

Substrate feeding Strategy Integrated with a Biomass Bayesian Estimator for a Biotechnological Process

Adriana Amicarelli+, Lucía Quintero Montoya*, Fernando di Sciascio+
+*Instituto de Automática (INAUT). Universidad Nacional de San Juan.*
Av. Libertador San Martín 1109 (oeste). J5400ARL. San Juan, Argentina.
amicarelli@inaut.unsj.edu.ar

* *Universidad EAFIT, Mathematical Sciences Department, School of Sciences, Cra.49 No.7sur-50*
oquintel@eafit.edu.co

Abstract— This work proposes a substrate feeding strategy for a bioprocess integrated with a biomass estimator based in nonlinear filtering techniques. The performance of the proposed estimator and the substrate strategy are illustrated for the δ -endotoxin production of *Bacillus thuringiensis* (*Bt*) in batch and fed batch cultures. Nonlinear filtering techniques constitutes an adequate option as estimation tool because of the strongly nonlinear dynamics of this bioprocess and also due to nature of the uncertainties and perturbations that cannot be supposed Gaussians distributed. Biomass estimation is performed from substrate and dissolved oxygen. Substrate feeding strategy is intended to obtain high product concentration. Simulations results along with their experimental verifications demonstrate the acceptable performance of the proposed biomass estimator and the substrate feeding strategy.

Keywords— **Biomass Estimation, Nonlinear Filtering, Bayes Theory, Fed Batch Process, Substrate Feeding Strategy.**

1. Introduction

Motivation of this paper arises from the experimental results of the previous work [1] where it can be seen that it is possible to increase biomass concentration of *B. thuringiensis subsp. medellin* by Intermittent Fed-Batch Culture (IFBC) with total cell retention (IFBC-TCR). In this mentioned work the objective was to achieve a fed batch bioprocess model. The process model belongs to so-called hybrid biochemical processes classes. These processes are being carried out at units under integrated operation such as membrane bioreactors. Hybrid operation indicates the coexistence of continuous and discrete event dynamics during the process operation. This way, allowing to improve yields and production efficiencies with reduced operating costs and initial investments [2, 3].

In the present work the authors propose a substrate feeding strategy according to the model previously developed and completed by the same authors [4] with an additional model of dissolved oxygen and propose a biomass estimator integrated with this feeding strategy with the objective to increment productivity with the adequate estimation of biomass concentration.

Frequently, process monitoring is a critical problem in many industrial applications. However, just a few of the considered process state variables (reactors instantaneous composition) are possible to measure by real physical sensors. In bioprocesses, the biomass concentration is rarely available online and it is generally the most important variable from the process control point of view and

hence, for its productivity. In the particular case of the δ -endotoxins production by *Bacillus thuringiensis* (*Bt*), the biomass constitutes the fermentation product itself. In this work, we propose a *Bt* δ -endotoxin estimator, i.e., a biomass observer design with the aim to give adequate and realizable online biomass estimation for the mentioned bioprocess. The constructed state could be used to control the process [4]. A survey of used techniques for control can be found in the work of [5]. The state estimation objective is to rebuild the largest possible amount of system information (since is not totally available online), from the all the available system states' measurements [6].

In [7] it is ensured that batch and semi batch process quality control problems can be reduced to estimation of process state from the initial conditions. In [8] some typical features of these processes are mentioned. There are many reports, in which state estimation in bioprocess is performed with different techniques ([9]; [10]; [11]; [12], [13] and so on. [14] proposed a state observer for a bioprocess with known kinetics structure and unknown kinetic parameters. The author proposed the use of these parameters as extra parameters design to guarantee reach stationary error equal to zero. In the mentioned work, Dochain divides the sates observers in two classes: classics observers, like the Luenberguer, Kalman, among others, and a second class of asymptotic observers, where the estimator's structure depends completely on the model's precision; the latter supposes that uncertainty of models rely on bioprocess kinetics. Recently [15] publicized an interesting work about the applications of recent observers to chemical process systems and classify them into six classes, which differentiate them with respect to their features and assists in the design of observers. In [9] and [16] biomass estimation in a bioprocess is performed by the use of different kinds of Neural Networks, and later using sensorial fusion techniques. [17] worked on variables prediction in batch processes, by the use of subspaces identifications techniques. Generally, not only in continuous, but also, in batch process, model and parameters uncertainty makes of the asymptotic observers a good option that does not require the perfect model knowledge but depends strongly of operation conditions [14]. In [4] and [13] several biomass estimators for the δ -endotoxins of *Bacillus thuringiensis* production process are presented. In the mentioned work, first principle based techniques are used; neural networks estimators and finally, fusion techniques to achieve better biomass estimation. It should be noted that except for the Kalman filter, those techniques are completely developed in a deterministic context. Modelling uncertainties are only taken into account through the variation of parameters, and the performance in presence of noise measurements is evaluated through numerical simulation.

Generally real processes have strongly nonlinear dynamics due to the interactions nature between their variables and the disturbances always present in the real systems. A common way to reflect the non linearities and disturbances is through the system representation in terms of state variable models. Nonlinear stochastic differential equations (SDE) are the mathematics object that characterises these kinds of models. If uncertainties or stochastic disturbances are neglected, then the SDE's are transformed to nonlinear differential equations. The form in which the information propagates through the system's evolution, determines the states values and their consequences over the output system; this way, propagation of information is determined by the probability distribution function of the system states. Sometimes, it is possible to determine the relationship between the system output and their states. However, due to its non-linear and non-Gaussian nature, the application of traditional information recovering techniques is difficult to apply in practice. Traditional information reconstruction is possible, if it is assumed that the way the system states are distributed is Gaussian and, in some cases, linearization around some equilibrium system points can be made.

From a stochastic point of view in bioprocess, [18] proposed a state estimation formulation like a filtering problem; in which, the stochastic terms are introduced into the dynamics, measurements and initial conditions. The object of interest is the full state conditional probability law respect to some noise measurements of some components. In [18] the mathematical modelling of the uncertainty is presented, through a stochastic differential system. Then, the authors designed a set of asymptotic observers used with along with the observed components, to approach the probability law of the unobserved ones. In addition, the authors established the relevance of stochastic modelling into the biotechnology field that allows using the “know how” existent in biology and optimization; also they have succeeded showing the feasibility of using this kind of approach in numerical solutions.

This work proposes the use of nonlinear filtering techniques for state estimation in a *Bt* δ -endotoxins production process with the aim to implement a substrate feeding strategy in order to increment the bioprocess productivity. Specifically a nonlinear non Gaussian filter is used, through the called Particle Filters [19, 20]. This technique is based on Bayesian theory and Monte Carlo Sequential Methods. The methodology has the advantage of not using system linearization, as well as having a great potential for parallel implementation in systems that require more complexity and significant computational load. A discrete formulation of the problem is made; it means that the SDE's that characterises the system in continuous time, it is discretized to obtain the difference stochastic equations. At this point, it is possible to pose a valid formulation in probabilistic form under Bayesian Theory in which an optimal solution is obtained from the states probability density function (pdf).

The δ -endotoxin production of *Bacillus turingiensis* is a batch or fed batch process that cannot be linearized around a unique equilibrium point because its behaviour is transient in the complete batch duration time. Hence, a piece-wise linearization doesn't bring an appropriate solution either. Another interesting feature is the irreversibility present in this kind of process [21];[22]. For this reason, there is a strong dependency on the initial conditions. Random nature of reactions at molecular scale has been mentioned and studied by other authors [23]. At macroscopic scale, the modelled the total effect of these individual reactions on global concentrations can be modelled through an additive noise term with variance proportional to the reaction kinetics [24]. In this context, two of the process states: biomass and substrate (X,S) were modelled as Markov process, satisfying the Langevin chemical equation [23]. All features previously mentioned, convert this system into an attractive application for the use of Nonlinear Filtering tools in state observers design.

The interest of this work is into monitoring the states of a biological reactor, emphasising on the biomass concentration, with the aim to increment productivity through a proposed substrate feeding strategy and their posterior application to bioprocess control. The paper is organised as follows, in Section II the batch fermentation process is presented; Section III describes details of the algorithms for state estimation design and feeding strategy design. Section IV presents and discusses the obtained results and finally conclusions are exposed.

2. Materials and methods

2.1. Production of δ - endotoxins from *bacillus thuringiensis*.

Bacillus thuringiensis is a bacteria that produces spores, and during this stage it also produces one or more adjacent crystalline protein bodies known as δ -endotoxins, used widely on industrial applications as an ecological bioinsecticide. It has two stages on its life span: a first stage characterised by its vegetative growth, and a second stage named sporulation phase. When the vegetative growth finalises, the beginning of the sporulation phase is induced when the mean exhaustion point had been reached. Normally the sporulation is accompanied by the δ -endotoxin synthesis. After sporulation, the process is completed with the cellular wall rupture (cellular lysis), and the consequent liberation of spores and crystals to the culture medium. The plant is located in Medellín Colombia, and consists of a reactor with a nominal volume of 20 liters.

The cultures were developed with an effective volume of 11 liters of culture medium, and they were inoculated to 10% (v/v) with the microorganism *Bt*. The temperature was maintained at 30 C° by using a ON/OFF control; whereas the pH was controlled between 6.5-8.5 with a proportional controller. The foam formation was avoided by aggregating adequate antifoam. The air flow rate was set up at 22 [l/min] and the agitation speed at 400 rpm. The system pressure was set at 6 psig by a pressure valve controller. The dissolved oxygen percentage values, substrate concentration, pH and temperature were registered during the fermentation time by using a data acquisition system (Advantech®). The reagents concentration used for the pH control and foam formation was nitric acid (5N), potassium hydroxide (2N) and antifoam (33% v/v).

2.2. Process model

For this bioprocess a model based on first principles was proposed by [1] and [25]. This model is a modification to the Rivera model [26]. The work of [4] has improved the process model by adding the dissolved oxygen (DO) dynamics, due to its importance in the biomass estimation problem and the posterior process control. Batch model was developed and validated using experimental data from four batch cultures with different initial glucose concentrations (8, 21, 32 and 40 g/L). Then, the four sets of parameters for batch models were identified and validated by considering that each batch model covers a region of an intermittent fed batch culture operation (see Table 2). The batch operation can be originated from any of the initial conditions mentioned above; Table 2 presents the sets of parameters according to the initial condition employed.

The model equations are:

$$\begin{bmatrix} \frac{dX_v(t)}{dt} \\ \frac{dX_s(t)}{dt} \\ \frac{dS(t)}{dt} \\ \frac{dC_{DO}(t)}{dt} \end{bmatrix} = \begin{bmatrix} (\mu(t) - k_s(t) - k_e(t)) X_v(t) \\ k_s(t) X_v(t) \\ -\left(\frac{\mu(t)}{Y_{x/s}} + m_s\right) X_v(t) \\ K_3 \cdot F_{air-in} (C_{DO*} - C_{DO}(t)) - K_1 \frac{dX}{dt} - K_2 \cdot X(t) \end{bmatrix} \quad (1)$$

where X_v is the vegetative cell concentration, X_s the sporulated cells concentration, S the substrate concentration and C_{DO} the dissolved oxygen concentration. The following equations define the specific growth speed μ , the spore formation rate k_s and the death cell specific rate k_e . For model parameters see Table 2.

Table 1: Variables in the bioprocess model.

Symbol	Description
S	Limiting substrate concentration [g/L]
T_s	Sampling time [h]
X_s	Sporulated cells concentration [g/L]
X_v	Vegetative cells concentration [g/L]
C_{DO}	Dissolved oxygen concentration [g/L]
μ	Specific growth rate [h^{-1}]
μ_{max}	Maximum specific growth rate [h^{-1}]
m_s	Maintenance constant [g subs.g cells/h]
k_s	Kinetic constant representing the spore formation [h^{-1}]
k_e	Death cell specific rate [h^{-1}]
$Y_{X/S}$	Growth yield [g cells/g subs.]
K_s	Saturation constant [g/L]
K_1	Oxygen consumption constant by growth (dimensionless)
K_2	Oxygen consumption constant for maintenance [h^{-1}]
K_3	Ventilation constant [L^{-1}]
C_{DO}^*	O2 saturation concentration (DO concentration in equilibrium with the oxygen partial pressure of the gaseous phase) [g/L]
$F_{air,in}$	Air flow that enters to the bioreactor [L/h]

Table 2: Model Parameters.

Symbol	$S_{max} < 10$	$10 < S_{max} < 20$	$20 < S_{max} < 32$	$S_{max} > 32$
μ_{max}	0,8	0,7	0,65	0,58
$Y_{X/S}$	0,7	0,58	0,37	0,5
K_s	0,5	2	3	4

m_s	0,005	0,005	0,005	0,005
$k_{s\max}$	0,5	0,5	0,5	0,5
G_s	1	1	1	1
P_s	1	1	1	1
$k_{e\max}$	0,1	0,1	0,1	0,1
G_e	5	5	5	5
P_e	4	4,7	4,9	6
K_1	$9,75 \cdot 10^{-4}$	$4,502 \cdot 10^{-3}$	$3,795 \cdot 10^{-3}$	$1,597 \cdot 10^{-3}$
K_2	$1,589 \cdot 10^{-4}$	$0,046 \cdot 10^{-3}$	$0,729 \cdot 10^{-3}$	$0,561 \cdot 10^{-3}$
K_3	$4,636 \cdot 10^{-4}$	$0,337 \cdot 10^{-3}$	$2,114 \cdot 10^{-3}$	$1,045 \cdot 10^{-3}$

3. Estimators design

3.1 – Estimator basics: Non Linear Filtering and Particle Filtering

3.1.1 Nonlinear filtering Problem

The main problem in the state estimation of a dynamic system can be interpreted as the nonlinear optimal filtering problem. In non-linear systems, a finite dimension filter that solves the optimal filtering problem does not exist (optimal state estimator), in the sense of Kalman filter for the linear case.

The usual form to model the disturbances and the uncertainties is through the use of stochastic differential equations (SDE). Then, the equations that describe the states and outputs time evolution can be expressed by a system of stochastic differential equations or the Itô diffusion equations, in which the disturbances or uncertainties are expressed through diffusion components.

To represent the dynamic process, suppose a system with states composed by an n - dimensional vector x_t , and the system evolution is represented by a nonlinear vector differential equation called states equation, with the form of:

$$dx_t = f(x_t, t)dt + \sigma_1(x_t, t)dw_t \quad (5)$$

Where the first term represents the deterministic drift part and the second is the diffusion term that models the process disturbances.

The functions $f(\cdot, \cdot)$ and $\sigma_1(\cdot, \cdot)$ are called respectively, the drift and diffusion functions associated with the stochastic differential equation.

The observations measurements y_t taken from the output system y_t can be expressed in the same way as the states, by a non-linear stochastic differential equation:

$$dy_t = g(x_t, t)dt + \sigma_2(x_t, t)dv_t \quad (6)$$

In Eqs. 5-6, f , g , σ_1 and σ_2 are matrix or vector functions of convenient dimension and the vectors v_t and w_t are independent Wiener process.

Basically, the nonlinear filtering problem, consists in obtaining the “best possible estimation” \hat{x}_t , of the state vector x_t for all time t , from the available information of the observations y_t until time t^- , this is $\{y_s; 0 \leq s \leq t^-\}$. The optimal state estimation, will be defined by the conditional expectation of the state vector applied to the available information, this is:

$$\hat{x}_t = E(x_t | y_s; 0 \leq s \leq t^-) \quad (7)$$

Obviously, to calculate the previous conditional expectation and consequently, the value for the best possible estimation \hat{x}_t , it is necessary to know each time the conditional probability distribution function (pdf) of the states vector.

This implies that, the optimal solution to the non-linear filtering is a problem of infinite dimension, due to the fact that linear stochastic partial differential equation for the un-normalized density of a hidden state (Zakai equation) must be solved [27].

3.1.2 Particle Filtering

As previously addresses, estimation methods uses the model of the system in order to calculate the best approach for the unknown states the next sampling time, and also, uses a probability function to reach that goal. For nonlinear Non Gaussian systems, this probability function is not easy to obtain and the following idea illustrate the procedure.

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 $q(x)$ defined as a probability distribution function that depends on the observations until discrete time k . This importance function will serve as a tool for the estimation of the probability density function of the states and has the advantage that can be easily sampled from observations.

Following this idea, the posterior density understood as the conditional probability of the states to the measurements until time $k-1$, can be approached by the following expression:

$$p(x_{0:k} / y_{1:k-1}) \approx \sum_{i=1}^{N_s} W_k^i \delta(x_{0:k} - x_{0:k}^i) \quad (8)$$

Where N_s is number of samples, $\delta(\cdot)$ is the Dirac delta function applied to states and the weights W_k^i are chosen by importance sampling under considerations of:

- i) The probability density of the evaluation samples is proportional to the importance function $p(x) \propto q(x)$ and
- ii) The states can be sampled from the importance density function $x^i \square q(x)$.

The algorithm consists on the recursive propagation of the weights W_k^i and support points when each measurement (output measurement) can be obtained sequentially. Some improvements are oriented to avoid the consideration of particles with no information contribution to the algorithm and the criteria to evolution will minimize the observer's variance. The latter is a very important fact because of the algorithm should minimize the variance of the estimator in order to improve the performance of the particle filter.

The resampling step should be performed by the observation of the degeneracy function and its comparison with a basis value. This step eliminates the particles with low weights and concentrates on the particles with large weights, it implies to generate a new set of samples generated by resampling or rejection with the form of $\left[\left\{ x_k^{j*} \right\}_{i=1}^{N_s} \right]$.

In this way, an approximate discrete representation of the posterior density function can be obtained.

The resampling step can be perform under several variants: deterministic sampling, multinomial sampling, stratified sampling, residual sampling and systematic sampling among others. The resampling step reduces the effects of degeneration of particles but introduces some practical issues, limiting the opportunity of parallelism in to the algorithm because it combines all the set of particles and in addition, the particles with large weights are selected statistically in many times. This phenomenon carries on to a sample impoverishment, which is a lost in the diversity due to that the particles contain repeated points. This is a very important outcome when a process noise is low; and finally due to the particles patterns diversity reduction, any smoothing technique can be degenerate.

Another possible solution to the weights degeneracy issue is to start the algorithm with an adequate importance function $q(x)$ or else, evaluating its behavior into the algorithm.

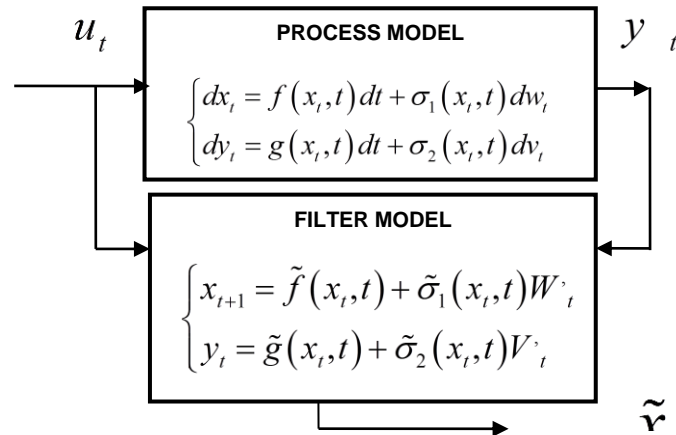
To make possible the use of this solution to the problem, some options arise like the ability to sample from the priori function $p(x_n / x_{n-1}^i, y_n)$; or to evaluate the integral to the weights calculation in each new state. The optimal importance function is not always easy to find and for this 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 sampling from $p(x_n / x_{n-1}^i, y_n)$ is possible in that way. This is the case of Jump Markov Linear System [28].

3.2- Filter Model

From the extended model equations (1-4) by [29], with dissolved oxygen dynamic incorporated, two filtering configurations were performed. First of all, it is assumed a stochastic component added to the deterministic models and the state space is now composed by stochastic differential equations.

For the test of the estimators, two experiments were carried out: Figure 1 shows the first filter configuration, the discrete model of the process it is considered as the filter model, with uncertainties is characterized as measurement and state noise. Also, the continuous model with

different uncertainties component is considered as the real plant. This approach was performed as a test previous to the second filter configuration.



The second configuration of the filter (Figure 2), replaces the continuous model used as real plant in the first approach and involves the data obtained from the real process. This way, the real data plays the role of real plant and the filter is improved by the use of a model with better uncertainties modeling. This is very important because by using real data, the problem solved is closer to the reality of estimation on line.

In both cases the measurable states considered were substrate concentration (S) and dissolved oxygen concentration (C_{DO}). It means substrate and dissolved oxygen data were used as information source for estimation of total biomass concentration during the fermentative process.

Discrete approach to continuous differential equations that modeled the batch fermentation behavior was made through Euler numerical method. This kind of approach allows selecting the appropriate sample time in order to accurately reproduce the process dynamics.

Fig.1. Estimation Scheme: First filter configuration

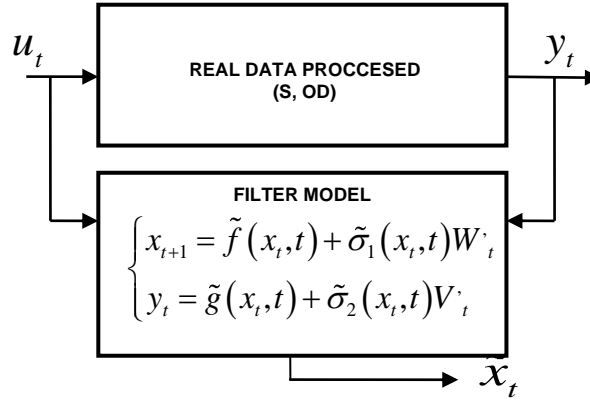


Fig.2. Estimation Scheme: Second filter configuration

In the figures 1 and 2, the input $u(t)$ and output $y(t)$ are signals corresponding to “process model block” in the first filter configuration or “real measurements of the plant” in the second configuration. These signals feed the “filter model” in both cases that make the estimation of the total cells and gives the information of total Biomass Concentration. These estimations are compared with the process model (considered the real process). Some tests at different sample times were performed, from 0.2 hours to 1 hour for batch fermentation; tests are done only during the first 20 hours.

3.2.1 First Filter Configuration

Due to the unavoidable errors of the mathematical models and the existence of unpredictable internal or external perturbations, uncertainty models were added to the states and the measurable outputs. This uncertainty models have the features to be an independent and identically distributed (iid) *diffusion terms*. In Fig. 1 was showed the estimator block diagram, for the first case of simulation presented.

Consequently, in the previously described model, equations (1-4), uncertainty and disturbances discrete time models were added under consideration of modeling uncertainties and measurements with noise of diverse nature. This way, the model presented became a stochastic differential equations model.

The mathematical treatment of SDE's has similarities, but also slight differences, to the usual theory of deterministic differential equations. One needs to be aware of these similarities and differences when considering numerical methods to solve them. Finding the exact solution of a (stochastic) nonlinear differential equation will usually be impossible; thus, only approximate sampled-data models will be obtained.

The interest of this part of the work, relays in the use of SDE's to model continuous-time stochastic systems (also called noise models), and from those models, to obtain sampled data which is

accurate in a well defined sense. In this particular case, authors are not interested into finding an exact solution of SDE, but are interested in the samples that can be obtained from its simulation behavior as a model. This procedure is done with the aim of calculating an approach to the probability density functions of the states, from Monte Carlo techniques for state estimation.

To reach the sampled data model, it was necessary to approach the model defined by Eqs. 1-4 by means of a numerical method compatible with stochastic approach. Due to this part it is a simple model, the discrete deterministic model will appear in section 3.2.2.

The discrete model used in the diffusion terms of both filter and process model in simulation was $\sigma_1(x_n, n) \in \mathfrak{R}^5$ from eq. (5) as a multidimensional Gaussian Noise, dependent on a uniformly distributed intern variable $z(n)$, updated each sample time particle by particle. The uncertainty model is a stochastic process defined by the following equations ([30]):

$$z(n) \sqsupseteq p(z(n)|z(n-1)), z(n) \sqsupseteq U(a, b) \quad (9)$$

$$\sigma_1(x_n, n) \sqsupseteq p(x(n)|z(n|n-1), x(n-1)) \quad (10)$$

$$\sigma_1(x_n, n) = p(x(n)|z(n|n-1), x(n-1))\Delta W_n \quad (11)$$

In Eqs. (9-11) the symbol \sqsupseteq means “sampled from” the conditional probability density function p . The function $\sigma_1(x_n, n) \in \mathfrak{R}^5$ corresponds to the second right term of eq. (5). Equation (12) illustrates the composition of the vectorial form of (11) that contains independent noise models for each state variable.

$$\sigma_1(x_n, n) = [\sigma_{11}(x_n, n) \quad \sigma_{12}(x_n, n) \quad \sigma_{13}(x_n, n) \quad \sigma_{14}(x_n, n) \quad \sigma_{15}(x_n, n)]^T. \quad (12)$$

Also, $z(n) \in \mathfrak{R}^5$ and $\Delta W_n \in \mathfrak{R}^5$ is a white noise generated with different mean and variance vectors. The values for uniformly distributed variable a and b are chosen by the designer. As a measurements model, the uncertainty term was selected as $\sigma_2(x, t) = 1$ combined with different Gaussian noise samples in the diffusion components. A simple white Gaussian Noise Δv_i was used, consistent to the Brownian motion discrete approximation, generated with statistics that vary, depending on the test performed.

3.2.2 Second Filter Configuration

The second proposed configuration consists in the use of a discretized deterministic model as an information model into the filter, and augmented or completed by a model that represents the uncertainty $\sigma_1(x_n, n)$ between the data set and the phenomenological results.

The uncertainty model tries to diminish the gap between model predictions and the reality of the process, with the aim of improving the biomass estimation. To reach the proposed objective, an adequate dynamical knowledge of the studied system is required. In this case, three sources bring reliable information: phenomenology, empirical input/output data and qualitative and quantitative

heuristic knowledge of the system. All available information, is combined, processed and presented in many ways, and through different models (phenomenological or white box models, ARMAX time series, fuzzy or neural networks, and so on).

Models themselves cannot represent completely the mechanisms that rule the dynamic behavior of systems and as a result their predictions are not always consistent with the real systems. In the works of ([31]) and ([32]) is suggested to pay special attention to the uncertainties modeling, when a physical system is modeled.

The purpose of representing the uncertainty as a model is to find answers to the following questions: What influences the uncertainty dimension? How it can be reduced or compensated? Ideal goals should compensate the uncertainty through an adequate model, incorporating it to the filtering scheme proposed.

The use of sample data models opens the discussion about the relationship between the discrete time samples and the original model described in continuous time obtained from the phenomenology modeled dynamics and differential equations. As a first approach, it is very common to sample frequently and then, replace the derivatives of the continuous time model by divided differences ([33]). Any sampled-data model for a nonlinear system will, in general, be an approximate description of the combination of two elements: the continuous-time system itself, together with the sample and hold devices.

An exact discrete-time description of such a hybrid nonlinear system is, in most cases, not known or impossible to compute ([34]). Thus, one needs to be clear about the potential accuracy achieved by any model. In fact, the accuracy of the approximate sampled-data plant model has been proven to be a key issue in the context of control design. In the specific case that a controller designed to stabilize a system (based on an approximate model) fails to stabilize the exact discrete-time model, no matter how small the sampling period has been chosen ([35]).

We supposed that the uncertainty can be modeled through the use of the first and second delay of the output and the first delayed error of the variable. This structure was used as an uncertainty model to complete the discrete version of the bioprocess model presented in equations (1-4) to build a reliable discrete model that describes properly the dynamics of the fermentation with the form (5) and (6) and also, completed by a noise model with the structure of equations (9-11). For this purpose we propose a set of Autoregressive Integrated Moving Average (ARIMA) models to model each fermentation uncertainties between phenomenological and real data A, B, C. An ARIMA model structure can be seen in equation 13.

$$\hat{Y}(t) = m + Y(t-1) + f(Y(t-1) - Y(t-2)) - qe(t-1) \quad (13)$$

In which m , f and q are parameters for estimation of the variable Y in terms of endogenous and exogenous variables.

As it is desired, a general uncertainties model for fermentation was obtained from the ARIMA models. In this way the uncertainties modeling was complete with ARIMA and noise models calculated with Matlab® to describe the three fermentation data used, named as Fermentation A, Fermentation B and Fermentation C. The complete model was applied into the filter to perform

estimations of biomass, being fed by the “observable” real data of each fermentation. Subsequently, the results were compared with the real data of biomass. Of these models, the best ones are presented in this paper to reach a generalized uncertainty model.

The model used for diffusion terms into the filter (for explanation see section 3.2.2) was a $\sigma_1(x_i, t)$ component, built as an ARIMA model (Auto Regressive Integrated Moving Average), identify the difference between real data and the output model ensuring the uncertainty modeling.

ARIMA models are, in theory, the most general class of models for forecasting a time series, which can be stationarized by transformations such as differencing and logging. In fact, the easiest way to think of ARIMA models is as fine-tuned versions of random-walk and random-trend models: the fine-tuning consists on adding lags of the differenced series (AR) and/or lags of the forecast errors (MA), to the prediction equation, as needed to remove any last traces of autocorrelation from the forecast errors. ARIMA models that include only AR terms are special cases of linear regression models; hence they can be fitted by ordinary least squares. AR forecasts are a linear function of the coefficients as well as a linear function of past data. In principle, least-squares estimates of AR coefficients can be exactly calculated from autocorrelations in a single "iteration".

In practice, one can incorporate an AR model in the Multiple Regression procedure on lags of itself (but you would get slightly different results from the ARIMA procedure) ARIMA models which include MA terms are similar to regression models, but can't be fitted by ordinary least squares: forecasts are a linear function of past data, but they are non-linear functions of coefficients--e.g., an ARIMA model without a constant is an exponentially weighted moving average in which the forecasts are a non-linear function of the MA parameter ("theta").

3.3 Proposed substrate feeding Strategy

As explained in section II, batch growth of Bt has two stages. In the first stage, nutrients are consumed and vegetative growth appears (each vegetative cell divides itself to form two new vegetative cells). When nutrients are consumed, the second stage begins and sporulation is initiated. In this stage each cell begins to form a spore and this cell cannot continue the cell division process. The process is completed with the rupture of the cell wall and the spore–crystal complex is released into the medium[1] and [25].

In this work the author proposes a substrate feeding strategy with the aim to optimize the production in the bioprocess. The strategy feed consist in to generate two pulses of fresh substrate medium and add to the bioreactor. Each pulse was added before the growth phase had finished. Something similar was used for the fed-batch bioprocess modeling, but initial concentrations used; the time when each pulse is applied and the final application are different. In this strategy the volumes added are controlled so that is does not exceed the permissible by the bioreactor. It is working with the maximum allowable concentrations of substrate to achieve a good cell production but without substrate inhibition occurs during fermentation.

The first pulse volume was 2 L and the second one was 5 L. A tangential filtration system containing 2 polyvinylidene fluoride membranes with pore size of 0.45 μm and a total filtration area of 1.0 m^2 (Millipore Co) can be used to eliminate the exhausted culture medium (permeate). After

each pulse addition the process continued as a batch culture. The point in the fermentation with maximum nutrients concentration is the optimization objective. The maximum glucose concentration in the medium (S_{pmax}) at any time of the process was used as the switching criteria among the estimated batch parameter sets (See Table 2). The Initial and operational conditions used to simulate the fed-batch operation of *Bacillus thuringiensis* subsp. *kurstaki* are exposed in Table 3.

Table 3: Initial and operational conditions used to simulate the fed-batch operation of *Bacillus thuringiensis* subsp. *kurstaki*.

Description value	
Initial glucose concentration (g/L)	25
Initial vegetative cell concentration (g/L)	0.645
Feed flow by each pulse (L/h)	60
Culture medium inlet flow to tangential filter (provided by pump) (L/h)	423.5
Permeate flow before the first pulse	60
Permeate flow before the second pulse (L/h)	27.27
Glucose concentration fed in the first pulse (g/L)	47
Glucose concentration fed in the second pulse (g/L)	40
Duration of filtration process before the first pulse (h)	0.2
Duration of filtration process before the second pulse (h)	0.6
Time at which the first pulse was added	5
Time at which the second pulse was added	6.9

4. Results analysis and discussion

4.1 – Simulation Results:

As follows, the results from the approach obtained through particle filtering technique in simulation will be presented, and will be compared with real biomass data. In the first performed configuration, a complete model of bioreactor was used, in which full states are included; they are biomass concentration (vegetative and sporulated cells concentrations), substrate concentration and dissolved oxygen concentration, $x = [X_v, X_s, S, C_{DO}]^T$. To the previously mentioned states, a simple diffusion component was added (Brownian movement) and the states considered as measurable outputs were substrate and dissolved Oxygen concentrations is $y = [S, C_{DO}]^T$.

For initial configuration, sample times from 1 to 0.01 hours were used.

A set of 10 to 1000 particles and variations of random components (statistics features) were tested. The set of tests was used to measure the filter effectiveness compared to the application computational cost and to observe the efficiency of the resampling schemes used in this configuration.

The test performed with the first filter configuration (previously described in section 3.2.1) are summarized in table 1, using the discretized model of [29]; the results presented in the table highlight the influence of particle numbers in the augmenting or diminishing of the approach error over interest variables.

Table 4 cites the estimation errors for some cases where reached estimation is accurate. In simulation (process model-filter model) different sample times were used to evaluate not only the dynamical filter behavior, but also the accuracy of the approach. The parameter to measure the estimator effectiveness is defined as:

$$error = \sup(\max_{resampling} \|x - \tilde{x}\|) \quad (14)$$

Table 4. Summary of state estimation error on initial tests, different diffusion features.

Estimation Errors				
Sample time	1 hour		0.5 hour	
Particle Number	<i>100</i>	<i>500</i>	<i>100</i>	<i>500</i>
Vegetative Cells (Xv)	0.0011	0.0005	0.0007	0.0024
Sporulate Cells (Xs)	0.0007	0.0008	0.0003	0.0011

From the information on Table 4, it can be observed that the estimation quality not only depends on resampling schemes involved, no matter how many particle numbers were used, but also, involves the measurements sample time. With the aim to illustrate the test performed, the simulation results will be shown for total biomass concentration (vegetative and sporulated cells) at sample time of 0.01 with deterministic resampling and 1 hour with residual resampling, for a 500 particles configuration. An estimation error (for the total biomass concentration) minor than 1g/L is more than acceptable for this process and then we can ensure that the results are satisfactory and the algorithm converges for this kind of estimation. It is important to remark that the resampling schemes have influence over simulation time, but in regard to the quality of the results obtained the variations are not relevant. The results of tests performed with short sample times of 0.2 hours, for a deterministic resampling scheme, give the shortest execution time of filtering algorithm, for high noise power configuration. Estimations reached are adequate and satisfactory according to the estimation error magnitude, in comparison with larger sample times.

4.2 – Experimental Results and Inputs Description:

It was necessary to use the experimental information of substrate and dissolved oxygen concentrations to use a particle filter as state estimator. Two fermentations were used: “A” and “B” of 18 hours duration. The sample time was fixed in $T_s = 1/10 \text{hours}$ by a Fourier frequency analysis. The dissolved oxygen data was acquired with a sample time of 0.01 hours. The previously selected sample time, as the most representative of this kind of dynamical system, was of 0.1 hours (6 minutes); because even the oxygen data can be obtained at less sample time, the dynamic behavior into this interval doesn't have important changes.

The available experimental data is variable measurements taken from the mentioned fermentations; namely: dissolved oxygen percentage (DO), substrate concentration (s), and total biomass concentration (X). The experimental data is sampled at different frequencies: 10 per hour for DO and Glucose S and 1 sample per hour for the biomass concentration (quantified through dry weight method). The experimental data of dissolved oxygen concentration collected from the process using a data acquisition system is immersed in high frequency noise. In order to remove these undesirable noise frequencies the original raw signal is filtered with a low-pass filter, as it is explained in the work of [29]

For the filter dynamical model the discrete approach of the process model was used but improving with uncertainty models, i.e., the model used internally in the filter has been corrected through an uncertainty modelling, looking for compensation between real data and the model used. In figure 2 the estimation scheme used for real data comparison is shown.

The inputs used in the estimation scheme are present in Figs. 3-6, i.e. the output completed with the ARIMA model previously described, and for comparison the output model from the original model for the fermentations A and B. The uncertainty measure model for sensors considered was the same that in the simulation test.

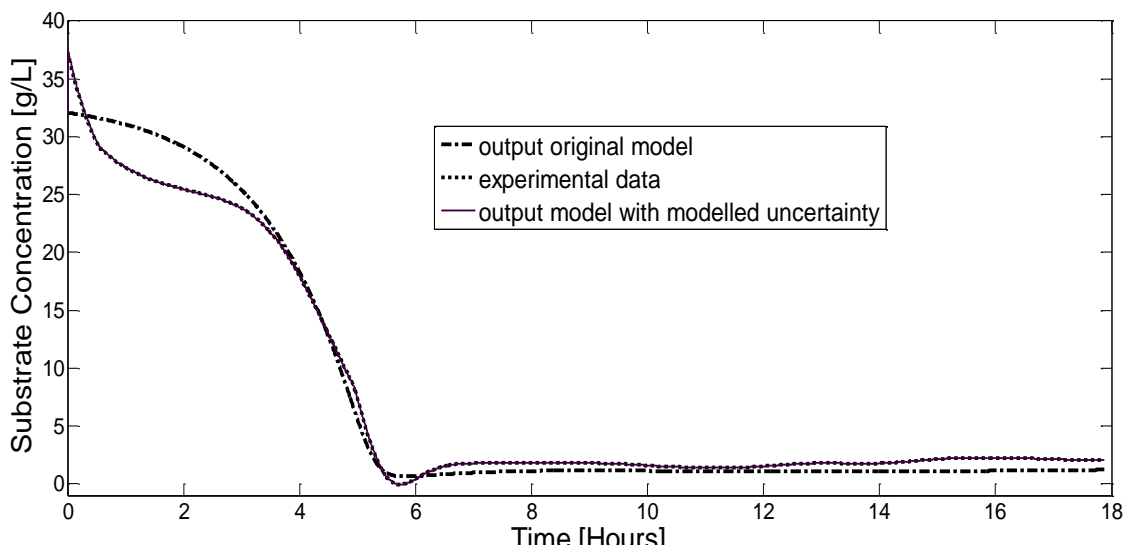


Fig. 3. Input for the Biomass Estimator (Substrate concentration for fermentation A): The dotted line represents the experimental data, the dashed line the output original model and the solid line represents the input to the particle filter (the discrete model plus ARIMA to model the uncertainties).

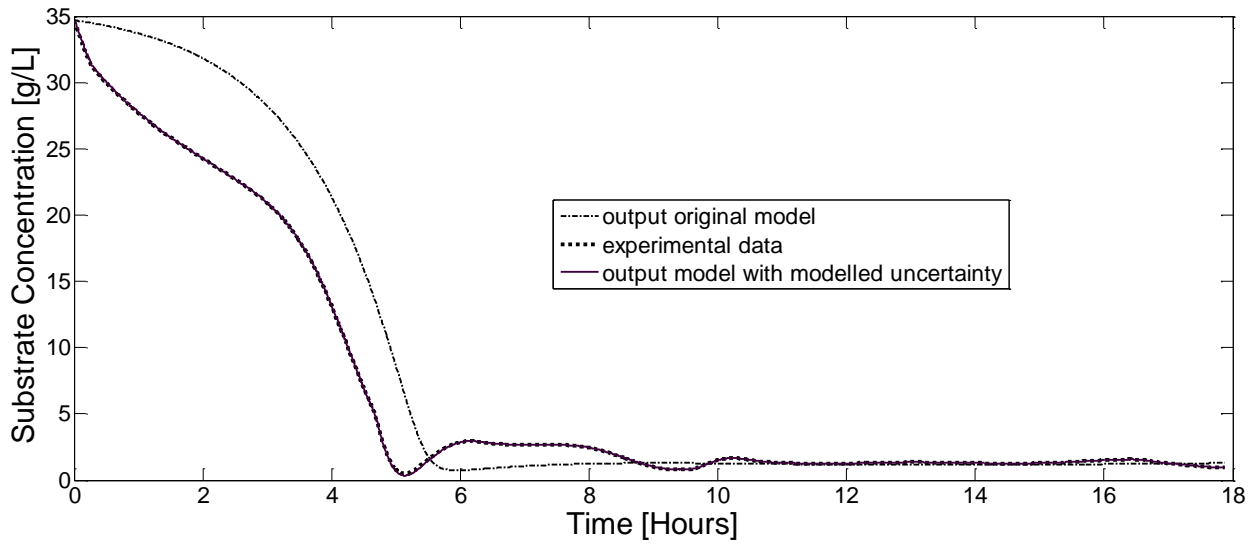


Fig. 4. Input for the Biomass Estimator (Substrate concentration for fermentation B): The dotted line is the experimental data, the dashed line represents the output original model and the solid line represents the input to the particle filter (the discrete model plus ARIMA to model the uncertainties).

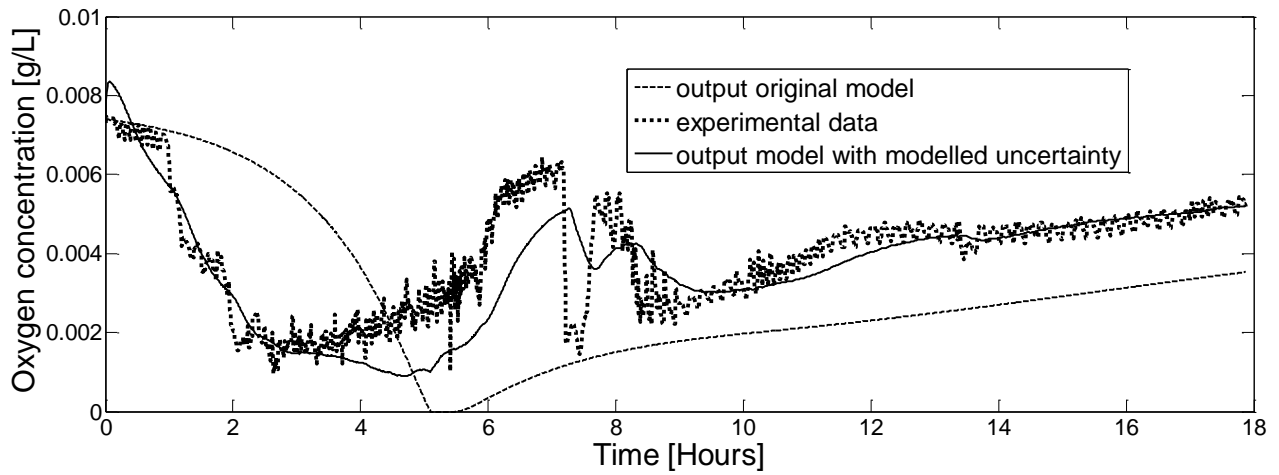


Fig. 5. Input for the Biomass Estimator (Dissolved Oxygen concentration for fermentation A): The dotted line represents the experimental data, the dashed line represents the output original model and the solid line represent the input to the particle filter (the discrete model plus ARIMA to model the uncertainties).

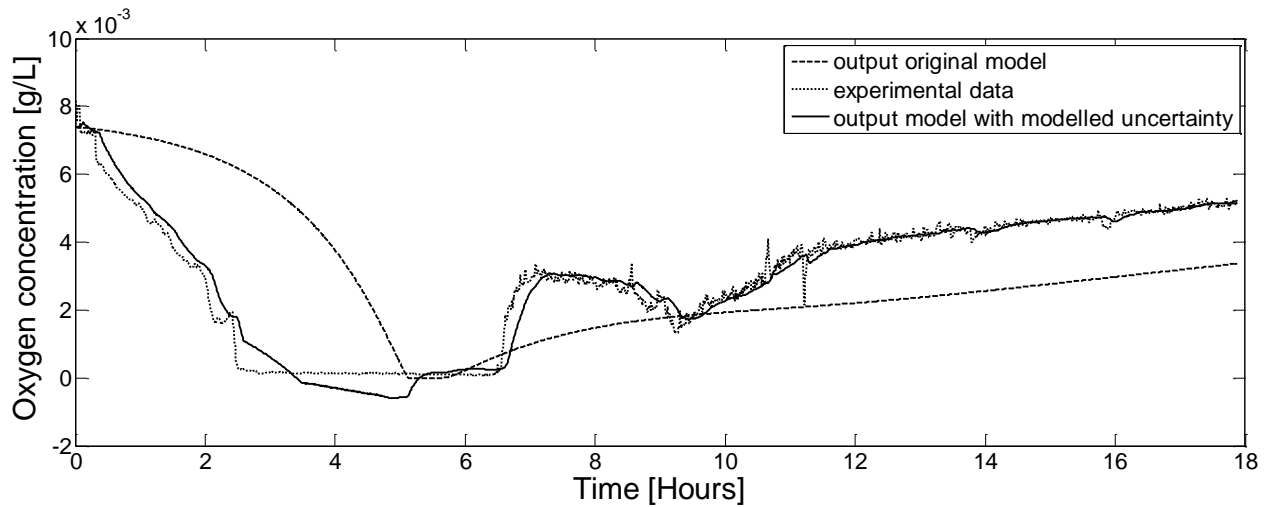


Fig. 6. Input for the Biomass Estimator (Dissolved Oxygen concentration for fermentation B): The dotted line represents the experimental data, the dashed line represents the output original model and the solid line represent the input to the particle filter (the discrete model plus ARIMA to model the uncertainties).

The biomass estimation results with a SIR filter with different resampling schemes are presented in Fig 7 and 8. The structure used condenses the Basic Model developed by [4] and the uncertainty model developed and proposed in this work (eqs 9-11). To validate those results, experimental data of biomass concentration was used, but due to the different sample frequency (1 per Hour), the biomass concentration's experimental data was completed using Bayesian Gaussian Regression (see the work of [12]).

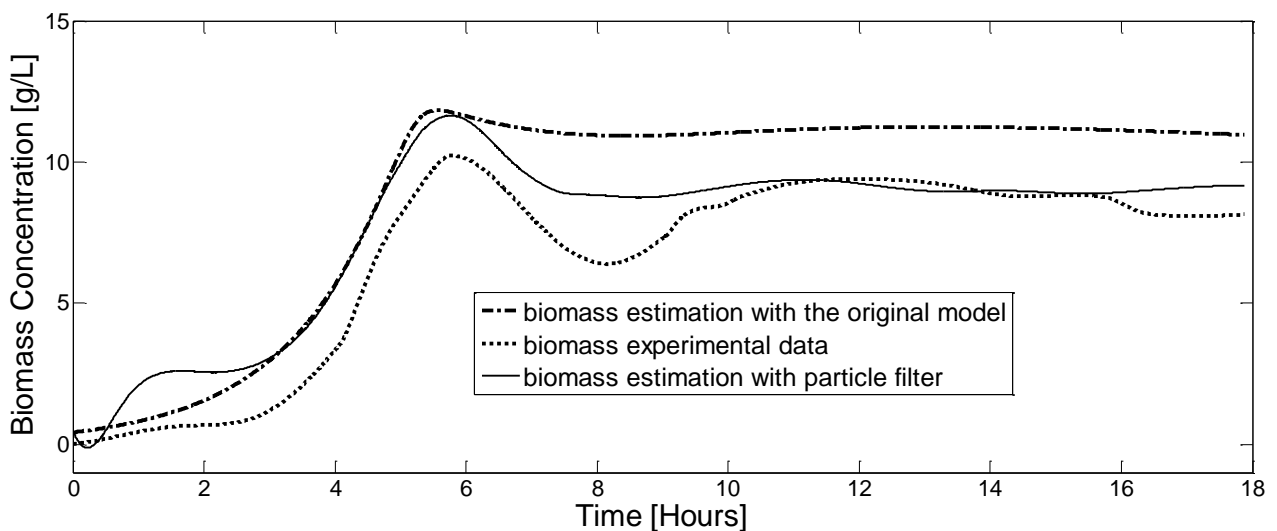


Fig. 7. Biomass estimation for batch process (Fermentation A) through the particle filter with the use of the discrete model complemented with uncertainties modeled through ARIMA model.

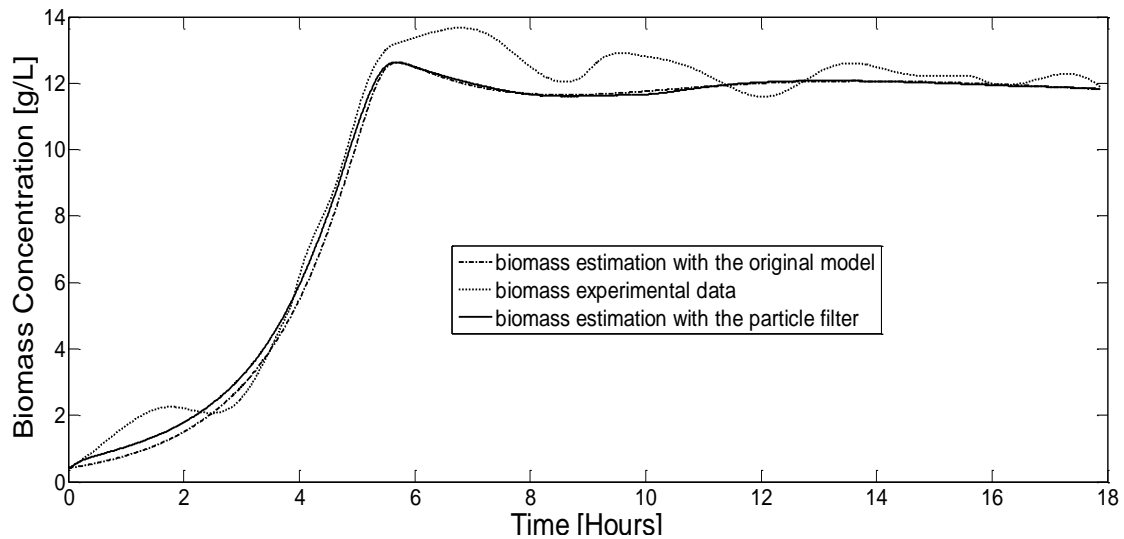


Fig. 8. Biomass estimation for batch process (Fermentation B) through the particle filter with the use of the discrete model complemented with uncertainties modeled through ARIMA model.

From the figures (7 and 8) can be observed the acceptable performance of the biomass estimator for the fermentations A and B. The results are satisfactory, it reproduces the behavior of batch fermentation even if the initial conditions are the same or not. These results are an improvement for previous estimators [11, 36] and can be used in closed loop with advanced control strategies [29] [37] for the optimal control of fermentation process.

The need to develop a methodology for the right implementation of particle filters in non-linear systems, rely on the particularities that these systems have; this features makes really difficult the task of characterizing them and makes difficult to find a correct solution path to these particular cases. Inside the methodology of particle filtering, by using the importance sampling, it is imperative to consider that, to reach the goal, it is necessary to propose an adequate *importance function*, and the filter designer should select it depending on each specific case. This causes that when doing the approach to the filter densities (objective densities functions) it will be extremely close to the good or bad choice of the importance function by the designer. An important remark is that the filter follows properly the model, and as it approaches to the real problem, the performance to online implementation was tested. To apply the SMC Particle Filtering methodology, the assumption of a sampled data model for the SDE's was necessary; this way the set of equations are posed as new and improved models that include 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, [38] [39] [40], [41].

The results of the feed substrate strategy integrated with the particle filter proposed shown in Figure 9. It can be seen that adding the proposed substrate pulses, the vegetative phase of the process becomes longer resulting in an increased productivity.

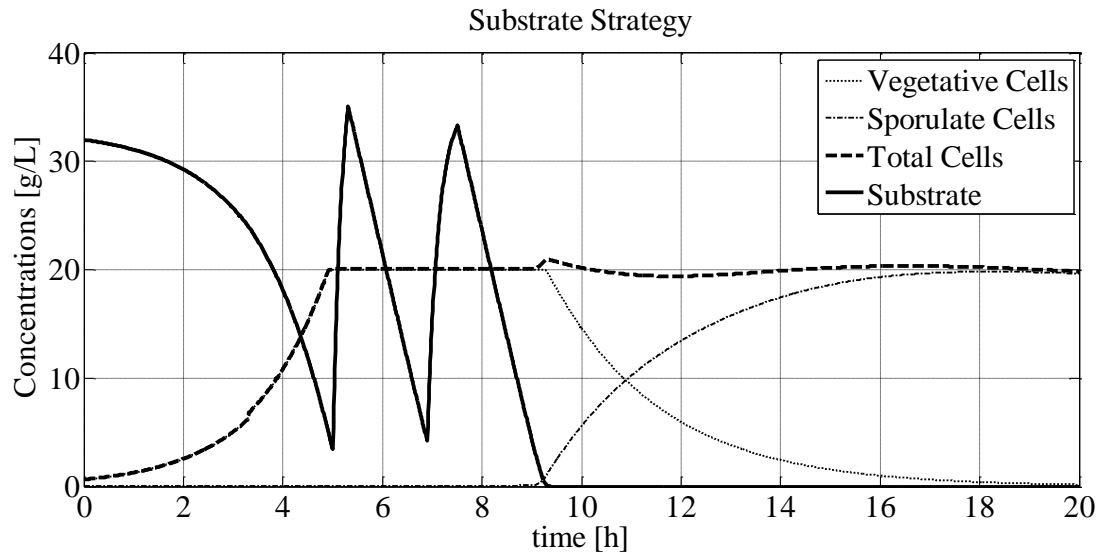


Fig. 9. Substrate feeding Strategy.

5. Conclusions

This work presented a state estimator based in nonlinear filtering techniques and its integration with a substrate feeding strategy. The application of the particle filters developed as state estimator can be used in nonlinear non Gaussian dynamical systems. The technique and methodology was illustrated in the case of a batch bio-reactor for production of δ -endotoxins of *Bacillus thuringiensis* (*Bt*). In this work the performance results for this methodology were presented first under a simulation setting for different conditions and then, contrasted with real experimental data.

Uncertainty modeling between the dynamical system and a basis model developed by [29] was also performed. The complete uncertainty model was developed through ARIMA models and diffusion components and this way, the existent model was completed, looking for the implementation of this particular estimator.

Results were compared with real data from two fermentations with initial conditions around the same operation point. It is important to remark the relevance of the development of uncertainty models for the use of Bayesian techniques into information extraction of dynamical nonlinear non Gaussian systems. The application of the particle filtering as state estimator is acceptable and feasible for its implementation to solve the problem of on line biomass estimation in a batch process, due to its reliability and admissible computational cost. The substrate feeding strategy transforms the batch process into a fed batch process with the consequent increment (duplication) of the bioprocess productivity. The feeding strategy resulted then in an increased productivity during the fed-batch cultivation.

Acknowledgments

The authors wish to thank the following persons and organizations who contributed to the completion of this work, and without whom, some of the information in this paper could not have been obtained. Particularly, Instituto de Automática - Universidad Nacional de San Juan,

Universidad EAFIT, Mathematical Sciences Department, School of Sciences, and CONICET Argentina.

References

1. Atehortúa P, Álvarez H, Orduz S. Modeling of growth and sporulation of *Bacillus thuringiensis* in an intermittent fed batch culture with total cell retention. *Bioprocess and Biosystems Engineering*. 2007, 30, 447-56.
2. Chekhova E, Barton P, Gorak A. Optimal operation processes of discrete-continuous biochemical processes. *Computers & Chemical Engineering*. 2000, 24, 1167-73.
3. Grøn S, Bierdermann K, Emborg C. Production of proteinase A by *Saccharomyces cerevisiae* in a cell-recycling fermentation system: experiments and computer simulations. *Appl Microbiol Biotechnol*. 1996, 44, 724-30.
4. Amicarelli A, di Sciascio F, Toibero J, Alvarez H. Including dissolved oxygen dynamics into the Bt δ -endotoxins production process model and its application to process control. *Brazilian Journal of Chemical Engineering*. 2010, 27, 41-62.
5. Bastin G, Dochain D. *On-line estimation and adaptive control of bioreactors*: Elsevier, Amsterdam, 1990 (ISBN 0-444-88430-0). xiv+ 379 pp. Price US \$146.25/Dfl. 285.00. *Analytica Chimica Acta*. 1991, 243, 324.
6. Ljung L. *System identification - Theory for the User*. Prentice-Hall 1999.
7. Russell SA, Robertson DG, Lee JH, Ogunnaike BA. Model-based quality monitoring of batch and semi-batch processes. *Journal of process control*. 2000, 10, 317-32.
8. Dondo R, Marqués D. Mass and energy balances as state-space models for aerobic batch fermentations. *Latin American applied research*. 2002, 32, 195-204.
9. Leal RR, Butler P, Lane P, Payne PA. Data fusion and artificial neural networks for biomass estimation. *Science, Measurement and Technology, IEE Proceedings -*. 1997, 144, 69-72.
10. Silveira MM, Molina MAB. Indirect estimation of *Bacillus thuringiensis* var. *israelensis* biomass concentration using oxygen balance data. *Brazilian Journal of Chemical Engineering*. 2005, 22, 495-500.
11. Quintero OL, Amicarelli AA, Di Sciascio F, Scaglia G. State estimation in alcoholic continuous fermentation of *Zymomonas mobilis* using recursive bayesian filtering: A simulation approach. *BioResources*. 2008, 3, 316-34.
12. di Sciascio F, Amicarelli AN. Biomass estimation in batch biotechnological processes by Bayesian Gaussian process regression. *Computers & Chemical Engineering*. 2008, 32, 3264-73.
13. Amicarelli A, Quintero O, di Sciascio F. Behavior comparison for biomass observers in batch processes. *Asia-Pacific Journal of Chemical Engineering*. 2014, 9, 81-92.
14. Dochain D. State and parameter estimation in chemical and biochemical processes: a tutorial. *Journal of process control*. 2003, 13, 801-18.
15. Mohd Ali J, Ha Hoang N, Hussain MA, Dochain D. Review and classification of recent observers applied in chemical process systems. *Computers & Chemical Engineering*. 2015, 76, 27-41.
16. Leal Ascencio RR, Reynaga F, Herrera E, Gschaedler A. Artificial neural networks as a biomass virtual sensor for a batch process. *Intelligent Control, 2001 (ISIC '01) Proceedings of the 2001 IEEE International Symposium on* 2001. p. 246-51.
17. Dorsey AW, Lee JH. Building inferential prediction models of batch processes using subspace identification. *Journal of process control*. 2003, 13, 397-406.
18. Rossi V, Vila JP. Bayesian multioutput feedforward neural networks comparison: a conjugate prior approach. *Neural Networks, IEEE Transactions on*. 2006, 17, 35-47.

19. Doucet A, Gordon NJ, Krishnamurthy V. Particle filters for state estimation of jump Markov linear systems. *Signal Processing, IEEE Transactions on*. 2001, 49, 613-24.
20. Doucet A, De Freitas N, Gordon N. An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo methods in practice*. Springer 2001. p. 3-14.
21. Sontag ED. *Mathematical control theory: deterministic finite dimensional systems* (2nd ed.). Springer-Verlag New York, Inc. 1998.
22. Haddad WM, Chellaboina V, Nersisov SG. Time-Reversal Symmetry, Poincaré Recurrence, Irreversibility, and the Entropic Arrow of Time: From Mechanics to System Thermodynamics. *Decision and Control, 2005 and 2005 European Control Conference CDC-ECC '05 44th IEEE Conference on* 2005. p. 5995-6002.
23. Gillespie DT. The Chemical Langevin and Fokker–Planck Equations for the Reversible Isomerization Reaction†. *The Journal of Physical Chemistry A*. 2002, 106, 5063-71.
24. Kurtz TG. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*. 1978, 6, 223-40.
25. Atehortúa P, Álvarez H, Orduz S. Comments on: “A Sporulation Kinetic Model for Batch Growth of *B. thuringiensis*”. *The Canadian Journal of Chemical Engineering*. 2006, 84, 386-8.
26. Rivera D, Margaritis A, Lasa Hd. A sporulation kinetic model for batch growth of *B. thuringiensis*. *Canadian Journal of Chemical Engineering*. 1999, 77, 903-10.
27. Lange C. Kallianpur, G., *Stochastic Filtering Theory, Applications of Mathematics* 13, Berlin-Heidelberg-New York, Springer-Verlag 1980, XVI, 316 S., DM 59,50, US \$ 35.10. ISBN 3-540-90445-X. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*. 1981, 61, 602-.
28. Andrieu C, Doucet A. Particle filtering for partially observed Gaussian state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002, 64, 827-36.
29. Amicarelli A, di Sciascio F, Toibero JM, Alvarez H. Including dissolved oxygen dynamics into the *Bt* δ -endotoxins production process model and its application to process control. *Brazilian Journal of Chemical Engineering*. 2010, 27, 41-62.
30. Quintero O, Amicarelli A, di Sciascio F. Biomass Estimation in Batch Process through Particle Filters.
31. Guergachi AA. Using statistical learning theory for modeling the uncertainty in business and engineering systems: a qualitative introduction. *Systems, Man, and Cybernetics, 2001 IEEE International Conference on* 2001. p. 423-8 vol.1.
32. Guergachi AA, Patry GG. Constructing a model hierarchy with background knowledge for structural risk minimization: application to biological treatment of wastewater. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*. 2006, 36, 373-83.
33. Yuz JI, Goodwin GC. On sampled-data models for nonlinear systems. *Automatic Control, IEEE Transactions on*. 2005, 50, 1477-89.
34. Nešić D, Teel AR, Kokotović PV. Sufficient conditions for stabilization of sampled-data nonlinear systems via discrete-time approximations. *Systems & Control Letters*. 1999, 38, 259-70.
35. Nes, x030C, ic, x, D., Teel AR. A framework for stabilization of nonlinear sampled-data systems based on their approximate discrete-time models. *Automatic Control, IEEE Transactions on*. 2004, 49, 1103-22.
36. Quintero O, Nieto J, Amicarelli A, Scaglia G, Luna T, di Sciascio F. *Control Engineering Perspective Of Fermentation Process From Zymomonas mobilis: Modeling, State Estimation And Control*. 2009.
37. Quintero OL, Amicarelli AA, Scaglia G, di Sciascio F. Control based on numerical methods and recursive Bayesian estimation in a continuous alcoholic fermentation process. *BioResources*. 2009, 4, 1372-95.

38. Briers M, Maskell SR, Wright R. A rao-blackwellised unscented Kalman filter. *Info Fusion*. 2003, 8-11.
39. Briers M, Doucet A, Singh SS. Sequential auxiliary particle belief propagation. *Information Fusion, 2005 8th International Conference on* 2005. p. 8 pp.
40. Briers M. Improved Monte Carlo methods for state-space models: University of Cambridge, 2007.
41. Klaas M, Briers M, De Freitas N, Doucet A, Maskell S, Lang D. Fast particle smoothing: If I had a million particles. *Proceedings of the 23rd international conference on Machine learning*. ACM2006. p. 481-8.