\partial\left(\mu_2\hat{X}\right)}{\partial\hat{O}}&\frac{\partial\left(\mu_2\hat{X}\right)}{\partial\hat{C}}\\\frac{\partial\left(\mu_3\hat{X}\right)}{\partial\hat{X}}&\frac{\partial\left(\mu_3\hat{X}\right)}{\partial\hat{S}}&\frac{\partial\left(\mu_3\hat{X}\right)}{\partial\hat{A}}&\frac{\partial\left(\mu_3\hat{X}\right)}{\partial\hat{O}}&\frac{\partial\left(\mu_3\hat{X}\right)}{\partial\hat{C}}\end{bmatrix} \tag{11}$$

The observer is then written from Eqs. (1) and (5) with the last term of Eq. (5) defined as:

$$\Omega\left(\hat{\xi},t\right)\left[\xi_1-\hat{\xi}_1\right]=\begin{bmatrix}\Omega_1 & \Omega_2\\ \Omega_3 & \Omega_4\\ \Omega_5 & \Omega_6\\ \Omega_7 & \Omega_8\\ \Omega_9 & \Omega_{10}\end{bmatrix}\begin{bmatrix}O-\hat{O}\\ C-\hat{C}\end{bmatrix} \tag{12}$$

The gain $\boldsymbol{\Omega}(\hat{X},\hat{S},\hat{A},\hat{O},\hat{C})$ is calculated from Eq. (8) as follows:

$$\Omega\left(\hat{\xi}\right)=\begin{bmatrix}\Omega_1 & \Omega_2\\ \Omega_3 & \Omega_4\\ \Omega_5 & \Omega_6\\ \Omega_7 & \Omega_8\\ \Omega_9 & \Omega_{10}\end{bmatrix}=\begin{bmatrix}R_8 & R_9\\ R_{11} & R_{12}\\ R_{13} & R_{14}\\ R_4 & R_{15}\\ R_{15} & R_5\end{bmatrix} \tag{13}$$

with the matrix R defined as:

$$R=\begin{bmatrix}R_1 & R_6 & R_7 & R_8 & R_9\\ R_6 & R_2 & R_{10} & R_{11} & R_{12}\\ R_7 & R_{10} & R_3 & R_{13} & R_{14}\\ R_8 & R_{11} & R_{13} & R_4 & R_{15}\\ R_9 & R_{12} & R_{14} & R_{15} & R_5\end{bmatrix} \tag{14}$$

The only tuning parameters for this observer are the initial values of the elements of this matrix, necessary for the numerical solution of Eq. (9).

It should be remarked that the performance of the EKO is highly dependent on the accuracy of the process model, requiring a large design effort. Moreover, numerical problems and convergence difficulties may exist due to approximations associated with model linearization. However, it is a useful algorithm for many practical estimation problems.

The performance of the observer was evaluated by calculating the quadratic difference between experimental and estimated data, according to the following equation:

$$dif_{\xi}=\sum_{j=1}^{np}\left(\frac{\xi_{exp,j}-\xi_{est,j}}{\bar{\xi}_{exp,j}}\right)^2 \tag{15}$$

where *np* is the number of experimental points and ξ_{exp} and ξ_{est} are experimental and estimated values of the state variable ξ.

4. Materials and Methods

4.1. Fermentation conditions

The experimental conditions for the fermentation process are described elsewhere (Rocha and Ferreira, 2002).

The gas transfer rates are calculated from gas analysis data obtained with a Tandem gas analyser (Adaptive Biosystems, UK) connected to the exhaust gas line of the fermenter and also to the inlet aeration line.

4.2. Hardware and software

On-line OTR and CTR calculations were performed through a C script imbedded in a LabVIEW (version 7.0) program that also acquired data from the fermenter Digital Control Unit.

The model simulations were performed by solving the differential equations of Eq. (2) using the MATLAB version 6 subroutine ODE23s. The implementation of the observers using both experimental and simulated data was conducted using the Euler integration method. The observability of the model, together with most of the mathematical operations behind the design of the state observers was performed using the Symbolic Math toolbox running in MATLAB 6.

5. Results and Discussion

In order to study the robustness of the developed observer algorithm, simulated “real” values of the state variables, obtained by integration of the differential equation of Eq. (2), were used. These “real” values were then corrupted with white noise, according to the standard deviations typically found in this process at the authors’ lab, originating “experimental” values. Then, the observer algorithm was used to obtain the “estimated” variables from the “experimental” data corresponding to the measured variables. Figure 1 presents a simulation result where the state variables (biomass, glucose and acetate) are well estimated, in spite of the introduction of noise, showing the robustness of the EKO.

The EKO was validated using experimental data. Although the objective of this work was to estimate the biomass concentration, the extended Kalman observer exhibited not only a good performance with a very low quadratic difference (Eq. 15) between experimental and estimated data for biomass (1.84) but also a satisfactory performance for the estimation of both glucose and acetate concentrations, giving quadratic differences between experimental and estimated data of 105.99 and 11.71, respectively.

Figure 2 shows the on-line data used to estimate the biomass, glucose and acetate concentrations (Figure 3). The variables measured on-line were O, C, CTR, OTR, F_{in} and W, while the estimated variables were X, S and A. Figure 3 shows a good agreement between the estimated values and off-line measurements obtained for biomass, glucose and acetate.

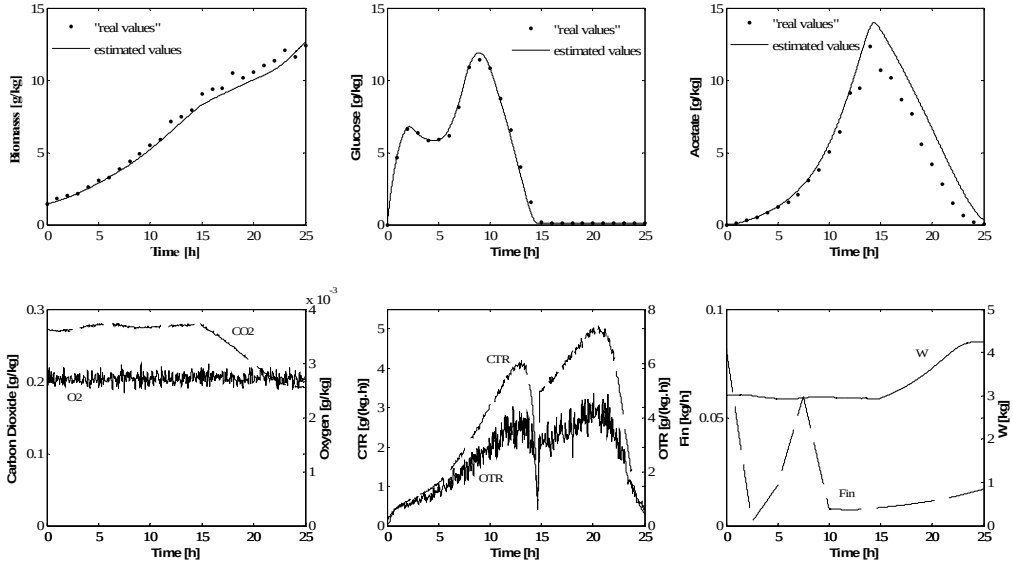

Fig. 1. Performance of the EKO regarding the time evolution of relevant variables in a fed-batch fermentation of *E. coli*.

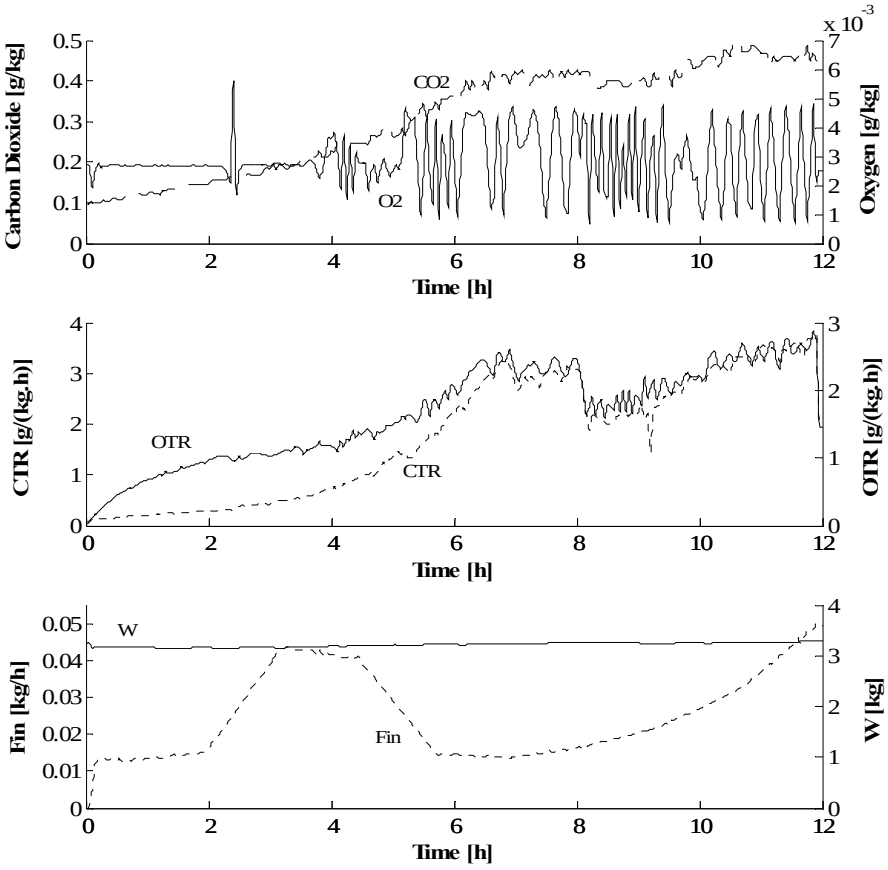

Fig. 2. State observation in a fed-batch fermentation of *E. coli*: on-line data.

A 20 h^{-1} sampling frequency for the measured variables allowed the observer algorithm to converge, which is compatible with the existing experimental acquisition data. This sampling frequency is also adequate if other sensors, like the developed FIA system (Rocha and Ferreira, 2002), are used to measure on-line other state variables, like acetate.

Moreover, although in most applications of this type of observers the tuning parameters (the initial values of the Riccati equation) represent a very time-consuming task, in this work zero initial values for all the tuning parameters allowed the convergence and stability of the EKO.

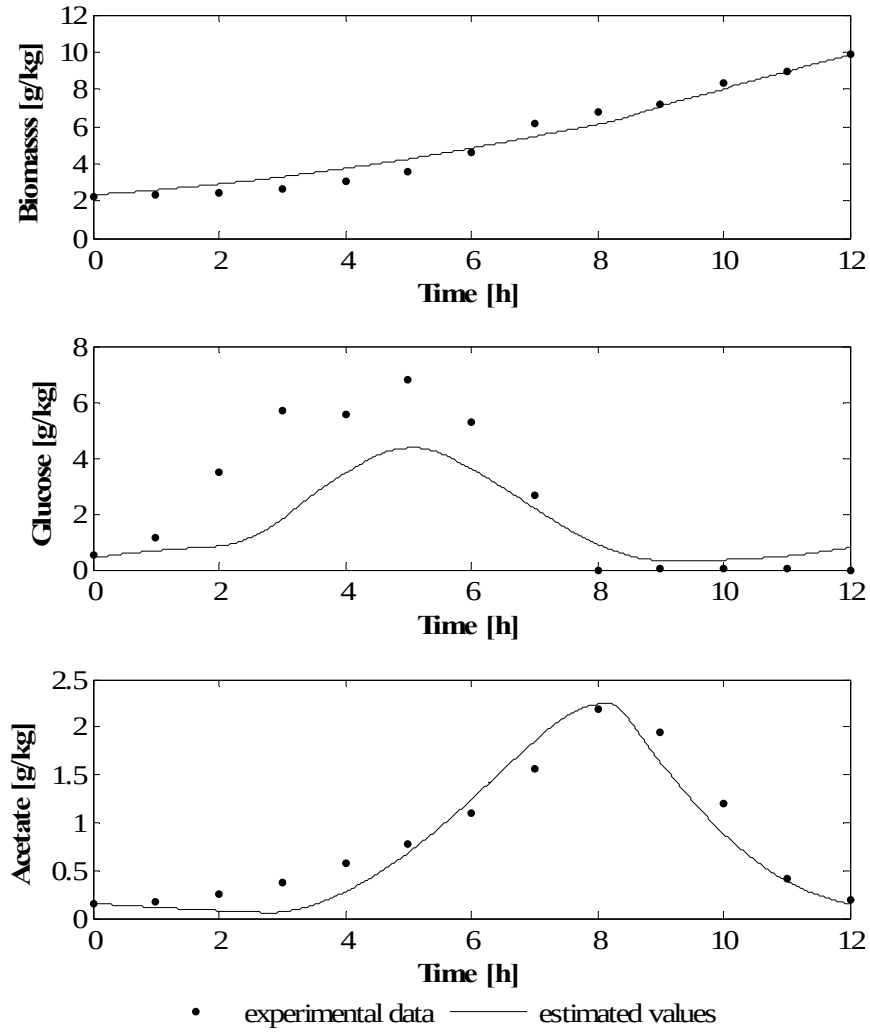

Fig. 3. State observation in a fed-batch fermentation of *E. coli*: estimated variables.

6. Conclusions

During a fed-batch *E. coli* fermentation process, variables such as biomass concentration are determined using off-line laboratory analysis, making them of limited use for control purposes. However, these variables can be on-line estimated using software sensors.

In this work, an Extended Kalman Observer algorithm was applied to the estimation of biomass, and its performance and flexibility was evaluated. The developed algorithm only requires on-line measurements of dissolved oxygen and carbon dioxide, together with the gaseous transfer rates, which represent common measurements both in industrial and academic facilities. The sampling frequency required is also compatible with most existing data acquisition systems.

In a first stage, the robustness of the algorithm regarding noise in the measured variables was checked with numerical simulations. The experimental validation was then performed, and a good agreement between estimated and experimental data was obtained, shown by the low quadratic error calculated.

Additionally, good results were obtained for the estimation of glucose and acetate concentrations, showing the flexibility of the method.

References

Assis, A.J., and Filho, R.M. (2000). Soft sensors development for on-line bioreactor state estimation. *Computers and Chemical Engineering*, 24, 1099-1103.

Bastin, G., and D. Dochain (1990). *On-line Estimation and Adaptive Control of Bioreactors*. Elsevier Science Publishers, Amsterdam.

Bernard, O. and Gouzé, J.-L. (2004). Closed loop observers bundle for uncertain biotechnological models. *Journal of Process Control*, 14, 765-774.

Biagiola, S.I. and Figueroa, J.L. (2004). Application of state estimation based NMPC to an unstable nonlinear process. *Chemical Engineering Science*, 59, 4601-4612.

Bogaerts, Ph. and Wouwer, A. V. (2004). Parameter identification for state estimation – application to bioprocess software sensors. *Chemical Engineering Science*, 59, 2465-2476.

Dochain, D. (2003). State and parameter estimation in chemical and biochemical processes: a tutorial. *Journal of Process Control*, 13, 801-818.

Rocha, I. and Ferreira, E.C. (2002). On-line simultaneous monitoring of glucose and acetate with flow-injection analysis during high-cell-density fermentation of recombinant *Escherichia coli*. *Anal. Chim. Acta*, 462, 293-304.

Rocha, I. and Ferreira, E.C. (2004). Yield and kinetic parameters estimation and model reduction in a recombinant *E. coli* fermentation". *ESCAPE-14*, Lisbon, Portugal, May 16-19, 2004 (CD-ROM).

Valdés, H., Flaus, J. and Acuña, G. (2003). Moving horizon state estimation with global convergence using interval techniques: application to biotechnological processes. *Journal of Process Control*, 13, 325-336.

Acknowledgements

A.C.A. Veloso is most grateful for the financial support provided by PRODEP.