

# Estimation of Substrate and Biomass Concentrations in a Chemostat using an Extended Kalman Filter

Oussama Hadj-Abdelkader<sup>1</sup>, Amine Hadj-Abdelkader<sup>2\*</sup>

<sup>1</sup>Laboratoire d'Automatique de Tlemcen  
Université de Tlemcen  
BP119 13000 Tlemcen Algeria  
E-mail: [hadjabdelkader.oussama@gmail.com](mailto:hadjabdelkader.oussama@gmail.com)

<sup>2</sup>AutoMed Laboratoire d'Automatique de Tlemcen  
Université de Tlemcen  
BP119 13000 Tlemcen, Algeria  
E-mail: [amine.hadj@gmail.com](mailto:amine.hadj@gmail.com)

\*Corresponding author

Received: December 19, 2017

Accepted: September 05, 2018

Published: June 30, 2019

**Abstract:** This paper presents the estimation of substrate and biomass concentrations inside a Chemostat used for waste-water treatment. These concentrations represent the state variables of the process model. Most research in this field used only deterministic models, not accounting for uncertainties and noises on the states and on the output. Hence, the estimation of these concentrations may not be sufficiently accurate. For a more realistic description, we used here a stochastic formulation. Unlike the other research works, we used a stochastic differential equations (SDE) model which provides a better representation of the system in his natural processing scale. This model also includes the aleatory effects in the process which had been discarded in the other works. We then deal with the state estimation problem using an Extended Kalman filter, which proceeds with a linearization of the model around a deterministic trajectory. The classical prediction and update steps of the filter are then carried-out and led to good results. Notice that the system in study has some interesting properties such as discrete-time observations, high noise intensities and slow-time evolution. Results are presented, discussed and compared with the related state-of-the-art researches.

**Keywords:** Extended Kalman filter, State estimation, Chemostat, Waste-water treatment, Stochastic differential equations.

## Introduction

Given the increasing need of the world population to clean water and the decline in water resources in a large part of the planet, one of the most promising solutions to this problem is the waste-water treatment. This treatment involves two steps: (i) a mechanical treatment including the filtration, the grit removal, the degreasing and the decantation and (ii) a biological treatment using mainly the bioreactor and the anaerobic digester [3]. We are more interested in the bioreactor: it is a vessel in which a chemical process is carried-out involving micro-organisms such as bacteria breaking-down polluting organic substances contained in the water. The laboratory device representing the bioreactor is the Chemostat [28] (Fig. 1), which we attempt to stabilize around an equilibrium in order to avoid the explosion or the extinction of the bacterial population.

Before the stabilization, it is necessary to know the biomass (bacteria) and the substrate (organic substances) concentrations inside the Chemostat. These quantities will represent the state variables in the Chemostat model. These concentrations are not directly measurable at the system's output, it is then necessary to use a software sensor to estimate their values. A first solution is the use of state observers: these methods require generally a relatively accurate knowledge of the system and also require the system to be observable. A detailed study about this last condition

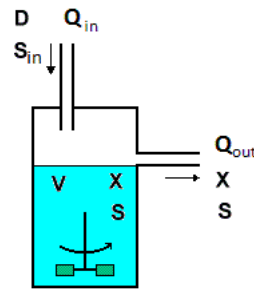


Fig. 1 The Chemostat (from [22])

is established by [10]. The observer based methods are applied at a macroscopic (large-size) scale i.e. in the case of large population size where the randomly-occurring (stochastic) effects are not visible on the system. However, at a mesoscopic (medium-size) or microscopic (small-size) scale, aleatory state variations appear on the system and need to be taken into account [5]. Hence, the system at this level of description is modeled by Stochastic Differential Equations (SDEs). In addition, the measurements are constantly subject to noises and uncertainties, which need to be modeled in the output equation.

We propose in this paper to use stochastic estimation methods in order to determine the state variables of the Chemostat. These methods allow to estimate both the state mean and covariance matrix. The Kalman Filters (KF) are the most common methods for this purpose. They use series of output measurements observed over time and containing statistical noise and other inaccuracies to generate estimates of the unmeasured system states. For nonlinear estimation problems, the Extended Kalman Filter (EKF) uses the same approach along with Taylor series development to linearize the system around a stable equilibrium. The EKF algorithm works in two steps: Prediction and Update. In the prediction step, the algorithm produces estimates of the current state variables means together with their covariance matrix (the estimation uncertainties). Notice that the EKF linearizes the system about an estimate of the mean before proceeding to this prediction step. In the update step, when the next measurement value becomes available (necessarily affected with some amount of noise), the previous estimations are updated using a weighted average called the Kalman gain. The EKF algorithm is recursive, it can run on-line, using only the present measurements and the previously calculated state estimations and their uncertainty matrix.

It is reported in some research work that the EKF may have some robustness issues against nonlinearities considering the abrupt changes in the system states that may occur during the system's evolution [15]. However, these incidents are not very common in this type of biological systems which makes this algorithm a good choice for the state estimation of this process.

This paper is organized as follows: in the next section, we carry out a brief and concise comparison of previous work related to this estimation problem. The section "The Chemostat model" contains a description of the Chemostat stochastic model and its simulation using the Euler-Maruyama scheme. In section "The Extended Kalman Filter", we explain the linearization procedure, the prediction and the update steps and we present the application of the EKF algorithm to the system in study. Finally, section "Results and Discussion" shows the simulation results and their discussion and comparison with the related researches. We conclude then about the efficiency of this algorithm in this context.

## Related work

There exist a tremendous amount of work on this topic from which we will mention the most similar research from the optimization based methods to the Bayesian methods of state estimation for the deterministic state-space model of the Chemostat. A state of the art on the most common methods for bioreactors is established by [7]. This paper gave a special attention to the adaptive observers, the EKF and artificial neural networks, without giving any concrete application of these methods on any particular model.

There exist, in fact, many applications of these techniques in the literature. For instance, in [8], the authors used an invariant observer and they proved its convergence for the model in hand. In [23], simple observer-based estimators were used for similar aerobic fermentation processes for which the models were linearized, thus ensuring convergence in the neighborhood of the equilibrium. The authors in [12] developed a KF based approach for a deterministic model of a Chemostat used for microbial cultivation, including zero-order exponential memory functions in the expressions of the specific growth rate.

The EKF algorithm appears in both [16, 26] and [24, 27]. In the first works, it was used for state and parameter estimation of a bioreactor using a deterministic model with a normal output noise, whereas in the lasts, it was used for state estimation of Fed-batch cultivation processes, taking place in a Chemostat. Finally, neural network based estimators were used in [25] for the estimation of the biomass concentration and the specific growth rate and were designed to work independently of each-other, while still being implemented on the deterministic versions of the Chemostat model.

On the stochastic version of the Chemostat model, the authors in [4] used a Bootstrap Particle Filter (PF) for the same state estimation purposes. In his implementation, the authors considered two cases: a high frequency observations test and a low frequency one. Good and similar results were found in both cases. However, this method may not be suited for practical implementations where the computation resources are usually very limited. In this paper we aim to obtain better/similar results as in [4] with less computational cost. Table 1 gives a summary of these related research works.

## The Chemostat model

We consider a stochastic model of the Chemostat given by two stochastic differential equations (1). This model was first introduced by [5]. More precisely, it is a pure jump model (continuous-time Markov chain) approximated by a diffusion (normal) process  $X_t = (B_t, S_t)^T$  which represents the solution of this SDE:

$$\begin{cases} dB_t = (\mu(S_t) - D)B_t dt + c_1 \sqrt{B_t} dW_t^1, \\ dS_t = D(s_{in} - S_t) dt - k_{sc} \mu(S_t) B_t dt + c_2 \sqrt{S_t} dW_t^2. \end{cases} \quad (1)$$

In this model, the state variables  $B_t$  and  $S_t$  are the biomass and substrate concentrations at time  $t$ ,  $\mu(S_t) = \mu_{max} \frac{S_t}{K_s + S_t}$  is a Monod type specific growth function [21], where  $\mu_{max}$  is the maximum growth rate and  $K_s$  is the half-saturation constant,  $s_{in}$  is the input substrate concentration,  $D$  is the dilution rate,  $k_{sc}$  is the yield coefficient,  $W_t^1$  and  $W_t^2$  are two mutually independent Brownian motions (Wiener processes) that are also independent of state's initial conditions  $X_0$ ,  $c_1$  and  $c_2$  are the noise intensities. We suppose  $B_t \geq 0$ ,  $S_t \geq 0$  for  $t \in [0, T]$ .

Table 1. Summary of Estimation Results in the Literature (NL : Nonlinear, D: Deterministic).

Author	Model	Estimation Method	Convergence Time	Estimation Error	Properties
[8]	NL. D. Chemostat Model	Invariant Observer	$t \approx 2$ days for both Substrate and Biomass	0.01	Adjustable convergence time Robust against parameter variations
[2]	NL. D. Chemostat Model with Contois Kinetics [6]	Asymptotic Observer	$t \approx 2$ days	$< 0.01$	Convergence speed dependent upon Dilution Rate Could perform in the case of partial Kinetics model knowledge
[11]	NL. D. Chemostat Model with Contois Kinetics [6]	High Gain Observer	$t \approx 3$ days	$< 0.01$	Performs badly in the presence of noisy measurements
[14] [13]	NL. D. Chemostat Model with Activated Sludge Process	Interval Observers	$t \approx 6$ days	$< 0.2$	Need the cooperativity property Initial intervals of the states must be known Convergence is obtained even for large initial intervals
[4]	NL. Stochastic Chemostat Model	Particle Filtering	$t \approx 1$ day	$< 0.1$	High computational cost
[26]	NL. D. Chemostat Model with normal output noises	EKF	$t < 1$ day	$< 0.01$	Very sensitive against far initial conditions

For the model output, we suppose that only the substrate concentration  $S_{t_k}$  is measured at discrete time instants  $t_k = k\Delta$  where  $\Delta$  is the observation time step. This output is subject to a measurement noise  $v_k$  of standard deviation  $\sigma$  which is supposed to be proportional to  $S_{t_k}$  [4]. This leads to the output Eq. (2).

$$y_k = S_{t_k} + \sigma S_{t_k} v_k, \quad (2)$$

where  $v_k \stackrel{iid}{\sim} \mathcal{N}(0, 1)$  (iid – independent and identically distributed) and  $\sigma$  is the noise intensity.

### System simulation

To simulate the System (1), we use an Euler-Maruyama scheme [19]. The simulation time step  $\delta$  is chosen small enough to have a good approximation of the integral in the system equations but not too small to avoid high computational cost. For a simulation time  $T$ , we perform  $N$  iterations of system simulation between every two observations  $y_k$  with the simulation step  $\delta = T / (NN_{obs})$ , and we compute an output value  $y_k$  for every step  $\Delta = T / N_{obs}$  where  $N_{obs}$  is the number of observations in the time interval  $[0, T]$ .

The Brownian motion terms are approximated by:

$$dW_{t_n} = W_{t_n} - W_{t_n - \delta} \quad (3)$$

thus

$$dW_{t_n} = \sqrt{\delta} w_{t_n}, \quad (4)$$

where  $w_{t_n} \stackrel{iid}{\sim} \mathcal{N}(0, Q)$  and  $Q$  is the identity matrix.

The Euler-Maruyama approximation of System (1) is given by (5):

$$\begin{cases} B_{t_n} = B_{t_{n-1}} + (\mu(S_{t_{n-1}}) - D)B_{t_{n-1}} \delta + c_1 \sqrt{B_{t_{n-1}}} \sqrt{\delta} w_{t_n}^1, \\ S_{t_n} = S_{t_{n-1}} + (D(s_{in} - S_{t_{n-1}}) - k_{sc}\mu(S_{t_{n-1}})B_{t_{n-1}}) \delta + c_2 \sqrt{S_{t_{n-1}}} \sqrt{\delta} w_{t_n}^2, \end{cases} \quad (5)$$

where  $t_n = n\delta$  and  $w_{t_n}^1, w_{t_n}^2$  are normally distributed random variables with 0 mean and variance 1.

Note that only the non-negative solutions are taken into account. The values are set to 0 whenever they cross the time axis. The simulation algorithm is given in Algorithm 1.

### The Extended Kalman filter

The Extended Kalman Filter algorithm contains two steps:

#### Prediction step

In order to carry out the EKF prediction for system (1), we need to linearize it around a nominal deterministic trajectory  $x(t)$  and to compute its predicted expectation  $\mathbb{E}[X_{t_{n+1}}]$  and its predicted variance  $\text{var}[X_{t_{n+1}}]$ .

This system is of the form (6):

$$dX_t = f(X_t) dt + g(X_t) dW_t. \quad (6)$$

---

**Algorithm 1** Simulation of the Chemostat model using Euler-Maruyama scheme.
 

---

```

 $v_0, \dots, v_{N_{obs}} \sim N(0, 1)$ 
 $w_0^1, \dots, w_{N^*N_{obs}}^1 \sim N(0, 1)$ 
 $w_0^2, \dots, w_{N^*N_{obs}}^2 \sim N(0, 1)$ 
# initialization
 $(B_{t_0}, S_{t_0}) \sim N(\mu_0, Q_0)$ 
# iterations
For  $k = 0, \dots, N_{obs}$  do
    For  $n = 1, \dots, N$  do
         $\mu = \mu_{max} \frac{S_{t_{n-1}}}{K_s + S_{t_{n-1}}}$ 
         $B_{t_n} = \max(0, B_{t_{n-1}} + (\mu - D)B_{t_{n-1}}\delta + c_1\sqrt{B_{t_{n-1}}}\sqrt{\delta}w_n^1)$ 
         $S_{t_n} = \max(0, S_{t_{n-1}} - k_{sc}\mu B_{t_{n-1}}\delta + D(s_{in} - S_{t_{n-1}})\delta + c_2\sqrt{S_{t_{n-1}}}\sqrt{\delta}w_n^2)$ 
    End For
     $S_k = S_{t_n}$ 
     $y_k = S_k + \sigma S_k v_k$ 
End For
    
```

---

The Euler-Maruyama discretization for system (6) is:

$$X_{t_{n+1}} = X_{t_n} + f(X_{t_n})\delta + g(X_{t_n})\sqrt{\delta}w_{t_n} \quad (7)$$

with  $\sqrt{\delta}w_{t_n} = W_{t_{n+1}} - W_{t_n}$ , this increment is normally distributed with 0 mean and variance  $\delta$ .

Suppose that the distribution of  $X_t$  is normal, and let:

$$\bar{X}_{t_n} = \mathbb{E}[X_{t_n}] \quad (8)$$

and

$$R_{t_n} = \text{Var}(X_{t_n}) = \mathbb{E}[(X_{t_n} - \bar{X}_{t_n})(X_{t_n} - \bar{X}_{t_n})^*]. \quad (9)$$

First, we compute the expectation  $\mathbb{E}[X_{t_{n+1}}]$  of  $X_{t_{n+1}}$  using Eq. (7):

$$\bar{X}_{t_{n+1}} = \mathbb{E}[X_{t_{n+1}}] = \bar{X}_{t_n} + \mathbb{E}[f(X_{t_n})\delta] + \mathbb{E}[g(X_{t_n})\sqrt{\delta}w_{t_n}] \quad (10)$$

since  $g(X_{t_n})$  and  $w_{t_n}$  are independent, we get

$$\bar{X}_{t_{n+1}} = \bar{X}_{t_n} + \mathbb{E}[f(X_{t_n})]\delta + \mathbb{E}[g(X_{t_n})]\mathbb{E}[w_{t_n}]\sqrt{\delta}. \quad (11)$$

Since the mean of  $w_{t_n}$  is 0, this last equation is reduced to:

$$\bar{X}_{t_{n+1}} = \bar{X}_{t_n} + \mathbb{E}[f(X_{t_n})]\delta \quad (12)$$

To compute  $\mathbb{E}[f(X_{t_n})]$ , we linearize  $f(X_{t_n})$  around a given nominal deterministic trajectory  $x(t_n)$  using the Taylor-series development:

$$f(X_{t_n}) \simeq f(x(t_n)) + (X_{t_n} - x(t_n))\nabla f(x(t_n)). \quad (13)$$

We thus get:

$$\mathbb{E}[f(X_{t_n})] \simeq f(x(t_n)) + (\bar{X}_{t_n} - x(t_n))\nabla f(x(t_n)). \quad (14)$$

By replacing Eq. (14) in Eq. (12), we get:

$$\bar{X}_{t_{n+1}} = \bar{X}_{t_n} + f(x(t_n)) \delta + (\bar{X}_{t_n} - x(t_n)) \nabla f(x(t_n)) \delta. \quad (15)$$

If we take the nominal trajectory  $x(t_n)$  as the mean of  $X_{t_n}$ , that is  $x(t_n) = \bar{X}_{t_n}$ , then Eq. (15) is reduced to :

$$\bar{X}_{t_{n+1}} = \bar{X}_{t_n} + f(\bar{X}_{t_n}) \delta. \quad (16)$$

The next step is to compute the variance  $R_{t_{n+1}} = \text{var}[X_{t_{n+1}}]$  of  $X_{t_{n+1}}$ :

$$R_{t_{n+1}} = \mathbb{E}[(X_{t_{n+1}} - \bar{X}_{t_{n+1}})(X_{t_{n+1}} - \bar{X}_{t_{n+1}})^*]. \quad (17)$$

First, we need to compute  $(X_{t_{n+1}} - \bar{X}_{t_{n+1}})$  and  $(X_{t_{n+1}} - \bar{X}_{t_{n+1}})^*$ :

$$\begin{aligned} (X_{t_{n+1}} - \bar{X}_{t_{n+1}}) &= (X_{t_n} + f(X_{t_n}) \delta + g(X_{t_n}) \sqrt{\delta} w_{t_n}) \\ &\quad - (\bar{X}_{t_n} + f(x(t_n)) \delta + (\bar{X}_{t_n} - x(t_n)) \nabla f(x(t_n)) \delta). \end{aligned} \quad (18)$$

This Eq. (18) is re-organized under the following form:

$$\begin{aligned} (X_{t_{n+1}} - \bar{X}_{t_{n+1}}) &= X_{t_n} - \bar{X}_{t_n} + (f(X_{t_n}) - f(x(t_n))) \delta \\ &\quad + \sqrt{\delta} g(X_{t_n}) w_{t_n} - (\bar{X}_{t_n} - x(t_n)) \nabla f(x(t_n)) \delta. \end{aligned} \quad (19)$$

In this equation,  $f(X_{t_n})$  is replaced by its linearization – Eq. (13), which leads to Eq. (20):

$$\begin{aligned} (X_{t_{n+1}} - \bar{X}_{t_{n+1}}) &= X_{t_n} - \bar{X}_{t_n} + ((X_{t_n} - x(t_n)) \nabla f(x(t_n))) \delta \\ &\quad + \sqrt{\delta} g(X_{t_n}) w_{t_n} - (\bar{X}_{t_n} - x(t_n)) \nabla f(x(t_n)) \delta. \end{aligned} \quad (20)$$

By taking the nominal trajectory  $x(t_n) = \bar{X}_{t_n}$  as above, we get :

$$(X_{t_{n+1}} - \bar{X}_{t_{n+1}}) = X_{t_n} - \bar{X}_{t_n} + (X_{t_n} - \bar{X}_{t_n}) \nabla f(\bar{X}_{t_n}) \delta + \sqrt{\delta} g(X_{t_n}) w_{t_n}. \quad (21)$$

Next, we compute the transposed of this last quantity:

$$(X_{t_{n+1}} - \bar{X}_{t_{n+1}})^* = (X_{t_n} - \bar{X}_{t_n})^* + \nabla f(\bar{X}_{t_n})^* (X_{t_n} - \bar{X}_{t_n})^* \delta + \sqrt{\delta} w_{t_n}^* g(X_{t_n})^*. \quad (22)$$

Finally, after computing the product of Eq. (21) and Eq. (22) and the expectation of the result, we get the expression of the variance  $R_{t_{n+1}}$ :

$$R_{t_{n+1}} = R_{t_n} + R_{t_n} \nabla f(\bar{X}_{t_n})^* \delta + \nabla f(\bar{X}_{t_n}) R_{t_n} \delta + g(\bar{X}_{t_n}) \mathbb{E}[w_{t_n} w_{t_n}^*] g(\bar{X}_{t_n})^* \delta. \quad (23)$$

Note that all the other terms are canceled because of either  $\mathbb{E}[(X_{t_n} - \bar{X}_{t_n})] = 0$  or  $\mathbb{E}[w_{t_n}] = 0$ .

Since  $\mathbb{E}[w_{t_n} w_{t_n}^*] = \text{var}(w_{t_n}) = 1$ , we get:

$$R_{t_{n+1}} = R_{t_n} + R_{t_n} F^* \delta + F R_{t_n} \delta + g(\bar{X}_{t_n}) g(\bar{X}_{t_n})^* \delta \quad (24)$$

with  $F = \nabla f(\bar{X}_{t_n})$ .

To perform the prediction step of the EKF algorithm for our system we use the following notation:

$$f(X_{t_n}) = \begin{pmatrix} f_1(B_{t_n}, S_{t_n}) \\ f_2(B_{t_n}, S_{t_n}) \end{pmatrix}, \quad (25)$$

where  $X_{t_n} = (B_{t_n}, S_{t_n})^*$  and  $f_1(B_{t_n}, S_{t_n}), f_2(B_{t_n}, S_{t_n})$  are given by:

$$f_1(B_{t_n}, S_{t_n}) = (\mu(S_{t_n}) - D)B_{t_n}, \quad (26)$$

$$f_2(B_{t_n}, S_{t_n}) = D(s_{in} - S_{t_n}) - k_{sc}\mu(S_{t_n})B_{t_n}. \quad (27)$$

Also let:

$$g(X_{t_n}) = \begin{pmatrix} g_1(B_{t_n}, S_{t_n}) \\ g_2(B_{t_n}, S_{t_n}) \end{pmatrix}, \quad (28)$$

where  $g_1(B_{t_n}, S_{t_n}), g_2(B_{t_n}, S_{t_n})$  are given by:

$$g_1(B_{t_n}, S_{t_n}) = c_1\sqrt{B_{t_n}}, \quad (29)$$

$$g_2(B_{t_n}, S_{t_n}) = c_2\sqrt{S_{t_n}}. \quad (30)$$

In the algorithms,  $\hat{X}_{t_n}^- = \begin{pmatrix} \hat{B}_{t_n}^- \\ \hat{S}_{t_n}^- \end{pmatrix}$  and  $R_{t_n}^-$  denote, respectively, the predicted mean values of  $X_{t_n}$  and their covariance matrix at time  $t_n$ . Whereas  $\hat{X}_{t_n} = \begin{pmatrix} \hat{B}_{t_n} \\ \hat{S}_{t_n} \end{pmatrix}$  and  $R_{t_n}$  denote the estimated mean values of  $X_{t_n}$  and their covariance at time  $t_n$ . Furthermore,  $Q^w = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  and  $Q^v = 1$  denote, respectively, the state noise covariance matrix and the output noise variance.

The following notation is also used in the algorithms:

$$F_{t_n} = \nabla f(X_{t_n}) = \begin{bmatrix} \mu(S_{t_n}) - D & \mu'(S_{t_n})B_{t_n} \\ -k_{cs}\mu(S_{t_n}) & -D - k_{cs}\mu'(S_{t_n})B_{t_n} \end{bmatrix}. \quad (31)$$

The EKF prediction step of system (1) is given by Algorithm 2.

### Update step

The update step of the EKF is given by Algorithm 3. In this step, the same method of the standard KF is used to calculate the updated values of the mean  $\mathbb{E}[X_{t_{n+1}}]$  and covariance matrix  $\text{var}[X_{t_{n+1}}]$ . Unfortunately, because of the nonlinearity between the state variable  $S_{t_k}$  and the noise variable  $v_k$  in the output Eq. (2), we can not apply the update step directly. We propose instead the following substitution for Eq. (2):

Let:

$$\tilde{y}_k = \log(y_k) \quad (32)$$

thus

$$\tilde{y}_k = \log(S_{t_k}(1 + \sigma v_k)) \quad (33)$$

by using the properties of the logarithm, we get:

$$\tilde{y}_k = \log(S_{t_k}) + \log(1 + \sigma v_k). \quad (34)$$



**Algorithm 2** Prediction step of the EKF for system (1) with output (2).

# initialization

$$\delta = T / (N * N_{obs})$$

$$X_{t_0} \sim N(\mu_0, Q_0)$$

$$\hat{X}_{t_0} \leftarrow \mu_0$$

$$R_{t_0} = Q_0$$

$$Q^w = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

# iterations

**For**  $k = 0, \dots, N_{obs}$  **do**

# prediction step

**For**  $n = 1, \dots, N$  **do**

$$\hat{B}_{t_n}^- \leftarrow \max(0, \hat{B}_{t_{n-1}} + f_1(\hat{B}_{t_{n-1}}, \hat{S}_{t_{n-1}}) \delta)$$

$$\hat{S}_{t_n}^- \leftarrow \max(0, \hat{S}_{t_{n-1}} + f_2(\hat{B}_{t_{n-1}}, \hat{S}_{t_{n-1}}) \delta)$$

$$R_{t_n}^- \leftarrow R_{t_{n-1}} + \left( R_{t_{n-1}} F_{t_{n-1}}^* + F_{t_{n-1}} R_{t_{n-1}} + g(X_{t_{n-1}}) Q^w g(X_{t_{n-1}})^* \right) \delta$$

$$\hat{B}_{t_n} \leftarrow \hat{B}_{t_n}^-$$

$$\hat{S}_{t_n} \leftarrow \hat{S}_{t_n}^-$$

$$R_{t_n} \leftarrow R_{t_n}^-$$

**End For**

**End For**

Using a first order Taylor series development of  $\log(1 + \sigma v_k)$  at point 1, we get:

$$\tilde{y}_k = \log(S_{t_k}) + \log(1) + \sigma v_k \log'(1). \quad (35)$$

Finally, we get:

$$\tilde{y}_k = \log(S_{t_k}) + \sigma v_k. \quad (36)$$

The following notation is used in the algorithms:

$$h(X_{t_k}) = \log(S_{t_k}) \quad (37)$$

and:

$$H_{t_k} = \nabla h(X_{t_k}) = \begin{bmatrix} 0 & \frac{1}{S_{t_k}} \end{bmatrix}. \quad (38)$$

In this update step, we use the equation  $\tilde{y}_k = h(X_{t_k}) + \sigma v_k$  instead of equation  $y_k = h(X_{t_k}, v_k)$  (that is Eq. (2)) to compute the matrix  $H_{t_k} = \nabla h(X_{t_k})$ . This last matrix is used later in the algorithm to compute the Kalman gain  $K_k$  and correct the predicted values of  $\hat{X}_{t_n}$  and  $R_{t_n}$ .

Notice that replacing  $y_k$  by  $\tilde{y}_k = \log(y_k)$  does not affect the estimation algorithm or its quality since the same information is acquired in both cases. This substitution is carried-out only in the update step: the system simulation still gives an output value  $y_k$ . However, this also implies considering  $\log(y_k)$  instead of  $y_k$  in the error term when updating the estimation of  $\hat{X}_{t_k}$ .

In order to present the correct structure of the EKF applied for this system, its complete listing is given by Algorithm 4.

---

**Algorithm 3** Update step of the EKF for system (1) with output (2).

---

 $Q^v = 1$  % output noise variance**For**  $k = 0, \dots, N_{obs}$  **do**

# update step

$$\hat{B}_{t_k}^- \leftarrow \hat{B}_{t_n}^-$$

$$\hat{S}_{t_k}^- \leftarrow \hat{S}_{t_n}^-$$

$$R_{t_k}^- \leftarrow R_{t_n}^-$$

$$K_k = R_{t_k}^- H_{t_k}^* (H_{t_k} R_{t_k}^- H_{t_k}^* + \sigma Q^v \sigma)^{-1} \# \text{ The Kalman gain}$$

$$\begin{bmatrix} \hat{B}_{t_k} \\ \hat{S}_{t_k} \end{bmatrix} \leftarrow \begin{bmatrix} \hat{B}_{t_k}^- \\ \hat{S}_{t_k}^- \end{bmatrix} + K_k (\log(y_k) - h(\hat{X}_{t_k}^-))$$

$$R_{t_k} \leftarrow (I - K_k H_{t_k}) R_{t_k}^-$$

**End For**

---

**Algorithm 4** State estimation for the Chemostat using the EKF algorithm for system (1) with output (2).

---

# initialization

$$\delta = T / (N * N_{obs})$$

$$X_{t_0} \sim N(\mu_0, Q_0)$$

$$\hat{X}_{t_0} \leftarrow \mu_0$$

$$R_{t_0} \leftarrow Q_0$$

$$Q^w = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Q^v = 1$$

# iterations

**For**  $k = 0, \dots, N_{obs}$  **do**

# prediction step

**For**  $n = 1, \dots, N$  **do**

$$\hat{B}_{t_n}^- \leftarrow \max(0, \hat{B}_{t_{n-1}} + f_1(\hat{B}_{t_{n-1}}, \hat{S}_{t_{n-1}}) \delta)$$

$$\hat{S}_{t_n}^- \leftarrow \max(0, \hat{S}_{t_{n-1}} + f_2(\hat{B}_{t_{n-1}}, \hat{S}_{t_{n-1}}) \delta)$$

$$R_{t_n}^- \leftarrow R_{t_{n-1}} + \left( R_{t_{n-1}} F_{t_{n-1}}^* + F_{t_{n-1}} R_{t_{n-1}} + g(X_{t_{n-1}}) Q^w g(X_{t_{n-1}})^* \right) \delta$$

$$\hat{B}_{t_n} \leftarrow \hat{B}_{t_n}^-$$

$$\hat{S}_{t_n} \leftarrow \hat{S}_{t_n}^-$$

$$R_{t_n} \leftarrow R_{t_n}^-$$

**End For**

# update step

$$\hat{B}_{t_k}^- \leftarrow \hat{B}_{t_n}^-$$

$$\hat{S}_{t_k}^- \leftarrow \hat{S}_{t_n}^-$$

$$R_{t_k}^- \leftarrow R_{t_n}^-$$

$$K_k = R_{t_k}^- H_{t_k}^* (H_{t_k} R_{t_k}^- H_{t_k}^* + \sigma Q^v \sigma)^{-1} \# \text{ The Kalman gain}$$

$$\begin{bmatrix} \hat{B}_{t_k} \\ \hat{S}_{t_k} \end{bmatrix} \leftarrow \begin{bmatrix} \hat{B}_{t_k}^- \\ \hat{S}_{t_k}^- \end{bmatrix} + K_k (\log(y_k) - h(\hat{X}_{t_k}^-))$$

$$R_{t_k} \leftarrow (I - K_k H_{t_k}) R_{t_k}^-$$

**End For**

---

## Results and discussion

For the simulations of the previously described algorithms, we took a simulation time interval of  $[0, 1000 \text{ h}]$  in which we had  $N_{obs} = 1000$  measurements, that is,  $N_{obs}$  iterations of update step in the EKF. Between every two successive output values, we had  $N = 10$  iterations of the system simulation and of the EKF prediction. Hence, the discretization time step is  $\delta = 0.1 \text{ h}$  and the final simulation time is  $T = 1000 \text{ h}$ . The model's parameters are:

- the substrate concentration at the input  $s_{in} = 100 \text{ g/l}$ ;
- the dilution rate  $D = 0.01 \text{ h}^{-1}$ ;
- the maximum growth rate  $\mu_{max} = 0.3 \text{ h}^{-1}$ ;
- the yield coefficient  $k_{sc} = 10$ ;
- the half saturation constant  $K_s = 10 \text{ mg/l}$ ;
- the state noise intensities  $c_1 = c_2 = 0.03$ ;
- the observation noise intensity  $\sigma = 0.2$ ;
- and the initial state distribution  $\mathbb{P}_{X_0}(dx) = \mathcal{N}(4, 2^2)$ .

The application of the EKF algorithm to the stochastic Chemostat model led to the estimation results of the biomass and substrate concentrations presented in Fig. 2. The EKF has, generally, a good estimation potential and, given reasonable noise intensities, high-frequency observations and close initial conditions, it will give an accurate estimation of the process variables. Similar results appear in related research work by [4] where a Bootstrap PF was applied to the same process in identical conditions and at low and high observation frequencies. The estimation errors are nearly the same (around 10%) and the convergence time is less than 1 day (see Fig. 3). Another study is currently in progress using the Unscented Kalman Filter (UKF) [1]: the results seem equivalent even though another approach (the unscented transform) is used for the approximation of the system's nonlinearities.

However, while these methods (EKF, UKF, PF) appear to be similar under regular conditions, they can present many differences in some particular situations where the system is subject to (i) high noise intensities, (ii) very distant measurement values and (iii) unknown initial conditions. In the literature, the KF methods are known to be robust to the last two issues, whereas the PF methods are known to be, furthermore, advantageous in the case of high noise variances. From another side, the EKF has a very low computational cost, as shown by the profiling results<sup>1</sup> in Fig. 4, yet it may diverge in the case of large variations of the states out of the linearization area. However, as mentioned above, for this system, the state variations are more likely to remain within the attractivity region of the considered equilibrium, so this issue is largely avoided. A more detailed comparison between these different estimation methods will be the subject of our future work.

The estimation errors are given by Fig. 5 and the mean square errors (MSE) of the state estimations have the values, Eqs. (39) and (40). Notice that these MSE values exceed the previously established 10% average because of the significant distance between the EKF initial values and the system's ones. These MSE values are reasonable for this class of biological models because of the relatively high noise intensities affecting this type of systems, the aleatory nature of its variables and the relatively distant observations. It is also obvious that the most significant estimation error comes from the biomass concentration, which is an unmeasurable variable in the output equation.

<sup>1</sup>Machine is : Core 2 Duo Intel Processor with 2GB RAM, running MATLAB 2012a on a Linux OS.

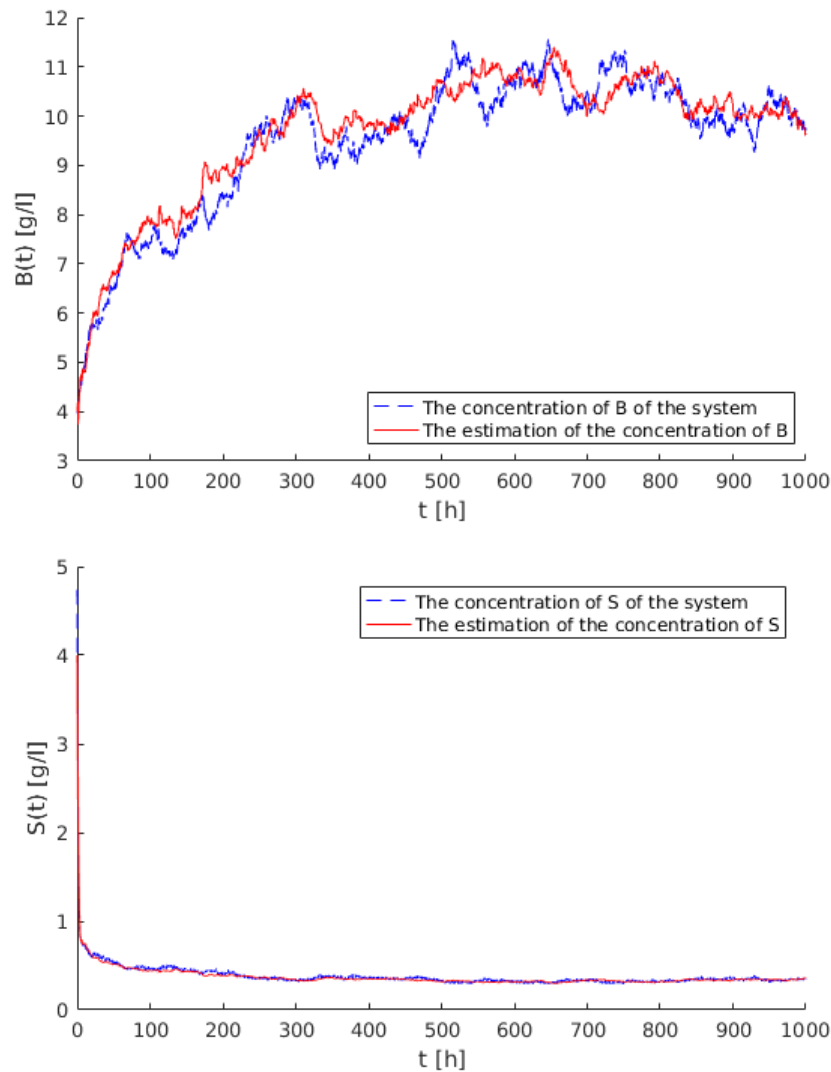


Fig. 2 Estimation of biomass  $B(t)$  and substrate  $S(t)$  concentrations using EKF

In order to inspect the efficiency of the algorithm in study, we performed additional tests in which we took the previously mentioned conditions to their limits. The first test was about small observation frequencies: we did three simulations with measurements every 10 h, 50 h and 100 h, respectively. In this case, the EKF updates were performed only at these periods. The obtained estimation errors are represented in Fig. 6. These errors are reasonable considering the width of the proposed time steps. The second test contained different measurement noise intensities, we simulated the EKF with three different intensities:  $\sigma = 0.2$ ,  $\sigma = 1$  and  $\sigma = 2$ . The obtained estimation errors are represented in Fig. 7 where the EKF does not appear to be robust against high output noises.

Finally, in view of the observation of [26] about the sensitivity of the EKF against far initial conditions as shown in Table 1, we proposed a third test where we performed ten independent simulations using ten different initial conditions (IC), randomly sampled from a uniform distribution in the interval  $[2, 20]$ . The obtained estimation errors are given in Fig. 8, the EKF converges in all the situations except where the IC is in the attractivity region of the other steady state of the system (the wash-out state) which represents the total extinction of the bacte-

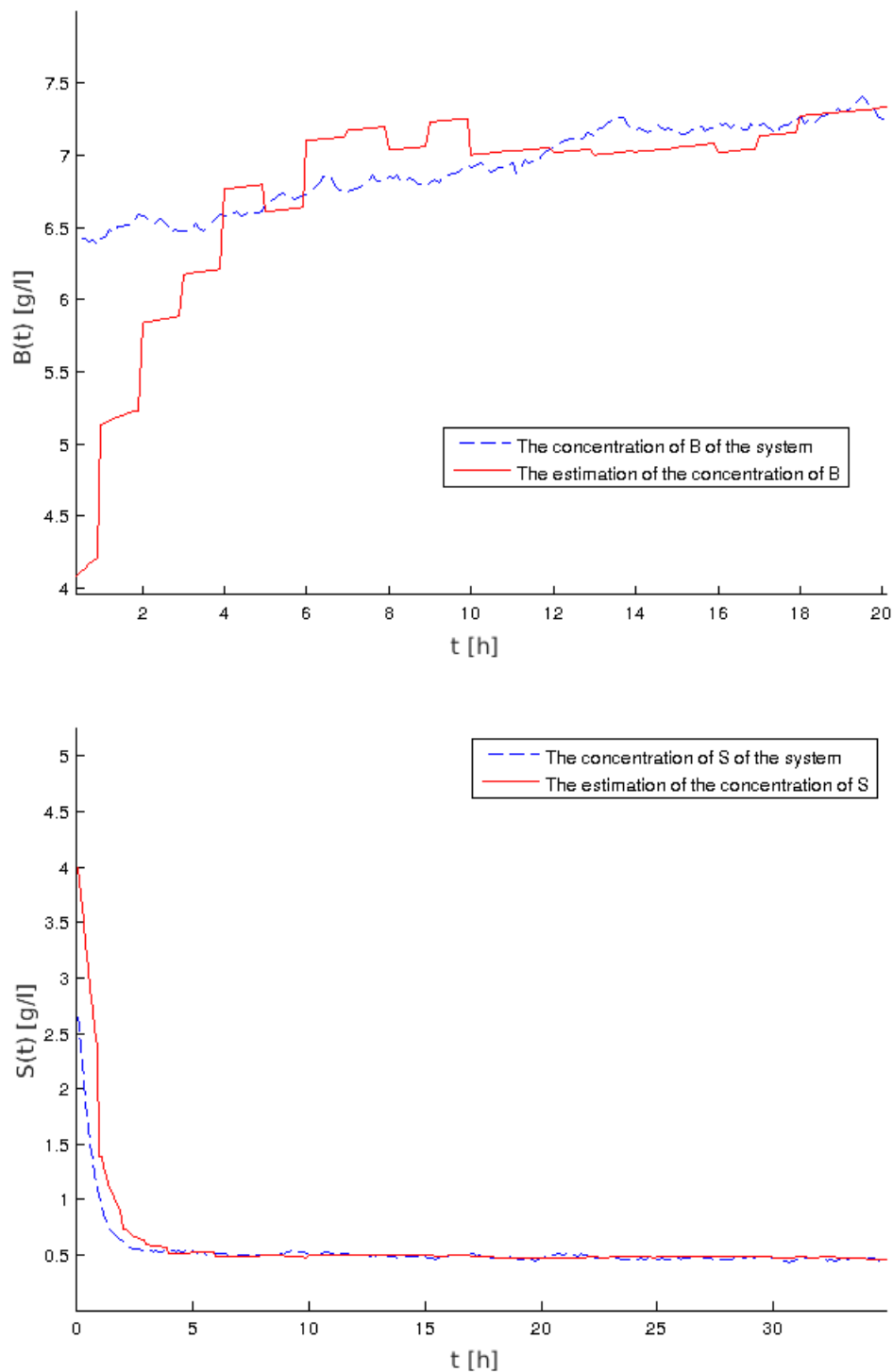




Fig. 3 Convergence time of the biomass and substrate concentrations

rial population. Furthermore, since the simulations in this last test are all independent, it could be considered as a stability check of the EKF using different series of the random numbers generated by the calculator. Similarly, the stability of the EKF algorithm was checked numerically by performing multiple and independent Monte Carlo runs (10000 simulations). All these simulations converged to the state values.

### Profile Summary

Generated 19-Dec-2016 10:48:55 using cpu time.

Function Name	Calls	Total Time	Self Time*	Total Time Plot (dark band = self time)
<a href="#">EKF</a>	1	0.014 s	0.007 s	
<a href="#">Chemostat_Simulation</a>	1	0.007 s	0.007 s	

**Self time** is the time spent in a function excluding the time spent in its child functions. Self time also includes overhead resulting from the process of profiling.

Fig. 4 Profiling Results using MATLAB 2012a on a Linux Machine

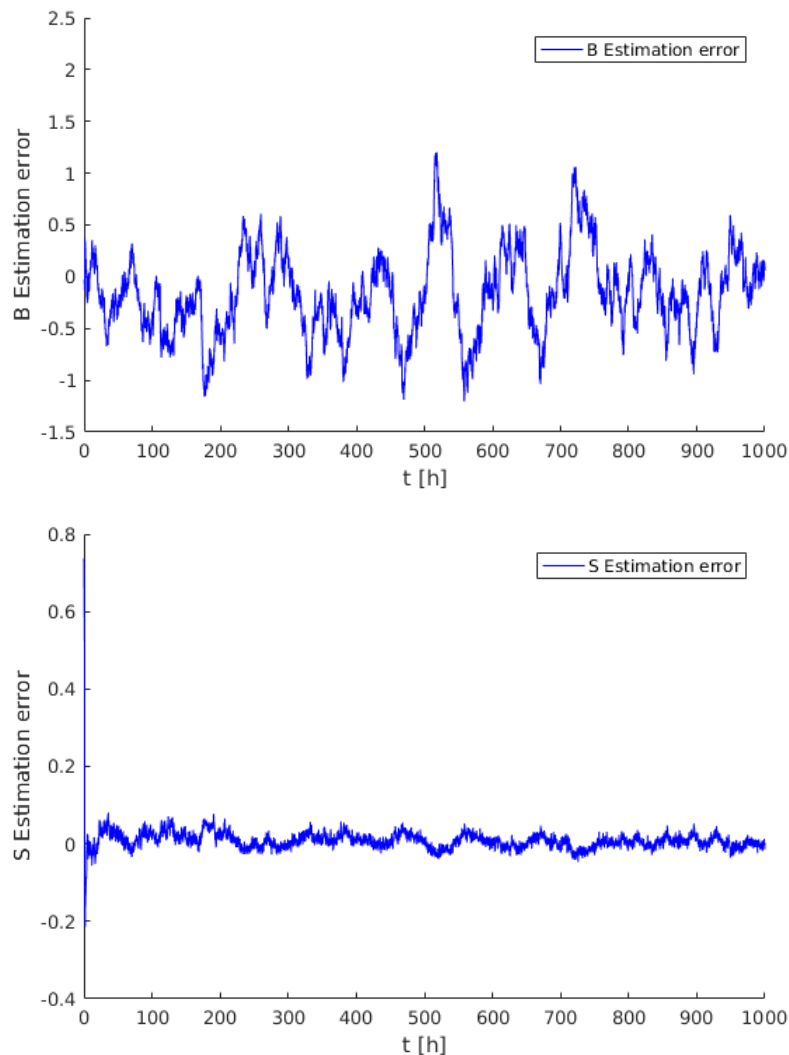


Fig. 5 Estimation error of biomass and substrate concentrations

This algorithm has a low computational cost: when used in an on-line estimation, it takes about  $7.10^{-3}$  s to compute one estimation value (Fig. 4). It is a fast algorithm compared to the UKF and to the PF. In this application, the EKF uses a single output second-order model with a time step of 0.1 h, hence the state estimates could be calculated on a low-cost micro-controller, unlike

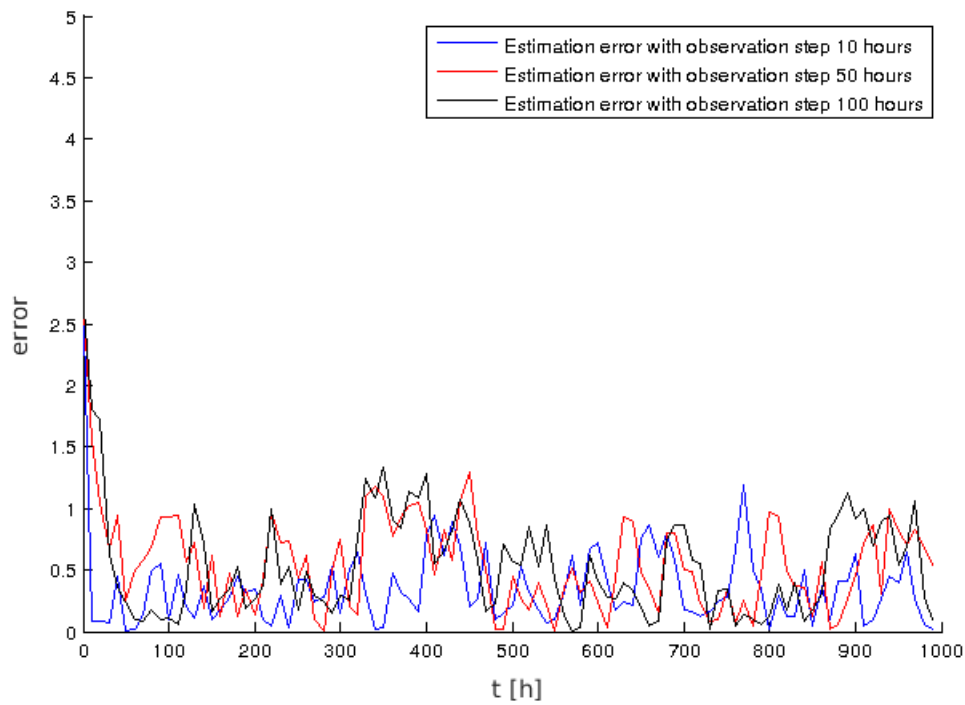


Fig. 6 Comparison of estimation errors with different observation frequencies

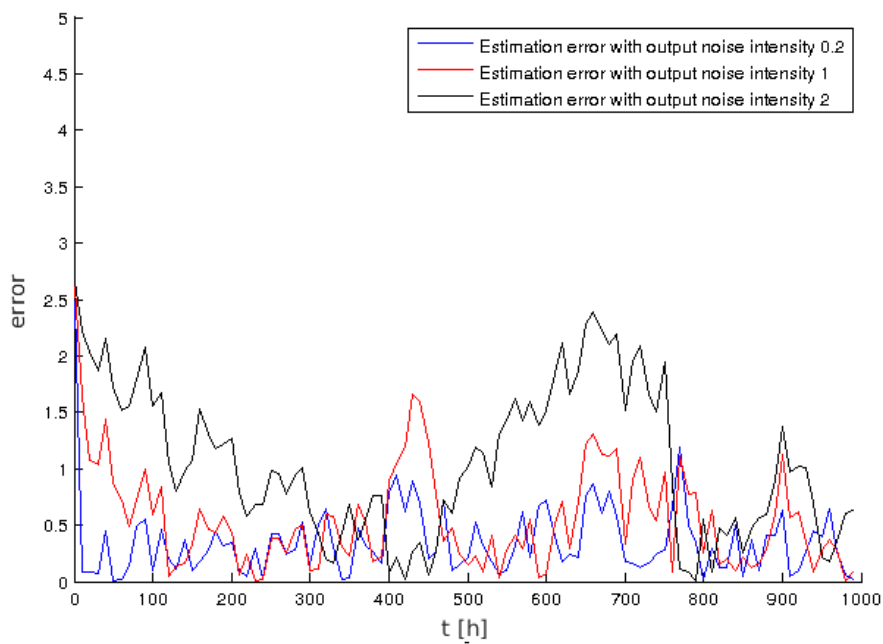


Fig. 7 Comparison of estimation errors with different output noise intensities

other methods in the literature which require heavy computations.

$$MSE_B = 0.1761, \tag{39}$$

$$MSE_S = 0.0021. \tag{40}$$

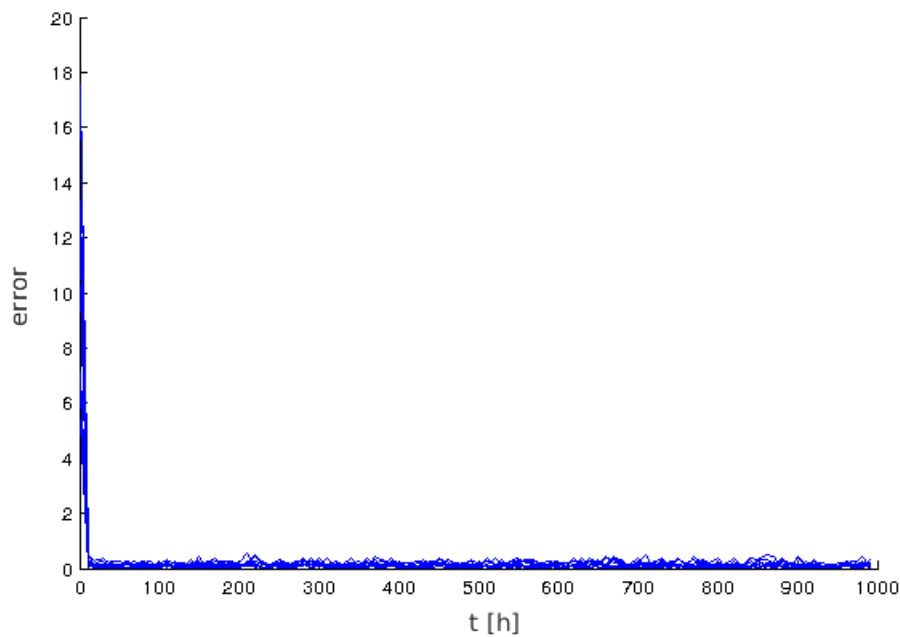


Fig. 8 Comparison of estimation errors with different and far initial conditions

## Conclusion

In this paper, we presented a method for the state estimation of a stochastic Chemostat model using the extended Kalman filtering approach. Some particular adaptations were required in this implementation, that are the linearization of the model around a deterministic trajectory and the substitution of the output equation by an equivalent one in order to use the EKF algorithm properly. These modifications will not require any additional calculations. The results obtained on this basis are good regarding the uncertainties acting on the system and compared to similar work in the literature. These results could perfectly be used for real-scale implementation using a mid-range computer board. In case of parameters uncertainties, it would be necessary to estimate the parameters values together with the state variables. This will be the subject of our future study. The authors in [9] gave a tutorial to deal with this problem for deterministic systems using observers. For stochastic systems, the most common parameter identification methods are the Maximum Likelihood Estimation (MLE) and the Markov Chain Monte Carlo (MCMC) methods described respectively in [17] and [20]. There exist also some KF based methods such as the dual or joint EKF methods [18]. The use of the EKF on higher-order models describing a two-reactions bioreactor is also in perspective.

## References

1. Abdelkader O. H., A. H. Abdelkader (2017). State Estimation for a Chemostat Model by the Unscented Kalman Filtering Approach, *Electrotehnica, Electronica, Automatica*, 65(2), 104-108.
2. Bastin G., D. Dochain (1990). *On-Line Estimation and Adaptive Control of Bioreactors*, Elsevier Science.
3. Benyahia B. (2012). *Modélisation et observation des bioprocédés à membranes: Application à la digestion anaérobie*, PhD Thesis, LBE, Narbonne.
4. Benyahia B., F. Campillo, B. Cherki, J. Harmand (2012). Particle Filtering for the Chemostat, *Proceedings of the 20<sup>th</sup> Mediterranean Conference on Control and Automation (MED 2012)*, Barcelona, Spain, 364-371.



5. Campillo F., M. Joannides, I. Larramendy-Valverde (2011). Stochastic Modeling of the Chemosta, *Ecological Modelling*, 222(15), 2676-2689.
6. Contois D. (1959). Kinetics of Bacterial Growth: Relationship between Population Density and Specific Growth Rate of Continuous Cultures, *Microbiology*, 21(1), 40-50.
7. de Assis A. J., R. Maciel Filho (2000). Soft Sensors Development for On-line Bioreactor State Estimation, *Computers & Chemical Engineering*, 24(2), 1099-1103.
8. Didi I., H. Dib, B. Cherki (2014). An Invariant Observer for a Chemostat Model, *Automatica*, 50(9), 2321-2326.
9. Dochain D. (2003). State and Parameter Estimation in Chemical and Biochemical Processes: A Tutorial, *Journal of Process Control*, 13(8), 801-818.
10. Dochain D., L. Chen (1992). Local Observability and Controllability of Stirred Tank Reactors, *Journal of Process Control*, 2(3), 139-144.
11. Gauthier J., H. Hammouri, S. Othman (1992). A Simple Observer for Nonlinear Systems Applications to Bioreactors, *IEEE Transactions on Automatic Control*, 37(6), 875-880.
12. Georgieva O., T. Patarinska (2002). State Observer Design Based on a Model of Chemostat Microbial Cultivation Accounting for the Memory Effects, *Systems Analysis Modelling Simulation*, 42(12), 1807-1827.
13. Gouzé J.-L., A. Rapaport, M. Z. Hadj-Sadok (2000). Interval Observers for Uncertain Biological Systems, *Ecological Modelling*, 133(1), 45-56.
14. Hadj-Sadok M. Z., J. L. Gouzé (2001). Estimation of Uncertain Models of Activated Sludge Processes with Interval Observers, *Journal of Process Control*, 11(3), 299-310.
15. Julier S. J., J. K. Uhlmann (1997). New Extension of the Kalman Filter to Nonlinear Systems, *Proceedings of the AeroSense'97*, International Society for Optics and Photonics, 182-193.
16. Kalchev B., I. Simeonov, N. Christov (2011). Kalman Filter Design for a Second-order Model of Anaerobic Digestion, *International Journal of Bioautomation*, 15(2), 85-100.
17. Kantas N., A. Doucet, S. S. Singh, J. Maciejowski, N. Chopin et al. (2015). On Particle Methods for Parameter Estimation in State-space Models, *Statistical Science*, 30(3), 328-351.
18. Khodadadi H., H. Jazayeri-Rad (2011). Applying a Dual Extended Kalman Filter for the Nonlinear State and Parameter Estimations of a Continuous Stirred Tank Reactor, *Computers & Chemical Engineering*, 35(11), 2426-2436.
19. Kloeden P. E., E. Platen (1992). *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag.
20. Leander J., T. Lundh, M. Jirstrand (2014). Stochastic Differential Equations as a Tool to Regularize the Parameter Estimation Problem for Continuous Time Dynamical Systems Given Discrete Time Measurements, *Mathematical Biosciences*, 251, 54-62.
21. Monod J. (1949). The Growth of Bacterial Cultures, *Annual Reviews in Microbiology*, 3(1), 371-394.
22. Novick A., L. Szilard (1950). Description of the Chemostat, *Science*, 112(2920), 715-716.
23. Patarinska T., V. Trenev, S. Popova (2010). Software Sensors Design for a Class of Aerobic Fermentation Processes, *International Journal of Bioautomation*, 99-118.
24. Patnaik P. R. (2008). Application of the Lyapunov Exponent to Evaluate Noise Filtering Methods for a Fed-batch Bioreactor for PHB Production, *International Journal of Bioautomation*, 9, 1-14.
25. Popova S., P. Koprinkova, T. Patarinska (2003). Neural Network Based Biomass and Growth Rate Estimation Aimed to Control of a Chemostat Microbial Cultivation, *Applied Artificial Intelligence*, 17(4), 345-360.
26. San K.-Y., G. Stephanopoulos (1984). Studies on On-line Bioreactor Identification. II. Numerical and Experimental Results, *Biotechnology and Bioengineering*, 26(10), 1189-1197.

27. Slavov T., O. Roeva (2011). Genetic Algorithm Tuning of PID Controller in Smith Predictor for Glucose Concentration Control, International Journal of Bioautomation, 15(2), 101-114.
28. Smith H. L., P. E. Waltman (1995). The Theory of the Chemostat: Dynamics of Microbial Competition, Cambridge University Press.

**Oussama Hadj-Abdelkader, Ph.D. Student**

Email: [hadjabdelkader.oussama@gmail.com](mailto:hadjabdelkader.oussama@gmail.com)



Oussama Hadj-Abdelkader was born in Tlemcen (Algeria), in 1991. He graduated from the Electrical Engineering Department in the AbouBekr Belkaid University in Tlemcen (Algeria), in 2013. He is currently a Ph.D. student at the AbouBekr Belkaid University of Tlemcen. His main research interests are Monte Carlo Methods and Stochastic Systems Identification.

**Assist. Prof. Amine Hadj-Abdelkader, Ph.D.**

Email: [amine.hadj@gmail.com](mailto:amine.hadj@gmail.com)



Amine Hadj-Abdelkader was born in Tlemcen (Algeria), in 1980. He graduated from the Electrical Engineering Department in the AbouBekr Belkaid University in Tlemcen (Algeria), in 2002. He received his Magister Degree from the same university in 2005 and later on, his Ph.D. degree in Automatic Control, Signal and Image Processing and Computer Engineering from the University of Lorraine (France), in 2011. He is currently an Assistant Professor at the AbouBekr Belkaid University of Tlemcen. His main research interests are System Identification and Human-Machine Interactions.



© 2019 by the authors. Licensee Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).