Journal of Process Control 23 (2013) 148–157

Contents lists available at SciVerse ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

# Experimental validation of an Extended Kalman Filter estimating acetate concentration in *E. coli* cultures

L. Dewasme[a,*], G. Goffaux[b], A.-L. Hantson[c], A. Vande Wouwer[a]

[a] *Service d'Automatique, Université de Mons, 31 Boulevard Dolez, 7000 Mons, Belgium*
[b] *Département de Génie Chimique, Ecole Polytechnique de Montréal, C.P. 6079 Montréal, Québec, Canada*
[c] *Service de Chimie et Biochimie Appliquées, Université de Mons, 56 Rue de l'Epargne, 7000 Mons, Belgium*

ARTICLE INFO

*Article history:*
Received 24 February 2012
Received in revised form 19 August 2012
Accepted 11 September 2012
Available online 25 October 2012

*Keywords:*
State estimation
Software sensors
Extended Kalman Filter
Fed-batch processes
Fermentation

ABSTRACT

Optimizing control of *Escherichia coli* cultures depends on the availability of appropriate on-line sensors for the main culture components. An efficient and practical way to maintain *E. coli* cultures in the neighborhood of the optimal operating conditions is to regulate the acetate concentration at a constant low level. Unfortunately, reliable acetate probes are currently not available on the market. A way to overcome this lack of information is to design software sensors. In this work, observability conditions are first examined, and Extended Kalman Filters (EKF) are developed for various hardware sensor configurations, taking account of their reliability and cost, and including a possible kinetic parameter estimation. The filters are validated using experimental data collected on a lab-scale bioreactor.



## 1. Introduction

Industrial recombinant protein production is usually achieved using fed-batch cultures of genetically modified yeast or bacteria strains. From an operational point of view, it is necessary to determine an optimal feeding strategy (i.e., the time evolution of the input flow rate to the fed-batch culture) in order to maximize the biomass productivity.

The main problem encountered comes from the metabolic changes in such strains in the case of a feeding excess. This "overflow metabolism", also called "short-term Crabtree effect" [1,2], is a metabolic phenomenon that is induced when the rate of glycolysis (i.e., the oxidation of glucose) exceeds a critical value expressing the maximum capability of energy production by glucose oxidation, leading to a generally inhibiting byproduct formation from pyruvate. This phenomenon occurs for instance in *Saccharomyces cerevisiae* cultures with aerobic ethanol formation, in *Pichia pastoris* with aerobic methanol formation, in *E. coli* cultures with aerobic acetate formation or in mammalian cell cultures with aerobic lactate formation. To avoid this undesirable effect, a closed-loop optimizing strategy is required, which can take various forms [3–6].

Two main strategies are generally considered: (a) controlling the substrate (generally glucose) concentration accurately at the critical level, i.e., the threshold separating two metabolic regimes (with and without feeding excess); (b) limiting the byproduct production by controlling its concentration at a very low level. Unfortunately, the sensitivity of currently available glucose probes is often insufficient to provide a correct measurement. Indeed, the critical level is very low (from $O(10^{-2})$ to $O(10^{-1})$). Consequently, the regulation of the byproduct concentration is preferred for obvious practical reasons. Nevertheless, even if ethanol and methanol are nowadays easily measurable, acetate and lactate probes are rare or non-existent on the market. On the other hand, biomass, $pO_2$ probes and gas-analyzers are now widely available, and can be the basis for the development of software sensors (see [7,8]).

In [9], the performances of an Extended Kalman Filter (EKF) and an Asymptotic Observer (AO) are assessed in an application to *E. coli* fed-batch cultures. The observability of the model is studied for different combinations of the measured and estimated state variables. However, biomass is always considered as unmeasurable, limiting the number of possible combinations. Moreover, from all the assumed unmeasurable states, only the biomass concentration is adequately estimated while the performances of both observers are not satisfactory for the estimation of the substrate and acetate concentrations.

In [10], off-line glucose measurements by flow injection analysis (FIA) are used by a Kalman Filter to reconstruct the biomass signal and the glucose signal between 2 measurements. The estimated state vector is also augmented by the addition of the maximal growth rate. This Kalman Filter is coupled to a model-based PI controller to regulate the glucose concentration at 0.2 g/l. The set-point is arbitrarily chosen by the user and considered as efficient

* Corresponding author.
*E-mail address:* laurent.dewasme@umons.ac.be (L. Dewasme).

0959-1524/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jprocont.2012.09.004

in the context of the application which aims at producing phytase. Moreover, a BL21(DE3) pPhyt109 *E. coli* strain is used. This strain is well-known for its property to produce small quantities of acetate, so that its accumulation only inhibits the biomass growth in the end of the culture.

In this work, a wild-type (B-11303, ATCC) strain is considered, which can produce larger amount of acetate, thus inhibiting the biomass growth in the early hours of the culture. This motivates the focus on the on-line estimation of the acetate concentration and, less importantly, the on-line estimation of the glucose concentration. In order to design the software sensors, biomass concentration is assumed to be measured on-line, so as the dissolved oxygen concentration and the outlet gas. This paper is organized as follows. The next section describes the process model and the EKF structure. Various sensor configurations are considered, and a global observability assessment is given, together with some simulation results through Sections 3 and 4. Section 5 presents the available materials and methods, from the bioreactor set-up to the available probes and analyzers. Experimental results are presented in Section 6. Conclusions are drawn in Section 7.

## 2. A generic mechanistic model

In this section, we first consider a generic mechanistic model that represents the culture of different strains presenting an overflow metabolism (yeasts, bacteria, animal cells, etc.).

In the following, three main catabolic reactions are considered:

$$\text{Substrate oxidation}: \quad k_{S1}S + k_{O1}\,O \overset{r_1 X}{\rightarrow} k_{X1}\,X + k_{CO_2 1}\,CO_2 \tag{1a}$$

$$\text{Substrate fermentation}: \quad k_{S2}S + k_{O2}\,O \overset{r_2 X}{\rightarrow} k_{X2}\,X + k_{P2}\,P + k_{CO_2 2}\,CO_2 \tag{1b}$$

$$\text{Byproduct oxidation}: \quad k_{P3}P + k_{O3}\,O \overset{r_3 X}{\rightarrow} k_{X3}\,X + k_{CO_2 3}\,CO_2 \tag{1c}$$

where $X$, $S$, $P$, $O$ and $CO_2$ are, respectively, the concentration in the culture medium of biomass, substrate (typically glucose), byproduct (i.e., ethanol or methanol in yeast cultures, acetate in bacteria cultures or lactate in animal cells cultures), dissolved oxygen and carbon dioxide. $k_{\xi i}$ ($i = 1, 2, 3$, $\xi = X, S, P, O, CO_2$) are the yield coefficients and $r_1$, $r_2$ and $r_3$ are the nonlinear specific consumption rates given by:

$$r_1 = \frac{\min(r_S, r_{S_{crit}})}{k_{S1}} \tag{2a}$$

$$r_2 = \frac{\max(0, r_S - r_{S_{crit}})}{k_{S2}} \tag{2b}$$

$$r_3 = \frac{\max(0, ((k_{O1}(r_{S_{crit}} - r_S))/(k_{O3}))(P/(P + K_P)))}{k_{P3}} \tag{2c}$$

Note that these specific consumption rates are divided, for each reaction, by the corresponding substrate yield coefficient (i.e., $k_{S1}$ and $k_{S2}$ for the main substrate, $S$, in the first two reactions and $k_{P3}$ for the consumed byproduct, $P$, in the third reaction) in order to normalize with respect to the consumed source. For instance, 1 mol of substrate will produce $k_{X1}/k_{S1}$ mole of biomass $X$ with the first reaction and $k_{X2}/k_{S2}$ mole of biomass $X$ with the second reaction. The kinetic terms associated with the substrate consumption $r_S$ and the critical substrate consumption $r_{S_{crit}}$ (function of the cells oxidative or respiratory capacity $r_O$) are given by:

$$r_S = \mu_S \frac{S}{S + K_S} \tag{3a}$$

$$r_{S_{crit}} = \frac{r_O}{k_{O1}} = \frac{\mu_O}{k_{O1}} \frac{O}{O + K_O} \frac{Ki_P}{Ki_P + P} \tag{3b}$$

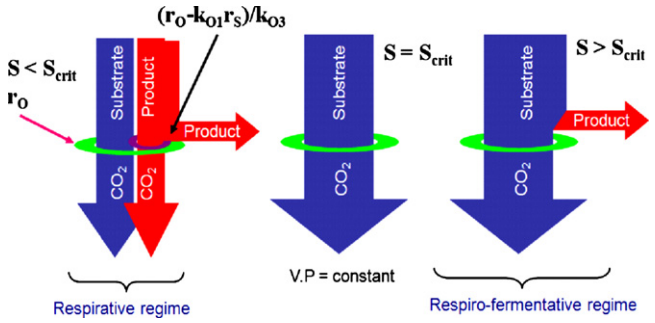


**Fig. 1.** Illustration of Sonnleitner's bottleneck assumption [11] for cells limited respiratory capacity.

These expressions take the classical form of Monod laws where $\mu_S$ and $\mu_O$ are the maximal values of specific growth rates and $K_S$ and $K_O$ are the saturation constants. Moreover, $P$ inhibits the cells oxidative capacity and $Ki_P$ represents the inhibition constant. The kinetic model (2) is based on Sonnleitner's bottleneck assumption [11] which was applied to a yeast strain *S. cerevisiae* (Fig. 1). During a culture, the cells are likely to change their metabolism because of their limited oxidative capacity. When the substrate is in excess (concentration $S > S_{crit}$ and the glucose consumption rate $r_S > r_{S_{crit}}$), the cells produce a byproduct $P$ through the fermentative pathway, and the culture is said in (respiro-) fermentative (RF) regime. On the other hand, when the substrate becomes limiting (concentration $S < S_{crit}$ and the glucose consumption rate $r_S < r_{S_{crit}}$), the available substrate (typically glucose), and possibly the byproduct $P$ (as a substitute carbon source), if present in the culture medium, are oxidized. The culture is then said in respirative (R) regime.

Component-wise mass balances give the following differential equations:

$$\frac{dX}{dt} = (k_{X1}r_1 + k_{X2}r_2 + k_{X3}r_3)X - DX \tag{4a}$$

$$\frac{dS}{dt} = -(k_{S1}r_1 + k_{S2}r_2)X + DS_{in} - DS \tag{4b}$$

$$\frac{dP}{dt} = (k_{P2}r_2 - k_{P3}r_3)X - DP \tag{4c}$$

$$\frac{dO}{dt} = -(k_{O1}r_1 + k_{O2}r_2 + k_{O3}r_3)X - DO + OTR \tag{4d}$$

$$\frac{dCO_2}{dt} = (k_{CO_2 1}r_1 + k_{CO_2 2}r_2 + k_{CO_2 3}r_3)X - DCO_2 - CTR \tag{4e}$$

$$\frac{dV}{dt} = F_{in} \tag{4f}$$

where $S_{in}$ is the substrate concentration in the feed, $F_{in}$ is the inlet feed rate, $V$ is the culture medium volume and $D$ is the dilution rate ($D = F_{in}/V$). $OTR$ and $CTR$ represent respectively the oxygen transfer rate from the gas phase to the liquid phase and the carbon transfer rate from the liquid phase to the gas phase. Classical models of $OTR$ and $CTR$ are given by:

$$OTR = k_L a_O (O_{sat} - O) \tag{5a}$$

$$CTR = k_L a_{CO_2} (CO_2 - CO_{2sat}) \tag{5b}$$

where $k_L a_O$ and $k_L a_{CO_2}$ are the volumetric transfer coefficients respectively of oxygen and carbon dioxide, and $O_{sat}$ and $C_{sat}$ are respectively the dissolved oxygen and carbon dioxide concentrations at saturation.

Several simplifications of (2) are possible, depending on the culture conditions. In controlled operation, the substrate concentration could be kept close to zero (in the following, this situation is denoted "regulation 1") leading to $S \approx 0, \dot{S} \approx 0$ or alternatively (and equivalently), the byproduct concentration can be regulated at a

low level ("regulation 2") leading to $\dot{P} = 0$, $P << K_{i_P}$. In addition, ample oxygenation is provided and thus, $\dot{O} = 0$ and $O \gg K_O$. These assumptions will be considered in the following, when analyzing system observability and designing an observer.

Note that (2) is a particular case of the following general representation of a mechanistic model resulting from mass balance equations [7]:

$$\frac{d\xi(t)}{dt} = K\varphi(\xi, t) - D(t)\xi(t) + F(t) - Q(t) \tag{6}$$

where $\xi$ is the vector of concentrations of the macroscopic components; $K$ is the pseudo-stoichiometric matrix; $\varphi$ is the vector of reaction rates; $F$ is the vector of feed rate of selected components ($F_j = D(t)\xi_{in,j}(t)$ if the component is diluted in the feed stream, or $F_j = F_j(t)$ if the component is introduced in the culture in gaseous form); $Q$ is the vector of outflow rates of the considered components in gaseous form.

## 3. Design of software sensors

### 3.1. Observability analysis

Observability is a structural system property that relates to the possibility of estimating the system state on the basis of the available measurement information, i.e., in spite of the lack of information about the true initial conditions. If the system under consideration is nonlinear (which is almost always the case in bioprocess applications), observability depends on the system inputs and can be defined as:

**Definition**. A system is observable if

$$\begin{aligned} &\forall t_0, \exists t_1 < \infty | y(t; t_0, \xi(0), u(t)) = y(t; t_0, \xi'(0), u(t)), \\ &\forall u(t), t_0 < t < t_1 \Rightarrow \xi(0) = \xi'(0). \end{aligned} \tag{7}$$

In other words, a system is observable if two identical output trajectories $y$ (function of time $t$, the initial states $\xi(0)$ and of the input command $u(t)$) over a given finite time horizon imply the equality of the initial states $\xi(0)$ and $\xi'(0)$.

A convenient way to analyze observability is provided by canonical observability forms as introduced in [12,13]:

$$\begin{aligned} &\forall i \in \{1, \ldots, q\}, x_i \in n_i\Re, n_1 \geq n_2 \geq \ldots \geq n_q, \\ &\sum_{1 \leq i \leq q} n_i = n \\ &\dot{x} = \begin{cases} \dot{x_1} = f_1(x_1, x_2) \\ \dot{x_2} = f_2(x_1, x_2, x_3) \\ \vdots \\ \dot{x_{q-1}} = f_{q-1}(x_1, \ldots, x_q) \\ \dot{x_q} = f_q(x_1, \ldots, x_q) \end{cases} \\ &y = x_1 = [x_{1,1}\, x_{1,2} \ldots x_{1,n_1}]^T \end{aligned} \tag{8}$$

where $x$ is the state vector, $y$ the vector of measured states, $f_i$ a partition of the nonlinear state equations, $q$ the number of partitions. Eq. (8) is also called a Lower Hessenberg System, i.e., a system where $((\partial f_i)/(\partial x_j)) = 0$ with $j > i+1$. To assess if the system is observable, one first checks if the bioprocess model can be put in the form of (8) by defining an appropriate partition, and then the following condition is evaluated:

$$rank \frac{\partial f_i}{\partial x_{i+1}} = n_{i+1} \quad \forall i \in \{1, \ldots, q-1\} \tag{9}$$

Eq. (9) simply translates the fact that a partition of states $x_{i+1}$ is only observable if any perturbation of these states propagates to partition $x_i$.

The only partition satisfying conditions (8) and (9) and relative to the observation of S and P is $y = x_1 = [X\, O]^T$ and $x_2 = [S\, P]^T$. In this case, the bioprocess model (2) can be cast in the form:

$$\begin{aligned} &\dot{x_1} = [\dot{X}\, \dot{O}]^T = f_1([X\, O]^T, [S\, P]^T) = f_1(x_1, x_2) \\ &\dot{x_2} = [\dot{S}\, \dot{P}]^T = f_2([X\, O]^T, [S\, P]^T) = f_2(x_1, x_2) \\ &y = x_1 = [X\, O]^T \end{aligned} \tag{10}$$

Note that the carbon dioxide differential equation is not considered as dissolved $CO_2$ is usually not measured, does not need to be estimated and does not influence the other states.

$$\begin{aligned} rank \frac{\partial f_1}{\partial x_2} &= rank \begin{bmatrix} \frac{\partial \dot{X}}{\partial S} & \frac{\partial \dot{X}}{\partial P} \\ \frac{\partial \dot{O}}{\partial S} & \frac{\partial \dot{O}}{\partial P} \end{bmatrix} \\ &= rank \begin{bmatrix} X\mu_S \frac{K_S}{k_{X2}(S+K_S)^2} & X \frac{(k_{X1} - k_{X2})\mu_O O K_{i_P}}{k_{O1}k_{X1}k_{X2}(O+K_O)(P+K_{i_P})^2} \\ -X\mu_S \frac{k_{O2}K_S}{k_{X2}(S+K_S)^2} & X \frac{(k_{X1}k_{O1} - k_{X2}k_{O2})\mu_O O K_{i_P}}{k_{O1}k_{X1}k_{X2}(O+K_O)(P+K_{i_P})^2} \end{bmatrix} = n_2 = 2 \\ X, O &\neq 0 \end{aligned} \tag{11}$$

As mentioned at the end of the previous section, several simplifying assumptions hold in controlled conditions, and in particular for regulation 1: $S \approx 0$, $\dot{S} \approx 0$, $\dot{O} = 0$ and $O \gg K_O$. System (10) can therefore be rewritten as follows:

$$\dot{x} = \begin{cases} \dot{X} = f_1(X, P) \\ \dot{P} = f_2(X, P) \\ y = X \end{cases} \tag{12}$$

In the particular case of (12), $n_i = 1 \quad \forall i$ and condition (9) reduces to $(\partial f_i)/(\partial x_{i+1}) \neq 0$ which is verified if $X, O \neq 0$ (note that in order to achieve an aerobic culture, the biomass and the oxygen concentrations can never vanish).

According to regulation 2: $\dot{P} = 0$ and $K_{i_P} >> P$, together with $\dot{O} = 0$ and $O \gg K_O$, so that system (10) reduces to:

$$\dot{x} = \begin{cases} \dot{X} = f_1(X, S) \\ \dot{S} = f_2(X, S) \\ y = X \end{cases} \tag{13}$$

and $(\partial X/\partial S) \neq 0$ if $X \neq 0$. Combining (12) and (13), it appears that the biomass measurement could be sufficient to reconstruct the byproduct and the substrate concentrations as long as those concentrations remain at respective sufficiently low levels (i.e., $S < O(10^{-1}$ g/l) and $0$ g/l $< A < 1$ g/l).

System observability can be assessed for various sensor configurations (also including the measurements of *OTR* or *S* for instance) using the bioprocess model (2) or a simplified version of it based on the assumptions detailed in Section 2. The best sensor configuration will depend on the previous observability analysis, economic considerations ($pO_2$ probes are generally cheaper than a gas analyser, which is cheaper than biomass and substrate probes), and practical observability as assessed from the performance of a software sensor running under realistic conditions of measurement noise (practical observability can also be investigated using a criterion defined in [14], as applied in [15]). To this end, we first introduce the Extended Kalman Filter (EKF).

### 3.2. The Extended Kalman Filter

The Kalman filter [16], which is by far the most popular state estimation technique used for bioprocess monitoring, is an

**Table 1**
Observability analysis for different sensor configurations.

| Combination | Canonical obs. | *P* estim. error (g/l) | *S* estim. error (g/l) |
|---|---|---|---|
| *X* | Simplified model | 0.1 | 0.01 |
| *O–OTR* | – | 0.1 | 0.01 |
| *S* | Simplified model | 1 | – |
| *X–O–OTR* | Full model | 0.1 | 0.01 |
| *S–O–OTR* | Full model | 0.1 | – |
| *X–S* | Simplified model | 0.1 | – |
| *X–S–O–OTR* | Full model | 0.01 | – |

exponential observer that minimizes the variance of the estimation error. The EKF is based on a first-order linearization of the process model along the estimated trajectory. In the context of bioprocess applications, the concentrations of the main species are measured at discrete times only and with relatively low sampling frequencies (the measurements are sometimes collected at different rates, i.e., resulting in an asynchronous measurement configuration). Therefore, the continuous-discrete EKF formulation, where continuous estimation is provided from discrete-time measurements, is of interest. The algorithm proceeds in two steps: a prediction step (corresponding to the time period between two measurement times) and a correction step occurring each time a new measurement is available. Using the notation of (6)

*Prediction* between $t_k$ and $t_{k+1}$:

$$\frac{d\xi(t)}{dt} = K\varphi(\xi(t), t) - D\xi(t) + D\xi^{in}; \quad \xi(t_k) = \xi(t_k^+), \quad t_k^+ \le t < t_{k+1}^- \tag{14a}$$

$$\frac{dC(t)}{dt} = A(\xi(t))C(t) + C(t)A(\xi(t))^T + R_\eta; \quad C(t_k) = C(t_k^+), \tag{14b}$$

*Correction* at time $t_{k+1}$:

$$\Omega(\xi(t_{k+1})) = C(t_{k+1}^-)L^T\left[LC(t_{k+1}^-)L^T + R_\epsilon(t_{k+1})\right]^{-1} \tag{15a}$$

$$\xi(t_{k+1}^+) = \xi(t_{k+1}^-) + \Omega(\xi(t_{k+1}))\left(y(t_{k+1}) - L\xi(t_{k+1}^-)\right) \tag{15b}$$

$$C(t_{k+1}^+) = C(t_{k+1}^-) - \Omega(\xi(t_{k+1}))LC(t_{k+1}^-) \tag{15c}$$

In these expressions, $L$ is the measurements matrix, $\Omega$ the correction gain, $C$ the covariance matrix of the state estimation errors, $R_\epsilon$ and $R_\eta$ the covariance matrices of respectively the measurement errors and state noises and $t_{k+1}^-$ and $t_{k+1}^+$ the time instants relative to the a priori and a posteriori estimation.

This filter is first applied to the bacterial culture model (2) with published parameter values, e.g., the parameter values from [17], using various sensor configurations, and realistic measurement noise. The noise standard deviation is chosen as $\sigma = 0.0001$ g/l for the dissolved oxygen measurement, $\sigma = 0.5$ g/l for the biomass measurement, $\sigma = 0.05$ g/l for the glucose measurement and $\sigma = 0.01$ g/l/s for the OTR measurement.

Table 1 summarizes the results of this analysis. The first column details the different measurement combinations, the second column specifies the model (the full model (2) or the simplified model obtained using the assumptions described in Section 2 for which canonical observability is verified) and the last 2 columns give a quantitative statement regarding the practical observability of the byproduct (*P* for acetate) and the substrate. This estimation error is assessed as

$$\epsilon = \sqrt{\frac{\sum_{i=1}^{N}(\xi_i - \hat{\xi}_i)^2}{N}} \tag{16}$$

where $N$ is the number of samples and $\hat{\xi}_i$ is the vector of the estimated states.

The cheapest way to estimate the byproduct concentration is to measure the dissolved oxygen concentration only, as shown in Table 1 and Fig. 2. However, in realistic conditions, the dissolved oxygen is regulated by the stirrer speed or the air flow and the measurement cannot be exploited. Hence, an additional measurement from the gas analyzer (i.e., the OTR) is required to retrieve information on the oxygen consumption. The byproduct and substrate estimation from the oxygen measurement can qualitatively be assessed as correct (see Fig. 2) and does not improve significantly if the biomass concentration is also available, as shown in Fig. 3 where biomass and dissolved oxygen concentrations are both measured.

An observability analysis using the simplifying assumptions of model (2) discussed in Section 2 shows that the biomass measurement could be sufficient to estimate the byproduct and the substrate concentrations as long as these concentrations remain sufficiently low. Indeed, Fig. 4 shows that acetate and substrate are both well estimated.

A sensor configuration using the substrate concentration measurement is not considered in the following as it is the most expensive measurement device (with questionable sensitivity at the considered substrate concentrations) and does not provide any improvement, as compared to the observer using the biomass and/or the dissolved oxygen measurements.

## 4. Parameter estimation

Kinetic parameters are often not estimated with high accuracy and the EKF estimation is particularly sensitive to model errors. To tackle this problem, an option is to augment the state vector with an uncertain parameter to be estimated. In this study, it would be of interest to estimate and monitor the critical substrate consumption level, $r_{S_{crit}}$, which is an indicator of the metabolic state of the bacteria (it is an image of their respiratory capacity). Considering that the process is operated in conditions where the produced acetate concentration is relatively low with respect to the inhibition constant ($K_{i_P} >> P$) and the oxygen is not limiting ($O \gg K_O$), the following expression can be derived from (3b):

$$r_{S_{crit}} \approx \frac{\mu_O}{k_{O1}} \tag{17}$$

This critical level is then approximately represented by $\mu_O$ in optimal operating conditions under the assumption of a perfectly known stoichiometry. The EKF state vector can therefore be augmented with $\mu_O$, and a sinusoidal dither signal of pulsation $\omega_d = \pi$ (frequency of 0.5 $h^{-1}$) can be added to the input in order to constantly excite the system and ensure the parameter convergence [18]. Simulation results are shown in Fig. 5 (with $\mu_O$ initialized at 1). In this simulation, no noise is added to the biomass measurement. Fig. 6 shows the detrimental effect of biomass measurement noise.

To alleviate the noise sensitivity problem, it is suggested to insert an anti-aliasing low-pass filter with a cut-off frequency of $\omega_{blpf} = 4\pi$ [19]. The cut-off frequency should be large enough so as to capture the effect of the dither signal. The results using low-pass filtering are shown in Fig. 7. Even if the estimation of the kinetic parameter is still perfectible, the current results provide a satisfactory estimation of the substrate and acetate concentrations since the confidence intervals are kept within a reasonable range compared to the estimated value (i.e., about 1 g/l). Getting the right trends is the most important for control purposes. Additional discussion is provided in the section on experimental results.

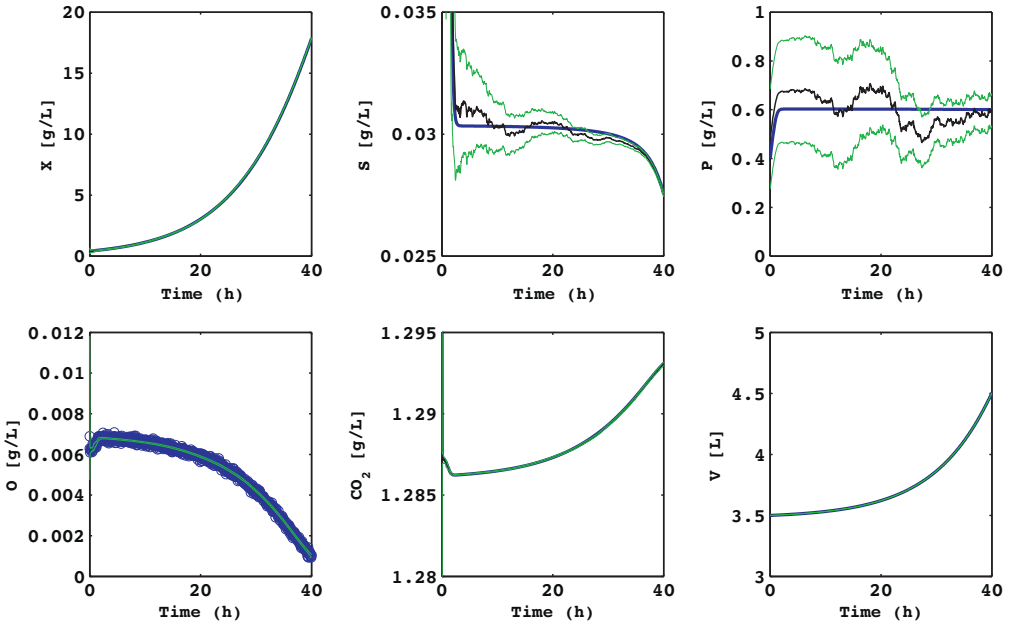


**Fig. 2.** Estimation of the acetate concentration using the dissolved oxygen measurement. In black: EKF estimation. In blue: model evolution. In green: confidence interval at 99%. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

## 5. Materials and methods

### 5.1. Bioreactor setup

Cultures of a wild-type strain (*B*) of *E. coli* (B-11303, ATCC) are performed in a 5 l compact laboratory scale bioreactor (Biostat B+– Sartorius). pH is regulated at 7 thanks to a PI controller (standard controller in the basic monitoring unit called Digital Fermenter Control or *DFC*, delivered with the vessel) injecting NaOH 2 M (base) and $H_3PO_4$ 0.5 M (acid). Oxygenation conditions are regulated by a 2-stage cascade controller (also provided in the *DFC*) acting first on the stirrer speed (from 200 to 1500 RPM) and, in case of stirrer saturation, on the aeration system supplying air flow (from 0.3 to 20 l/min). The temperature is controlled by the *DFC* at 37 °C using a heating water jacket.

### 5.2. Media preparation and composition

The preculture, culture and feeding media are prepared in accordance with the recipes of [17]. The general culture medium used in the following experimental results has the composition: 1–5 g/l of glucose (depending on the application), 6 g/l of $Na_2HPO_4$, 3 g/l of $KH_2PO_4$, 1 g/l of $NH_4Cl$, 0.5 g/l of NaCl, 0.12 g/l of $MgSO_4{\cdot}7H_2O$ and 0.34 g/l of thiamin. Note that the culture medium used in [17] is

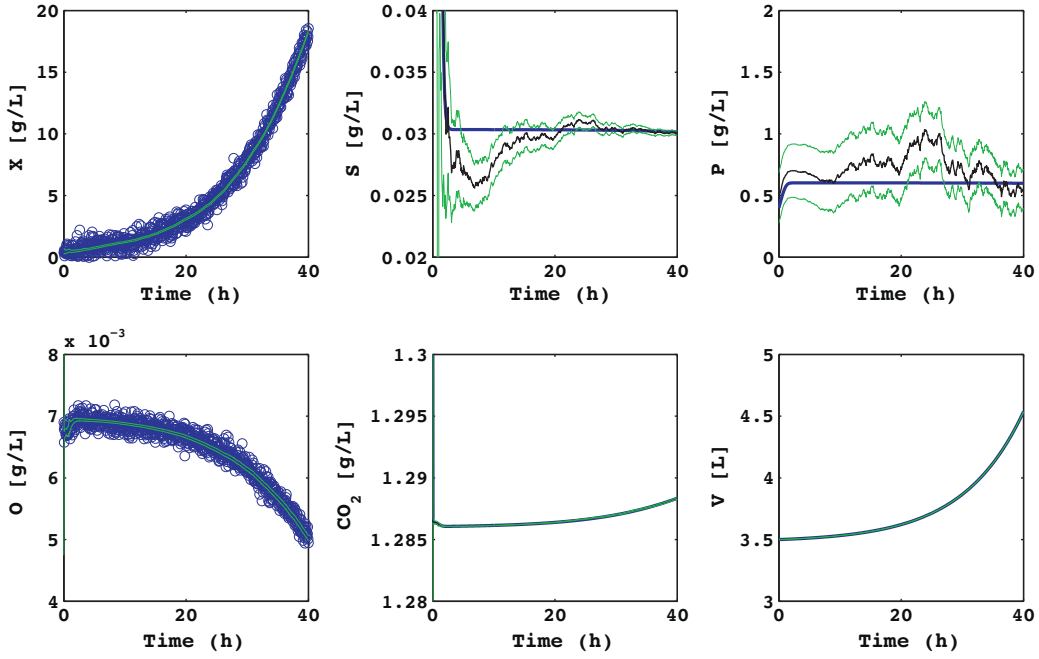


**Fig. 3.** Estimation of the acetate concentration using the biomass and the oxygen measurement. In black: EKF estimation. In blue: model evolution. In green: confidence interval at 99%. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

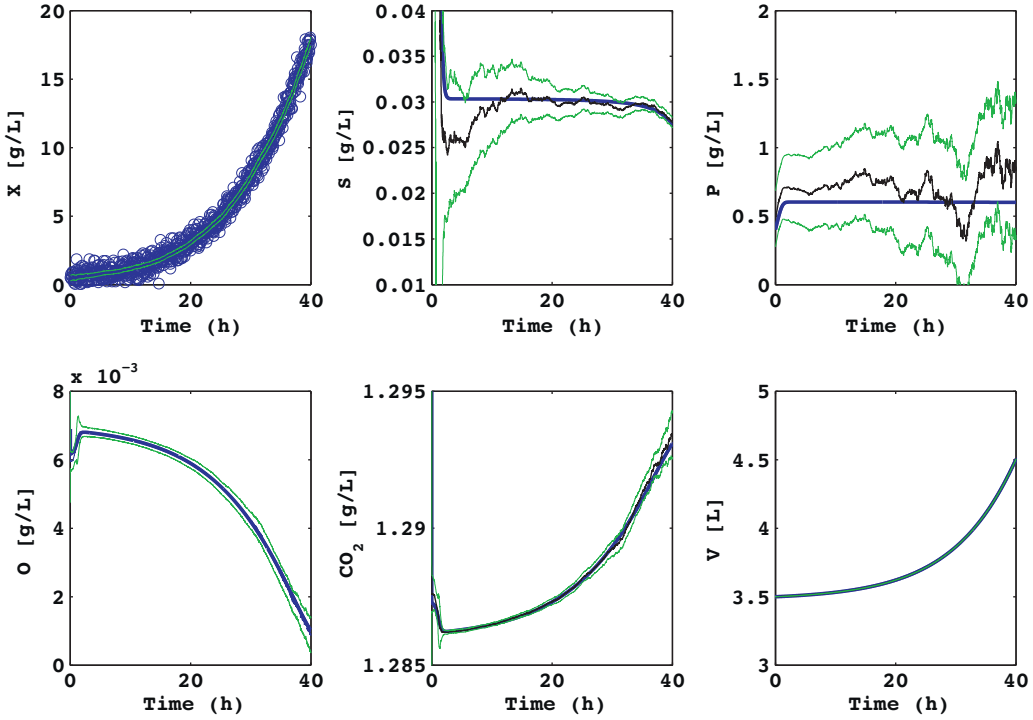


**Fig. 4.** Estimation of the acetate concentration using the biomass measurement. In black: EKF estimation. In blue: model evolution. In green: confidence interval at 99%. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

completed with $CaCl_2{\cdot}2H_2O$, kanamicin, trace metals solution and vitamin solution but these are not considered as necessary in the present experiments. The preculture is achieved in 2 erlenmeyers, each of them containing 500 ml of medium inoculated after 24 h in the bioreactor already filled with 2 l of medium so that a starting batch phase is always performed in a volume of 3 l before any addition of feeding. The feeding medium is composed of 200 g/l of glucose, 10 g/l of $NH_4Cl$ and 4 g/l of $MgSO_4{\cdot}7H_2O$.

### 5.3. Biomass measurement

Off-line biomass measurement is performed thanks to a Shimadzu UV-vis mini spectrophotometer measuring the optical density (*OD*) at 560 nm. On-line biomass measurement is also available through the use of the turbidity measurement system Fundalux II (also from Sartorius) also delivering an *OD* measurement at a very high sampling rate (5 s).

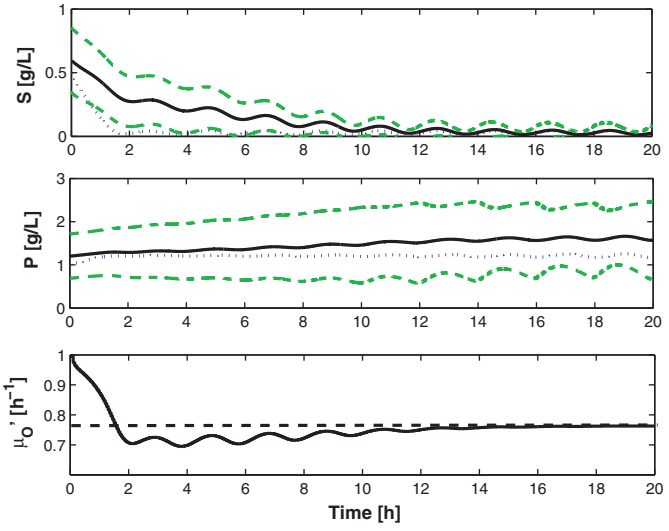


**Fig. 5.** Estimation of the substrate and acetate concentrations using the biomass measurement (without noise) and the estimation of the kinetic parameter $\mu_O$. In black: EKF estimation. In dotted-blue: model evolution. In dashed-green: confidence intervals at 99%. In dashed-black: $\mu_O$ nominal value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

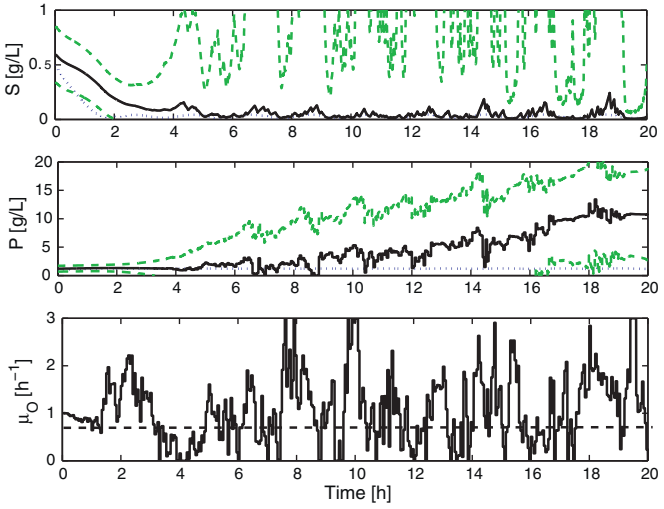


**Fig. 6.** Estimation of the substrate and acetate concentrations using the biomass measurement (with noise) and the estimation of the kinetic parameter $\mu_O$. In black: EKF estimation. In dotted-blue: model evolution. In dashed-green: confidence intervals at 99%. In dashed-black: $\mu_O$ nominal value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

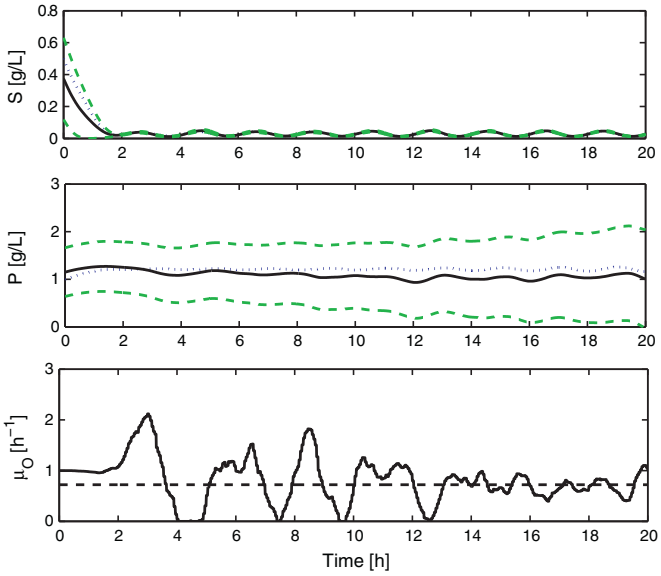


**Fig. 7.** Estimation of the substrate and acetate concentrations using the low-pass filtered biomass measurement (with noise) and the estimation of the kinetic parameter $\mu_O$. In black: EKF estimation. In dotted-blue: model evolution. In dashed-green: confidence intervals at 99%. In dashed-black: $\mu_O$ nominal value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

### 5.4. Glucose measurement

Off-line glucose measurements are performed by a Process Trace enzymatic system. Note that on-line measurements can also be achieved through the same system using a dialysis probe or a microfiltration probe. However, the highest sensitivity of available measurement kits is 0.01 g/l. Unfortunately, the level of the critical glucose concentration is in this range. This measurement is then useful in batch operating conditions where glucose concentrations are sufficiently high but less interpretable in fed-batch conditions where the glucose concentration is too small to allow an accurate measurement.

### 5.5. Off-line acetate measurement

Off-line acetate measurements can be achieved by an Alliance HPLC (Waters, USA) using a 3 μm Atlantis C18 column (4, 6 × 150 mm, Waters, USA) at 30 °C and a UV detector 486 (Waters, USA) set at 210 nm in isocratic mode with a $NaH_2PO_4$ 20 mM solution as mobile phase (using a flow rate of 0.5 ml/min). To ensure reproducibility and stability of the measurement, the samples are buffered in the mobile phase prior to injection. For comparison purposes, an acetic acid kit purchased from Megazyme (Ireland) is also used.

### 5.6. On-line gas measurement

Molar fractions of $O_2$ and $CO_2$ in the outlet gas are available thanks to gas analysis using a DUET – gas analyser from System-C-Industry. As explained in [17], *OTR* and *CTR* can be measured from the knowledge of molar fractions of $O_2$, $CO_2$ and $N_2$ in the outlet gas, the volume and the inlet air flow. Indeed, using the ideal gas law, assuming that temperature and pressure are the same for the inlet and outlet gas, that no nitrogen is consumed and a quasi steady-state of oxygen in the gas phase (i.e., no dramatic change in oxygen), the following relation holds [17]:

$$OTR = \frac{Air_{in} M_{O_2}}{V}(O_{2_{in}} - O_{2_{out}}) \tag{18}$$

where, $Air_{in}$ is the volumetric air inflow rate; $M_{O_2}$ is the molar mass of oxygen; $V$ is the medium volume; $O_{2_{in}}$ is the molar fraction of $O_2$ in the inlet gas; $O_{2_{out}}$ is the molar fraction of $O_2$ in the outlet gas (measured by the gas analyser).

Similarly, the *CTR* can be estimated by the following expression:

$$CTR = \frac{Air_{in} M_{CO_2}}{V}(CO_{2_{in}} - CO_{2_{out}}) \tag{19}$$

where $M_{CO_2}$ is the molar mass of carbon dioxide; $CO_{2_{in}}$ is the molar fraction of $CO_2$ in the inlet gas; $CO_{2_{out}}$ is the molar fraction of $CO_2$ in the outlet gas (measured by the gas analyser).

In the next experimental applications, the inlet air flow is chosen equal to 1 l/min and $O_{2_{in}}$ and $CO_{2_{in}}$ are measured in the outlet gas before starting the culture.

## 6. Experimental results

### 6.1. Model simplifications

In order to gain a priori knowledge about this particular bioprocess, before applying any observation or control technique, a simple experimental model identification exercise is performed. Starting from the general model (2), a simplified reaction scheme taking only the respiro-fermentative pathway into account (i.e., oxidation and fermentation of glucose) and considering that oxygen conditions are not limiting the cells growth is derived as follows:

$$\text{Substrate oxidation :} \quad k_{S1} S + k_{O1} O \xrightarrow{r_1 X} k_{X1} X \tag{20a}$$

$$\text{Substrate fermentation :} \quad k_{S2} S + k_{O2} O \xrightarrow{r_2 X} k_{X2} X + k_{P2} P \tag{20b}$$

The specific growth rates are:

$$r_1 = \frac{\min(r_S, r_{S_{crit}})}{k_{S1}} \tag{21}$$

$$r_2 = \frac{\max(0, r_S - r_{S_{crit}})}{k_{S2}} \tag{22}$$

As the oxygen is assumed not limiting, $K_O$ is neglected in (3b) so that $r_{S_{crit}} = (\mu_O/k_{O1})(O/(O+K_O))(Ki_P/(Ki_P+P)) \approx (\mu_O/k_{O1})(Ki_P/(Ki_P+P))$. The kinetic terms associated with the substrate consumption $r_S$ and the critical substrate consumption $r_{S_{crit}}$ are:

$$r_S = \mu_S \frac{S}{S+K_S} \tag{23a}$$

$$r_{S_{crit}} = \frac{r_O}{k_{O1}} \approx \frac{\mu_O}{k_{O1}} \frac{Ki_P}{Ki_P+P} \tag{23b}$$

Therefore, assuming the existence of the fermentation reaction and taking (Section 6.1) into account, the specific growth rates take the following forms:

$$r_1 = \frac{r_{S_{crit}}}{k_{S1}} \tag{24}$$

$$r_2 = \frac{r_S - r_{S_{crit}}}{k_{S2}} \tag{25}$$

Finally, normalizing the yield coefficients with respect to the substrate ($k_{S1} = k_{S2} = 1$), the simplified reaction scheme gives the following component-wise mass balances differential equation system:

$$\frac{dX}{dt} = (k_{X1} r_{S_{crit}} + k_{X2}(r_S - r_{S_{crit}}))X - DX \tag{26a}$$

$$\frac{dS}{dt} = -r_S X - D(S - S_{in}) \tag{26b}$$

$$\frac{dA}{dt} = k_{P2}(r_S - r_{S_{crit}})X - DA \tag{26c}$$

### 6.2. Experimental scheme and parameter identification

Three semi fed-batch (a batch phase followed by a fed-batch phase) experiments are performed. For each run, biomass, glucose and acetate are measured off-line. The biomass concentration is obtained through the *OD* measurement, the glucose through the Process TRACE enzymatic measurement and acetate through HPLC measurement. The initial concentrations are different from one experiment to another.

The identification is achieved through the following steps:

- The identification of $k_{X1}$, $k_{X2}$, $k_{P2}$, $\mu_S$ and $(\mu_O)/(k_{O1})$ is achieved on the basis of the batch phases only ($K_S$ and $Ki_P$ can therefore be neglected as, in batch mode, the glucose concentration is assumed to be much larger than the saturation constant $K_S$ of (3a) so that $\mu_S(S/(S+K_S)) \approx \mu_S$ and, moreover, assuming that the byproduct concentration is smaller than $Ki_P$, $r_{S_{crit}} \approx (\mu_O)/(k_{O1})$).
- Using the corresponding fed-batch phases, a new identification of the previously determined parameters coupled to $K_S$ and $Ki_P$ is performed (using the previous values as starting point so as to alleviate local minima).

For this identification procedure, a simplex method (Nelder–Mead algorithm used in the MATLAB™ function "fminsearch") minimizing, in the Gauss–Markov sense, the following cost function, is used:

$$F_c = \sum_{i=1}^{N} (\xi_{sim}(i) - \xi(i))^T Q^{-1} (\xi_{sim}(i) - \xi(i)) \tag{27}$$

where $\xi = \begin{bmatrix} X & S & A \end{bmatrix}$ is the measured state vector; $\xi_{sim} = \begin{bmatrix} X_{sim} & S_{sim} & A_{sim} \end{bmatrix}$ is the state vector generated by the model; $N$ is the number of measurements; $Q$ is the measurement error covariance (assuming a gaussian measurement noise) of the following form:

$$Q = \begin{pmatrix} \sigma_X^2 & 0 & 0 \\ 0 & \sigma_S^2 & 0 \\ 0 & 0 & \sigma_A^2 \end{pmatrix}$$

where $\sigma_X^2$, $\sigma_S^2$ and $\sigma_A^2$ are respectively the variances of the biomass, substrate and acetate measurement errors. In the following identification procedure, these variances are fixed to 10% of the relative $X$, $S$ or $A$ value.

The first identification uses 2 batch data sets only. The second identification is performed using $\theta_1$ as initial conditions and considers $K_S$ and $Ki_P$ (respectively initialized at 0.1 and 5).

The following parameter values are obtained: $k_{X1} = 1.742$, $k_{X2} = 0.305$, $k_{P2} = 1.196$, $k_{O1} = 0.974$, $k_{O2} = 0.197$, $\mu_S = 0.72$, $K_S = 0.05$, $\mu_O = 0.0935$ and $K_{i_P} = 4$.

Taking results from the literature, e.g., [20], into account, the stoichiometry appears reasonable as the amount of biomass and acetate produced, so as substrate and oxygen consumed in the two main reactions of the fermentative regime are relatively equivalent. Kinetic parameters also take values comparable to the ones reported in the published literature, which suggests that the previous identification lead to a correct set of identified parameters.

### 6.3. Acetate and substrate estimations using the EKF

The EKF presented in Section 3.2 is used in order to estimate the acetate and glucose concentrations on the basis of on-line biomass concentration measurements from a Sartorius™ turbidimetric probe and the identified model (Section 6.1) with identified parameter values from Section 6.2.

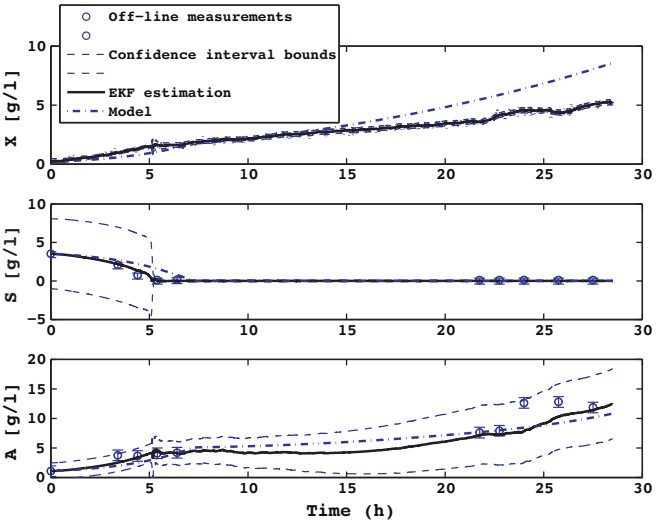


**Fig. 8.** Experimental test of the EKF for the acetate and glucose concentrations using the biomass on-line measurements. In black: EKF estimation. In dash-dot blue: model evolution. In dash blue: confidence interval at 95%. Blue circles: off-line measurements. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

**Table 2**
Observation errors: $\bar{e}$ represents the mean error.

| Experience | Without parameter estimation | | With $\mu_O$ estimation | |
|---|---|---|---|---|
| | $\overline{e_S}$ (g/l) | $\overline{e_A}$ (g/l) | $\overline{e_S}$ (g/l) | $\overline{e_A}$ (g/l) |
| 1 | 0.0027 | 1.2286 | 0.0004 | 0.163 |
| 2 | 0.5567 | 0.7006 | 0.4081 | 0.6145 |

The biomass concentration signal is generally affected by a white noise with a standard deviation around 0.2 g/l. Consequently, the covariance matrix of the measurement noise is designed as $R_\epsilon = 10^{-2}$. Moreover, the confidence degree in the model is chosen through the covariance matrix of the model noise, which in this case is taken as

$$R_\eta = \begin{bmatrix} 10^{-2} & 0 & 0 \\ 0 & 10^{-2} & 0 \\ 0 & 0 & 10^{-2} \end{bmatrix},$$

i.e., reflecting an equal level of confidence in the measurements and the model prediction. Note that biomass on-line measurements present a very high sampling (the sampling period is 5 s) so that they appear as continuous in the next results.

For the first experiment (see Fig. 8), the observer obviously performs better than the identified model (which, as the biomass grows, drifts away from the turbidimetric on-line measurements) testifying of a certain level of robustness against model errors thanks to the biomass measurement. Indeed, glucose and acetate estimations fit very well with the off-line measurements. However, note that the confidence interval increases with time, indicating a decreasing accuracy of the estimation. During the fed-batch phase (i.e., when the glucose concentration becomes close to 0 g/l), off-line measurements also lack accuracy as they approach the sensitivity level (0.1 g/l). One can consider that the estimation of the EKF around 0.01 g/l is more reliable as the confidence interval is tight.

During the second experiment (see Fig. 9), model prediction and EKF estimation are both close to the off-line measurements and the quality of the results is almost comparable to the previous experiment. In the last hours of the culture, the feed profile is not

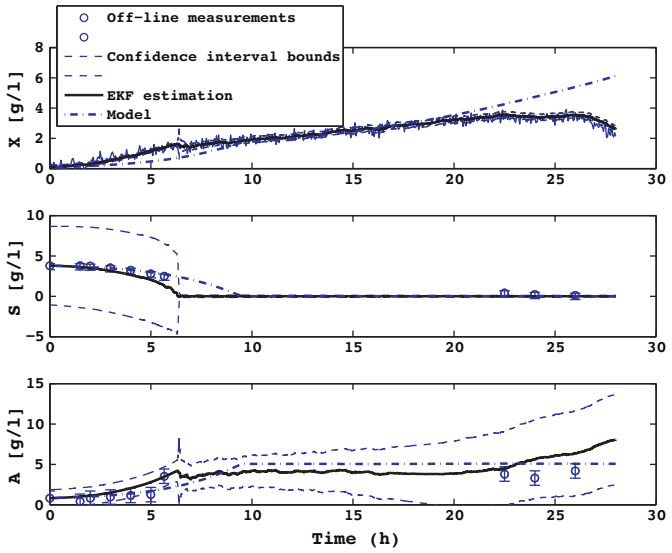


**Fig. 9.** Experimental test of the EKF for the acetate and glucose concentrations using the biomass on-line measurements. In black: EKF estimation. In dash-dot blue: model evolution. In dash blue: confidence interval at 95%. Blue circles: off-line measurements. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

sufficient to keep the cells in the respiro-fermentative regime so that acetate is slightly reconsumed. This explains why neither the model nor the observer succeed in following the last two off-line measurements. Further results in the identification of a respirative regime model could improve these results. However, as the goal of further research is to couple this observer to an acetate concentration regulator, a model-based observer like the EKF, taking only the respiro-fermentative regime into account, could be sufficient. Indeed, as the acetate concentration is controlled, the quantity of acetate must always increase to counteract the dilution effect all along the fed-batch process.

### 6.4. Acetate, substrate and kinetic parameter estimations using the EKF

As introduced in Section 4, augmenting the estimated state vector with the kinetic parameter $\mu_O$ could be interesting in order to make the substrate and acetate estimations more robust to model uncertainties. The EKF with parameter estimation can now be tested with the same experimental data sets as in Section 6.3. Qualitatively, the results are similar to Figs. 8 and 9, and Table 2 shows the estimation mean error values for each estimated state (i.e., substrate and acetate) during the experiments with and without parameter estimation. Obviously, the estimation of $\mu_O$ is beneficial to state estimation.

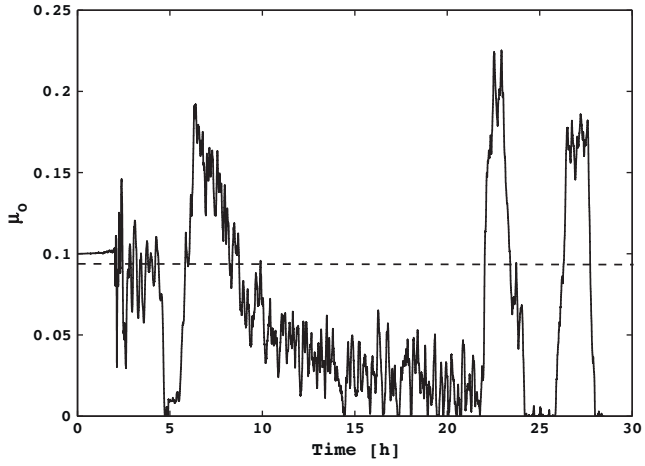


**Fig. 10.** Experimental test of the EKF for the acetate and glucose concentrations together with kinetic parameter estimation, using the biomass on-line measurements. Evolution of $\mu_O$. Solid line: EKF estimation. Dashed line: nominal value.

Moreover, the evolution of the parameter $\mu_O$ during the first experiment (presented in Fig. 10), shows that the nominal value 0.0935 is only valid in the first few hours of the culture. This is an important observation suggesting that the respiratory capacity of the biomass decreases due to the accumulation of a metabolic product such as acetate or nutrient exhaustion. The evolution of $\mu_O$ during the second experiment is comparable. Note that the same low-pass filter as described in Section 4 was used during the experiments.

## 7. Conclusion

The availability of biomass probes offers new possibilities for estimating the acetate and glucose concentrations in fed-batch cultures of *E. coli*. Theoretical and practical observability analyses show that the best sensor configuration consists of a biomass probe together with the measurement of dissolved oxygen and off-gas analysis. However, the sole measurement of biomass already provides satisfactory results, when the underlying bioprocess model is accurate enough (i.e., has been carefully identified from experimental data). Robustness of the estimator can be enhanced by adaptating a kinetic parameter on-line.

## Acknowledgments

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

## References

[1] H. Crabtree, Observations on the carbohydrate metabolism of tumors, Biochemical Journal 23 (1929) 536–545.
[2] R.H.D. Deken, The crabtree effect: a regulatory system in yeast, Journal of General Microbiology 44 (1966) 149–156.
[3] Y. Pomerleau, Modélisation et commande d'un procédé fed-batch de culture des levures à pain, Ph.D. Thesis, Département de génie chimique, Université de Montréal, 1990.
[4] L. Chen, G. Bastin, V. van Breusegem, A case study of adaptive nonlinear regulation of fed-batch biological reactors, Automatica 31 (1) (1995) 55–65.
[5] F. Renard, A. Vande Wouwer, S. Valentinotti, D. Dumur, A practical robust control scheme for yeast fed-batch cultures – an experimental validation, Journal of Process Control 16 (2006) 855–864.
[6] L. Dewasme, A. Richelle, P. Dehottay, P. Georges, M. Remy, P. Bogaerts, A. Vande Wouwer, Linear robust control of *S. cerevisiae* fed-batch cultures at different scales, Biochemical Engineering Journal 53 (1) (2010) 26–37.
[7] G. Bastin, D. Dochain, On-Line Estimation and Adaptive Control of Bioreactors Process Measurement and Control, vol. 1, Elsevier, Amsterdam, 1990.
[8] P. Bogaerts, A. Vande Wouwer, Software sensors for bioprocesses, ISA Transactions 42 (2003) 547–558.
[9] A.C.A. Veloso, I. Rocha, E.C. Ferreira, Monitoring of fed-batch *E. coli* fermentations with software sensors, Bioprocess and Biosystems Engineering 32 (2009) 381–388.
[10] M. Arndt, S. Kleist, G. Miksch, K. Friehs, E. Flaschel, J. Trierweiler, B. Hitzmann, A feedforward–feedback substrate controller based on a Kalman filter for a fed-batch cultivation of *Escherichia coli* producing phytase, Computers and Chemical Engineering 29 (2005) 1113–1120.
[11] B. Sonnleitner, O. Käppeli, Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: formulation and verification of a hypothesis, Biotechnology and Bioengineering 28 (1986) 927–937.
[12] J.-P. Gauthier, I. Kupka, Observability and observers for nonlinear systems, SIAM Journal Control and Optimization 32 (4) (1994) 975–994.
[13] M. Zeitz, Observability canonical (phase-variable) form for nonlinear time-variable systems, International Journal of System Science 15 (9) (1984) 949–958.

[14] P. Bogaerts, A. Vande Wouwer, Parameter identification for state estimation – application to bioprocess software sensors, Chemical Engineering Science 59 (12) (2004) 2465–2476.
[15] L. Dewasme, P. Bogaerts, A. Vande Wouwer, Monitoring of bioprocesses: mechanistic and data-driven approaches, in: Maria do Carmo Nicoletti, Lakhmi C. JainIn (Eds.), Computational Intelligent Techniques for Bioprocess Modelling, Supervision and Control, Studies in Computational Intelligence, Springer Verlag, Heidelberg, 2009, pp. 57–97.
[16] A. Gelb, The Analytic Sciences Corporation Technical Staff: Applied Optimal Estimation, MIT Press, Cambridge, 1974.
[17] I. Rocha, Model-based strategies for computer-aided operation of recombinant *E. coli* fermentation, Ph.D. Thesis, Universidade do Minho, 2003.
[18] L. Dewasme, B. Srinivasan, M. Perrier, A. Vande Wouwer, Extremum-seeking algorithm design for fed-batch cultures of microorganisms with overflow metabolism, Journal of Process Control 21 (7) (2011) 1092–1104.
[19] M.S. Grewal, A.P. Andrews, Kalman Filtering: Theory and Practice Using MATLAB, third ed., Wiley, New York, 2008.
[20] M. Akesson, P. Hagander, J.-P. Axelsson, Avoiding acetate accumulation in *Escherichia coli* cultures using feedback control of glucose feeding, Biotechnology and Bioengineering 73 (3) (2001) 223–230.