# STATE ESTIMATION IN ALCOHOLIC CONTINUOUS FERMENTATION OF *ZYMOMONAS MOBILIS* USING RECURSIVE BAYESIAN FILTERING: A SIMULATION APPROACH

Olga L. Quintero,* Adriana A. Amicarelli,[a] Fernando di Sciascio,[a] and Gustavo Scaglia[b]

This work presents a state estimator for a continuous bioprocess. To this aim, the Non Linear Filtering theory based on the recursive application of Bayes rule and Monte Carlo techniques is used. Recursive Bayesian Filters Sampling Importance Resampling (SIR) is employed, including different kinds of resampling. Generally, bio-processes have strong non-linear and non-Gaussian characteristics, and this tool becomes attractive. The estimator behavior and performance are illustrated with the continuous process of alcoholic fermentation of *Zymomonas mobilis*. Not too many applications with this tool have been reported in the biotechnological area.

*Contact information:  Instituto de Automatica, Facultad de Ingenieria, Universidad Nacional de San Juan, Av. Libertador San martin 1109 (oeste) San Juan, Argentina; b: Instituto de Ingenieria Quimica Facultad de Ingenieria, Universidad Nacional de San Juan, Av. Libertador San martin 1109 (oeste) San Juan, Argentina. *Corresponding author: olquinte@inaut.unsj.edu.ar*

## INTRODUCTION

Over the last few years several authors have contributed to the important progress in the technology used for monitoring, sensing, and process control. In spite of this progress, the biotechnological area has a lack of online process information that characterize chemical and biological variables such as biomass concentration, specific bacterial activity, and intermediate products concentration, among others. Frequently, these variables constitute the bioprocess states, and they are very necessary for its monitoring and control. A state estimator can be defined as a set of calculations that provides an estimate of a critical parameter, the value of which cannot be directly measured, based on information obtained from other variables in a dynamic system. The choice of an observer or a state estimator depends inherently on the particular problem. In practice, this choice is mainly influenced by the availability of a sufficiently representative model of the process, as well as by the reliability of the experimental data. When an adequate model is available, an Extended Kalman Filter (EFK) can be used, or a kind of High Gain observer or several estimators that use the process model (generally based on first principles) to perform the estimation of different variables (Jahanmiri and Rasooli 2005). On the other hand, if a model is not representative enough, asymptotic observers are implemented because the dependency of the model is not too strict, but their convergence depends on the operating conditions. Many times, observers are based on artificial

intelligence, such as neural networks. Approaches using support vector machines and fuzzy logic might be explored (Karatuzu et al. 2006; Yan et al. 2004).

Simulation and modeling are intimately related. Modeling is the activity of building mathematical expressions of a process by describing its fundamental physical and/or chemical relationships. Process behavior can be reproduced by simulations, through a process model. The construction of high-accuracy models requires deep modeling and process expertise, and is usually performed by an experienced specialist. By contrast, much process simulation is carried out using "off-the-shelf" models that provide little competitive advantage, or purely steady-state models that do not always capture the complexity of process operation. Instead of that, process simulation is a very important tool and valuable and essential activity, which must be significantly enhanced by using high-accuracy customer models of the process to capture corporate knowledge. From a control engineering point of view, simulation is a useful tool for the development and optimization of complex and nonlinear processes. Particularly we are focusing on the complex dynamic of enzymatic hydrolysis of lignocellulosic material and its subsequent fermentation to produce ethanol from *Zymomonas mobilis (Z. m)*.

When the model is incomplete or is not available by way of a-priori knowledge about the process, then by using different techniques it is possible to look in the literature for the state estimation from the input/output information. Several approaches for state estimation of biotechnological process have been developed, the most representative of which are the works of (Dochain 2002, 2003; Boillereaux 2000; Ascencio Leal 2001; Adilson and Rubens 2000; Rallo et al. 2002). Not many applications in the biotechnological area have been reported with nonlinear filtering tools.

*Zymomonas mobilis* is a gram-negative bacterium that attracts the researcher's attention due to the use of the Entner-Doudoroff pathway to produce energy from glucose catabolism. This pathway is inefficient because one ATP molecule per glucose molecule is consumed. This inefficiency is compensated by the bacterium's ability to metabolise glucose at a high rate (Parker et al. 1997). There is an industrial interest in the use of *Z. m* due to its capability to produce ethanol and sorbitol (Oliveira et al. 2005). There exists a requirement of more competitive ethanologenic microorganisms to expand the ethanol industry. *Z. m* has attracted attention as a promising bacterium for improving ethanol production (Daugulis et al. 1997). These micro-organisms show a highly nonlinear and oscillatory kinetic behaviour; besides, some states of the process are difficult or impossible to measure, and these include the biomass concentration and intermediate variables that represent the rate of ethanol production and to determine the inhibition effect.

In *Z. m* there are two possible mechanisms involving lipids to explain the ethanol tolerance. In the first mechanism, it is postulated that the high levels of cyclic lipids in the cell membranes protect the bacterium from the toxic effects of ethanol. In the second mechanism, it is postulated that the high levels of cis-vaccenic acid in the phospholipids of the bacterial membrane protect the bacterium from ethanol toxicity (Tano et al. 2000). The alcoholic fermentation of the *Z. m* has a few advantages when compared to other micro-organisms; these advantages are: *Z. m* provides ethanol levels production near to those that are achieved theoretically; it has a low biomass wastage; there are not any requirements respect to oxygen; the fermentation might be produced at low value of pH;

and most importantly, it shows a high specific productivity (98%), high efficiency in ethanol production, and a doubling of the specific rate (Mullan et al. 1991). The most important benefit of the process is the last characteristic: high ethanol yields and higher specific productivities (Kesava and Panda 196l; Kesava et al. 1996).

Recently another benefit has been observed, in a broad context. Biomass is a term for all organic material that stems from plants, trees, and crops. Also, organic waste and agricultural and forest residues are considered as biomass. By means of hydrolysis of lignocellulosic materials, as a pretreatment for ethanol production by *Zymomonas mobilis* fermentation, it is possible to add this fermentable substrate to the process. There have been many publications associated with this concept (OlssonHahn-Hiigerdal 1996; Szczodrak and Fiedurek 1996; Sun and Cheng 2002). The fermentation with *Zymomonas mobilis* has increased recently in importance because there exists the possibility of adding fermentable substrate to the process, reducing the cost of production and giving characteristics of lower environmental impact. The acid hydrolysis of lignocellulosic biomass produces reducing sugar in a suitable concentration for the mentioned fermentation process. Simulation is useful as tool to find the optimal actions to use *Z. m* properly for fuel ethanol and higher value products.

Quintero et al. (2004, 2005) explored the possibility of using the Kalman filter and the extended Kalman filter to perform the biomass estimation in this fermentation. The estimations obtained were not satisfactory due to the strong nonlinearity present in all the process states. The authors present in this work a comparison between particles filtering techniques and the classical tools.

In general, the optimum filtering techniques are used to reach the states estimation of a dynamical system with inputs and outputs that are observed by measurements disturbed by noise. "System states" is defined as the minimum information requirements in time that, in conjunction to the inputs values defined in all time from t≥t0, make it possible to determine the behavior of the system to any time t≥t0. The measurements are in general uncertain; they are called "noise measurements" and, even if the real states system are known, the measurements are not a deterministic function of the states mentioned, and also have a random component. In this context, the time evolution of the states is modeled through a dynamical system perturbed by a stochastic process (state noise), by using a stochastic differential equation. The noise or states disturbance, which is incorporated into the model to represent the uncertainties of the dynamic system, can arise not only from the random nature of the system, but also from signals or dynamics not considered in the model. In accordance with the Bayesian paradigm, the solution of the optimal filtering problem in time consists into obtaining the conditional probability distribution of the states with respect to the information obtained from the available measurements.

In this work, a state estimator to the continuous alcoholic fermentation process of *Z. m* is developed. In this area a nonlinear filtering, based on the recursive application of Bayes rule and Monte Carlo techniques, constitutes a novel tool. There are several variations of this method, proposed initially by (Gordon et al. 1993), and it is known by different names in the literature: particle filters, recursive Bayesian filters, Monte Carlo recursive filters, and simulation based filters (see Doucet 1998; Crisan and Doucet 2002). Spec-ifically, in this work the authors use variations of a Bayesian recursive filter SIR

(Sam-pling Importance Resampling), and different resampling schemes were applied to reduce the effect of the "sampling impoverishment" (Doucet 1998; Doucet et al. 2001, 2006).

Biotechnological processes have pronounced nonlinear and non-Gaussian features, and basically for this reason it is justifiable the use of the Bayes theory. In addition, as mentioned previously, other estimation techniques applied to the mentioned process, have not provided satisfactory results. In this work, the system states are estimated (Biomass concentration and the intermediate variables) from input/output information and an available process model. In addition, the random nature of a biochemical reaction at the molecular scale has been mentioned and studied by Gillespie (2000). At a macroscopic scale, Kurtz (1978) modeled the overall effect of these individual reactions on the global concentrations, by an additive noise term of variance proportional to the reaction kinetics (or propensity function) $r$. In this context, the states (biomass, substrate and product) are a Markov process, satisfying the Langevin chemical equation (Joannides 2004). All the previously mentioned features convert the bioprocess into an attractive application to use nonlinear filtering tools for state observer design. The results are based on simulation and numerical results. The use of such results as the basis for this work should not have a negative impact, since it is pioneering the use of particle filters in the biotechnological area.

The paper is organized as follows: First, the main features of the continuous fermentation process of *Zymomonas mobilis* are presented, then a brief summary of the estimation tools used in this work. Subsequently, a description of algorithms and the structure used to the state estimation is addressed; also, the results obtained and the corresponding analyses are presented.


## CASE STUDY AND ESTIMATION TOOLS

### Alcoholic Continuous Fermentation of *Zymomonas mobilis*

The continuous alcoholic fermentation process of *Z. m* can provide high ethanol performance, but it has an oscillatory behavior on the state variables of the process. From the control perspective, it represents a challenge due to the difficulties to measure some of these states, with the aim to be used as feedback signals. A model of the process (Oliveira 2005; Daugulis et al. 1997; Tano et al. 2000) is represented by the following differential and algebraic equations:

The change in biomass concentration can be obtained from,

$$\frac{dX}{dt} = \left( \mu + \left( R\frac{F_4}{V} + D_s \right) \left( \left( 4\frac{F_6}{F_2} \right) - 1 \right) \right) X$$

(1)

where $\mu$ is the specific speed growth, $R$ is the micro organisms recycling rate, and $F_2$, $F_4$, and $F_6$ are the bio reactor output, separator output and biomass feedback flow, respectively. $D_s$ is the substrate dilution rate, and $V$ the volume (see Fig. 1).

The change in substrate concentration is given by,

$$\frac{dS}{dt} = \left(-\frac{1}{Y_{p/s}}\right)(Q_p X) + DS_{in} - DS$$

(2)

where $Y_{p/s}$ is the substract/product performance coeficient, $Q_p$ is the specific ethanol production rate, $D$ is the total dilution rate, and $S_{in}$ is the substrate concentration on the input flow.
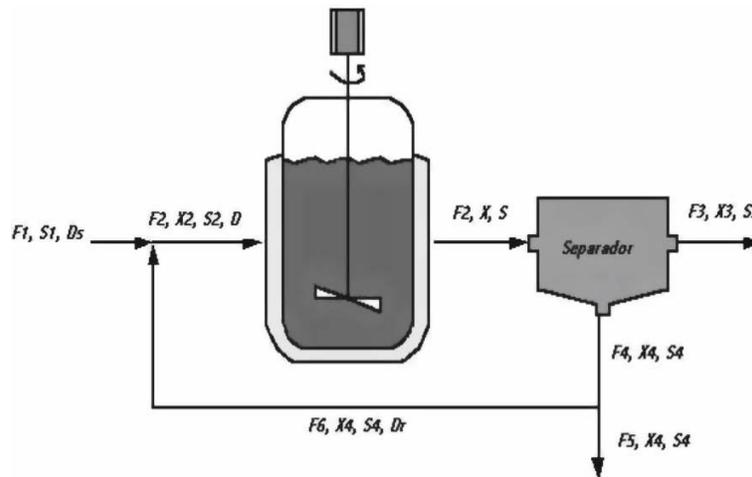


Figure. 1 Continuous fermentation process scheme

The change in product concentration is given by

$$\frac{dP}{dt} = Q_p X - DP$$

(3)

The weighted average of the ethanol concentration rate is,

$$\frac{dZ}{dt} = \beta(I - Z)$$

(4)

where $\beta$ is a weighted historic parameter for the ethanol concentration rate and $I$ is an intermediate variable auxiliary for the inhibition effect determination:

$$\frac{dI}{dt} = \beta(Q_p X - DP - I)$$

(5)

For further information about the physical meaning of inhibition variables see (Daugulis et al. 1997). The dynamic effect of the ethanol concentration rate on the biomass growth is given by,

$$f\mu = \frac{1}{2}\left(1 - \frac{e^{\lambda Z - \delta} - e^{-\lambda Z - \delta}}{e^{\lambda Z - \delta} + e^{-\lambda Z - \delta}}\right)$$

(6)

where $\delta$ and $\lambda$ are the parameters associated with the inhibition factor of the ethanol concentration rate.

The biomass growth rate is given by,

$$\mu e = \frac{\mu_{max} S \left( 1 - \left( \frac{P}{P_{ma}} \right)^a \right) \left( 1 - \left( \frac{P - P_{ob}}{P_{mb} - P_{ob}} \right)^b \right)}{K_s + S + \frac{S(S - S_i)}{K_i - S_i}} \tag{7}$$

where $\mu_{max}$ is the maximum value of the specific growth speed, $P_{ma}$ and $P_{ob}$ are factors of the ethanol inhibition for the specific growth rate expressed in (g/L), $P_{mb}$ is the factor related to the maximum ethanol inhibition for the cells growth expressed in (g/L), a and b are inhibition exponents for the ethanol production rate, $K_s$ is the substrate saturation coefficient, and $K_i$ is a substrate inhibition. The following conditions are considered:

$$\left( \frac{P - P_{ob}}{P_{mb} - P_{ob}} \right) = 0, P \le P_{ob}, S - S_i = 0, S \le S_i$$
$$\left( \frac{P - P_{ob}}{P_{mb} - P_{ob}} \right) = 1, P > P_{mb} \tag{8}$$

The dynamic growth speed is defined by:

$$\mu = f \mu * \mu e \tag{9}$$

And finally, the specific rate to the ethanol production is given by:

$$Qp = Qp_{max} \left( \frac{S}{K_{mp} + S} \right) \left( 1 - \left( \frac{P}{P_{me}} \right)^\alpha \right) \tag{10}$$

For this process, it is very important to reach an accurate estimation of the non-measurable system states with the purpose of using them for control.

It is necessary to put all equations in form of state variables, to see all interactions between variables involved. From Eqs (1)-(5),

$$\dot{x}_1 = \left( \mu + \left( R \frac{F_4}{V} + D_S \right) \left( \left( 4 \frac{F_6}{F_2} \right) - 1 \right) \right) x_1$$
$$\dot{x}_2 = \left( -\frac{1}{Y_{p/s}} \right) (Q_p x_1) + D * S_{in} - Dx_2$$
$$\dot{x}_3 = Q_p x_1 - Dx_3 \tag{11}$$
$$\dot{x}_4 = \beta (x_5 - x_4)$$
$$\dot{x}_5 = \beta (Q_p x_1 - Dx_3 - x_5)$$

By working on the equations and replacing the Eqs. (6), (7), (9), and (10), one can obtain the following:

$$\dot{x}_1 = \left(\left[\left[\left(\frac{1}{2}\left(1-\frac{e^{\lambda x_4-\delta}-e^{-\lambda x_4-\delta}}{e^{\lambda x_4-\delta}+e^{-\lambda x_4-\delta}}\right)\right)\right]\left(\frac{\mu_{max}x_2\left(1-\left(\frac{x_3}{P_{ma}}\right)^a\right)\left(1-\left(\frac{x_3-P_{ob}}{P_{mb}-P_{ob}}\right)^b\right)}{K_s+x_2+\frac{x_2(x_2-S_i)}{K_i-S_i}}\right)\right]+\left(R\frac{F_4}{V}+D_S\right)\left(\left(4\frac{F_6}{F_2}\right)-1\right)\right)x_1$$

$$\dot{x}_2 = \left(-\frac{1}{Y_{p/s}}\right)\left(\left[Qp_{max}\left(\frac{x_2}{K_{mp}+x_2}\right)\left(1-\left(\frac{x_3}{P_{me}}\right)^\alpha\right)\right]x_1\right)+D*S_{in}-Dx_2$$

$$\dot{x}_3 = \left[Qp_{max}\left(\frac{x_2}{K_{mp}+x_2}\right)\left(1-\left(\frac{x_3}{P_{me}}\right)^\alpha\right)\right]x_1-Dx_3 \tag{12}$$

$$\dot{x}_4 = \beta(x_5-x_4)$$

$$\dot{x}_5 = \beta\left(\left[Qp_{max}\left(\frac{x_2}{K_{mp}+x_2}\right)\left(1-\left(\frac{x_3}{P_{me}}\right)^\alpha\right)\right]x_1-Dx_3-x_5\right)$$

By a set of algebraic considerations from Eq. (1), and working with the defined flows in (Echeverry et al. 2003), the biomass term will be expressed as function of the mentioned $D$ that is the total dilution rate. The total dilution rate is $D_s + D_r$, in which $D_r$ is the dilution rate associated to biomass recycle $R$ and substrate dilution rate $D_s$. Then let us consider the following set of differential equations,

$$\dot{x}_1 = \left[\left[\left(\frac{1}{2}\left(1-\frac{e^{\lambda x_4-\delta}-e^{-\lambda x_4-\delta}}{e^{\lambda x_4-\delta}+e^{-\lambda x_4-\delta}}\right)\right)\right]\left(\frac{\mu_{max}x_2\left(1-\left(\frac{x_3}{P_{ma}}\right)^a\right)\left(1-\left(\frac{x_3-P_{ob}}{P_{mb}-P_{ob}}\right)^b\right)}{K_s+x_2+\frac{x_2(x_2-S_i)}{K_i-S_i}}\right)\right]x_1+[RD/4+D_s](R-1)x_1$$

$$\dot{x}_2 = \left(-\frac{1}{Y_{p/s}}\right)\left(\left[Qp_{max}\left(\frac{x_2}{K_{mp}+x_2}\right)\left(1-\left(\frac{x_3}{P_{me}}\right)^\alpha\right)\right]x_1\right)+D*S_{in}-Dx_2$$

$$\dot{x}_3 = \left[Qp_{max}\left(\frac{x_2}{K_{mp}+x_2}\right)\left(1-\left(\frac{x_3}{P_{me}}\right)^\alpha\right)\right]x_1-Dx_3 \tag{13}$$

$$\dot{x}_4 = \beta(x_5-x_4)$$

$$\dot{x}_5 = \beta\left(\left[Qp_{max}\left(\frac{x_2}{K_{mp}+x_2}\right)\left(1-\left(\frac{x_3}{P_{me}}\right)^\alpha\right)\right]x_1-Dx_3-x_5\right)$$

The set of parameters used to this simulation are the defined in (Daugulis et al., 1997; Daugulis et al. 1999; Echeverry et al, 2003). The parameters used to simulate oscillatory behavior of the *Z. m* are presented in Table 1.

Table 1. Parameters Used for Simulation of Fermentation.

| Parameters | Oscilatory behavior |
|---|---|
| $\alpha$ | 8.77 |
| $\beta$ | 0.0366 |
| $\delta$ | 0.824 |
| $\lambda$ | 21.05 |
| $a$ | 0.3142 |
| $b$ | 1.415 |
| $Qp_{max}$ | 2.613 |
| $\mu_{max}$ | 0.41 |
| $Y_{p/s}$ | 0.495 |
| $P_{ma}$ | 217 g/l |
| $P_{ob}$ | 50 g/l |
| $P_{mb}$ | 108.0 g/l |
| $K_s$ | 0.5 g/l |
| $K_i$ | 200 g/l |
| $S_i$ | 80.0 g/l |
| $P_{me}$ | 127 g/l |
| $K_{mp}$ | 0.5 g/l |

The solution proposed to simulate the real behavior of the continuous fermentation was performed by the use of Matlab, through the Euler and Runge Kutta numerical methods in self developed functions to be used as part of the filtering simulation. The sample time used for simulation, was that which has been considered most appropriate for real plants. The sampling time values selected were 0.05 and 0.1 hours.

**Estimator Design**

The method that establishes the basis for many developed filters is known as bootstrap, condensation algorithm, particle filtering, and survival fittest, among other names. This method uses the importance function, defined as a probability distribution function which depends on the observations until time $k$, when it will be easily sampled. In this way, the posterior density (the conditional probability of the states to the measurements until time k-1) can be approached by

$$p(x_{0:k} / y_{1:k-1}) \approx \sum_{i=1}^{Ns} W_k^i \delta(x_{0:k} - x_{0:k}^i)$$

(14)

where $N_s$ is the number of particles used, $\delta(.)$ is the dirac delta function, the weights $W_k^i$ are chosen by importance sampling, under the following considerations i) the probability density of the evaluation samples is proportional to the importance function $q(x)$, $p(x) \propto q(x)$, and ii) the states should be sampled from the importance density function $x^i \sim q(x)$

Then, the importance function is chosen and the weights are factorised to make it dependent only on $x_{k-1}$; this way the weights can be expressed as,

$$W_k^i \propto W_{k-1}^i \frac{p(y_k / x_k^i) \, p(x_k^i / x_{k-1}^i)}{q(x_k^i / x_{k-1}^i, y_k)}$$

(15)

and the posterior density, by:

$$p(x_k / y_{1:k}) \approx \sum_{i=1}^{Ns} W_k^i \delta(x_k - x_k^i)$$

(16)

The algorithm consists of the recursive propagation of the weights and support points when each measurement (output measurement) is obtained sequentially. The algorithm itself, in its original conception, had many issues with respect to the variance reduction and presented some problems of particles degeneration. Consequently, some improvements oriented to avoiding the consideration of particles with no information contribution, and to minimise the observer's variance, were performed. The resampling step should be performed by the observation of the degeneracy of the function and its comparison with a basis value. This step eliminates the particles with low weights and concentrates the particles with large weights.

This implies generating a new set of samples generated by resampling or rejection with the form of $\left[ \{ x_k^{i*} \}_{i=1}^{Ns} \right]$. In this way, an approximate discrete representation of the posterior density function can be obtained:

$$p(x_k / y_{1:k}) \approx \sum_{i=1}^{Ns} W_k^i \delta(x_k - x_k^i)$$

(17)

The resampling step is obtained by using different variants: stratified sampling, residual sampling and systematic sampling among others. The resampling step reduces the effects of degeneration but introduces some practical issues, limiting the opportunity of parallelism in the algorithm, because it combines set of particles, and in addition, the particles with large weights are selected statistically in many times. This phenomenon leads to a sample impoverishment, which is a loss in the diversity as the particles contain repeated points. This is a very important outcome when a process noise is low, and finally, due to the particles pattern diversity reduction, any application of smoothing technique must be degenerated. Another possible solution to the weights degeneracy issue is to start the algorithm with a good importance function, or else, evaluate its behaviour into the algorithm. The choice of an importance function, *q(x)*, is shown in the literature as a key factor in the SIS-SIR algorithm. This function must be chosen to minimise the variance of the optimal weights $W_k^*$, maximising the degeneracy or maximum effort measure, $\hat{N}eff$. (Doucet 1998; Crisan and Doucet 2002; Doucet et al., 2001, 2006). The importance function will be expressed as:

$$q\left(x_{k/}x_{k-1}^i, y_k\right)_{opt} = p(x_k / x_{k-1}^i, y_k)$$

$$q\left(x_{k/}x_{k-1}^i, y_k\right)_{opt} = \frac{p(y_k / x_k / x_{k-1}^i)\, p(x_k^i / x_{k-1}^i)}{p(y_k / x_{k-1}^i,)}$$

(18)

by some operations on weights expressions,

$$W_n^i \propto W_{n-1}^i \int p(y_n / x_n^i)\, p(x_n^i / x_{n-1}^i) dx_n^i$$

(19)

See Appendix A for the detailed algorithm. In order to use this solution to the problem, there exist some issues, such as the ability to sample from the a-priori function $p(x_n / x_{n-1}^i, y_n)$, or to evaluate the integral to the weights calculation in each new state. The optimal importance function is not always easy to find; for this reason, just a few cases can be performed as optimal. One of these is the case when states $x_k$ belong to a finite set; in that case, the integral becomes tractable and might be handled as summatory; the sampling from $p(x_n / x_{n-1}^i, y_n)$ is possible in that way. This is the case of the Jump Markov Linear System. Figure 2 shows the estimation scheme. The input signals $u(t)$ (input substrate concentration $S_{in}$, dilution rate $D_s$, and microorganisms recycle term $R$), and the output signals $y(t)$ (outflow Substrate $S$ and product $P$) corresponding to the model used as real process, feed the filter block that makes the state estimation. The estimations are compared with the model considered as patron. The bioreactor simulation was made in open loop (oscillatory behavior), and the inputs $u(t)$ were considered constants (Daugulis et al. 1997). The dynamics and the delays on the inhibition variable that make this process highly nonlinear easily observed.

$$u_t$$

**PROCESS MODEL**

$$\begin{cases} dx_t = f\left(x_t, t\right) dt + \sigma_1\left(x_t, t\right) dw_t \\ dy_t = g\left(x_t, t\right) dt + \sigma_2\left(x_t, t\right) dv_t \end{cases}$$

$$y_t$$

**FILTER MODEL**

$$\begin{cases} x_{t+1} = f\left(x_t, t\right) dt + \sigma_1\left(x_t, t\right) dw_t \\ y_t = g\left(x_t, t\right) dt + \sigma_2\left(x_t, t\right) dv_t \end{cases}$$
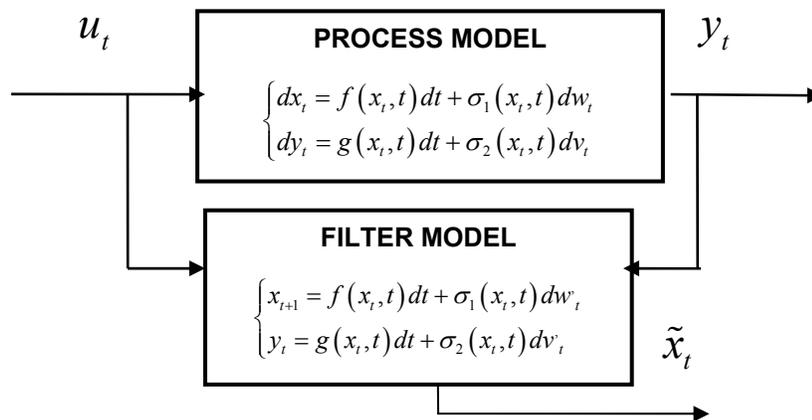
$$\tilde{x}_t$$

Figure. 2. Estimation scheme used

The model used to the diffusion terms of both filter and process model was $\sigma_1(x_t, t)$ as a Gaussian noise that was dependent on a uniformly distributed internal variable,

updated each sample time particle by particle; and $\sigma_2(x_t,t)=1$ combined with different Gaussian noise samples in the components $dw_t, dv_t, dw_t$ and $dv_t$, to simulate unknown disturbances of measurement equipment.

It is important to remark that the measurements used to feed the filter as external information were taken from the model corresponding to the plant. They included the diffusion term built with the Gaussian noise and the corresponding diffusion term for states simulation (because both Substrate and Product are also states of the system). These values of Substrate and Product updated each simple time are also added to the value calculated from the corresponding diffusion term by simulating the effects of electronic noise in sensors, as discussed briefly above.

## RESULTS AND DYNAMIC ANALYSIS

The results given in this section are based on simulation and numerical results of particle filtering technique, using from 100 to 1000 particles. Variations of the random components added to the dynamical system were made. A set of tests to measure the effectiveness of the filter proposed in front of the computational application cost were used; and the efficiency resampling schemes used in this configuration was analysed. The results showed that to the intermediate inhibition variables, which depend on the rates into product concentrations, the errors presented were, in some cases less than the other variables. It will be used as a monitoring variable into a continuous fermentation, to know how the fermentation process is possible to be handled. It is important to remark that this information is relevant, in the sense of information for control. The model has the enough accuracy to be validated and used for a process, and the estimation obtained by the filter can be easily used for real control purposes. Figure 3 shows the test results performed with 500 particles by the use of the deterministic resampling to the continuous bioprocess. A measurable output of substrate and product (*S* and *P*) was considered. As additive model to the measures, a normal Gaussian noise was used, with mean and variance adequate and of considerable magnitude.

Computational time to reach a good approach was smaller than the sample time appropriate for this application. Sample time used for simulation of *Z.m* fermentation was 0.1 hours (6 minutes), because the dynamic behaviour doesn't change considerably within this time interval. It means that algorithms developed can be easily implemented into a real plant.

It is important to remark that at the end estimation time, the error magnitude tends to increase. That is, the quality of the estimation tends to decrease because the deterministic resampling scheme is not efficient in this case, in addition to the effect of the meaningful noise features, used as test to the robustness in the filter estimation. It is established as a hypothesis that the increase in the error magnitude is dependent on the choice of the importance function to this particular case, and the possibility to use another one is considered.
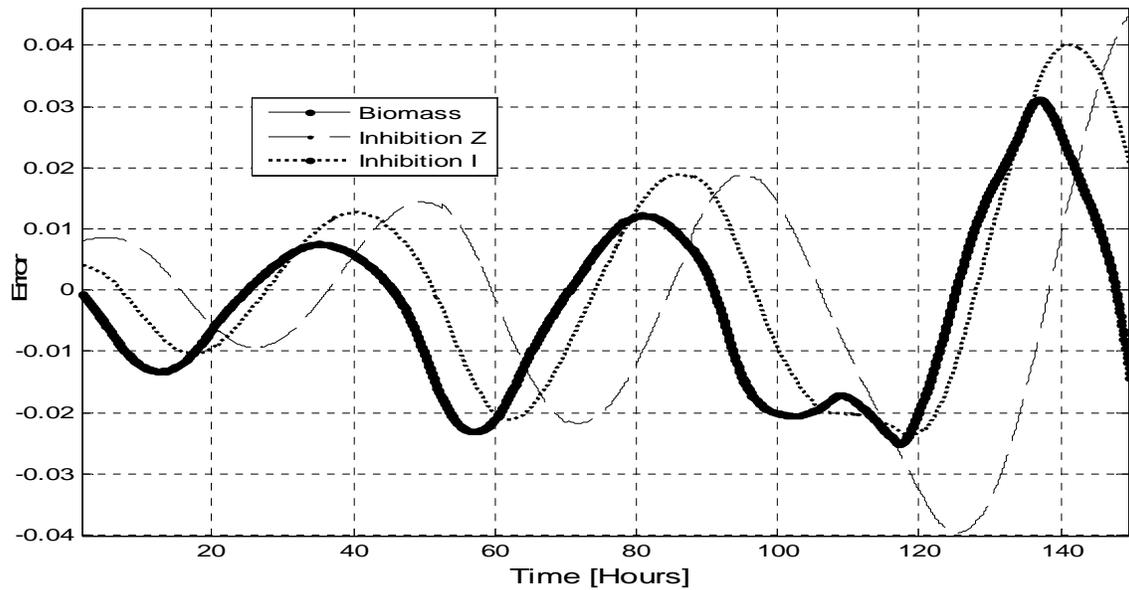
Figure 3. Estimation errors of process states during the 150 hours test in simulation with a sample time of 0.1 hours.

In Fig. 4 it can be observed from results that the particle filter performance is adequate for simulated data. These results are remarkable from the perspective of the level of accuracy of the model used (Daugulis et al. 1997; Raposo et al. 2005; Daugulis et al. 1999), and its closeness to the data reported in the literature. This figure also illustrates the fact that for biotechnological variables, the measurements cannot be easily sampled, and this approach provides an approximated estimation value that can replace the real measurement over a time interval. In the same way, the filter performance makes it possible to assume the uncertainties on measures and disturbances in the measurement methods.

Figure 5 presents the inhibition variables I and Z present in the dynamic behavior of the *Z. m* bacteria. These variables are relevant in the sense of the information they provide about the effect of the ethanol concentration rate, not only as a weighted average, but also as an intermediate variable for the determination of the inhibition effect from ethanol. The filter performs these estimations of variables satisfactorily and handles the possible influence of noised measurements of product, and its possible effects on the estimation properly.

Future work will be performed by the use of real data from measurable variables, and the estimation scheme will be developed looking for the real application and control purposes. An important remark is that the filter follows the model properly, and as approach to the real problem, the performance to online implementation was tested. The modeled dynamics are according to the real behavior and the robustness against disturbances of modeling and uncertainties was shown in simulation. It was necessary to apply the SMC Particle Filtering methodology the assumption of a sampled data model for the SDE´s; this way the set of equations are posed as a new and improved model that includes uncertainties and disturbances. SIR Filters are satisfactory, but even if this is a

novel application of the SMC, it may require a more advanced SMC method to the real data problem solution (Briers et al. 2004, 2005; Briers 2006).
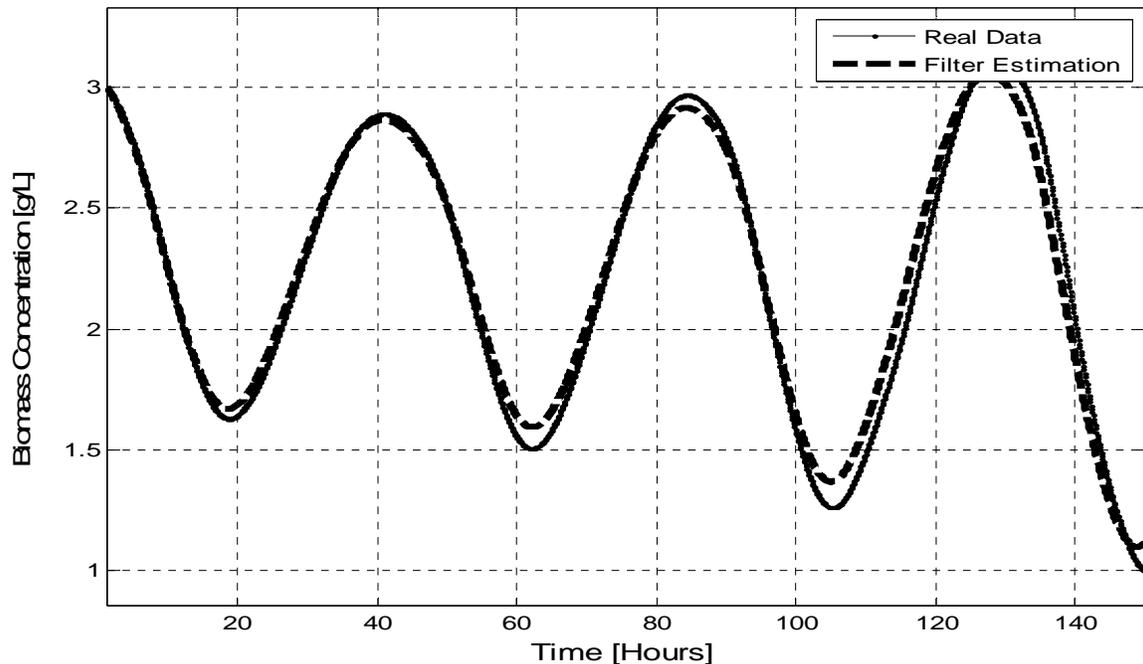


Figure 4. Estimation performance for the biomass concentration in the fermentation process of *Zymomonas mobilis*. The dotted line describes the estimated biomass concentration by the SIR filter while the solid line describes the real value.

In previous works, some techniques for state estimation in *Z.m* have been explored. The work of Quintero et al. (2004) presented a control scheme in closed loop with a virtual sensor based on a fuzzy model. The estimator performs well in simulation, but despite that, it is not reliable, and its performance depends on the data used for training. After that, Quintero et al. (2005) used the Kalman Filter and Extended Kalman Filter for the same purpose. These results are inferior to those presented in this work, even if they are obtained in simulation, by the use of models developed and validated with real data by (Raposso et al. 2005).

The implementation on a real fermentation should be the logical continuation of research in the subject. Currently the authors are working on the improvement of this work based son real data fermentation both in oscillatory and steady state behavior.
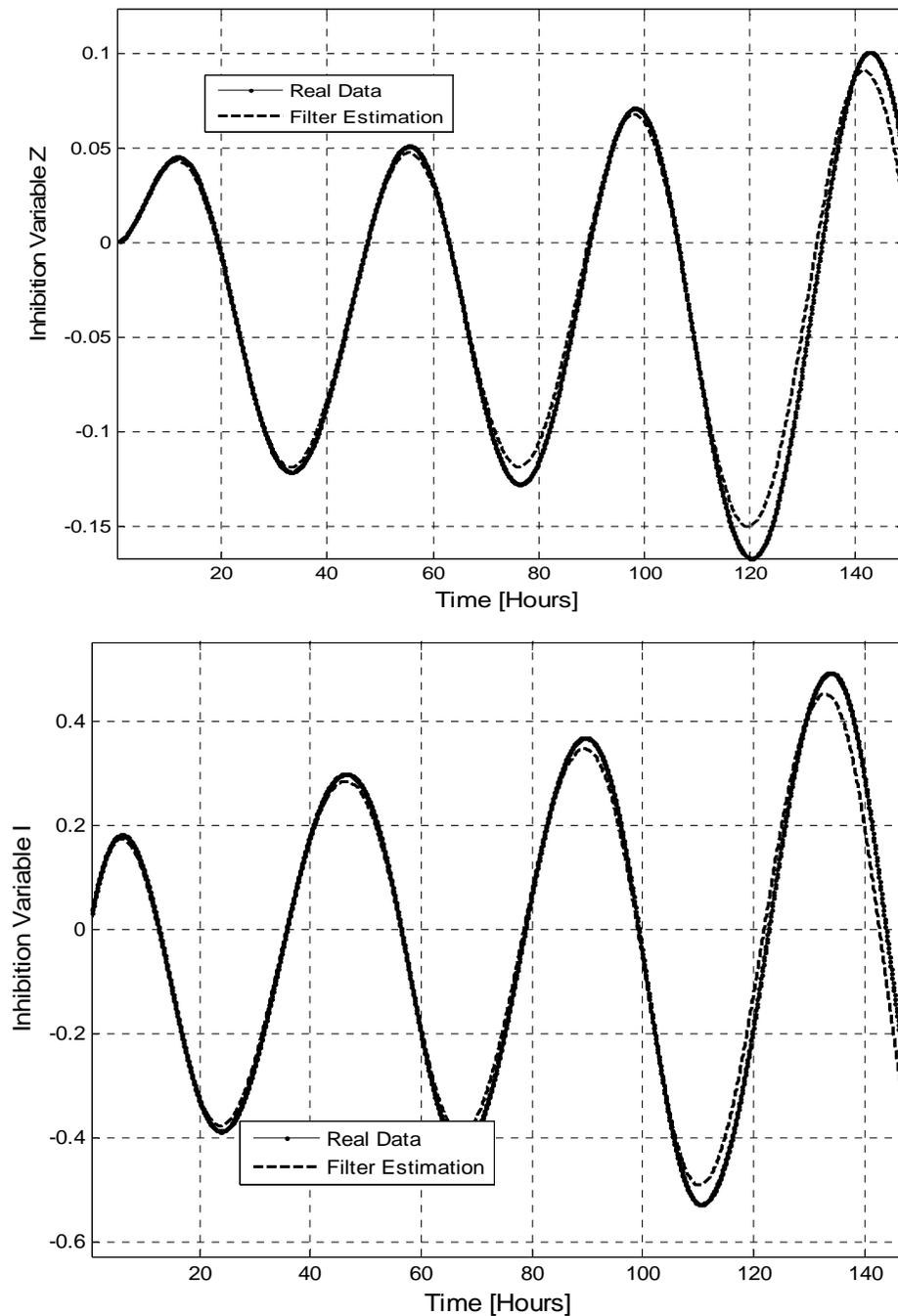
Figure 5a. 5b. Estimation performance for the inhibition variables. The dotted lines represent the dynamics of Inhibition Z and I estimated by the SIR filter and while the solid lines are the considered real system dynamics

## CONCLUSIONS

1. A state estimator based on non-linear filtering techniques was presented. The application of the particle filters developed as state estimators can be applied to non-linear non-Gaussian dynamical systems. The technique and methodology were applied to the case of a bioreactor for the continuous alcoholic fermentation process of *Zymomonas mobilis*, one of the most promising micro-organisms for genetic engineering envisaging the development of strains for lignocellulose fermentation.
2. An approach for a SDE´s model for a *Zymomonas mobilis* continuous fermentation was presented, looking for the correct modeling of uncertainties and the right implementation of SMC methodology for state estimation. It was observed that the application of the use of particle filtering as state estimator is acceptable and feasible, and the implementation to solve the problem of online biomass estimation, in a continuous process, is viable due to its reliability and admissible computational cost relative to the real problem sample times.
3. In the same way, the filter showed a good performance with the inhibition variables, of relevant importance for the dynamic behaviour of bacteria in open and closed loop. Currently, we are first proving this method with real data and later, the performance of the estimation tool into a numerical methods control closed loop will be studied.
4. Results were satisfactory beyond the authors' expectations. Compared with other techniques previously developed, such as fuzzy logic and Kalman filter and Extended Kalman filter, the performance of this estimator was better.

## APPENDIX A

**Filtering general algorithm detail**, Sequential Importance Sampling *SIS* (Arumpalam et al. 2002):

$$\left[\left\{x_k^i, W_k^i\right\}_{i=1}^{Np}\right] = SIS\left[\left\{x_{k-1}^i, W_{k-1}^i\right\}_{i=1}^{Np}, y_k\right]$$

*For* $i = 1 : Np$

$$x_k^i \ \Box \ q\left(x_{k/x_{k-1}^i}, y_k\right)$$

$$W_k^i \propto W_{k-1}^i \frac{p(y_k / x_k^i) \, p(x_k^i / x_{k-1}^i)}{q(x_k^i / x_{k-1}^i, y_k)}$$

*end*

Sequential importance sampling with resampling step *SIR* (Doucet, de Freitas, 2000):

$$\left[ \left\{ x_k^{j*}, W_k^j, i^j \right\}_{j=1}^{Np} \right] = RESAMPLE\left[ \left\{ x_k^i, W_k^i \right\}_{i=1}^{Np} \right]$$

Initialize counter

$c_1 = 0$

*For* $i = 2 : Np$

Construye la CDF

$c_i = c_{i-1} + W_k^i$

*end*

Starts at the beginning of CDF

$i = 1$

Generates the start point

$u_1 \ \Box \ U\left[ 0, \dfrac{1}{Np} \right]$

*For* $j = 1 : Np$

Go through CDF

$u_j = u_i + \dfrac{(j-1)}{Np}$

*While* $u_j > c_i$

$i = i + 1$

*end While*

Asigns the sample

$x_k^{i*} = x_k^i$

Asigns the weight

$W_k^j = \dfrac{1}{Np}$

Asigns parents

$i^j = i$

*end*

The resampling methods used are the residual, deterministic, and multinomial.


## ACKNOWLEDGMENTS

## REFERENCES CITED

Adilson, J., and Rubens, M. (2000). "Soft sensors development for on-line bio-reactor state estimation," *Computers and Chemical Engineering* 24, 1099-1103.

Arulampalam M. S., Maskell S., Gordon N., and Clapp T., "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian tracking," IEEE Trans. Signal Processing, vol. 50, no. 2, pp. 174–188, Feb. 2002.

Ascencio Leal, R. (2001). "Artificial neural networks as a biomass virtual sensor for a batch process," Proceedings of the IEEE International Symposium on Intelligent Control, September, Mexico City.

Boillereaux, L., and Flaus, J. (2000). "A New Approach for Designing Model-Based Indirect Sensors". IEEE Transaction on Control Systems Technology 8, 4.

Briers, M. (2006). Improved Monte Carlo Methods for State-Space Models, Cambridge University PhD Thesis.

Briers, M., Doucet, A, and Maskell, S. (2004). "Smoothing algorithms for state-space models," Cambridge University Engineering Department Technical Report, CUED/F-INFENG/TR.498.

Briers, M., Doucet, A., and Singh, S. (2005). "Sequential auxiliary particle belief propagation," International Conference on Information Fusion.

Chen T, Morris J and Martin E. "Particle Filters for the Estimation of a State Space Model" Centre for Process Analytics and Control Technology School of Chemical Engineering and Advanced Materials University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK 2003.

Crisan, D., and Doucet, A. (2002). "A survey of convergence results on particle filters for practitioners," IEEE Transaction on Signal Processing.

Daugulis, A. J., McLellan, P. J., and Li, J. (1997). "Experimental investigation and modeling of oscillatory behavior in the continuous culture of Zymomonas mobilis," Biotech. and Bioeng. 99-105, October.

Daugulis, A., McLellan, P., and Li, J. (1999). "The incidence of oscillatory behavior in the continuous fermentation of ZM," Biotechnology Progress.

de Freitas N, J. F. G. "Bayesian methods for neural networks," Ph.D. thesis, Dept. Eng., Univ. Cambridge, 1998. Available on line in: http://www.cs.ubc.ca/_nando/publications.html. Department of Electrical Engineering Linkoping University, SE{581 83 Linkoping, Sweden, 2003

de Freitas website http://www.cs.ubc.ca/~nando/software.html

Dochain, D. (2002). "State A state observer for (bio)processes with uncertain kinetics," Proceedings of the American Control Conference. Anchorage, AK May 8-1 0.

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

Doucet. A. (1998). "On sequential simulation-based methods for Bayesian filtering," Technical report University of Cambridge.

Doucet A., de Freitas N., K. Murphy, and S. Russell, "Rao-Blackwellised particle filtering for dynamic Bayesian networks" in Proc. UAI2000, pp. 176–183, 2000.

Doucet, A., Briers, M., and Senecal, S. (2006). "Efficient block sampling strategies for sequential Monte Carlo methods," Journal of Computational and Graphical Statistics 15(3), 693-711.

Doucet, A., de Freitas, N., and Gordon, N. (2001). Sequential Monte Carlo in Practice, Springer-Verlag. ISBN 0-387-95146-6.

Echeverry, N., Quintero, O., Ramírez, M. (2003) Modelamiento Simulación y pruebas de control para un biorreactor alcohólico en continuo. Trabajo Dirigido de Grado Universidad Nacional de Colombia Sede Medellín, Facultad de Minas.

Gillespie, D. (2000). "The chemical Langevin equation," J. Chem. Phys. 113, 297–306.

Gordon, N., Salmond, D., and Smith A. (1993). "Novel approach to non linear-non Gaussian Bayesian state estimation," IEE Proceedings F Radar, Sonar Navig. 140, 107–113.

Jahanmiri A., and Rasooli H. (2005) "On-line states and parameter identification of acetone–butanol–ethanol fermentation process," *Biochemical Engineering Journal* 24(2), 115-123.

Joannides, M., Larramendy–Valverde, I. and Rossi, V. (2000). "Monte–Carlo observer for a stochastic model of bioreactors," Technical report.

Karakuzu, C., Türker, M., and Öztürk, S. (2006). "Modelling, on-line state estimation and fuzzy control of production scale fed-batch baker's yeast fermentation," *Control Engineering Practice* 14(8), 959-974.

Kesava, S. S., and Panda, T. (1996). "Ethanol production by immobilized whole cells of Zymomonas mobilis in a continuous flow expanded bed bioreactor and a continuous flow stirred tank bioreactor," Journal of Industrial Microbiology 17, 11–14.

Kesava, S. S, Panda, T., and Rakshit, S. K. (1996). "Production of ethanol by immobilised whole cells of Zymomonas mobilis in an expanded bed bioreactor," Process Biochemistry 31, 449–456.

Kurtz, T. (1978). "Strong approximation theorems for density dependent Markov chains," Stoch Proc Appl. 6, 223–240.

Oliveira, I., Rapôso, A. C., Souto-Maior, A. M., and Lopes, C. E. (2005). "Oscillatory behavior of zymomonas mobilis in continuous stirred tank bioreactor," 2nd Mercosur Congress on Chem. Eng., Rio de Janeiro Brasil.

Olsson L., and Hahn-Hiigerdal, B. (1996). "Fermentation of lignocellulosic hydrolysates for ethanol production," Enzyme and Microbial Technology 18, 312-331.

O'Mullan, P., Szakacs-Dobozi, M., and Eveleigh, D. E. (1991). "Identification of saccharolytic enzymes of Zymomonas mobilis CP4," Biotechnology Letters 13(2), 137–142.

Parker, C., Peekhaus, N., Zhang, X., and Conway, T. (1997). "Kinetics of sugar transport and phosphorylation influence glucosa and fructose cometabolism by Zymomonas mobilis,". Applied and Environmental Microbiology 63(9), 3519–3525.

Quintero, O., di Sciascio, F., and Álvarez, H., (2004). "Estructura de dos grados de libertad y sensor virtual para controlar un bioproceso," XIX Congreso Argentino de Control Automático AADECA, Bs. As., Argentina.

Quintero, O., and di Sciascio, F. (2005). "Aplicación del filtro de Kalman extendido a un proceso de fermentación alcohólica en continuo," XI Reunión de Trabajo en Procesamiento de la Información y Control RPIC.

Quintero, O., Amicarelli, A. di Sciascio, F., "Recursive Bayesian Filtering for States Estimation: An Application Case in Biotechnological Processes", MCMSki (3rd IMS-ISBA) meeting in Bormio, Italy Jan 9-11, 2008.

Rallo, R., Ferre-Gine, J., Arena, A., and Giralt, F. (2002). "Neural virtual sensor for the inferential prediction of product quality from process variables," Computers and Chemical Engineering 26, 1735/1754.

Raposo, A., Pinto, J., and Pinheiro, I. (2005). "Parameter estimation of continuous fermentations using particle swarm," 2nd Mercosur Cong. on Chem. Eng., Rio Brasil.

Sun, Y., and Cheng, J. (2002). "Hydrolysis of lignocellulosic materials for ethanol production: A review," Bioresource Technol. 83, 1–11.

Szczodrak J., and Fiedurek, J. (1996). "Technology for conversion of lignocellulosic biomass to ethanol," Biomass and Bioenergy 10(S/6), 367-375.

Tano, M., Buzato, J., and Celliogoi, M. A. (2000). "Sugar cane juice fermentation by Zymomonas mobilis CP4 subjected to inhibition by ethanol and high initial concentration of substrate," Brasilian Arch. Biol. Tech. 43(April).

Yan, W., Shao H., and Wang., X. (2004). "Soft sensing modeling based on support vector machine and Bayesian model selection," *Computers & Chemical Engineering* 28(8), 1489-1498.