4

On-line Biomass Estimation in a Batch Bio-technological Process: Bacillus Thuringiensis δ - Endotoxins Production.

Adriana Amicarelli, Fernando di Sciascio, Olga Quintero and Oscar Ortiz
Universidad Nacional de San Juan-Instituto de Automática (INAUT)
Argentina

1. Introduction

Biomass concentration in a biotechnological process is one of the states that characterize a bioprocess. Moreover, it is generally the main direct or indirectly desired product. It is well known that the biomass concentration is not normally measured online because this measurement is not possible or this is economically unprofitable. Therefore, for control purposes it is necessary to replace the unavailable biomass concentration measurements with reliable and robust online estimations. To this aim, several states observers can be found in the literature. A review of commonly used techniques can be found in (Bastin & Dochain, 1990; Dochain, 2003) and references therein. Observers can be coarsely divided into two broad classes: first principles or phenomenological estimators and empirical estimators. The phenomenological estimators can be also subdivided into classical observers and asymptotic observers. Classical observers include extended Kalman filter (EKF), extended Luenberger observer, high gain observer, nonlinear observers, and full horizon observer. In this class of estimators, a detailed knowledge of the reaction kinetics and associated transport phenomena are required to represent the balance equations. Modeling the biological kinetics reactions is a difficult and time-consuming task, and therefore the model used by the estimators could differ significantly from reality. This is the main disadvantage of these phenomenological estimators, i.e., their efficiency strongly relies on the model quality. Asymptotic observers are based on the idea that uncertainty in bioprocess models lies in the process kinetics models. The design of these observers is based on a state transformation performed to provide a model which is independent of the kinetics. A potential drawback of the asymptotic observers is that the rate of convergence is completely determined by the operating conditions, i.e., the rate of convergence can be very slow or the observer may not converge. Empirical estimators are based on constructing appropriate nonlinear models of biotechnological processes exclusively from the process input–output data without considering the functional or phenomenological relations between the bioprocess variables.

However, the conventional empirical modeling approach is based on the knowledge of the structure (functional form) of the data-fitting model (in advance). This is a difficult task since it involves the heuristic selection of an appropriate nonlinear model structure from numerous alternatives.

Source: Biomass, Book edited by: Maggie Momba and Faizal Bux,
ISBN 978-953-307-113-8, pp. 202, September 2010, Sciyo, Croatia, downloaded from SCIYO.COM

For the machine learning community, the data-based modeling of the biomass concentration from a finite number of noisy samples (the training dataset) is a supervised learning problem. From this area, in recent years, the artificial neural network methodology has become one of the most important techniques applied to biomass estimation, e.g. (Leal, 2001; Li, 2003; Amicarelli *et al.*, 2006) and references therein. Neal's work on Bayesian learning for neural networks (Neal, 1996) shows that many Bayesian regression models based on neural networks converge to a class of probability distributions known as Gaussian Processes according as the number of hidden neurons tends to infinity. Furthermore, Neal argued that in the Bayesian approach for real-world complex problems, neural network models should not be limited to nets containing only a small number of hidden units. Neal's observation motivates the idea of replacing parameterized neural networks and work directly with Gaussian Process models for the high-dimensional applications to which neural networks are typically applied (Neal, 1997).

This Chapter addresses the problem of the biomass estimation in a batch biotechnological process: the *Bacillus thuringiensis* (*Bt*) δ-endotoxins production process, and presents different alternatives that can be successfully used in this sense. The development of the Chapter includes the design of various biomass estimators, namely: a phenomenological biomass estimator, a standard EKF biomass estimator, a biomass estimator based on ANN, a decentralized Kalman Filter and a biomass concentration estimator based on Bayesian regression with Gaussian Process.

Finally, conclusions about the estimators are presented and the results show the techniques for the *Bacillus thuringiensis* δ-endotoxins production process on the basis of experimental data from a set of various fermentations.

2. *Bacillus thuringiensis* δ-endotoxins production process

2.1 Bioprocess description

In the last years, due to environmental reasons the interest in biological agents for their use in ecological insecticides (bioinsecticides) has notably increased. *Bacillus thuringiensis* is one of the microorganisms most frequently studied as toxin producer. *Bt* is an aerobic spore - former bacterium which, during the sporulation; also produces insecticidal crystal proteins known as δ-endotoxins. It has two stages on its life span: a first stage characterized by its vegetative growth, and a second stage named sporulation phase. When the vegetative growth finalizes, the beginning of the sporulation phase is induced when the mean exhaustion point has been reached. Normally the sporulation is accompanied by the δ-endotoxin synthesis. After the sporulation, the process is completed with the cellular wall rupture (cellular lysis), and the consequent liberation of spores and crystals to the culture medium (Starzak & Bajpai, 1991; Aronson, 1993, Liu & Tzeng, 2000).

This research has been conducted with the same process and fermentation conditions as the work of Atehortúa *et al.* (2007). The microorganisms used in this work were *Bacillus thuringiensis serovar. kurstaki* strain 172-0451 isolated in Colombia and stored in the culture collection of Biotechnology and Biological Control Unit (CIB), (Vallejo *et al.*, 1999). The medium (CIB-1) contained: $MnSO_4.H_2O$ (0.03 $g.L^{-1}$), $CaCl_2.2H20$ (0.041 $g.L^{-1}$), KH_2PO_4 (0.5 $g.L^{-1}$), K_2HPO_4 (0.5 $g.L^{-1}$), $(NH_4)_2SO_4$ (1 $g.L^{-1}$), yeast extract (8 $g.L^{-1}$), $MgSO_4.7H_2O$ (4 $g.L^{-1}$) and glucose (8 $g.L^{-1}$). Growth experiments of the fermentation process with *Bacillus thuringiensis* were performed in a reactor with a nominal volume of 20 liters (Fig.1). The fermentations were developed with an effective volume of 11 liters of cultivation medium,

and they were inoculated to 10% (v/v) with the microorganism *Bt* culture. The inoculum added consisted of a vegetative phase culture: 5 mL spore suspension with 1 ·107 UFC/mL (stored at –20 °C) was used to inoculate a 500 mL flask containing 100 mL of CIB-1, and incubated with shaking at 250 rpm at 30 °C during 13 h. Fifty milliliters of this culture were aseptically transferred to each one of two 2 L flasks containing 500 mL of CIB-1 and incubated as above for 5 h. The pH medium was adjusted to 7.0 with KOH before its heat sterilization. Culture conditions at harvest are typified by 90% free spores and δ-endotoxins crystals.

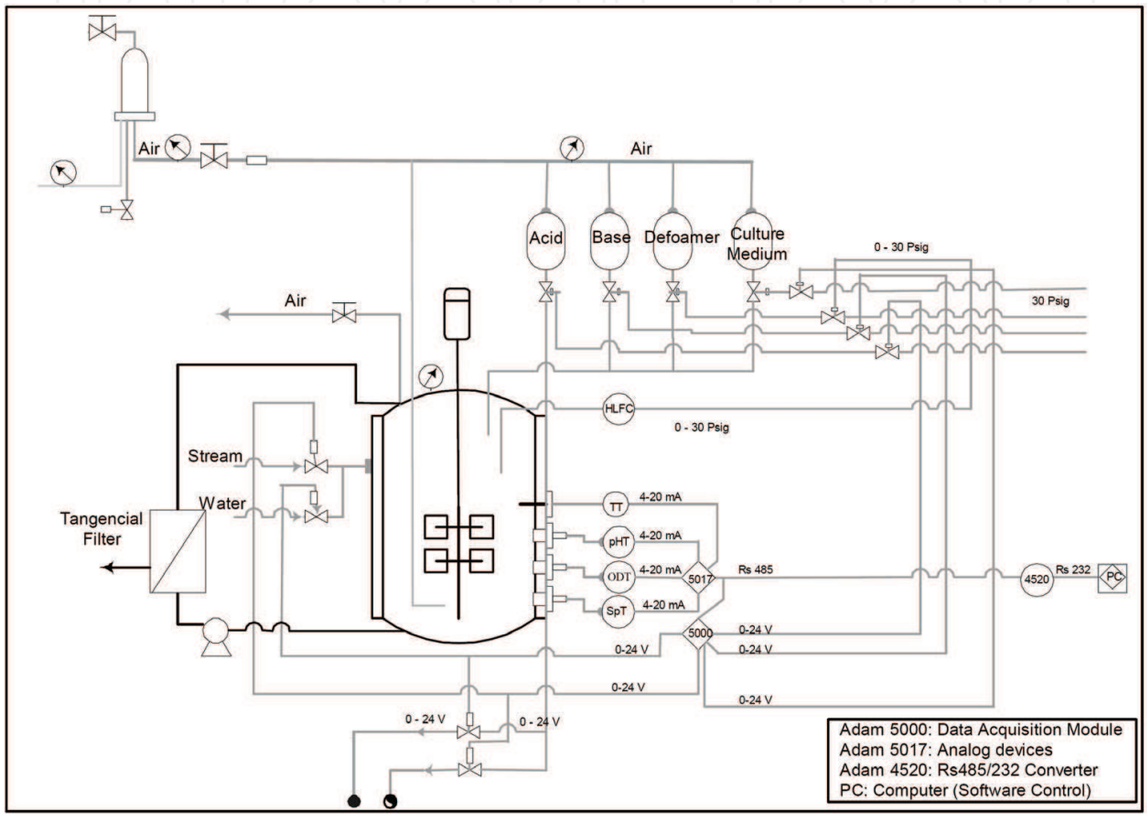


Fig. 1. Fermentation pilot plant scheme.

The temperature was maintained around 30 ºC by using an ON/OFF control; whereas the pH was fixed between 6.5 and 8.5. The air flow was set up at 1320 [L.h^{-1}] and the agitation speed at 400 rpm. Manometric pressure in the reactor was set at 41,368 Pa using a pressure controller. Temperature, pH, dissolved oxygen, and glucose concentration were registered by a data acquisition system using an Advantech® PCL card. Dissolved oxygen was measured by a polarographic oxygen sensor InPro 6000 (Mettler Toledo, Switzerland), and glucose concentration was determined with a rapid off-line measurement method through a glucose analyzer (YSI 2700).

The reagents concentrations used for the pH control and foam formation were nitric acid (5N), potassium hydroxide (2N) and defoamer (33% v/v). Cell growth was determined as dry cell weight (Dry cell weight (DCW, g /L) = (final weight - initial weight)/(volume of microbial suspension filtered). The foam formation was avoided by manually aggregating a defoamer.

Bacillus thuringiensis δ-endotoxins production is an aerobic operation, i.e., the cells require oxygen as a substrate to achieve cell growth and product formation (Ghribi *et al*., 2007).

2.2 Bioprocess model

The phenomenological estimator and the standard EKF presented in this Chapter are based on the phenomenological model presented in this Section, i.e. the model is necessary for its design. As pointed out in the introductory Section, the EKF is a classical nonlinear state estimator, and it's implemented for comparison purposes with the phenomenological biomass concentration estimator.

A first principle based model for *Bt* δ-endotoxins production process consists of a set of differential and algebraic equations (DAE system) in the continuous-time case, and a set of difference and algebraic equations in the discrete-time case. A simple phenomenological model was proposed by Rivera *et al*., (1999), a modification to the Rivera model was given by Atehortúa *et al*., (2006, 2007). Afterwards, Amicarelli *et al.* (2006, 2010) improved the model process adding the dissolved oxygen (DO) dynamics due to its importance in the biomass estimation problem and the posterior process control. The following state-space model is a discrete-time version of the continuous-time counterpart developed by Amicarelli *et al.* (2010).

$$\begin{bmatrix} X_v(k+1) \\ X_s(k+1) \\ S(k+1) \\ DO(k+1) \end{bmatrix} = \begin{bmatrix} \left((\mu(k) - k_s(k) - k_e(k))Ts + 1\right) X_v(k) \\ k_s(k) X_v(k)\, Ts + X_s(k) \\ -\left(\dfrac{\mu(k)}{Y_{x/s}} + m_s\right) X_v(k)\, Ts + S(k) \\ (K_1 - K_2 Ts)\, X(k) - K_1\, X(k+1) + DO(k) + K_3 Q_A\, Ts\left(DO^* - DO(k)\right) \end{bmatrix} \tag{1}$$

Where X_v is the vegetative cell concentration, X_S the sporulated cell concentration, $X = X_v + X_S$ is the total cell concentration ($X(k+1) = \left(\mu(k) - k_e(k)\right) Ts X_v(k) + X(k)$), S is the limiting substrate concentration and DO is the dissolved oxygen concentration.

The following algebraic equations define the specific growth speed μ (model based on Monod equation for each limiting nutrient S and DO), the spore formation rate k_S, and the death cell specific rate k_e.

$$\mu(k) = \mu_{max}\left(\frac{S(k)}{(K_S + S(k))}\frac{DO(k)}{(K_O + DO(k))}\right) \tag{2}$$

$$k_s(k) = k_{smax}\left(\frac{1}{1+e^{Gs(S(k) - Ps)}}\right) - k_{smax}\left(\frac{1}{1+e^{Gs(S_{initial} - Ps)}}\right) \tag{3}$$

$$k_e(k) = k_{emax}\left(\frac{1}{1+e^{Ge(Tsk - Pe)}}\right) - k_{emax}\left(\frac{1}{1+e^{Ge(t_{initial} - Pe)}}\right) \tag{4}$$

The complete notation and model parameter's values are presented in Tables 1 and 2.

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

Table 1. Phenomenological model variables.

Four batch cultures with different initial glucose concentration (8, 21, 32 and 40 $g.L^{-1}$) were carried out to generate experimental data for model validation and parameters tuning. In this context, four parameter sets guarantee a representative covering of an intermittent fed batch culture (IFBC) with total cell retention (TCR) in the operation space according to the work of Atehortúa *et al.* (2007), see Table 2.
Maximum glucose concentration in the medium (S_{max}) was used as the switching criteria among the estimated batch parameter sets.

	S max <10 $g.L^{-1}$	10$g.L^{-1}$< S max <20 $g.L^{-1}$	20$g.L^{-1}$< S max <32$g.L^{-1}$	S max >32 $g.L^{-1}$
μ_{max} $[h^{-1}]$	0.8	0.7	0.65	0.58
Yx/s $[g.g^{-1}]$	0.7	0.58	0.37	0.5
K_s $[g.L^{-1}]$	0.5	2	3	4
K_o $[g.L^{-1}]$	1×10^{-4}	1×10^{-4}	1×10^{-4}	1×10^{-4}
ms $\left[g.\left[g.h^{-1}\right]^{-1}\right]$	5×10^{-3}	5×10^{-3}	5×10^{-3}	5×10^{-3}
ksmax $[h^{-1}]$	0.5	0.5	0.5	0.5
Gs $\left[g.L^{-1}\right]^{-1}$	1	1	1	1
Ps $[g.L^{-1}]$	1	1	1	1
kemax $[h^{-1}]$	0.1	0.1	0.1	0.1
Ge $[h^{-1}]$	5	5	5	5
Pe [h]	4	4.7	4.9	6
K_1 dimensionless	9.725×10^{-4}	4.502×10^{-3}	3.795×10^{-3}	1.597×10^{-3}
K_2 $[h^{-1}]$	1.589×10^{-4}	0.046×10^{-3}	0.729×10^{-3}	0.561×10^{-3}
K_3 $[L^{-1}]$	4.636×10^{-4}	0.337×10^{-3}	2.114×10^{-3}	1.045×10^{-3}
Ts [h]	0.1	0.1	0.1	0.1

Table 2. Model parameters for the intermittent fed batch culture with total cell retention of *Bacillus thuringiensis serovar. Kurstaki.*

3. Biomass concentration estimators design.

The duration of the batch fermentation is limited and depends on the initial conditions of the microorganism culture. All the fermentations used in this work were initialized with the same inoculate and different substrate concentration conditions (Atehortúa et al, 2007). When the medium is inoculated, the biomass concentration increases at expense of the

nutrients, and the fermentation concludes when the glucose that limits its growth was consumed, or when 90% or more of cellular lysis is presented. After that, the latency period was removed (the bioprocess dead time is not considered), and the duration of each experiment is approximately 16 hours in this case.

The collected data from the fermentations is a set of concentrations measurements of dissolved oxygen (*DO*), primary substrate (*S*), and biomass (*X*) which have been sampled at different speed, 10 samples per hour for the concentrations of dissolved oxygen and glucose and 1 per hour for the biomass concentration, that was quantified by cell dry weight method. Practically, *DO* could be continuously measured whereas *S* can be measured up to 20 times per hour. From the bandwidth estimation of system signals by using Fourier frequency analysis, the sampling time $Ts = 1/10$ hours has been selected for dissolved oxygen and substrate measurements (di Sciascio & Amicarelli, 2008; Amicarelli, 2009).

In order to design biomass estimators for the *Bacillus thuringiensis* δ-endotoxins production process, it is proposed a two-stage method (di Sciascio & Amicarelli, 2008). In the first stage, the biomass concentrations data set is completed to have the same size as the dissolved oxygen concentration and primary substrate (glucose) concentration data sets. For this missing data problem (Little & Rubin, 2002), it was considered a Bayesian Gaussian Process Regression as an imputation strategy for filling the missing values. In the second stage, different biomass estimators are designed.

3.1 First stage design for all estimators- filling the biomass missing data

For the theory of Bayesian Regression Framework and Gaussian Process see Appendix C.

Suppose that we have a noisy training data set *D* which consists of *m* pairs of *n*-dimensional input vectors $\{x_i\}$ (regression vector) joined in a n×m matrix $\boldsymbol{X}$, and *m* scalar noisy observed outputs $\{y_i\}$ collected in a vector $\boldsymbol{y}$.

$$D = \left\{ \left(x_i , y_i \right) \middle| \, i = 1, L, m \right\} = \left\{ X, y \right\} \tag{5}$$

In order to construct a probabilistic statistical model for *D*, the following data-generating process is assumed:

$$y_i = f\left(x_i \right) + \varepsilon_i \tag{6}$$

where the latent real-valued function *f* is the deterministic or systematic component of the model, and the additive random term ε is the observation error. The aim of regression is to identify the systematic component f from the empirical observations *D*.

In this section, the biomass concentration data vector is completed with virtual filtered measurements to have the same size as dissolved oxygen and substrate data vectors. This is a missing data problem, and the Gaussian Process Regression will be used as imputation method for filling the missing values (note that this task in a deterministic framework which can be viewed as a curve-fitting or interpolation problem).

For all experimental fermentations, the data-generating model for biomass concentration is:

$$X(tk) = \hat{X}(tk) + \varepsilon(tk) \tag{7}$$

The training data set *D* consists of 18 pairs of time inputs $t = \{tk\} = \{1,...,18\}$ (in hours), and noisy biomass measurements outputs $X = \{X_k\} = \{X(t_1),...,X(t_{18})\}$. The latent functions $\hat{X} = \{\hat{X}_k\} = \{\hat{X}(t_1),...,\hat{X}(t_{18})\}$ are the estimated biomass concentrations.

The expression "Gaussian Process Regression Model" refers to the use of a Gaussian Process as a prior on **f**. This means that every finite-dimensional marginal joint distributions of function values **f** associated to any input subset of **X** is multivariate Gaussian.

$$p\left(f|X,\theta_P\right) = N\left(m(X), K\left(X,\theta_P\right)\right) \tag{8}$$

A Gaussian Process is fully specified by a mean function $m(X) = [m(x_1), L, m(x_m)]^T$ and a positive-definite covariance matrix $K(X, \theta_P)$, and it can be viewed as a generalization of the multivariate Gaussian distribution to infinite dimensional objects. Choosing a particular form of covariance function, the hyperparameters θ_P may be introduced to the Gaussian Process prior. Depending on the actual form of the covariance function $K(X, \theta_P)$ the hyperparameters θ_P can control various aspects of the Gaussian Process.

In this work, the elements of the parameterized covariance matrix, $C(X, \theta_P, \sigma^2)$, are denoted $C_{ij} = C(x_i, x_j)$, and they are functions of the training input data **X**, because these data determine the correlation between the training data outputs y. A suitable parametric form of the covariance function is:

$$Cij = \theta_0 + \theta_1 \exp\left[-\frac{1}{2} \sum_{l=1}^{n} \frac{\left(x_i^{(l)} - x_j^{(l)}\right)^2}{r_l^2} \right] + \theta_2\, \delta(i,j) + \sum_{l=1}^{n} \alpha_l\, x_i^{(l)}\, x_j^{(l)} \tag{9}$$

where $x_i^{(l)}$ is the l^{th} dimension of the input vector, x_i.

From the training data *D*, and by means of a conjugate gradient routine $\#\theta = 5$ hyperparameters, and the matrix **C** are determined recursively through:

$$\log\theta = [\log\theta_0, \log\theta_1, \log r_1, .., \log r_n, \log\theta_2, \log\alpha_1, .., \log\alpha_n]^T \tag{10}$$

and

$$\begin{cases} L = -\frac{1}{2}\log|C| - \frac{1}{2} y^T C^{-1} y - \frac{m}{2}\log 2\pi + \log p(\theta) + c \\ \frac{\partial L}{\partial \theta_i} = -\frac{1}{2}\operatorname{trace}\left(C^{-1} \frac{\partial C}{\partial \theta_i} \right) + \frac{1}{2} y^T C^{-1} \frac{\partial C}{\partial \theta_i} C^{-1} y + \frac{\partial \log p(\theta)}{\partial \theta_i} \end{cases} \tag{11}$$

Afterwards, at different times, $t_* = 0.1, 0.2, ..., 17.9, 18$ by (12)

$$\hat{f}_* = E\left(f_* \middle| D, x_*, K, \sigma^2\right) = k_*^T C^{-1} y$$
$$\sigma^2_{\hat{f}_*} = k_{**} - k_*^T C^{-1} k_* \quad (12)$$

the latent functions $\hat{X}_* = \left\{\hat{X}_*\right\} = \left\{\hat{X}(t*)\right\}$ and the variance $\sigma^2_{\hat{X}_*}$ are estimated. The expression "virtual filtered measurements" refers to the latent functions $\hat{X}_*$, because the additive normal noise ε has been removed (filtered) from the "virtual measurement" X_* in the data-generating model (7). Figure 2 gives an example of completion of biomass missing data for two fermentations (Fermentation 1, and Fermentation 2).

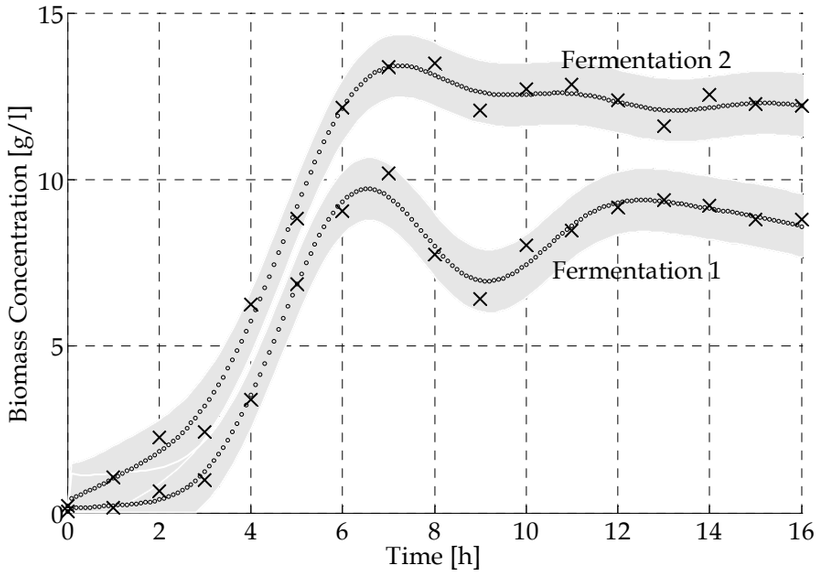


Fig. 2. Example of completion of biomass missing data for Fermentation 1 and Fermentation 2. The crosses being the biomass concentration measurements (training data), the small circles represent the biomass estimated (virtual filtered biomass measurements), and the grey region depicts the 95% confidence interval for the estimations (±2 standard deviations) (from di Sciascio & Amicarelli, 2008).

3.2 Phenomenological observer

In order to design a biomass phenomenological estimator, the dissolved oxygen balance from the nonlinear state-space model (1) presented before is employed in this Section

$$X(k) = \frac{1}{K_1}\left[\left(K_1 - K_2 Ts\right) X(k-1) - DO(k) + DO(k-1) + K_3 Q_A Ts\left[DO^* - DO(k-1)\right]\right] \quad (13)$$

From (13) it can be inferred that online, the total biomass concentration can be estimated with experimental data of dissolved oxygen concentration (*DO*) and with biomass past values (X(k-1)) for the current estimation. The remaining constants and parameters are known for this estimator. As the biomass is normally measured using an offline method, in this case the dry

weight method, the mentioned past values are not available for online estimation at the instant *k*. For this reason, it is not realistic to use the biomass measurements obtained by dry weight method and consequently, for online biomass estimation the values provided from the phenomenological model (1), $X(k-1) = Xv(k-1) + Xs(k-1)$ were used. Figure 3 shows the model structure for the phenomenological biomass estimator.

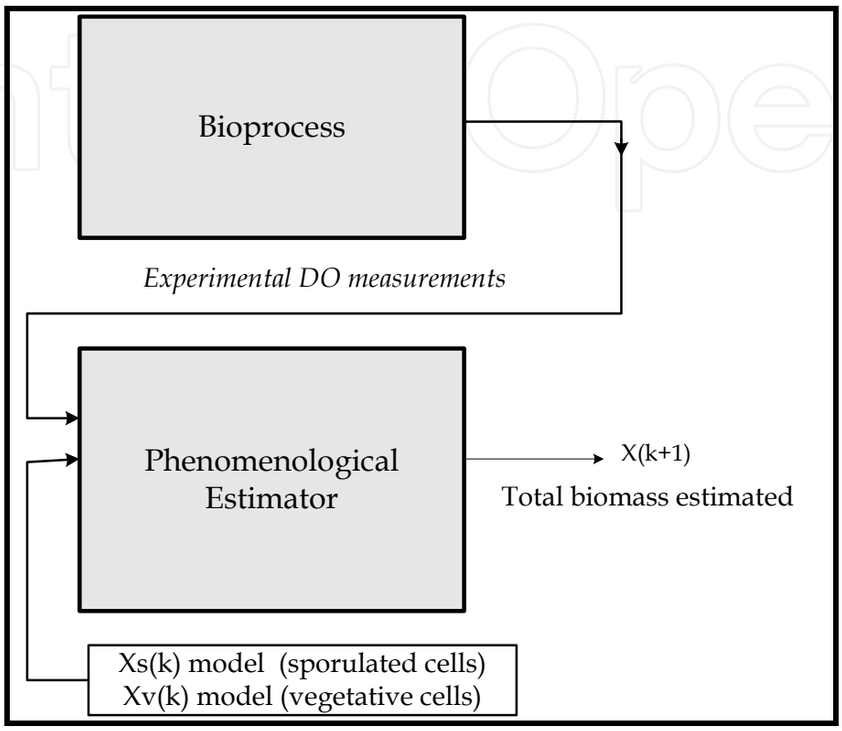


Fig. 3. Simulated output model structure for the phenomenological biomass estimator.

Figure 4 shows the phenomenological estimation results. This observer can approximate the biomass concentration better than the first model proposed by Atehortúa *et al.* (2007).
This is because, this estimator includes the dissolved oxygen consumption for growth and maintenance of the microorganism on its structure and through the experimental data of dissolved oxygen available online (Fig.3). In Fig 5. it can be seen the dissolved oxygen percentages time evolution for both fermentations.
Moreover, Fig. 4 shows satisfactory results and a correct behavior of the phenomenological estimator for two different fermentations. Estimated biomass follows closely the real biomass measurements. Similar results can be obtained for almost all fermentations. It can be noted that this performance is achieved by a phenomenological observer derived from the dissolved oxygen model available for this process. It is important to remark that the estimator involves in its structure the original model of vegetative and sporulated cells, whereas the consideration of the dissolved oxygen influence on the microorganism concentration improves the biomass estimation performance. It is important to remark that when the DO influence is not significant, the biomass estimation achieved with the model without the dissolved oxygen dynamics and the phenomenological estimator are comparable (show Fermentation 1 in Fig.4.). However, for those cases in which the DO approaches critical values (see Fermentation 2 in Fig. 4), the phenomenological observer gives better estimations (Fermentation 2).

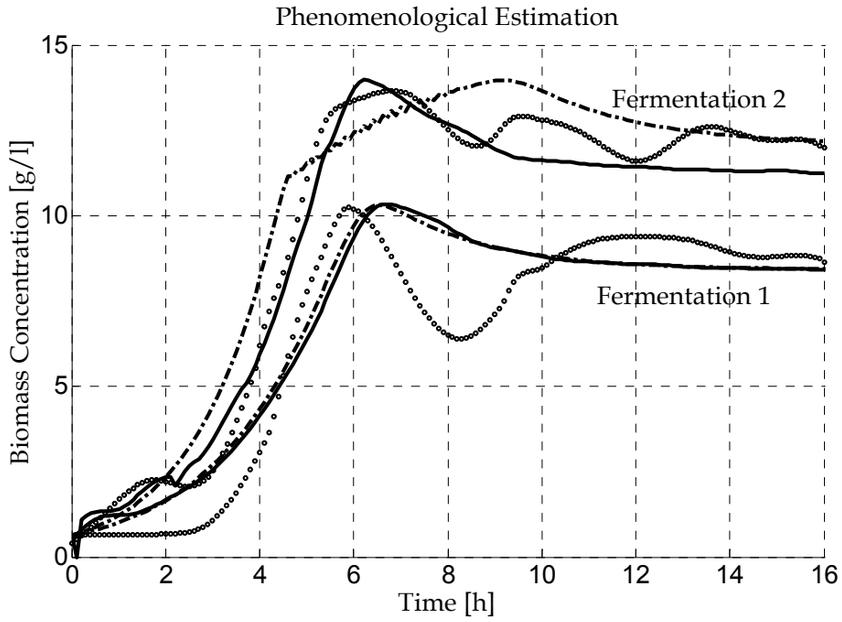


Fig. 4. Biomass estimator performance. The dash-dot line describes the behavior of biomass when considering the model (1); the solid line depicts the phenomenological estimator behavior based on DO dynamics; and the real biomass measurements are represented by small circles.

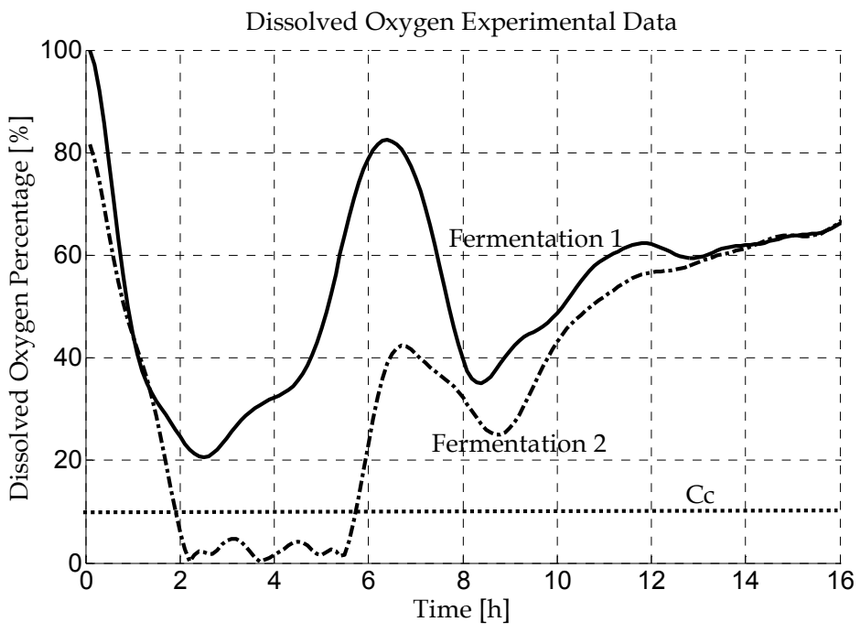


Fig. 5. Dissolved Oxygen experimental data. The solid line describes the Dissolved Oxygen behavior for the Fermentation 1; the dash-dot line depicts the Dissolved Oxygen behavior for the Fermentation 2 and the dotted line corresponds to the percentage of Dissolved Oxygen for the critical Dissolved Oxygen concentration for this process.

3.3 EKF standard estimator

Kalman filters are a widely useful tool used in biomass estimation due to its convergence and reliability properties. The estimation achieved from Kalman filters depends on the process model accuracy as well as on the available state measurements. The suitability of this estimation method can be concluded for the *Bt* fermentation process. Furthermore, this Section proposes a biomass concentration estimator for the mentioned biotechnological batch process through an Extended Kalman Filter (EKF) implementation.

The underlying theory of the EKF is largely known in the literature devoted to filtering, estimation, and control; see, for example, the classic books by Jazwinski (1970), Anderson & Moore (1979), or most recently, the book by Simon (2006). Therefore, in this work only brief explanations of the specific EKF implementation are given. In the EKF framework, the state transition and observation models are nonlinear differentiable states functions.

State transition model:

$$x(k+1) = f(x(k), u(k), k) + w(k) \tag{13}$$

Measurements model:

$$y(k) = h(x(k), k) + v(k) \tag{14}$$

Where $f(\times,\times)$ is the state transition function; $h(\times,\times)$ is the measurement function; $x(k)$ is the system state vector with initial condition $x(0) \sim N(x_0, Q_0)$ (as is usual in statistical literature the symbol (~) means "distributed according to"); $u(k)$ is the input or control vector; $y(k)$ is the observation vector; $w(k)$ is a discrete-time normal white noise process (process noise) with null mean and covariance matrix $\boldsymbol{Q}$, i.e., $w(k) \sim N(0, Q)$; and $v(k)$ is a discrete-time normal white noise process (measurements noise) with null mean and covariance matrix $\boldsymbol{R}$, i.e., $v(k) \sim N(0, R)$. The initial condition $x(0)$, and the sequences $w(k)$, and $v(k)$ are uncorrelated for all time shifts.

In our case the nominal State transition model (without the process noise $w(k)$) is obtained by introducing (2), (3) and (4) in (1).

$$x(k+1) = f(x(k), k) \tag{15}$$

The system state vector is $x(k) = [X_V(k) \quad X_S(k) \quad S(k) \quad DO(k)]^T$, the input vector is $u(k) = 0$ (the bioprocess has no external input), and the bioprocess outputs (observation vector) is $y(k) = [S(k) \quad DO(k)]^T$ (Fig.7). The experimental dissolved oxygen percentages and substrate concentration data employed are shown in Fig. 5 and 6.

The measurement model is linear in the states:

$$y(k) = H x(k) \tag{16}$$

where $H = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Taking into account the scales of the outputs, a balanced linear combination of $S(k)$ and $DO(k)$ can be considered as an alternative measurement model.

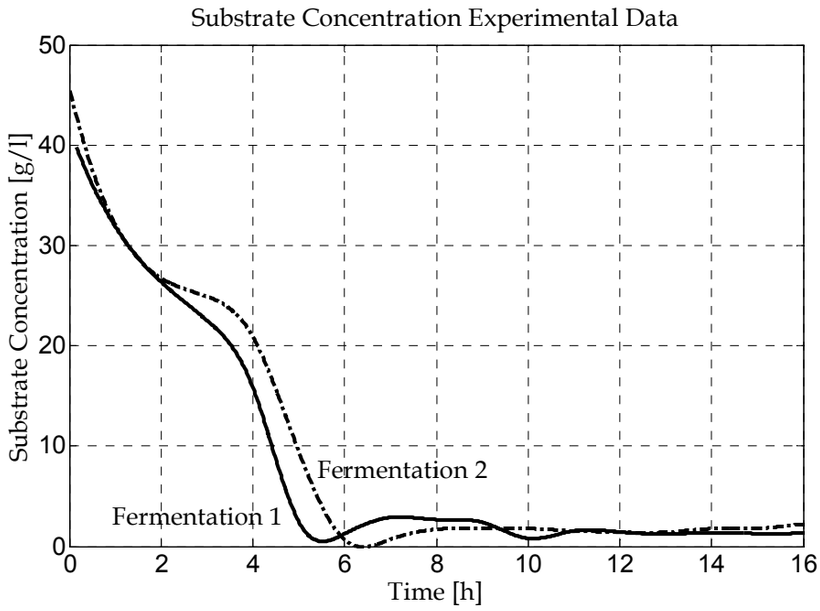


Fig. 6. Substrate Concentration experimental data. The solid line describes the Substrate Concentration for the Fermentation 1 and the dash-dot line depicts the Substrate Concentration for the Fermentation 2.

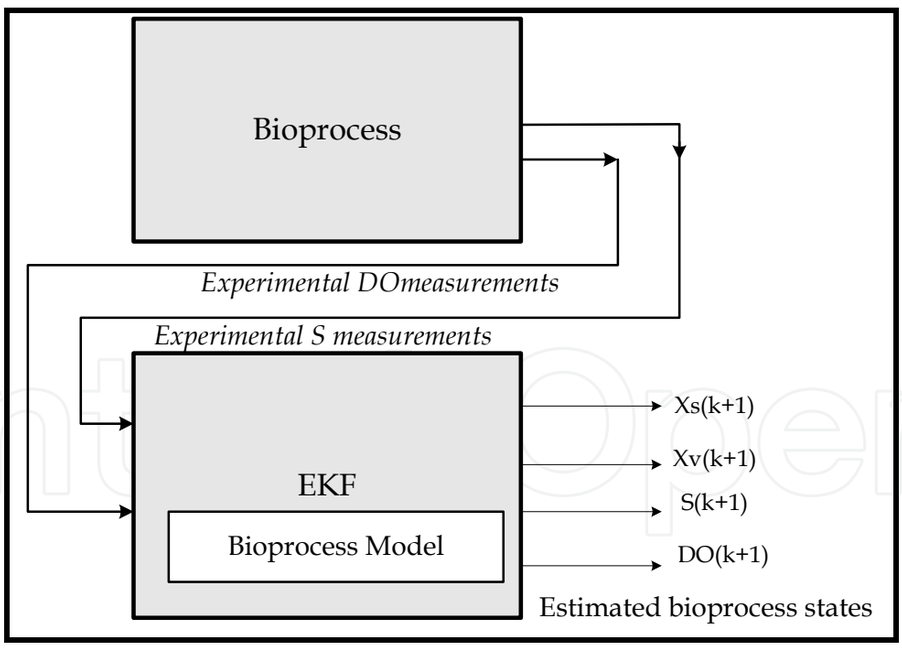


Fig. 7. Simulated output model structure for EKF biomass estimator.

$$y'(k) = H'x(k) = \alpha S(k) + \beta DO(k) \tag{17}$$

In this measurement model $H' = [0 \quad 0 \quad \alpha \quad \beta]$
where:

$$\alpha = DOmax/(S\ max + DOmax) \qquad \beta = S\ max/(Smax + DOmax)$$

The next step is to obtain the Jacobian matrices $\frac{\partial f(x(k),k)}{\partial x}$, and $\frac{\partial h(x(k),k)}{\partial x}$ evaluated at $\hat{x}(k\text{-}1|k-1)$.

$$A(k) = \left.\frac{\partial f(x(k),k)}{\partial x}\right|_{\hat{x}(k\text{-}1|k\text{-}1)}$$

$$A(k) = \begin{bmatrix} \left.\frac{\partial f_1(x(k),k)}{\partial x_1}\right|_{\hat{x}(k\text{-}1|k\text{-}1)} & \cdots & \left.\frac{\partial f_1(x(k),k)}{\partial x_4}\right|_{\hat{x}(k\text{-}1|k\text{-}1)} \\ . & . & . \\ \left.\frac{\partial f_4(x(k),k)}{\partial x_1}\right|_{\hat{x}(k\text{-}1|k\text{-}1)} & \cdots & \left.\frac{\partial f_4(x(k),k)}{\partial x_4}\right|_{\hat{x}(k\text{-}1|k\text{-}1)} \end{bmatrix} \tag{18}$$

$$H(k) = \left.\frac{\partial h(x(k),k)}{\partial x}\right|_{\hat{x}(k|k\text{-}1)} = \left.\frac{\partial H x(k))}{\partial x}\right|_{\hat{x}(k|k\text{-}1)} = H \tag{19}$$

The entries of the matrix $\boldsymbol{A}(k)$ and the EKF algorithm can be seen in Appendix A.

Finally, initializing the elements of the matrices $\boldsymbol{P}$, $\boldsymbol{Q}$ and $\boldsymbol{R}$, we have all the components of the EKF algorithm (see Table 3 in Appendix A). In order to obtain the best possible fit of the EKF to the experimental data, the elements of the matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ are empirically adjusted by simulations. Figure 8 shows results for two different fermentations. It is performed a comparison between this estimator and the phenomenological observer based on dissolved oxygen dynamics (DO) previously presented. The aim of this investigation is to remark the relevance of the information used for both observers.

It can be concluded that the performance of the standard EKF estimator is adequate. This of course does not mean that the performance of the EKF cannot be meaningfully enhanced by using a better model of the bioprocess or by some of the numerous improvements to the basic EKF scheme. In particular, different EKFs can be designed using a long list of engineering tricks: different coordinate systems; different factorizations of the covariance matrix; combinations of all of the above, as well as other bells and whistles invented by engineers in the hope of improving higher order Taylor series corrections to the state vector EKF performance (Daum, 2005).

The phenomenological estimator presents an adequate behavior, but their efficiency strongly relies on the model quality for this dissolved oxygen dynamics. It should be noticed that both estimators highlight the importance of the DO dynamics for this process.

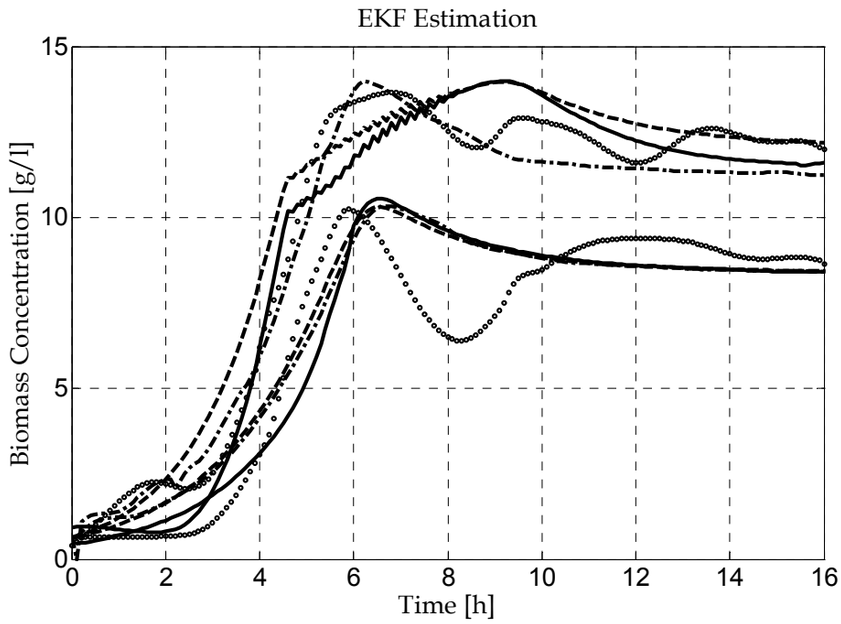


Fig. 8. Biomass estimator performance. The dashed line describes the biomass evolution obtained from the original model (1); the solid line depicts the EKF behavior, the dashed-dotted line depicts the phenomenological estimator behavior based on DO dynamics; and the real biomass measurements are represented by small circles.

3.4 ANN based estimator

Through artificial neural networks (ANN) the empirical knowledge (set of measurements) that characterizes a phenomenon of interest can be adequately codified. Due to the high degree of parallelism, the high generalization capability and the possibility to use an architecture of multiple inputs and outputs, the ANNs can provide a satisfactory solution to the problems of models identification, variables estimation, pattern recognition, functions approximation, among others. ANNs have the ability to abstract automatically essential characteristics of the experimental data, and to generalize from the previous experience; this allows the identification of the model process at lower cost.

Supervision and control techniques require optimizing the fermenter operation and the monitoring of all variables online is the best solution, since the methods offline delay the possibility of getting results and generally require more effort.

The ANN employed in this work is a multilayer perceptron with a hidden layer of 30 neurons and one output layer. For the training stage the Back Propagation algorithm (Haykin, 1999) was employed. The network was trained with data from a fermentation identified as "Fermentation 1" (See Fig. 9) and was generalized with other set of experimental data "Fermentation 2" (See Fig. 10). The activation functions of the hidden layer were hyperbolic tangent and a linear function for the output layer.

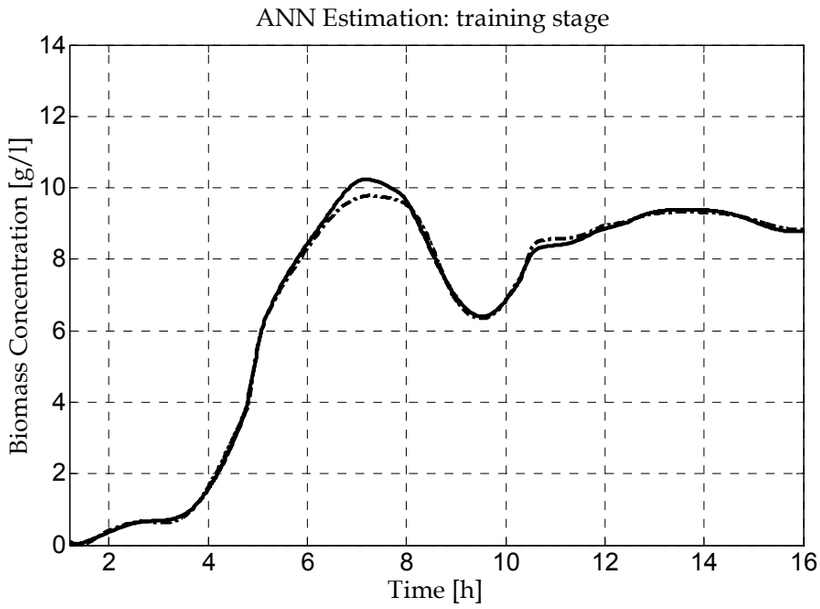


Fig. 9. Biomass estimator performance. The dashed line describes the biomass evolution obtained by the ANN in the training stage and the real biomass measurements are represented by the solid line. The perceptual training error e= 0.16 %.

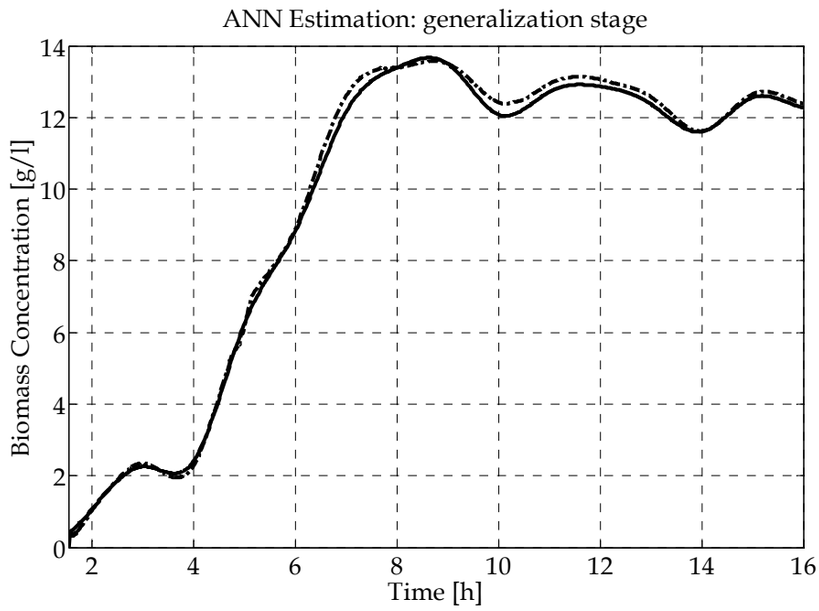


Fig. 10. Biomass estimator performance. The dashed line describes the biomass evolution obtained from the ANN in the generalization stage and the real biomass measurements are represented by the solid line. The perceptual generalization error e= 0.25 %.

3.5 Fusion through decentralized Kalman filter

The aim of this Section is to obtain an optimal biomass value for the process of *Bt.* To do this, two measurements (estimates) sequences are considered: the biomass estimation

available from the phenomenological observer and the biomass estimation provided by the ANN-based observer. Assuming that the estimations are the optimum value for each sequence in time and the relationship between these values is given by:

$$X^i = X^{iOPT} + v^i \tag{20}$$

where v^i is a random variable with zero mean and covariance R^i. In order to obtain an optimum value for biomass estimation, it was considered a decentralized Kalman filter (Brawn, 1997). In a basic approach of the decentralized Kalman Filter, each local filter operates autonomously. Each local filter has its own set of measurements, and there is no sharing of measurements. Note that this is inherently a cascaded operation mode, because the outputs of one or more of the local filters are acting as inputs to the master filter. The local filters (one for each sequence of measurements), the master filter and the different variables involved can be appreciated in Fig. 11.

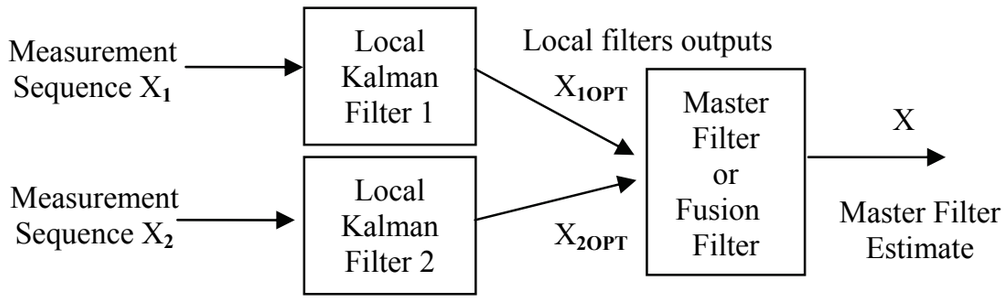


Fig. 11. Fusion scheme through a Decentralized Kalman Filter.

The mean and covariance for each sequence of measurements are calculated recursively according to:

$$\hat{X}^i(k+1) = \hat{X}^i(k) + \mu \left(X^i(k) - \hat{X}^i(k) \right) \tag{21}$$

$$R^i = R^i + \mu \left(\left(X^i - \hat{X}^i \right)^2 - R^i \right) \tag{21}$$

where $\hat{X}^i$ is the average sequence value of X^i and $0 < \mu < 1$ is a design constant. Then each sequence is individually filtered:

$$\left(P^i \right)^{-1} = \left(M^i \right)^{-1} + \left(R^i \right)^{-1} \tag{23}$$

$$X_v^{i.OPT} = P^i \left[m^i \left(M^i \right)^{-1} + \left(R^i \right)^{-1} X_v^i \right] \tag{24}$$

Equation (23) provides the updated information matrix and Eq. (24) are the states estimated updates, M^i and m^i are the covariance error and the previous estimation values for the

measurements sequences X^i respectively. All values are merged to obtain the optimum value of the estimated biomass. In Fig. 12 the results achieved with this approach can be observed.

$$X = P\left[\frac{m}{M} + \sum_i \left(\frac{X_l^{i.OPT}}{P^i} - \frac{m^i}{M^i}\right)\right] \tag{25}$$

$$P^{-1} = M^{-1} + \sum_i \left[\left(P^i\right)^{-1} - \left(M^i\right)^{-1}\right] \tag{26}$$

$$\begin{aligned} M^i &= P^i \\ m^i &= X^{i.OPT} \\ m &= X \\ M &= P \end{aligned} \tag{27}$$

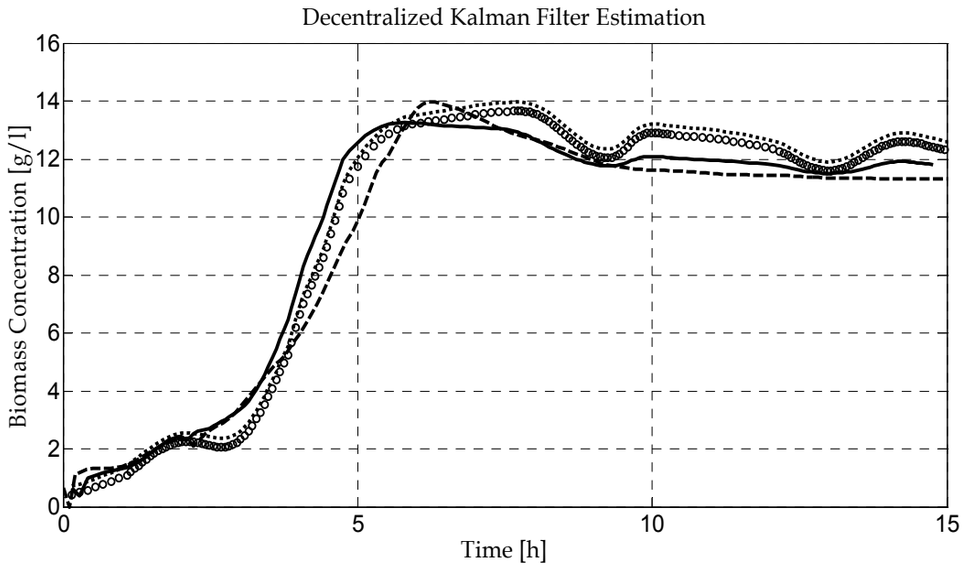


Fig. 12. Biomass estimator performance. The dashed line describes the biomass evolution obtained from the phenomenological estimator and the dot line describes a biomass estimation obtained from the ANN. The solid line describes the biomass evolution obtained through the decentralized Kalman Filter. The real biomass measurements are represented by small circles.

This architecture allows the complete autonomy of the local filters. The system achieves optimality in each individual local filter and global optimality in the primary filter.

3.6 Estimator based on Bayesian Regression through Gaussian Process

The first step in this design, is selecting the regressors, i.e., the components of the input vector $\boldsymbol{x}$. This is a laborious task, and has been done heuristically, chosen from numerous alternatives. The best empirical results have been achieved with:

$$x(kTS) = [DO(kTs), S(kTs), \hat{X}((k-1)Ts), \hat{X}((k-2)Ts)]^T$$

where $k = \{1, L, 180\}$ is the time index, $Ts = 1/10$ hours is the sampling time, $DO(.)$ is the dissolved oxygen concentration, $S(.)$ is the substrate concentration, and $\hat{X}(.)$ is the virtual filtered biomass measurement. In this case, the training data set D consists of 180 pairs of input vectors $\{x(kTs)\} = \{x_k\} \in R^4$ collected in a matrix $X \in R^{4\times 180}$, and scalars outputs $\{\hat{X}(kTs)\} = \{\hat{X}_k\}$ collected in a vector $\hat{X} \in R^{180}$ (note that in this section, the virtual filtered biomass measurements $\{\hat{X}_k\}$ are considered as true observed measurements). For the theory of Bayesian Regression Framework and Gaussian Process see Appendix C.

The data-generating process is $\hat{X}_k = \hat{\hat{X}}_k + \varepsilon_k$, being the latent function $\hat{\hat{X}}_k(.)$, and the additive normal noise ε. Once again, the $\#\theta = 11$ hyperparameters, and the new covariance matrix $\mathbf{C}$ eq. (9) are determined via a conjugate gradient routine from (11) and:

$$Cij = \theta_0 + \theta_1 \exp\left[-\frac{1}{2} \sum_{l=1}^{n} \frac{\left(x_i^{(l)} - x_j^{(l)} \right)^2}{r_l^2} \right] + \theta_2 \delta(i,j) + \sum_{l=1}^{n} \alpha_l x_i^{(l)} x_j^{(l)} \qquad (28)$$

Furthermore, by (12) the biomass concentration $\hat{\hat{X}}_* = \hat{\hat{X}}(t_*)$ and the variance $\sigma^2_{\hat{\hat{X}}_*}$ are estimated for a set of different times $\{t*\}$, $0 < t_* < 18\,\text{hours}$.

For the training stage, a one-step ahead predicted output schema is performed, i.e., the input measurements, $DO(k)$ $S(k)$, and the previous output measurements $\hat{X}(k-1)$, $\hat{X}(k-2)$ are used as repressors in:

$$\hat{\hat{X}}(k) = \hat{\hat{X}}(DO(k), S(k), \hat{X}(k-1), \hat{X}(k-2))$$

For on-line estimation the implemented estimator is the simulated output schema, i.e., only input measurements DO(k), S(k) are used. The simulated output is obtained as above, by replacing the measured outputs by the simulated output from the previous steps, i.e., previous outputs from the model have to be fed back into the model computations on-line (Fig.13).

$$\hat{\hat{X}}(k) = \hat{\hat{X}}(DO(k), S(k), \hat{\hat{X}}(k-1), \hat{\hat{X}}(k-2))$$

The one-step ahead predicted output scheme is also known as Nonlinear Auto Regressive with Exogenous input model (NARX), or as Series-Parallel model. Furthermore, the simulated output schema is known as Nonlinear Output Error model (NOE), or as Parallel model (Narendra & Parthasarathy, 1990; Ljung, 2006).

The biomass concentration of fermentations Fermentation 1 and Fermentation 2 from the preceding section (see Fig.2) has been adopted as training, and validation data respectively. Figures 5 and 6 show the measurements of dissolved oxygen percentages (*DO*) and glucose concentration (*S*) respectively. Both signals have been filtered with a low-pass filter with a 1/36 Hz corner frequency.

Figure 14 shows the results for the proposed biomass estimator and the results from the previous section, i.e., the true biomass measurements, the virtual filtered biomass measurements, and the 95% confidence intervals.

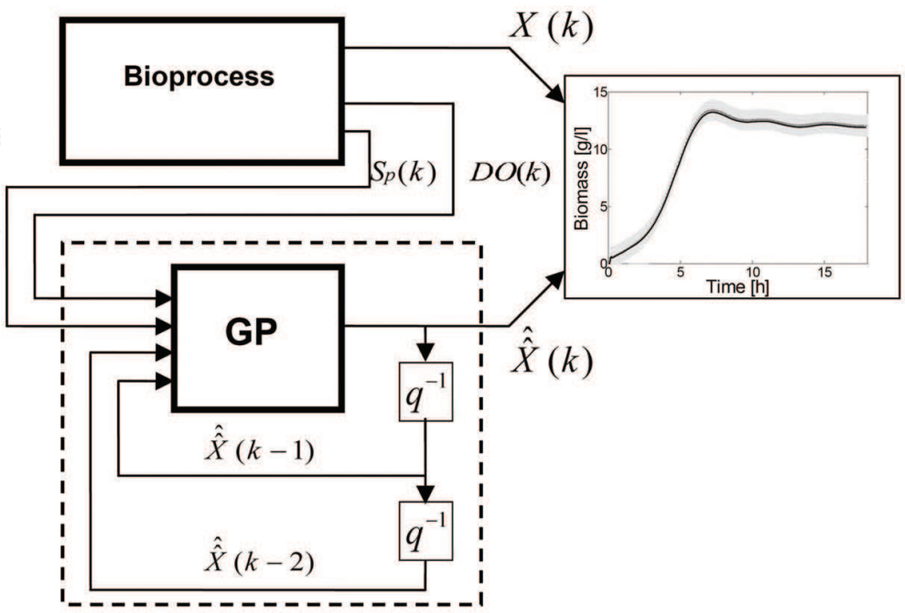


Fig. 13. Simulated output model structure of proposed biomass estimator (from di Sciascio & Amicarelli, 2008).

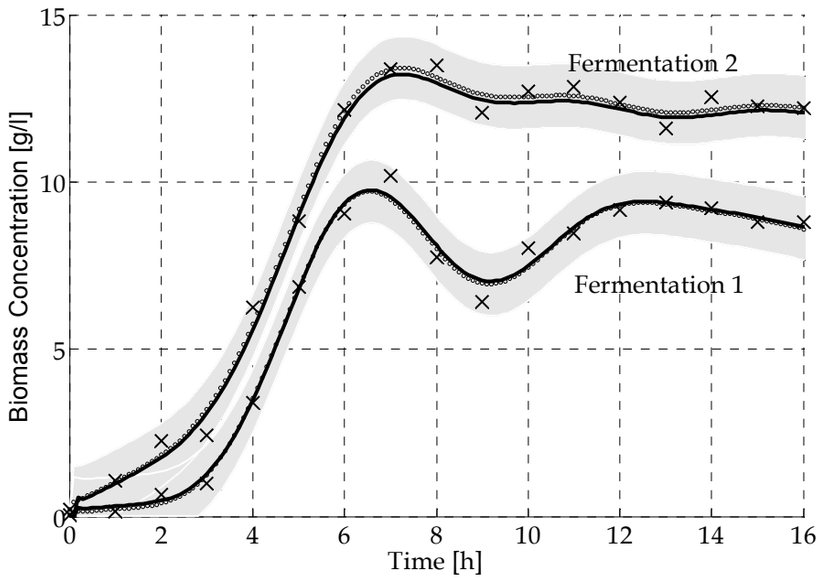


Fig. 14. Biomass estimator performance. The bold solid-line describes the behavior of the proposed biomass estimator ($\hat{X}(kTs)$), the crosses are the true biomass measurements, the virtual filtered biomass measurements ($\hat{X}(kTs)$) are represented by small circles, and the grey region depicts the 95% confidence interval (from di Sciascio & Amicarelli, 2008).

From Fig. 14, the correct behavior of the proposed biomass estimator can be clearly seen. The estimated biomass follows closely the true and the virtual filtered biomass measurements, and similar results can be obtained for almost all (except for some atypical) fermentations. This performance is achieved settling the 11 hyperparameters of the covariance function. For example, in order to obtain a similar performance with a multilayer feedforward neural network with one hidden layer (Haykin, 1999; Bishop, 1995) it can be shown that more than 30 neurons are necessary (Amicarelli *et al.*, 2006; Amicarelli, 2009). This means that hundreds of parameters must be calculated during the training phase.
The most probable explanation for poor results in some atypical fermentations is the aggregation of antifoam throughout the fermentations; this operation can be viewed as an unmodelled perturbation. If the dosage isn't correct, i.e., too much defoamer is aggregated, then DO concentration decreases markedly, and due to the aerobic nature of the *Bacillus thuringiensis* production process, this affects the estimator performance adversely. This undesirable behavior of the estimator can be avoided or at least minimized by an adequate control of the antifoam dosage.
The time evolution of the biomass concentration is considered as an uncertain dynamic system perturbed by process noise, i.e., a stochastic process, and the evolution of biomass concentration for a particular fermentation is a realization of the stochastic process. Furthermore, *Bacillus thuringiensis* has two stages on its life span (vegetative growth and sporulation) with very different dynamics clearly distinguishable in Fig. 14. This means that the time evolution of biomass concentration is a non-stationary stochastic process, moreover the last term of the covariance function (9) capture the non-stationarity of the biomass evolution.

4. Discussion and conclusions

In this Chapter it has been addressed the problem of the biomass estimation in a batch biotechnological process: the *Bacillus thuringiensis* (*Bt*) δ-endotoxins production process. Different alternatives that can be successfully used in this sense were presented. It has been exposed the design of various biomass estimators, namely: a phenomenological biomass estimator, a standard EKF biomass estimator, a biomass estimator based on ANN, a decentralized Kalman Filter, and a biomass concentration estimator based on Bayesian regression with Gaussian Process.
Each estimation method has its own advantages and drawbacks according to their ability to take into account the model uncertainties and the measurement errors. For all the proposed estimators, at the first design stage, the biomass concentrations data set was completed to have the same size as in dissolved oxygen percentage and substrate (glucose) concentration data sets.
The phenomenological biomass estimator is based on information from the dissolved oxygen balance for the process. The biomass concentration does not depend only on the mentioned variable; therefore, the proposed observer does not take into account biomass variations produced due to variations of temperature, pH, antifoam aggregation and inherent conditions of inocula. This estimator has achieved satisfactory results and a correct behavior for a set of different fermentations. This performance was reached by a phenomenological observer derived from a dissolved oxygen model available for this process. This observer considered the DO influence on the microorganism concentration, which can improve the biomass estimation performance when the DO influence is significant.

The observation vector of the standard extended Kalman Filter observer was built with experimental dissolved oxygen and substrate data. This estimator also provides an adequate biomass estimation. However the phenomenological estimator has a better behavior. On the other hand, the phenomenological estimator efficiency strongly relies on the model quality for this process. It should be noted that both estimators highlight the importance of dissolved oxygen for this process and are based on the quality of the process model.

The ANN based estimator was a Recurrent Multilayer Perceptron. The proposed virtual sensor provided satisfactory results for the biomass estimation showing acceptable performance. The choice of the input variables is important, since it is a batch process which has "infinite memory". Next, the result provided by the phenomenological and the ANN observers was compared with a new estimation given by its fusion through a Decentralized Kalman Filter. This new estimation is useful when redundant and comparable information from different sensors exists.

In designing a biomass concentration estimator based on Bayesian regression with Gaussian Process; the time evolution of biomass is conceived as a dynamic system perturbed by a certain process noise. Although the bioprocess is not truly stochastic, this noise is used for modeling the uncertainties in the system dynamics, i.e., the stochasticity is only used for representing the model uncertainties. Biomass concentration estimation is obtained indirectly through observed noisy measurements. Noise in the measurements refers to a disturbance in the sense that the measurements are uncertain, i.e., even in the hypothetical case that the true biomass concentration is known, the measurements would not be deterministic functions of this true biomass, but would have a certain distribution of possible values. The major difficulty when the biomass estimation is implemented is related to the uncertainty of the models used to describe their dynamics.

The proposed biomass concentration estimator based on Bayesian regression with Gaussian Process was formulated like a filtering problem. In the design, a Gaussian Process Regression schema was used for biomass estimation. The regressors selection and the estimator stability are important. Regressors selection is a problem related to the choice of states in a state space representation of the system. Finding a set of "good" regressors for biomass estimation is a non trivial task, because this set is characteristic of each specific bioprocess. In this work, this task has been done heuristically by trial and error between numerous alternatives; however, more elaborated methods can be used, for example, optimization-based regressors selection, Analysis of Variance (ANOVA), and so on (Hastie *et al.*, 2001; Lind, 2006; Mannale, 2006). The first option is the set of regressors $\{DO(kTs), DO((k-1)Ts), \hat{X}((k-1)Ts)\}$ that can be derived from the phenomenological model of Section 2.2, (Amicarelli *et al.*, 2006, Amicarelli, 2009). However, the best empirical results have been achieved with the regressors $\{DO(kTs), S(kTs), \hat{X}((k-1)Ts), \hat{X}((k-2)Ts)\}$.

As a result of the fed back of preceding outputs, the proposed on-line biomass estimator is a dynamic system, therefore potential instability problems can appear, even if the original off-line predicted estimator is stable. It is very difficult to analyze analytically the on-line estimator stability properties and only simulations analyses have been carried out previously to implementation. In this specific application such potential instability problems have never appeared.

It is important to remark that, the performance achieved with the Bayesian estimator is achieved setting 11 hyperparameters of the covariance function and in order to obtain a similar performance with a multilayer feed-forward neural network it can be shown that at least 30 neurons are necessary.

Appendix A. Extended Kalman Filter implementation

The entries of the matrix A(k) are:

$$A(k) = \left.\frac{\partial f(x(k),k)}{\partial x}\right|_{\hat{x}(k-1|k-1)}$$

$$A(k) = \begin{bmatrix} \left.\frac{\partial f_1(x(k),k)}{\partial x_1}\right|_{\hat{x}(k-1|k-1)} & \cdots & \left.\frac{\partial f_1(x(k),k)}{\partial x_4}\right|_{\hat{x}(k-1|k-1)} \\ . & . & . \\ \left.\frac{\partial f_4(x(k),k)}{\partial x_1}\right|_{\hat{x}(k-1|k-1)} & \cdots & \left.\frac{\partial f_4(x(k),k)}{\partial x_4}\right|_{\hat{x}(k-1|k-1)} \end{bmatrix}$$

$$a_{11} = 1$$

$$a_{12} = \frac{k_{smax}\,Ts}{1+e^{Gs(S(k)-Ps)}} - \frac{k_{smax}\,Ts}{1+e^{Gs(S_{initial}-Ps)}}$$

$$a_{13} = -\frac{k_{smax}\,Ts\,X_v(k)\,Gs\,e^{Gs(S(k)-Ps)}}{\left(1+e^{Gs(S(k)-Ps)}\right)^2}$$

$$a_{14} = 0$$

$$a_{21} = 0$$

$$a_{22} = 1 + \frac{k_{emax}\,Ts}{1+e^{-Ge(t_{initial}-Pe)}} - \frac{k_{emax}\,Ts}{1+e^{-Ge(Tsk-Pe)}} + \frac{k_{smax}\,Ts}{1+e^{Gs(S_{initial}-Ps)}} - \frac{k_{smax}\,Ts}{1+e^{Gs(S(k)-Ps)}} + \frac{Ts\,S(k)\,DO(k)\,\mu_{max}}{(K_d+DO(k))(K_s+S(k))}$$

$$a_{23} = \frac{k_{smax}\,Ts\,X_v(k)\,Gs\,e^{Gs(S(k)-Ps)}}{\left(1+e^{Gs(S(k)-Ps)}\right)^2} - \frac{Ts\,X_v(k)\,S(k)\,DO(k)\,\mu_{max}}{\left(K_d+DO(k)\right)\left(K_s+S(k)\right)^2} + \frac{Ts\,X_v(k)\,DO(k)\,\mu_{max}}{\left(K_d+DO(k)\right)(K_s+S(k))}$$

$$a_{24} = -\frac{Ts\,X_v(k)\,S(k)\,DO(k)\,\mu_{max}}{\left(K_d+DO(k)\right)^2\left(K_s+S(k)\right)} + -\frac{Ts\,X_v(k)\,S(k)\,\mu_{max}}{\left(K_d+DO(k)\right)\left(K_s+S(k)\right)}$$

$$a_{31} = 0$$

$$a_{32} = -\frac{Ts\,S(k)\,DO(k)\,\mu_{max}}{Y_{x/s}\left(K_d + DO(k)\right)\left(K_s + S(k)\right)} + Ts\,m_s$$

$$a_{33} = 1 + \frac{Ts\,X_v(k)\,S(k)\,DO(k)\,\mu_{max}}{Y_{x/s}\left(K_d + DO(k)\right)\left(K_s + S(k)\right)^2} - \frac{Ts\,X_v(k)\,DO(k)\,\mu_{max}}{Y_{x/s}\left(K_d + DO(k)\right)\left(K_s + S(k)\right)}$$

$$a_{34} = \frac{Ts\,X_v(k)\,S(k)\,DO(k)\,\mu_{max}}{Y_{x/s}\left(K_d + DO(k)\right)^2\left(K_s + S(k)\right)} - \frac{Ts\,X_v(k)\,S(k)\,\mu_{max}}{Y_{x/s}\left(K_d + DO(k)\right)\left(K_s + S(k)\right)}$$

$$a_{41} = K_2\,Ts$$

$$a_{42} = K_2\,Ts - \frac{k_{emax}\,Ts\,S(k)\,DO(k)\,u_{max}}{(1+e^{Ge(Pe-Tsk)})(K_d + DO(k))(K_s + S(k))} + \frac{k_{emax}\,Ts\,S(k)\,DO(k)\,u_{max}}{(1+e^{Ge(Pe-Tsk)})(K_d + DO(k))(K_s + S(k))}$$

$$a_{43} = \frac{k_{emax}\,Ts\,X_v(k)\,S(k)\,DO(k)\,u_{max}}{(1+e^{Ge(Pe-Tsk)})(K_d + DO(k))(K_s + S(k))^2} - \frac{k_{emax}\,Ts\,X_v(k)\,S(k)\,DO(k)\,u_{max}}{(1+e^{Ge(Pe-Tsk)})(K_d + DO(k))(K_s + S(k))^2} +$$

$$\frac{k_{emax}\,Ts\,X_v(k)\,DO(k)\,u_{max}}{(1+e^{Ge(Pe-Tsk)})(K_d + DO(k))(K_s + S(k))} - \frac{k_{emax}\,Ts\,X_v(k)\,DO(k)\,u_{max}}{(1+e^{Ge(Pe-Tsk)})(K_d + DO(k))(K_s + S(k))}$$

$$a_{44} = 1 - K_3\,Q_A\,Ts + \frac{k_{emax}\,Ts\,X_v(k)\,S(k)\,DO(k)\,u_{max}}{(1+e^{Ge(Pe-Tsk)})(K_d + DO(k))^2(K_s + S(k))} - \frac{k_{emax}\,Ts\,X_v(k)\,S(k)\,DO(k)\,u_{max}}{(1+e^{Ge(Pe-Tsk)})(K_d + DO(k))^2(K_s + S(k))} +$$

$$+ \frac{k_{emax}\,Ts\,X_v(k)\,S(k)\,u_{max}}{(1+e^{Ge(Pe-Tsk)})(K_d + DO(k))(K_s + S(k))} - \frac{k_{emax}\,Ts\,X_v(k)\,S(k)\,u_{max}}{(1+e^{Ge(Pe-Tsk)})(K_d + DO(k))(K_s + S(k))}$$

Predict Step

$$\hat{x}(k|k-1) = f(\hat{x}(k-1|k-1), k)$$

$$P(k|k-1) = A(k)\,P(k|k-1)\,A^T(k) + Q$$

Update Step

$$\tilde{y}(k) = y(k) - H\hat{x}(k|k-1)$$

$$K(k) = P(k|k-1)\,H^T\left[H\,P(k|k-1)\,H^T + R\right]^{-1}$$

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)\,\tilde{y}(k)$$

$$P(k|k) = \left(I + K(k)\,H\right)P(k|k-1)$$

Table 3. EKF algorithm

Appendix B. Decentralized Kalman Filter

Decentralized Filtering (no feedback to the local filters)

i. Information matrix update

$$P^{-1} = \left(P^{-}\right)^{-1} + H^{T} \quad R^{-1} \quad H \tag{B1}$$

ii. Gain computation

$$K = P\,H^{T}R^{-1} \tag{B2}$$

iii. Estimate update

$$\hat{x} = \hat{x}^{-} + K \quad \left(x \; - H \quad \hat{x}^{-}\right) \tag{B3}$$

iv. Project ahead to next step

$$\hat{x}_{k+1}^{-} = \varphi_k \quad \hat{x}_k \tag{B4}$$

$$P_{k+1}^{-} = \varphi_k \quad P_k \quad \varphi_k^{T} + Q_k \tag{B5}$$

Recall that $\mathbf{P^{-1}}$ is called the *information matrix*. In terms of information, i) says that the updated information is equal to the prior information plus the additional information obtained from the measurement at time t_k.. Furthermore, if $\mathbf{R_k}$ is block diagonal, the total "added" information can be divided into separate components, each representing the contribution from the respective measurement blocks. That is, we have (omitting the k subscripts for convenience).

$$H^{T}R^{-1}H = H_1^{T}R_1^{-1}H_1 + H_2^{T}R_2^{-1}H_2 + \ldots + H_N^{T}R_N^{-1}H_N \tag{B6}$$

We also note that the estimate update equation al time t_k can be written in a different form as follows:

$$\begin{aligned}
\hat{x} &= \left(I - K \quad H\right) \quad \hat{x}^{-} + K \quad X \\
\hat{x} &= \left(I - P \quad H^{T} \quad R^{-1} \quad H\right) \quad \hat{x}^{-} + P \quad H^{T} \quad R^{-1} \quad X \\
\hat{x} &= \left(P \quad P^{-1} - P \quad H^{T} \quad R^{-1} \quad H\right) \quad \hat{x}^{-} + P \quad H^{T} \quad R^{-1} \quad X \\
\hat{x} &= P \quad \left[\left(P^{-1} - H^{T} \quad R^{-1} \quad H\right) \quad \hat{x}^{-} + H^{T} \quad R^{-1} \quad X\right] \\
\hat{x} &= P \quad \left[\left(P^{-}\right)^{-1} \quad \hat{x}^{-} + H^{T} \quad R^{-1} \quad X\right]
\end{aligned} \tag{B7}$$

When writing in this form, it is clear that the updated estimate is a linear blend of the old information with the new information.

For simplicity, we will start with just two local filters in our decentralized system, and we will continue to omit the k subscripts to save writing. Both filters are assumed to implement the full-order state vector, and at step k both are assumed to have available their respective prior estimates $\mathbf{m_1}$ and $\mathbf{m_2}$ and their associated error covariances $\mathbf{M_1}$ and $\mathbf{M_2}$. For Gaussian Process, m1 and m2 will be the means of x conditioned on their respective measurement streams up to, but not including, time t_k. The measurements presented to filters 1 and 2 at time t_k are X_1 and X_2 and they have the usual relationship to x:

$$X_1 = H_1 x + v_1 \tag{B8}$$

$$X_2 = H_2 x + v_2 \tag{B9}$$

Where $\mathbf{v_1}$ and $\mathbf{v_2}$ are zero mean random variables with covariances $\mathbf{R_1}$ and $\mathbf{R_2}$. The estate **x** and noises $\mathbf{v_1}$ and $\mathbf{v_2}$ are assumed to be mutually uncorrelated as usual.
If we assume now that local filters 1 and 2 do not have access to each other's measurements, the filters will form their respective error covariances and estimates according to (B1) and (B7).

Local filter 1

$$P_1^{-1} = M_1^{-1} + H_1^T \quad R_1^{-1} \quad H_1 \tag{B10}$$

$$\hat{X}_{1OPT} = P_1 \left(M_1^{-1} \quad m_1 + H_1^T \quad R_1^{-1} \quad X_1 \right) \tag{B11}$$

Local filter 2

$$P_2^{-1} = M_2^{-1} + H_2^T \quad R_2^{-1} \quad H_2 \tag{B12}$$

$$\hat{X}_{2OPT} = P_2 \left(M_2^{-1} \quad m_2 + H_2^T \quad R_2^{-1} \quad X_2 \right) \tag{B13}$$

Note that the local estimates will be optimal, conditioned on their respective measurement streams, but not with respect to the combined measurements.
Now consider the master filter. It is looking for an optimal global estimate of **x** conditioned on both measurement streams 1 and 2.
m = optimal estimate of **x** conditioned on both measurement streams up to but not including t_k
M = covariance matrix associated with **m**
The optimal global estimate and associated error covariance are then

$$P^{-1} = \begin{bmatrix} H_1^T & H_2^T \end{bmatrix} \begin{bmatrix} R_1^{-1} & 0 \\ 0 & R_2^{-1} \end{bmatrix} \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} + M^{-1}$$
$$P^{-1} = M^{-1} + H_1^T \quad R_1^{-1} \quad H_1 + H_2^T \quad R_2^{-1} \quad H_2 \tag{B14}$$

$$\hat{x} = P \left(M_{-1} \quad m + H_1^T \quad R_1^{-1} \quad H_1 + H_2^T \quad R_2^{-1} \quad H_2 \right) \tag{B15}$$

However, the master filter does not have direct access to X**1** and X**2**, so we will rewrite (B14) and (B15) in terms of the local filter's computed estimates and covariances. The result is:

$$P^{-1} = \left(P_1^{-1} - M_1^{-1}\right) + \left(P_2^{-1} - M_2^{-1}\right) + M^{-1} \tag{B16}$$

$$\hat{x} = P\left[\left(P_1^{-1}\ \hat{x}_1 - M_1^{-1}\ m_1\right) + \left(P_2^{-1}\ \hat{x}_2 - M_2^{-1}\ m_2\right) + M^{-1}\ m\right] \tag{B17}$$

It can now be seen that the local filters can pass their respective $\hat{x}_i$, P_i^{-1}, m_i and M_i^{-1} (i=1,2) on to the master filter, which, in turn, can then compute its global estimate. The local filters can, of course, do their own local projections and then repeat the cycle at step (k+1). Likewise, the master filter can project its global estimate and get a new **m** and **M** for the next step. Thus, it can be seen that this architecture permits complete autonomy of the local filters, and it yields local optimality with respect to the respective measurement stream. The system also achieves global optimality in the master filter.

Appendix C: Bayesian Regression framework and Gaussian Process Regression

Suppose that we have a noisy training data set D which consists of m pairs of n-dimensional input vectors $\{x_i\}$ (regression vector) joined in a $n \times m$ matrix X, and m scalar noisy observed outputs $\{y_i\}$ collected in a vector $\boldsymbol{y}$.

$$D = \left\{\left(x_i, y_i\right) \middle| i = 1, \ldots, m\right\} = \left\{X, y\right\} \tag{C1}$$

In order to construct a probabilistic statistical model for D, the following data-generating process is assumed:

$$y_i = f\left(x_i\right) + \varepsilon_i \tag{C2}$$

where the latent real-valued function f is the deterministic or systematic component of the model, and the additive random term ε is the observation error. The aim of regression is to identify the systematic component f from the empirical observations D. The Bayes' rule (C3) shows the components of Bayesian Inference (Bernardo & Smith, 2006): the joint likelihood, the prior distribution, and the posterior distribution. Bayesian inference alludes to the process of updating our beliefs according to Bayes' rule, i.e. computing the posterior from likelihood and the prior, integrating the information contained in the observed data.

$$\overbrace{p\left(f \middle| D, \theta_L, \theta_P\right)}^{\text{Posterior}} = \frac{\overbrace{p\left(y \middle| f, \theta_L\right)}^{\text{Likelihood}}\overbrace{p\left(f \middle| X, \theta_P\right)}^{\text{Prior}}}{\underbrace{p\left(D \middle| \theta_P, \theta_L\right)}_{\text{Evidence}}} \tag{C3}$$

Where $P(D|.) = P(y|X,.)$, $P(.|D,.) = P(.|y,X,.)$, $y_i = y(x_i)$ $y = [y_1, L, y_m]^T$, denotes the observed outputs, $f_i = f(x_i)$ $f = [f_1, L, f_m]^T$, are the latent function values, and θ_L, θ_P denote additional parameters (hyperparameters) of the likelihood and prior distribution respectively.

The evidence or marginal likelihood is the normalising constant appearing in the denominator of Bayes' rule. This quantity is one of the most useful quantities in the Bayesian framework (e.g., in hypothesis testing applications) (O'Hagan, 2004; Bernardo & Smith, 2006), however the evidence is not considered in the remainder of the exposition.

If the measurements of the training data set *D* are independent, and the observation error ε it is assumed that is normal, independent and identically distributed with mean zero and variance σ^2, then, in this case the joint likelihood is:

$$p(y|f,\sigma^2) = \prod_{i=1}^{m} N(f_i,\sigma^2) = N(f,\sigma^2 I_{m\times m}) \tag{C4}$$

In the Bayesian non-parametric approach, a prior is put directly on the space of functions and the inference is carried out on *f*. The prior distribution is usually chosen from a parametric family of distributions or a mixture of these. The expression "Gaussian Process Regression model" refers to using a Gaussian Process as a prior on **f**. This means that every finite-dimensional marginal joint distributions of function values **f** associated to any input subset of **X** is multivariate Gaussian.

$$p\left(f|X,\theta_P\right) = N\left(m\left(X\right), K\left(X,\theta_P\right)\right) \tag{C5}$$

A Gaussian Process is fully specified by a mean function $m(X) = [m(x_1), L, m(x_m)]^T$ and a positive-definite covariance matrix $K(X, \theta_P)$, and it can be viewed as a generalization of the multivariate Gaussian distribution to infinite dimensional objects. Choosing a particular form of covariance function, the hyperparameters θ_P may be introduced to the Gaussian Process prior. Depending on the actual form of the covariance function $K(X, \theta_P)$ the hyperparameters θ_P can control various aspects of the Gaussian Process. For simplicity the prior mean function is set to be zero $\mu(X)=0$, this is completely general provided that a constant term is included in the covariance function (Williams & Rasmussen, 1996; Kuss, 2006).

The posterior distribution over function values is obtained introducing (C4) and (C5) in (C3)

$$p\left(f|D,\sigma^2,K\right) \propto N\left(f,\sigma^2 I_{m\times m}\right) N\left(0,K\right) \propto N\left(K\left(K+\sigma^2 I_{m\times m}\right)^{-1} y, \left(K^{-1}+\sigma^{-2} I_{m\times m}\right)^{-1}\right) \tag{C6}$$

The distribution of the latent function value $f_* = f(x_*)$ for an arbitrary new input x_* conditioned on the training function outputs is (Kuss, 2006):

$$p\left(f_* \middle| f, x_*, X, K\right) \propto N\left(k_*^T K^{-1} f, k_{**} - k_*^T K^{-1} k_*\right) \tag{C7}$$

where $k_* = [K(x_*, x_1), L, K(x_*, x_m)]^T$ is a vector of prior covariances between x_* and the training inputs $\mathbf{X}$, and $k_{**} = K(x_*, x_1)$.

The predictive distribution of f_* is obtained by integrating out the training function values $\mathbf{f}$ from (B6) over the posterior distribution (C7). The predictive distribution is again multivariate normal:

$$p\left(f_* \middle| D, x_*, K, \sigma^2\right) = \int p\left(f_* \middle| f, x_*, X, K\right) p\left(f \middle| D, \sigma^2, K\right) df \propto N\left(k_*^T C^{-1} y, k_{**} - k_*^T C^{-1} k_*\right) = N(\hat{f}_*, \sigma^2_{\hat{f}_*}) \tag{C8}$$

The predictive uncertainty, i.e. the covariance matrix of f_*, does not depend on y, but only on the dependencies induced by the covariance as a function of x_* and $\mathbf{X}$. This can be generalized to an arbitrary set of new inputs $\mathbf{X}_*$, meaning that the posterior process $f \mid D$ is again a Gaussian Process with posterior mean and covariance function

$$\begin{aligned} \hat{f}_* &= E\left(f_* \middle| D, x_*, K, \sigma^2\right) = k_*^T C^{-1} y \\ \sigma^2_{\hat{f}_*} &= k_{**} - k_*^T C^{-1} k_* \end{aligned} \tag{C9}$$

where $C = C(X, \theta_P, \sigma^2) = K(X, \theta_P) + \sigma^2 I_{m \times m}$.

The resulting posterior (C6) and the predictive distribution (C8) are of the same family of distributions as the prior (C5). The class of prior distributions with this property is called conjugate to a likelihood model (O'Hagan, 2004; Bernardo & Smith, 2006). The calculations are analytically tractable only for conjugate models with normal noise. For all other models the posterior and the predictive distribution cannot be computed analytically, so techniques for approximate inference have been used, for example, Markov Chain Monte Carlo (MCMC) sampling techniques (Williams & Rasmussen, 1996; Neal, 1997; Kuss, 2006).

Covariance Function

The elements of the parameterized covariance matrix, $C(X, \theta_P, \sigma^2)$, are denoted $C_{ij} = C(x_i, x_j)$, and they are functions of the training input data $\mathbf{X}$, because these data determine the correlation between the training data outputs y. A suitable parametric form of the covariance function is:

$$C_{ij} = \theta_0 + \theta_1 \exp\left[-\frac{1}{2} \sum_{l=1}^{n} \frac{\left(x_i^{(l)} - x_j^{(l)}\right)^2}{r_l^2} \right] + \theta_2 \delta(i, j) + \sum_{l=1}^{n} \alpha_l x_i^{(l)} x_j^{(l)} \tag{C10}$$

where $x_i^{(l)}$ is the l^{th} dimension of the input vector, x_i.

The four terms in this equation are now briefly described.

i- A bias term θ_0 controlling the scale of the bias contribution to the covariance. The constant term θ_0, adds a constant offset to the estimated latent function value $f_* = f(x_*)$. This justifies assigning $m(X) = 0$ to the prior mean function (C5) without loss generality.

ii- The exponential term (involving θ_1 and *rl*) expresses our belief that inputs which are close to each other give rise to outputs which are close to each other or that are highly correlated; the *rl* hyperparameters allow a different distance measure for each input dimension, and θ_1 gives the overall scale of variations in the output space (local correlations).

iii- The third term involving the hyperparameter $\theta_2 = \sigma^2$, i.e., the variance of the noise model for the outputs and therefore only occurs in $Ci\,j$ when $i = j$.

iv- The fourth and last term characterizes the nonstationarity of the covariance function. It is a linear regression term and involve α_l, $l = 1, L, n$, these hyperparameters controlling the scale of the linear trends to the covariance.

Besides (C10), there are other forms of the covariance function which could be used. The only restriction is that the covariance matrix be positive definite. Abrahamsen has written a comprehensive survey on numerous valid covariance functions (Abrahamsen, 1997).

Hyperparameters Determination

The hyperparameters of the covariance function are not known in advance, and they must be determined using the training data. The literature has reported several approaches to hyperparameter estimation: Cross-Validation method (Wahba, 1990), Evidence Maximization (MacKay, 1992; Gibbs, 1997), Monte Carlo methods (Neal, 1997), Maximum Likelihood, and Maximum a Posteriori method (Rasmussen, 1996). In this work the last method is used, i.e., the Maximum a Posteriori (MAP) approach.

For implementation purpose, the hyperparameters vector is defined as:

$$\log\theta = [\log\theta_0, \log\theta_1, \log r_1, .., \log r_n, \log\theta_2, \log\alpha_1, .., \log\alpha_n]^T$$

The number of hyperparameters of the covariance function (C10) increases linearly with n, the dimension of the input space, i.e., $\#\theta = 2n + 3$. The likelihood of the parameters is $L(\theta) = p(D \mid C(\theta)) = p(D \mid \theta, C(.))$, where $C(.)$ specifies the form of covariance function. The maximum likelihood approach is to maximize $L(\theta)$ to yield the optimum hyperparameters. An improvement over this is to incorporate a prior, $p(\theta)$, on the hyperparameters. By Bayes theorem, the posterior probability of the hyperparameters, given the training data, is:

$$p(\theta \mid D, C(.)) = \frac{p(D \mid \theta, C(.))p(\theta \mid X, C(.))}{p(D \mid C(.))} \qquad \text{(C11)}$$

Maximization of $p(\theta \mid D, C(.))$ is known as the MAP approach, which is a Bayesian version of maximum likelihood estimation. It is possible to analytically express the posterior $p(\theta \mid D, C(.))$ and its partial derivatives with respect to hyperparameters $\{\theta i\}$ as derived, for example, in Mardia & Marshall (1984). Let $L=\log p(\theta \mid D, C(.))$, then:

$$\begin{cases} L = -\frac{1}{2}\log|C| - \frac{1}{2}y^T C^{-1} y - \frac{m}{2}\log 2\pi + \log p(\theta) + c \\ \frac{\partial L}{\partial \theta_i} = -\frac{1}{2}\text{trace}\left(C^{-1}\frac{\partial C}{\partial \theta_i}\right) + \frac{1}{2}y^T C^{-1}\frac{\partial C}{\partial \theta_i}C^{-1}y + \frac{\partial \log p(\theta)}{\partial \theta_i} \end{cases} \quad \text{(C12)}$$

where the prior on $\boldsymbol{\theta}$, $p(\theta)$ is assumed to be independent of the other priors.
The MAP estimates are found using (C12) in a gradient descent, or conjugate gradient optimization techniques to locate a local maximum of the posterior distribution. This approach has the advantage that a reasonable approximation to a local maximum can be found with relatively few function and gradient evaluations. On the other hand, the posterior distribution is often multimodal, and the risk of using gradient optimization techniques is that the algorithm can keep trapped in bad local maxima. In order to reduce this problem, suitable priors can be assigned and multiple random restarts for the optimization routines can be fulfilled. The initialization of the hyperparameters is important, because improper initial values will make the partial derivatives of the likelihood very small, thus creating problems for the optimization algorithm.
In order to implement the algorithm (equations (C9), (C10), and (C12)) it was necessary to invert the covariance matrix $\mathbf{C}$. Any exact inversion method has an associated computational cost that is $O(m^3)$, moreover, direct inversion implementation can run into numerical problems because $\mathbf{C}$ is generally ill-conditioned, i.e., the condition number is large. In order to improve the condition number of the matrix inversion operation, C^{-1} can be computed indirectly by using Cholesky, LU, or SVD decomposition (note that the positive definiteness property of the covariance matrix is guaranteed). All these methods also require $O(m^3)$ operations (Golub & Van Loan, 1990).
It is important to observe that the optimization routines require in each step the evaluation of the gradient of the log likelihood, i.e., the computation of C^{-1} and so calculating gradients becomes time consuming for large training data sets. If m is large ($m > 10^3$), Skilling's approximate inversion methods, which are $O(m^2)$, can be used (Skilling, 1993).

5. References

Abrahamsen, P. (1997). A review of Gaussian random fields and correlation functions. *Technical Report 917*, Norwegian Computing Center, Oslo, Norway, 2nd edition.

Amicarelli A. Modelado, Identificación y Control de Bioprocesos. (2009). (In Spanish) *PhD thesis.* Universidad Nacional de San Juan Argentina – Instituto de Automática. ISBN978-987-05-7087-5, pp. 1-191.

Amicarelli, A.; di Sciascio, F.; Álvarez, H. & Ortiz, O. (2006). (In Spanish) Estimación de Biomasa en un proceso Batch: Aplicación a la producción de δ-endotoxinas de Bt. In: *XXII Interamerican congress of chemical engineering*.

Amicarelli, A., di Sciascio F., Toibero J. M. & Álvarez, H. (2010). Including dissolved oxygen dynamics to the Bacillus thuringiensis δ -endotoxins production process. *Brazilian Journal of Chemical Engineering*, Vol. 27, No. 01, pp. 41 – 62.

Anderson, B.D.O., & Moore, J.B. (1979). Optimal Filtering. Prentice-Hall, Englewood Cliffs, New Jersey. ISBN 0136381227 : 0136381227.

Aronson, A. I. (1993). The two faces of Bacillus thuringiensis: insecticidal proteins and post exponential survival. *Molecular Microbiology*, Vol. 7, pp. 489–496.

Atehortúa, P., Álvarez, H., & Orduz, S. (2006). Comments on: A sporulation kinetic model for batch growth of B. thuringiensis. *The Canadian Journal of Chemical Engineering*, Vol. 84, No. 03, pp. 386-388.

Atehortúa, P., Alvarez, H., & Orduz, S. (2007). Modeling of growth and sporulation of Bacillus thuringiensis in an intermittent fed batchculturewith total cell retention. *Bioprocess and Biosystems Engineering*, Vol. 30, pp. 447–456.

Bastin, G. & Dochain, D. (1990). On-line estimation and adaptive control of bioreactors. Amsterdam: Elsevier. ISBN-13: 978-0444884305.

Bernardo, J. M., & Smith, A. F. (2006). Bayesian Theory (2nd ed.) Chichester: Wiley.

Bishop C.M. (1995, reprint 2005). Neural Networks for Pattern Recognition, Oxford University Press.

Brawn R. , Hwang P. (1997) .Introduction to Random Signals and Applied Kalman Filtering. 3rd ed. John Wiley & Sons: New York, pp. 371-375.

Daum, F. (2005). Nonlinear Filters: Beyond the Kalman Filter. IEEE *Aerospace and Electronic Systems Magazine,* Vol. 20, No 8, Part 2, pp. 57- 69.

di Sciascio, F. & Amicarelli, A. (2008). Biomass Estimation in Batch Biotechnological Processes by Bayesian Gaussian Process Regression, *Computers and Chemical Engineering* Vol. 32, pp. 3264 – 3273.

Dochain, D. (2003). State and Parameter Estimation in Chemical and Biochemical Processes: A Tutorial. *Journal of Process Control*, Vol. 13, No. 8, pp. 801–818.

Ghribi, D., Zouari, N., Trabelsi, H., & Jaoua, S. (2007). Improvement of Bacillus thuringiensis delta-endotoxin production by overcome of carbon catabolite repression through adequate control of aeration. *Enzyme and Microbial Technology*, Vol. 40, No. 4, pp. 614–622.

Gibbs, M.N. (1997). Bayesian Gaussian Processes for Regression and Classification, PhD thesis, University of Cambridge.

Golub, G.H., Van Loan, C.F. (1990). Matrix Computation, Baltimore, John Hopkins University Press.

Hastie, T., Tibshirani, R., & Friedman, J.H. (2001). The Elements of Statistical Learning. Springer Verlag.

Haykin, S. (1999). Neural Networks: A comprehensive Foundation. Second Edition by Prentice-Hall, Inc., New Jersey.

Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory, Academic Press, Inc. New York.

Kuss, M. (2006). Gaussian Process Models for Robust Regression, Classification, and Reinforcement Learning. *PhD thesis*, Technische Universität Darmstadt.

Leal Ascencio, R. (2001). Artificial neural networks as a biomass virtual sensor for a batch process. *In Proceedings of the 2001 IEEE interational symposium on intelligent control*.

Li, B. (2003). Artificial neural network based software sensor for biomass during microorganism cultivation. *PhD thesis*. South China University of Technology.

Lind, I. (2006). Regressor and Structure Selection Uses of ANOVA in System Identification. *PhD thesis,* Department of Electrical Engineering, University of Linköping, Sweden.

Little, R., & Rubin, D. (2002). Statistical Analysis with Missing Data. New York, Wiley.

Liu, B. L., & Tzeng, Y. M. (2000). Caracterization study of the sporulation kinetics of Bacillus thuringiensis. *Biotechnology and Bioengineering*, Vol. 68, No 1, pp. 11–17.

Ljung, L. (2006). Some Aspects on Nonlinear System Identification. *14th IFAC Symposium on System Identification*, SYSID-2006, Newcastle, Australia.

MacKay, D.J. (1992). Bayesian interpolation. Neural Computation 4-33, pp. 415–447.

Mardia, K.V., & Marshall, R.J. (1984). Maximum Likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, Vol. 71, No 1, pp. 135-146.

Mannale, R. (2006). Comparison of Regressor Selection Methods in System Identification. PhD thesis, Department of Electrical Engineering, University of Linköping, Sweden.

Narendra, K., & Parthasarathy, K. (1990). Identification and Control of Dynamical Systems Using Neural Networks. *IEEE Transaction On Neural Networks.*

Neal, R. M. (1996). Bayesian learning for neural networks. Springer-Verlag New York. ISBN 0-387-94724-8.

Neal, R. M. (1997). Monte Carlo implementation of GP models for Bayesian regression and classification. *Tech. Rep*. No. 9702. Toronto: Department of Statistics, University of Toronto.

O'Hagan, A. (2004). Bayesian Inference (2nd ed.). Kendall's Advanced Theory of Statistics 2B. London: Edward Arnold.

Rasmussen, C.E. (1996).Evaluation of Gaussian Processes and Other Methods for Nonlinear Regression. *PhD thesis* University of Toronto.

Rivera, D., Margaritis, A., & De Lasa, H. (1999). A sporulation kinetic model for batch growth of B. thuringiensis. *Canadian Journal of Chemical Engineering*, Vol. 77, pp. 903–910.

Simon, D. (2006), Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches. John Wiley & Sons, Inc., Hoboken, New Jersey.

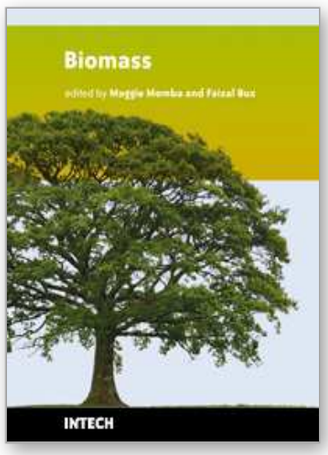
Skilling, J. (1993) Bayesian numerical analysis, In Physics and Probability, ed. by W. Grandy and P. Milonni, Cambridge University Press.

Starzak, M. & Bajpai, R. (1991). A structured model for vegetative growth and sporulation in Bacillus thuringiensis. *Applied Biochemistry and Biotechnology*, Vol. 28/29, pp. 699–718.

Vallejo, F., González, A., Posada, A., Restrepo, A., & Orduz, S. (1999). Production of Bacillus thuringiensis subsp. Medellín by batch and fed-batch culture. *Biotechnology Techniques*, Vol. 13, pp. 279–281.

Williams, K. I,. & Rasmussen, C. E. (1996) Gaussian Processes for Regression. MIT Press.

Wahba, G. (1990) Spline Models for Observational Data. Volume 59 of Series in Applied Mathematics. SIAM, Philadelphia.



Biomass

Edited by Maggy Ndombo Benteke Momba

ISBN 978-953-307-113-8
Hard cover, 202 pages
Publisher Sciyo
Published online 12, August, 2010
Published in print edition August, 2010

Due to demands placed on natural resources globally and subsequent deterioration of the environment, there is a need to source and develop appropriate technology to satisfy this requirement. For decades mankind has largely depended on natural resources such as fossil fuels to meet the ever increasing energy demands. Realizing the finite nature of these resources, emphasis is now shifting to investigating alternate energy source governed by environmentally friendly principles. The abundance of biomass and associated favorable techno-economics has recently changed global perceptions of harnessing biomass as a valuable resource rather than a waste. To this end this book aims to make a contribution to exploring further this area of biomass research and development in the form of a compilation of chapters and covering areas of ecological status of different types of biomass and the roles they play in ecosystems, current status of biomass utilization and deriving energy and other value added products from biomass. In this context biomass can be defined as large plants and trees and different groups of microorganisms. This book will serve as an invaluable resource for scientists and environmental managers in planning solutions for sustainable development.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Adriana Amicarelli (2010). On-line Biomass Estimation in a Batch Biotechnological Process: Bacillus thuringiensis δ - endotoxins production., Biomass, Maggy Ndombo Benteke Momba (Ed.), ISBN: 978-953-307-113-8, InTech, Available from: http://www.intechopen.com/books/biomass/on-line-biomass-estimation-in-a-batch-biotechnological-process-bacillus-thuringiensis-endotoxins-pro

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen