

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: $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ (0.03 g.L⁻¹), $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ (0.041 g.L⁻¹), KH_2PO_4 (0.5 g.L⁻¹), K_2HPO_4 (0.5 g.L⁻¹), $(\text{NH}_4)_2\text{SO}_4$ (1 g.L⁻¹), yeast extract (8 g.L⁻¹), $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ (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 $1 \cdot 10^7$ 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.

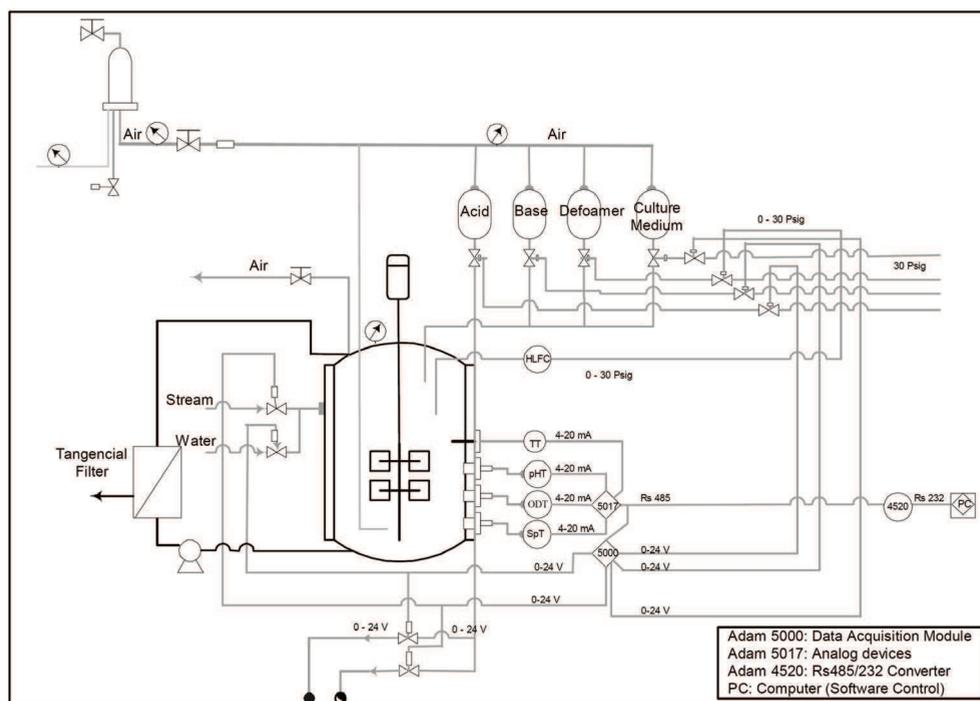


Fig. 1. Fermentation pilot plant scheme.

The temperature was maintained around 30°C by using an ON/OFF control; whereas the pH was fixed between 6.5 and 8.5. The air flow was set up at $1320\text{ [L}\cdot\text{h}^{-1}]$ and the agitation speed at 400 rpm. Manometric pressure in the reactor was set at $41,368\text{ Pa}$ 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).

$$\begin{bmatrix} X_V(k+1) \\ X_S(k+1) \\ S(k+1) \\ DO(k+1) \end{bmatrix} = \begin{bmatrix} ((\mu(k) - k_S(k) - k_e(k))Ts + 1) X_V(k) \\ k_S(k)X_V(k) Ts + X_S(k) \\ - \left(\frac{\mu(k)}{Y_{x/s}} + m_S \right) X_V(k) Ts + S(k) \\ (K_1 - K_2 Ts) X(k) - K_1 X(k+1) + DO(k) + K_3 Q_A Ts (DO^* - DO(k)) \end{bmatrix} \quad (1)$$

Where X_V is the vegetative cell concentration, X_S the sporulated cell concentration, $X = X_V + X_S$ is the total cell concentration ($X(k+1) = (\mu(k) - k_e(k)) Ts X_V(k) + X(k)$), 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 .

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

$$k_S(k) = k_{S\max} \left(\frac{1}{1 + e^{G_S(S(k) - P_S)}} \right) - k_{S\max} \left(\frac{1}{1 + e^{G_S(S_{\text{initial}} - P_S)}} \right) \quad (3)$$

$$k_e(k) = k_{e\max} \left(\frac{1}{1 + e^{G_e(Tsk - Pe)}} \right) - k_{e\max} \left(\frac{1}{1 + e^{G_e(t_{\text{initial}} - Pe)}} \right) \quad (4)$$

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

| Symbol | Description |
|--------------|---|
| s | Limiting substrate concentration $\left[\text{g. L}^{-1} \right]$ |
| T_s | Sampling time [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 $\left[\text{h}^{-1} \right]$ |
| K_3 | Ventilation constant $\left[\text{L}^{-1} \right]$ |
| DO^* | O ₂ 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.

| | $S_{\max} < 10 \text{ g.L}^{-1}$ | $10 \text{ g.L}^{-1} < S_{\max} < 20 \text{ g.L}^{-1}$ | $20 \text{ g.L}^{-1} < S_{\max} < 32 \text{ g.L}^{-1}$ | $S_{\max} > 32 \text{ g.L}^{-1}$ |
|--|----------------------------------|--|--|----------------------------------|
| $\mu_{\max} [\text{h}^{-1}]$ | 0.8 | 0.7 | 0.65 | 0.58 |
| $Y_{x/s} [\text{g.g}^{-1}]$ | 0.7 | 0.58 | 0.37 | 0.5 |
| $K_s [\text{g.L}^{-1}]$ | 0.5 | 2 | 3 | 4 |
| $K_o [\text{g.L}^{-1}]$ | 1×10^{-4} | 1×10^{-4} | 1×10^{-4} | 1×10^{-4} |
| ms $\left[\text{g} \cdot \left[\text{g} \cdot \text{h}^{-1} \right]^{-1} \right]$ | 5×10^{-3} | 5×10^{-3} | 5×10^{-3} | 5×10^{-3} |
| $k_{s\max} [\text{h}^{-1}]$ | 0.5 | 0.5 | 0.5 | 0.5 |
| $G_s \left[\text{g.L}^{-1} \right]^{-1}$ | 1 | 1 | 1 | 1 |
| $P_s [\text{g.L}^{-1}]$ | 1 | 1 | 1 | 1 |
| $k_{e\max} [\text{h}^{-1}]$ | 0.1 | 0.1 | 0.1 | 0.1 |
| $G_e [\text{h}^{-1}]$ | 5 | 5 | 5 | 5 |
| $P_e [\text{h}]$ | 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} |
| $K_2 [\text{h}^{-1}]$ | 1.589×10^{-4} | 0.046×10^{-3} | 0.729×10^{-3} | 0.561×10^{-3} |
| $K_3 [\text{L}^{-1}]$ | 4.636×10^{-4} | 0.337×10^{-3} | 2.114×10^{-3} | 1.045×10^{-3} |
| $T_s [\text{h}]$ | 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 $T_s = 1/10$ 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 $n \times m$ matrix X , and m scalar noisy observed outputs $\{y_i\}$ collected in a vector y .

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

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

$$y_i = f(x_i) + \varepsilon_i \quad (6)$$

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:

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

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

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

A Gaussian Process is fully specified by a mean function $m(X) = [m(x_1), \dots, m(x_m)]^T$ 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:

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

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 $\#0 = 5$ hyperparameters, and the matrix C are determined recursively through:

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

and

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

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

$$\begin{aligned} \hat{f}_* &= E\left(f_* \mid D, x_*, K, \sigma^2\right) = k_*^T C^{-1} y \\ \sigma_{\hat{f}_*}^2 &= k_{**} - k_*^T C^{-1} k_* \end{aligned} \quad (12)$$

the latent functions $\hat{X}_* = \{\hat{X}_*\} = \{\hat{X}(t^*)\}$ and the variance $\sigma_{\hat{X}_*}^2$ 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).

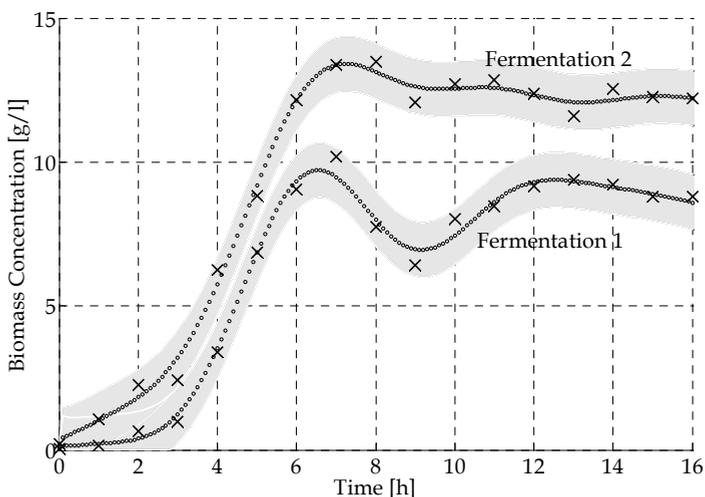


Fig. 2. Example of completion of biomass missing data for Fermentation 1 and Fermentation 2. The crosses being the biomass concentration measurements (training data), the small circles represent the biomass estimated (virtual filtered biomass measurements), and the grey region depicts the 95% confidence 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

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

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), $X(k-1) = X_v(k-1) + X_s(k-1)$ were used. Figure 3 shows the model structure for the phenomenological biomass estimator.

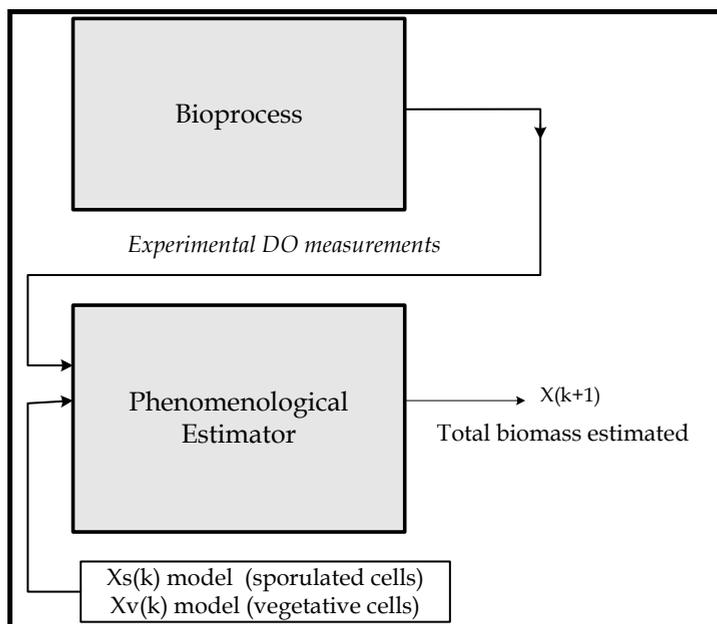


Fig. 3. Simulated output model structure for the phenomenological biomass estimator.

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).

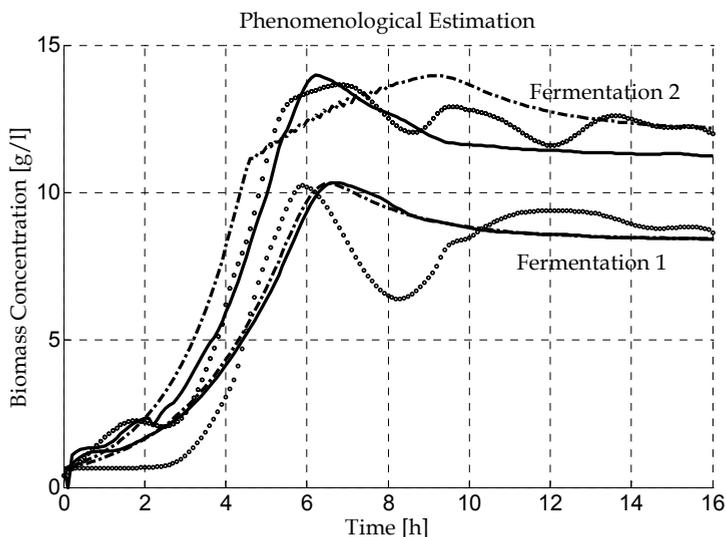


Fig. 4. Biomass estimator performance. The dash-dot line describes the behavior of biomass when considering the model (1); the solid line depicts the phenomenological estimator behavior based on DO dynamics; and the real biomass measurements are represented by small circles.

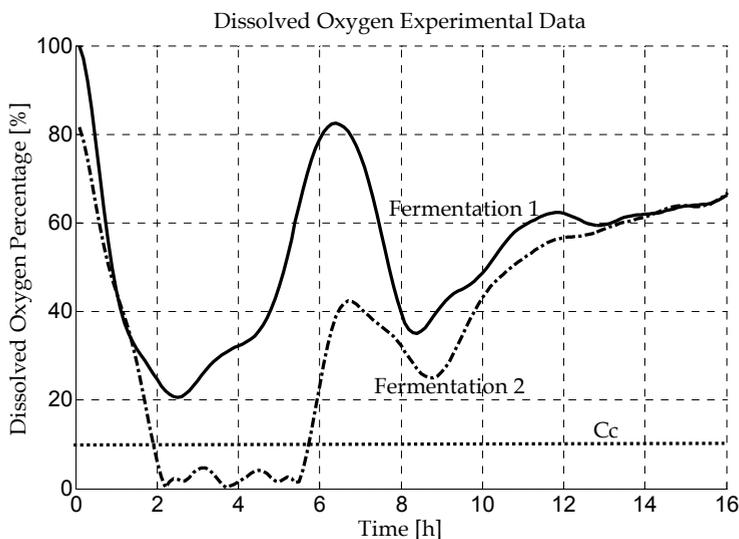


Fig. 5. Dissolved Oxygen experimental data. The solid line describes the Dissolved Oxygen behavior for the Fermentation 1; the dash-dot line depicts the Dissolved Oxygen behavior for the Fermentation 2 and the dotted line corresponds to the percentage of Dissolved Oxygen for the critical Dissolved Oxygen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

$$x(k+1) = f(x(k), u(k), k) + w(k) \quad (13)$$

Measurements model:

$$y(k) = h(x(k), k) + v(k) \quad (14)$$

Where $f(x, \times)$ is the state transition function; $h(x, \times)$ 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 (\sim) 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 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 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).

$$x(k+1) = f(x(k), k) \quad (15)$$

The system state vector is $x(k) = [X_v(k) \ X_s(k) \ S(k) \ DO(k)]^T$, the input vector is $u(k) = 0$ (the bioprocess has no external input), and the bioprocess outputs (observation vector) is $y(k) = [S(k) \ DO(k)]^T$ (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:

$$y(k) = Hx(k) \quad (16)$$

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

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.

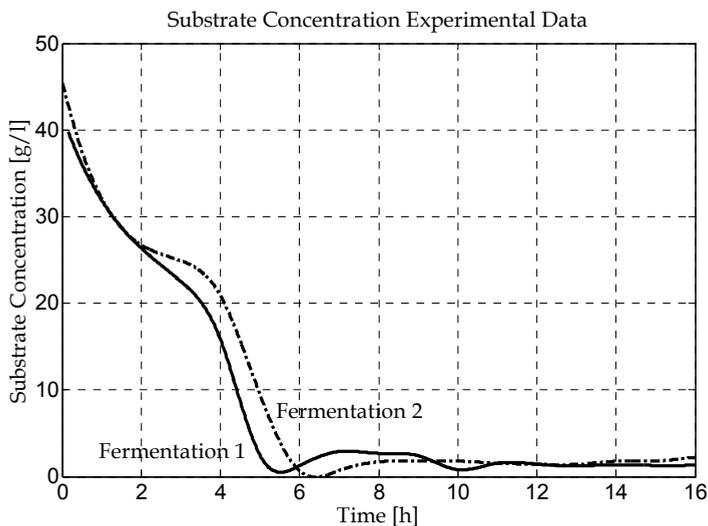


Fig. 6. Substrate Concentration experimental data. The solid line describes the Substrate Concentration for the Fermentation 1 and the dash-dot line depicts the Substrate Concentration for the Fermentation 2.

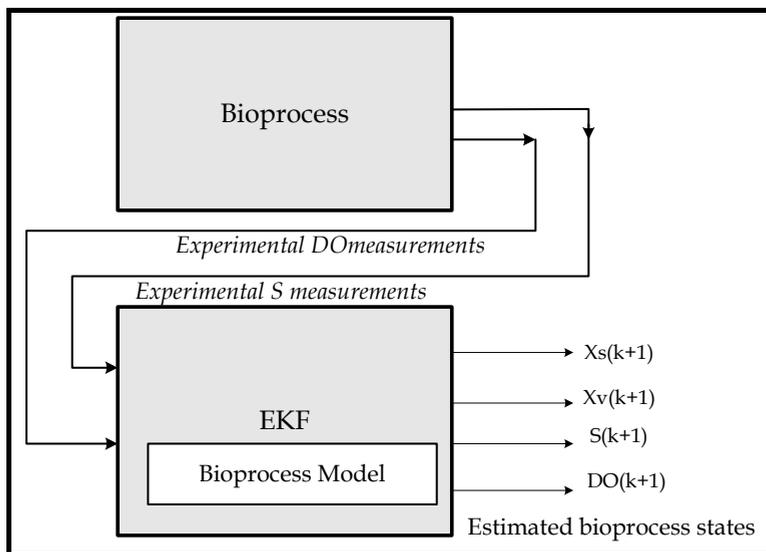


Fig. 7. Simulated output model structure for EKF biomass estimator.

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

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

where:

$$\alpha = \text{DOmax} / (\text{S max} + \text{DOmax}) \quad \beta = \text{S max} / (\text{Smax} + \text{DOmax})$$

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 $\hat{x}(k-1|k-1)$.

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

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

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

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

Finally, initializing the elements of the matrices P , Q and 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 Q and 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.

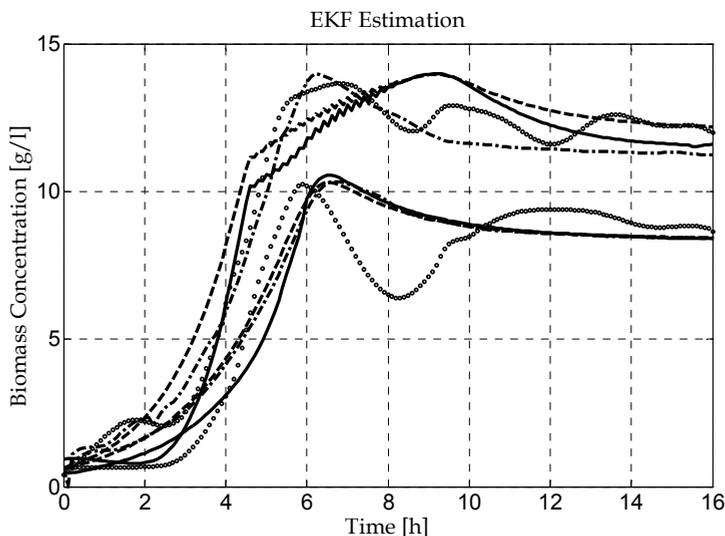


Fig. 8. Biomass estimator performance. The dashed line describes the biomass evolution obtained from the original model (1); the solid line depicts the EKF behavior, the dashed-dotted line depicts the phenomenological estimator behavior based on DO dynamics; and the real biomass measurements are represented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

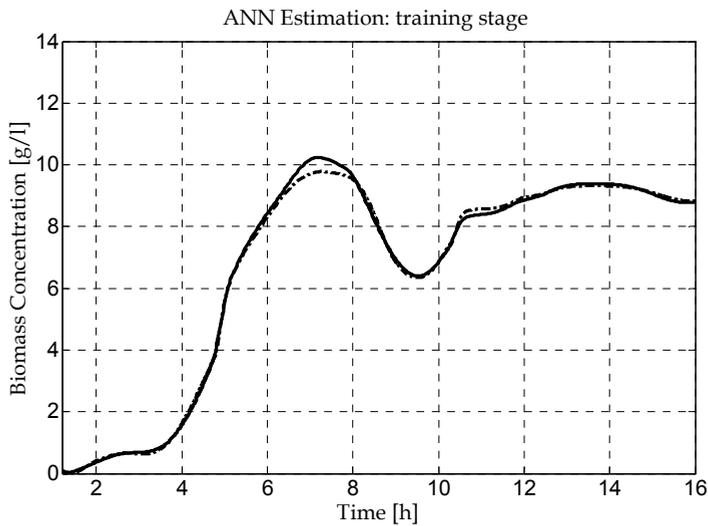


Fig. 9. Biomass estimator performance. The dashed line describes the biomass evolution obtained by the ANN in the training stage and the real biomass measurements are represented by the solid line. The perceptual training error $e = 0.16\%$.

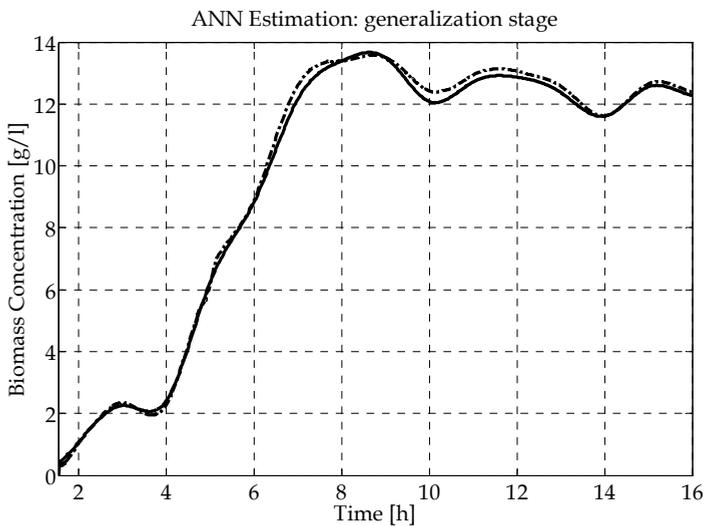


Fig. 10. Biomass estimator performance. The dashed line describes the biomass evolution obtained from the ANN in the generalization stage and the real biomass measurements are represented by the solid line. The perceptual generalization error $e = 0.25\%$.

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:

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

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.

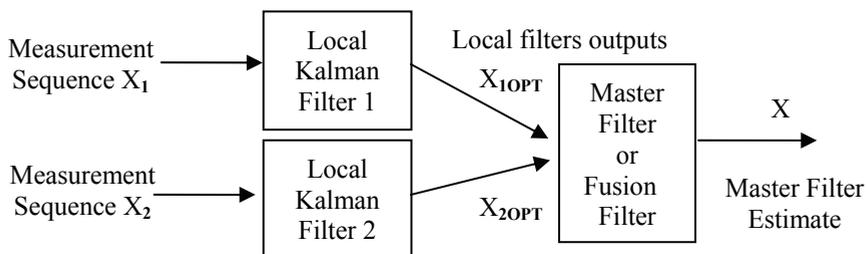


Fig. 11. Fusion scheme through a Decentralized Kalman Filter.

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

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

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

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

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

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

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

Thank You for previewing this eBook

You can read the full version of this eBook in different formats:

- HTML (Free /Available to everyone)
- PDF / TXT (Available to V.I.P. members. Free Standard members can access up to 5 PDF/TXT eBooks per month each month)
- Epub & Mobipocket (Exclusive to V.I.P. members)

To download this full book, simply select the format you desire below

