Biochemical Engineering Journal 135 (2018) 22–35

Contents lists available at ScienceDirect

Biochemical Engineering Journal

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

Regular article

Parameter estimation of a dynamic model of *Escherichia coli* fed-batch cultures

C. Retamal [a], L. Dewasme [a,*], A.-L. Hantson [b], A. Vande Wouwer [a]

[a] Automatic Control Laboratory, University of Mons, 31, Boulevard Dolez, 7000 Mons, Belgium
[b] Department of Chemical and Biochemical Engineering, University of Mons, Belgium

ARTICLE INFO

Article history:
Received 19 September 2017
Received in revised form 5 February 2018
Accepted 12 March 2018
Available online 21 March 2018

Keywords:
Overflow metabolism
Mathematical modeling
Parameter estimation
Biotechnology

ABSTRACT

This work proposes an original sequential parameter identification method of a dynamic model of *Escherichia coli BL*21(*DE*3) fed-batch cultures. The proposed macroscopic model is based on the overflow (or bottleneck) metabolism assumption of Sonnleitner and Kappeli, suggesting two metabolic pathways (respirative and respiro-fermentative), and consists of a set of nonlinear mass balance differential equations. Model unknown parameters are estimated from dedicated experimental data collected from a 5-liter pilot bioreactor. Experiments are designed in order to force switches from one metabolic pathway to another. A sequential identification procedure, based on a specific data partitioning, is achieved and results are qualitatively and quantitatively assessed using Fisher information. The resulting dynamic model is in good agreement with the experimental data, and could be used for process optimization and/or control. The identification procedure could also easily be transposed to other processes.
© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Escherichia coli is one of the most popular host microorganisms for the production of several biopharmaceuticals, such as recombinant proteins, insulin, hormones, among others. *E. coli* has several interesting features, including rapid growth, possibly high cell densities, known genome sequence and simple nutritional requirements. To reach high cell densities, a rich medium supplemented with a high glucose concentration appears to be a straightforward approach. However, an excessive concentration of this carbon source can lead, under aerobic conditions, to acetate production which inhibits cell growth. This phenomenon is known as overflow metabolism or short-term Crabtree effect [1].

It has been suggested that the acetate formation could be due to an imbalance between glucose metabolism and respiration, NADH excess, repression of the TCA cycle via an enzymatic limitation [2–4] and a lack of regulation in the maximum glucose uptake rate velocity by the phosphotransferase system [5]. It is likely that all these causes are interrelated. In practice, it is observed that both rapid growth and high glucose levels may contribute to the production of acetate [6], and a critical level of glucose can be defined, which depends on the strain, culture conditions and medium composition.

Several strategies based on the process operating conditions have been used to reduce the acetate production, where cells are forced to grow below the critical specific growth rate using a fed-batch culture mode. Different feeding strategies can be used such as exponential feeding, pH-stat, DO-stat, among others [7,8]. These strategies require the development of control algorithms which rely on the measurement (or estimation) of the key components. Controllers also require a minimum of process a priori knowledge, in the form of a dynamic model. For this purpose, grey-box models are often developed, as opposed to descriptive white-box models derived from metabolic fluxes (as in [9,10]). These models should predict the real system behaviour with reasonable accuracy, while remaining as simple as possible to avoid parameter identifiability issues (too many parameters cannot be identified from a set of limited experimental data), complexity issues (the relative complexity of the models precludes stability and performance analysis), as well as computational issues (the model might be used in a model-based control strategy).

Sonnleitner and Kappeli [11] proposed a mechanistic model of overflow in baker yeast cultures, based on the bottleneck assumption. It assumes that aerobic glucose metabolism is ruled by the cell respiratory capacity and carried out through two pathways: one fraction of glucose is metabolized through oxidation (respiration) while the other one is used in a reductive pathway (fermentation), producing ethanol. On the other hand, ethanol can also be metabolized by oxidation as substitute carbon source. This catabolism is

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

https://doi.org/10.1016/j.bej.2018.03.012
1369-703X/© 2018 Elsevier B.V. All rights reserved.

assumed to be governed by the respiratory capacity of the cells represented by a bottleneck. When glucose flux is low enough (subcritical) and entirely contained into the bottleneck, only oxidative metabolism is activated, i.e. glucose and/or ethanol oxidation. Growth on glucose and/or ethanol occurs indeed in parallel. Conversely, if the glucose flux exceeds (supracritical) the respiration bottleneck, part of glucose is oxidized and what remains is reduced in ethanol production.

Regarding *E. coli*, several authors consider Sonnleitner bottleneck assumption as a basis for the elaboration of their models [12,5,13,14,6,8,15]. One of the most detailed models is presented in [5] where each pathway is subdivided into a flux used for anabolism and another one used for energy purposes. However, this leads to a more complex representation with additional parameters. Recently, [16] proposed an original moving-horizon on-line estimation method aiming at validating quickly an identified model of new strains, resulting for instance, of genetic modifications or mutations).

The bottleneck assumption is of course not the only interpretation of the Crabtree effect. For instance, [17] considers overflow metabolism as a consequence of acetate cycling, as observed in recent proteomic studies. Also, in-depth studies of physiological aspects are reported in the literature, such as [18–22]), but they go beyond the level of description which is adequate for process control, in a large sense.

In this latter connection, the vast majority of the dynamic models consider the critical level triggering acetate formation as constant. However, some recent models assume that the critical glucose uptake rate decreases with the increase of acetate concentration. Thus, acetate inhibition has to be taken into account in the respiratory capacity, assuming a varying critical glucose specific uptake rate. It is the case for the *E. coli* model developed in [8] and the generic overflow model presented in [23].

In this context, the objective of this study is to establish a dynamic model of fed-batch cultures of *E. coli* BL21(DE3), and to estimate model parameters based on a relatively large set of experimental data. The proposed model is derived in the same spirit as [8] and [15], i.e. with state estimation (software sensor) and control applications in view (such as in [24–26]).

Experiment design is used in order to define informative operating conditions, and a sequential identification procedure based on data partitioning (inspired by [27]) is developed. Parameter accuracy is assessed using the Fisher Information Matrix (FIM) and the evaluation of confidence intervals for the parameters and the model prediction.

This paper is organized as follows. Section 2 introduces the formulation of a macroscopic model of fed-batch cultures of *E. coli*. Section 3 presents parameter estimators applied in this study. The experimental setup is described in Section 4 and experiment design in Section 4.4. Data partitioning, parameter identifiability and the related sequential parameter identification procedure are respectively presented in sections 5, 5.1 and 6. Conclusions are drawn in Section 7.

2. Macroscopic model of *E. coli*

The macroscopic model considered in this study describes the metabolism of *E. coli* through the following reaction scheme:

$$\text{Substrate oxidation:}\quad k_{S1}\mathrm{S} + k_{O1}\ \mathrm{O} \overset{\mu_1 X}{\rightarrow} k_{X1}\ \mathrm{X} + k_{C1}\ \mathrm{C} \tag{1a}$$

$$\text{Substrate fermentation:}\quad k_{S2}\mathrm{S} + k_{O2}\ \mathrm{O} \overset{\mu_2 X}{\rightarrow} k_{X2}\ \mathrm{X} + k_{A2}\ \mathrm{A} + k_{C2}\ \mathrm{C} \tag{1b}$$

$$\text{Acetate oxidation:}\quad k_{A3}\mathrm{A} + k_{O3}\ \mathrm{O} \overset{\mu_3 X}{\rightarrow} k_{X3}\ \mathrm{X} + k_{C3}\ \mathrm{C} \tag{1c}$$

where X, S, A, O and C are, respectively, the concentration in the culture medium of biomass, glucose, acetate, dissolved oxygen and carbon dioxide. $k_{\xi i}$ (i = 1, 2, 3, ξ = X, S, P, O, C) are the yield coefficients and μ_1, μ_2 and μ_3 are the specific growth rates which are based on Sonnleitner's bottleneck assumption [11]: the cells are likely to change their metabolism because of their limited oxidative capacity. In other words, when the concentration of glucose is larger than the critical concentration S_{crit} and likewise the glucose consumption rate is larger than the critical consumption rate μ_{Scrit}, the cells will produce acetate. This respiro-fermentative mode is represented by reactions (1a) and (1b). On the other hand, when the substrate becomes limiting, i.e. the substrate concentration is lower than the critical level, then the available substrate is oxidized. Moreover, acetate, if present in the culture medium, is oxidized as a secondary carbon source. This respirative mode is explained by reactions (1a) and (1c).

Hence, the specific growth rates are defined by:

$$\mu_1 = \frac{\min(q_S, q_{Scrit})}{k_{S1}} \tag{2a}$$

$$\mu_2 = \frac{\max(0, q_S - q_{Scrit})}{k_{S2}} \tag{2b}$$

$$\mu_3 = \frac{\max(0, q_{AC})}{k_{A3}} \tag{2c}$$

and the consumption rates are given by:

$$q_S = q_{Smax}\frac{S}{K_S + S} \tag{3a}$$

$$q_{Scrit} = \frac{q_O}{k_{OS}} = {q_{Omax}}^* \frac{K_{iA}}{K_{iA} + A} \quad {q_{Omax}}^* = \frac{q_{Omax}}{k_{OS}} \tag{3b}$$

$$q_{AC} = \frac{k_{OS}(q_{Scrit} - q_S)}{k_{OA}}\frac{A}{A + K_A} \tag{3c}$$

where q_S and q_{Smax} are the glucose consumption rate and its maximal value, respectively. q_{Scrit} is the critical consumption rate associated to the respiratory capacity q_O and q_{Omax} is the maximal value. q_{AC} is the acetate consumption rate. k_{OS} and k_{OA} are respectively the yield coefficients between the oxygen and substrate consumptions and between oxygen and acetate consumptions [8]. Their function is to normalize the oxygen consumption rate in order to compare the glucose consumption rate and the acetate consumption rate, respectively. K_S, K_A and K_O are the saturation constants for glucose, acetate and oxygen, respectively. K_{iA} is the acetate inhibition constant. It is important to note that cultures are assumed to be well oxygenated (i.e. no limitation of oxygen is considered).

Mass balance equations for a bioreactor operated in batch-mode are defined by (4)–(6). Note that the mass balances are normalized with respect to the substrates (k_{S1} = k_{S2} = k_{A3} = 1).

$$\frac{d\xi}{dt} = K\varphi - D\xi + F + Q \tag{4}$$

$$\xi = \begin{pmatrix} X \\ S \\ A \\ O \\ C \end{pmatrix},\quad K = \begin{pmatrix} k_{X1} & k_{X2} & k_{X3} \\ -1 & -1 & 0 \\ 0 & k_{A2} & -1 \\ -k_{O1} & -k_{O2} & -k_{O3} \\ k_{C1} & k_{C2} & k_{C3} \end{pmatrix},\quad \varphi = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} X \tag{5}$$

$$F = \begin{pmatrix} 0 \\ DS_{in} \\ 0 \\ 0 \end{pmatrix}, \quad Q = \begin{pmatrix} 0 \\ 0 \\ OTR \\ -CTR \end{pmatrix} \tag{6}$$

OTR is the oxygen transfer rate which is defined as the amount of oxygen transferred from the gas phase to the liquid phase. The solubility of oxygen in the culture medium is low, and one can assume that all the dissolved oxygen is almost instantaneously consumed, leading to a slow-fast approximation with $dO/dt \approx O$. As dilution effects are negligible, *OTR* can be expressed as:

$$OTR = (k_{O1}\mu_1 + k_{O2}\mu_2 + k_{O3}\mu_3)X \tag{7}$$

CTR represents the carbon dioxide transfer rate from the liquid to the gas phase. Compared to the oxygen, the solubility is much higher and is enhanced by the production of bicarbonate at the operating pH (pH 7). Hence, the dissolved concentration cannot be considered constant.

3. Parameter estimation procedure

3.1. Model parametrization

A bioprocess model is developed from the a priori knowledge of the biological system, i.e. the appropriate choice of a reaction scheme (1), kinetics (2) and mass balance equations of the key macroscopic components (5) and (6). The next logical step is the estimation of the model parameters from experimental data. However, accurate estimation of model parameters is a difficult task in relation with the collection of data from experiments carried out in different conditions, varying for instance the initial concentrations in batch experiments or the dilution rate profile in fed-batch experiments. The selection of a range of operating conditions is particularly important to ensure a sufficient amount of information about the several underlying biological phenomena included in the model, and in turn about the parameters describing these phenomena (for instance, inhibition can only be detected when concentrations are driven in specific ranges). Information is moreover impacted by the level of noise corrupting the data.

Parameter estimation is based on the minimization of a cost function measuring the distance between the experimental data and the model prediction. In the following, the parameter identification methods used in this study are presented as well as the quantification of the uncertainty on the estimation.

3.2. Non-linear least squares estimator

A non-linear dynamical model can be written in a state-space form as:

$$\dot{\xi} = f(\xi, \theta, t), \quad \xi(0) = \xi_0(\theta) \tag{8}$$

where $\xi(t, \theta)$ is the state space vector and θ is the vector of parameters. The model output and the vector of measurements are represented, respectively, by:

$$y_m(t, \theta) = h(\xi, \theta, t); \quad y(k) = y_m(k, \theta^*) + \varepsilon_k, \quad k = 1, \ldots, n_t \tag{9}$$

where n_t is the number of measurement times, θ^* is the true value of the parameters (assuming that the model structure is correct, i.e. there is no error of characterization), and ε_k represents the measurement errors which are assumed to be independent, white with a zero-mean Gaussian distribution, i.e. $\varepsilon_k \sim N(0, \Sigma)$.

The parameter estimates are solutions of the following minimization problem:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} J(\theta) \tag{10}$$

A weighted least-square cost function is given by:

$$J(\theta) = \sum_{k=1}^{n_t} (y(t_k) - y_m(t_k, \theta))^T \Sigma^{-1} (y(t_k) - y_m(t_k, \theta)) \tag{11}$$

The definition of the cost function $J(\theta)$ depends on the selection of the weighting matrix Σ. Assuming that measurements are corrupted by independent and identically distributed white noises, Σ is a diagonal matrix with identical elements

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma^2} & 0 & 0 \\ 0 & \hat{\sigma^2} & 0 \\ 0 & 0 & \hat{\sigma^2} \end{pmatrix} \quad \hat{\sigma^2} = \frac{J(\theta)}{n_t - p} \tag{12}$$

and the measurement variance can be estimated a posteriori whenever needed. In (12), p represents the number of parameters and n_t the number of measurements. In this work, the matrix Σ is chosen to represent independent, but not identically distributed noises. Σ includes an a priori chosen scaling matrix, e.g. a diagonal matrix S with the square of the maximum output values could be used to scale the variables, i.e.

$$\hat{\Sigma} = \frac{J(\theta)}{n_t - p} S_{meas} \tag{13}$$

and possibly, a posterior estimate of the covariance of the measurement errors.

In the particular case where the model is linear in the parameters or can be linearized, minimization (10) has an analytic solution. Otherwise, the Nelder-Mead method, as implemented in function 'fminsearch' is used in this study to solve the minimization problem. This method minimizes a function of several variables without using derivatives, iterating transformations of a simplex in parameter space, i.e. of a polytope with $p+1$ dimension vertices. The advantage of this method is its relative robustness to noise in the cost function, but it suffers from a slow convergence. It is important to note the presence of local minima if the initial guesses are far away from the true parameters since the method tends to converge to the closest minimum with respect to the initial guess. A solution is to apply a multi-start procedure using a large number of different initial guesses.

3.3. Parametric sensitivity and uncertainty

A parameter will be accurately estimated if it has a strong influence on the model output. Thus, the larger the influence, the more accurate the estimated value. This influence can be quantified through parametric sensitivities computed by:

$$\frac{\partial y_\theta}{\partial t} = \frac{\partial f}{\partial y} y_\theta + \frac{\partial f}{\partial \theta}; \quad y_\theta(0) = \frac{\partial y(0)}{\partial \theta} \tag{14}$$

For the ease of comparison, each sensitivity function is normalized by multiplying it with the parameter value under study and dividing it by the mean value y_{mean} of the state variable as follows:

$$\frac{\partial y_i}{\partial \theta_i} \frac{\theta_i}{y_{imean}} \tag{15}$$

In order to characterize the uncertainty in the estimated parameters, a bound on the covariance matrix can be obtained based on the inverse of the Fisher information matrix (FIM):

$$\hat{P} \succ FIM^{-1}(\theta, \Sigma) \tag{16}$$

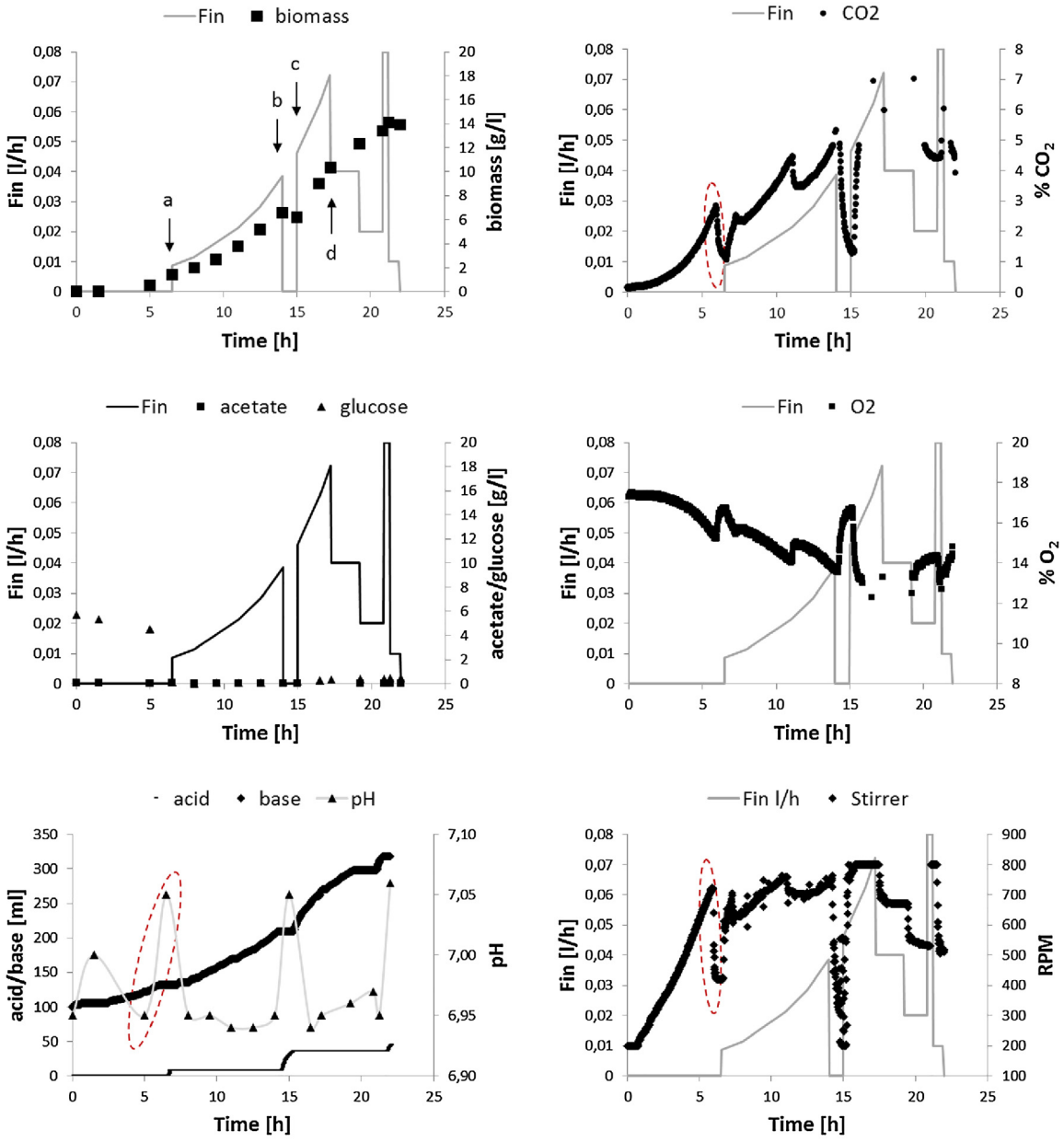

Fig. 1. Experiment 1. Time profiles of feed-rate, biomass, acetate, glucose, added acid and base, pH, volumetric fractions of CO_2 and O_2 and stirrer speed. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

Under the assumption of Gaussian noise, the FIM can be computed by:

$$FIM = \sum_{i=1}^{n_t} y_\theta(\xi(i), \hat{\theta})^T \Sigma^{-1} y_\theta(\xi(i), \hat{\theta}) \quad (17)$$

The 95% of confidence interval for θ_j is given by: $\hat{\theta}_j \pm 2\sigma_{\theta_j}$, with σ_{θ_j} being the standard deviation of $\hat{\theta}_j$ and corresponds to the square root of the j_{th} diagonal element of $\hat{P}$.

4. Materials and methods

4.1. Strain and culture medium

The considered strain is *E. coli* BL21(DE3), which is widely used in industry and produces less acetate than other *E. coli* strains [2].

The medium composition is based on the work of [28], defined for high cell density (DHCD medium) and composed by (per liter of deionized water): $MgSO_4 \cdot 7H_2O$ (1.2 g), $(NH_4)_2HPO_4$ (6.6 g), citric acid (1.7 g), KH_2PO_4 (13.3 g), trace metal solution (10 mL). One liter of trace metal solution contains Fe(III)citrate (6 g), $MnCl_2{\cdot}4H_2O$ (1.5 g), $Zn(COO)_2{\cdot}2H_2O$ (0.8 g), $Na_2MoO_4{\cdot}2H_2O$ (0.25 g), $CuCl_2{\cdot}2H_2O$ (0.15 g) and EDTA (0.8 g). In the current batch cultures, DHCD medium is supplemented with 5 g/L of glucose. In fed-batch cultures, feeding solution contains glucose (500 g/L) and $MgSO_4$ (13 g/L). During the preparation, solutions were filtered to avoid contamination during cultures.

4.2. Experimental set-up

Cell cultures are performed in a 5 L Biostat B+ from Sartorius-Stedim. This bioreactor is equipped with a water jacket for

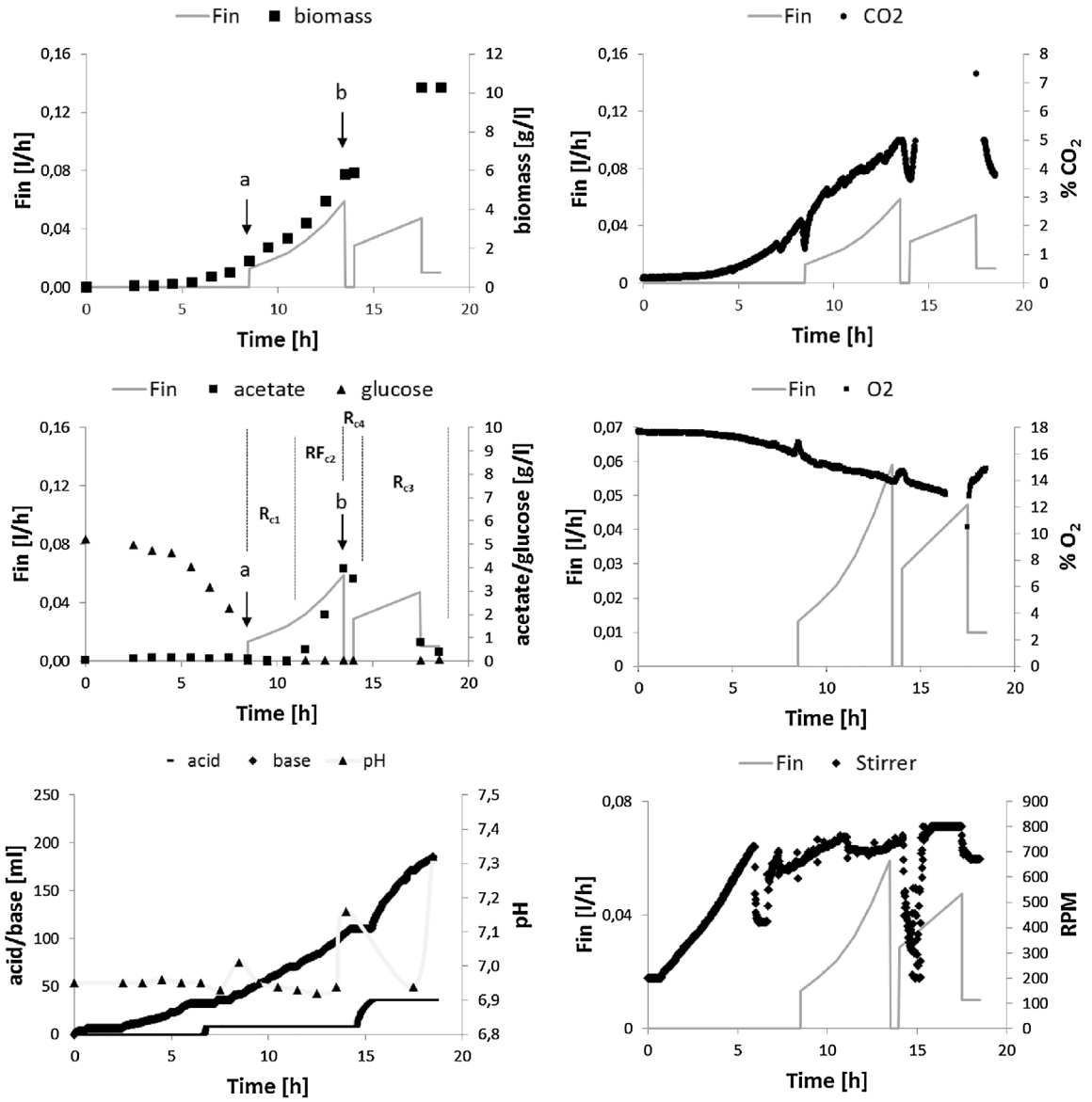

Fig. 2. Experiment 2. Time profiles of feed-rate, biomass, acetate, glucose, added acid and base, pH, volumetric fractions of CO_2 and O_2 and stirrer speed. R_{c1}: respirative regime from scenario 1. RF_{c2}: respiro-fermentative regime from scenario 2. R_{c3}: Respirative regime from scenario 3. R_{c4}: respirative regime from scenario 4.

temperature control, agitation rotor, sensors and regulation for dissolved oxygen, pH and temperature.

Biomass concentration is measured off-line using optical density (at 600 nm) from a Shimadzu UV Spectrophotometer, correlated with dry cell weight. Glucose in the supernatant is measured using the DNS method which is based on the presence of reducing sugars. Acetate in the culture supernatant is detected using an enzymatic kit from Megazyme.

It is important to note that biomass determination takes around 10 minutes and it is carried out after every sampling. In the case of glucose measurements, the results are obtained within 45 min and carried out every 2 sampling times. However, acetate determination takes much longer. Consequently, acetate samples are frozen and analysis are achieved, generally, at the end of the culture.

During the fed-batch phase, samples are taken almost every hour. However, in some cases, the sampling is shorter, to catch the dynamics induced by quick feed-rate changes, or larger, for instance, in the case of an overnight batch.

The off-gas stream is dehumidified by a condenser and directed towards a gas analyzer DUET from system C-industry which measures the volumetric fractions of oxygen and carbon dioxide. This information is used for computing the *OTR* (oxygen transfer rate) and *CTR* (carbon transfer rate) as follows:

$$OTR = \frac{Q_{air} M_{O_2} (O_{2in} - O_{2out})}{22.4 V_M}, \quad CTR = \frac{Q_{air} M_{CO_2} (CO_{2in} - CO_{2out})}{22.4 V_M} \tag{18}$$

with Q_{air} being the volumetric air inflow rate (l/h), M_{O_2} the molar mass of oxygen, M_{CO_2} the molar mass of CO_2, O_{2in} (CO_{2in}) the molar fraction of oxygen (CO_2) in the inlet gas, O_{2out} (CO_{2out}) the volumetric fraction of oxygen (CO_2) in the outlet gas which is measured by the gas analyser and V_M the working volume in the bioreactor. Finally, the factor 22.4 refers to the molar volume of a gas (l_{gas}/mol_{gas}) in standard conditions of temperature and pressure.

Four experiments are performed. The cell cultures consist in a batch phase followed by a fed-batch one. Previously, a pre-culture containing DHCD medium and 10 g/L of glucose is performed in a shake flask at 37°C and an agitation of 200 rotations per minute (RPM). For experiments 1 and 2, pre-cultures are incubated overnight for 14 h. The bioreactor is inoculated with the pre-culture

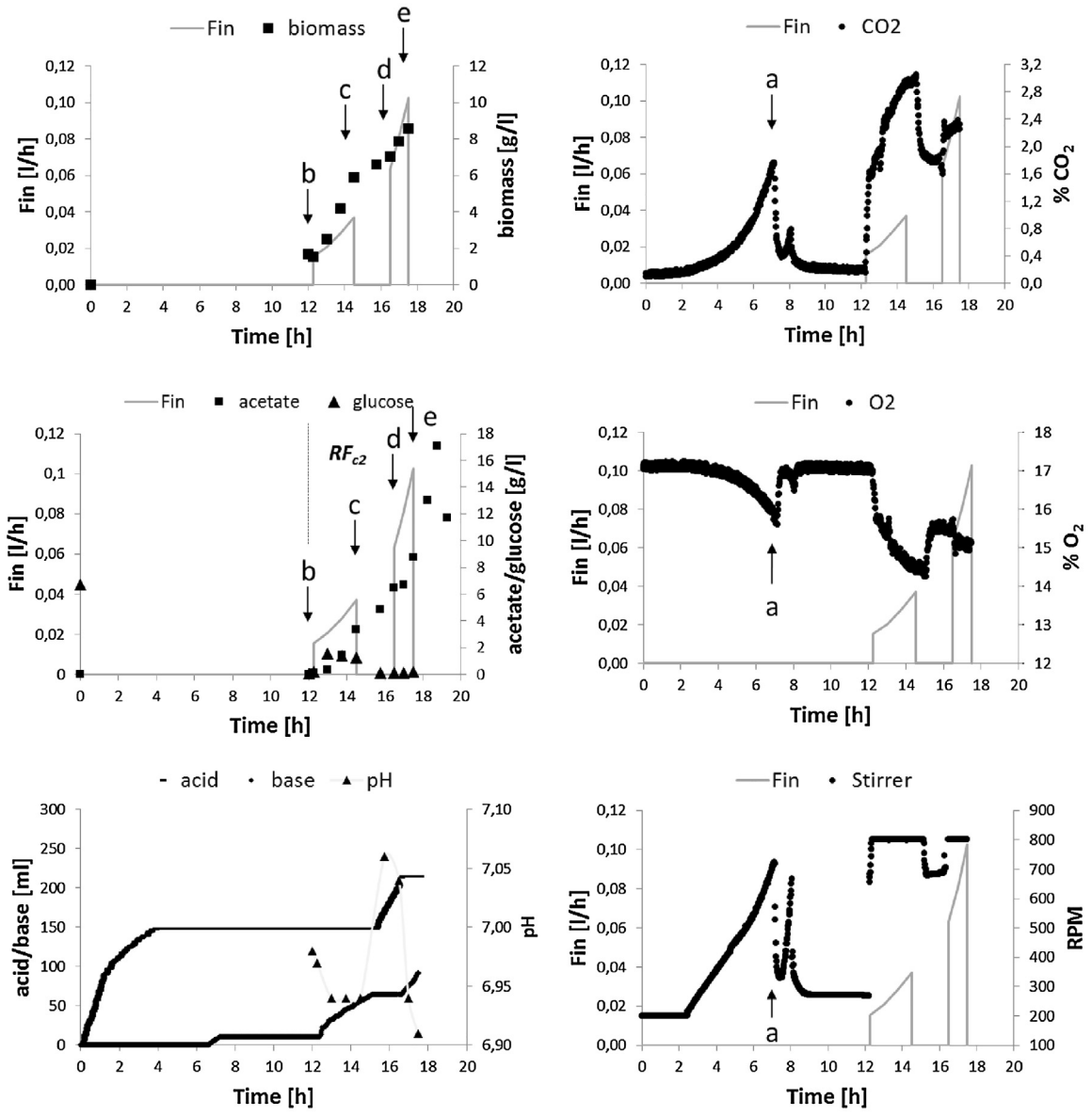

Fig. 3. Experiment 3. Time profiles of feed-rate, biomass, acetate, glucose, added acid and base, pH, volumetric fractions of CO_2 and O_2 and stirrer speed. RF_{c2}: respiro-fermentative regime from scenario 2.

and the initial optical density (OD) reaches typically 0.3 to 0.6. The batch phase is monitored during the day. Once the glucose is depleted, the feeding solution is added and the fed-batch starts.

For experiments 3 and 4, both batch phases, pre-cultures and reactor cultures, differ from the previous experiments. Indeed, pre-culture is performed during the day and the incubation lasts 6 h. The batch phase in the bioreactor is performed overnight, lasts 14 h and is not monitored. The initial optical density (OD) in batch phase is very low in order to reach glucose depletion later than experiments 1 and 2. Nevertheless, cells are facing starvation for a few hours. However, it was reported that *E. coli* remains metabolically active and can quickly resume growth when the required nutrients are available [29,30].

4.3. Experimental strategy

The operating conditions affect the uncertainty of the parameter estimation (available sensors, inputs, measurement times, among others). Moreover, it is very important to design experiments taking the final objective into account [31]: in the current case, parameter identification. The fed-batch experiments are carried out setting different feed-rate profiles in order to provide informative measurements with respect to the parameters associated to both metabolic regimes: respirative and respiro-fermentative. Moreover, the following scenarios might be observed during fed-batch cultures:

1. Scenario 1: Cells consume glucose as sole carbon source, i.e. no acetate is present in the culture, and no acetate production is observed. In this case, only the first reaction (1a) is concerned and cells are in respirative regime. It follows that:

$$q_S \leq q_{Scrit}; \quad q_{AC} = 0 \rightarrow \mu_1 = q_S = q_{Smax}\frac{S}{K_S + S}; \quad \mu_2 = \mu_3 = 0 \tag{19}$$

2. Scenario 2: Cells are overflowed with glucose (consequently as sole carbon source) and acetate is produced. Cells are in respiro-

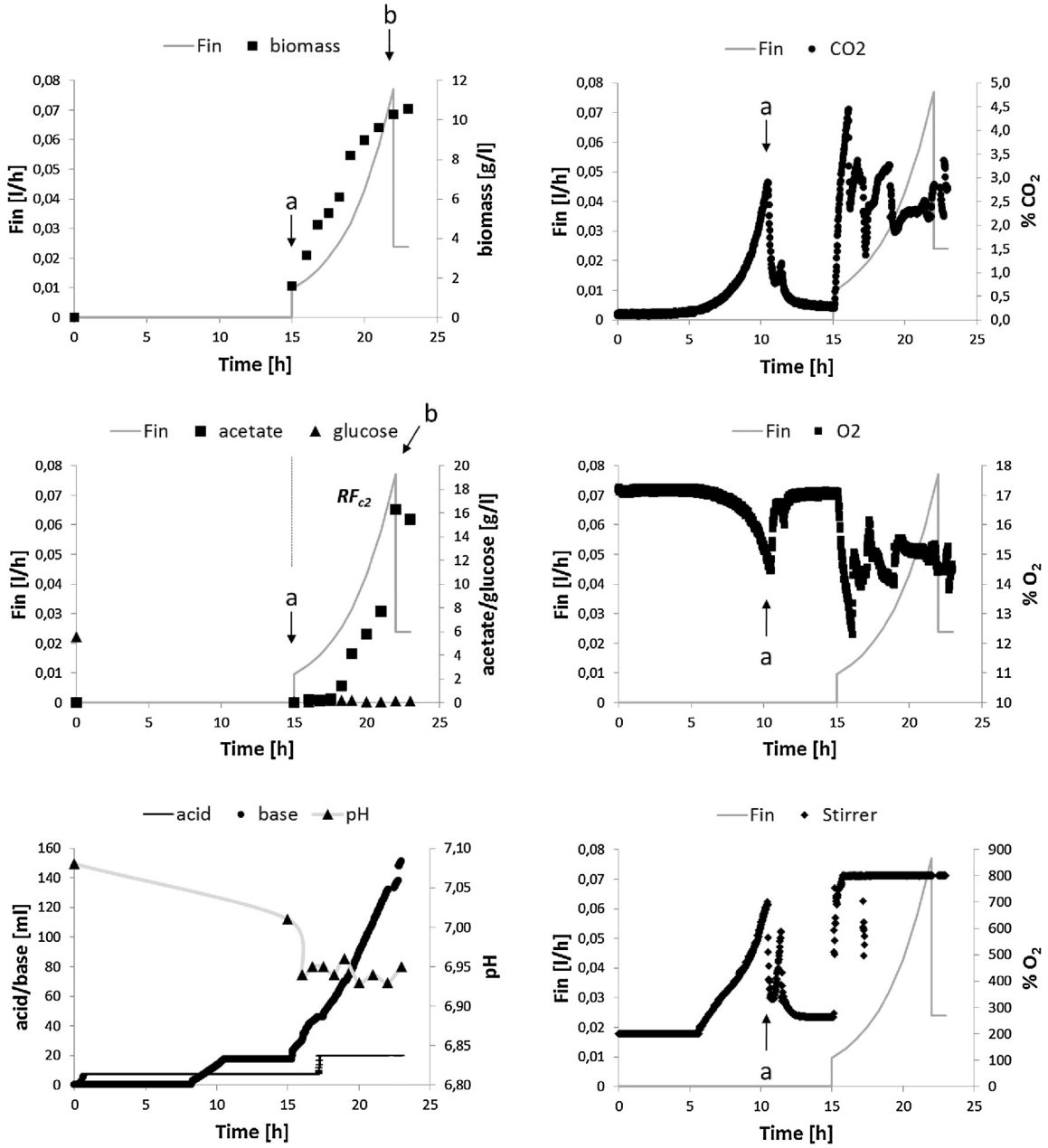

Fig. 4. Experiment 4. Time profiles of feed-rate, biomass, acetate, glucose, added acid and base, pH, volumetric fractions of CO_2 and O_2 and stirrer speed. RF_{c2}: respiro-fermetative regime from scenario 2.

fermentative regime and associated reactions are (1a) and (1b) so that reaction rates become:

$$q_S > q_{Scrit};\quad q_{AC} = 0 \rightarrow \mu_1 = q_{Scrit} = {q_{Omax}}^* \frac{K_{iA}}{K_{iA} + A};\quad \mu_2 = (q_S - q_{Scrit});\quad \mu_3 = 0 \tag{20}$$

3. Scenario 3: Glucose and acetate are present in the medium and cells consume both of them. Cells are therefore in respirative regime and reactions (1a) and (1c) are concerned. It comes that:

$$q_S + q_A \le q_{Scrit} \rightarrow \mu_1 = q_S = q_{Smax} \frac{S}{K_S + S};\quad \mu_2 = 0;\quad \mu_3 = q_{AC} = \frac{k_{OS}(q_{Scrit} - q_S)}{k_{OA}} \frac{A}{A + K_A} \tag{21}$$

4. Scenario 4: Only acetate is present in the medium and is consumed by the cells. In this particular case, only acetate oxidation is concerned and the related consumption rate reaches its maximal value:

$$q_S = 0;\quad q_A \le q_{Scrit};\quad \mu_1 = \mu_2 = 0;\quad \mu_3 = q_{AC} = \frac{k_{OS}(q_{Scrit} - q_S)}{k_{OA}} \frac{A}{A + K_A} \tag{22}$$

Some other important issues must be taken into account for the design of the experimental conditions: (i) Concerning the identification of the half-saturation constant in (3a), substrate concentration measurements in the range of plausible values of K_S should be used. However, K_S is generally so small (the order of magnitude is generally ranging from 0.01 to 0.1 g/L) that its identifiability is poor, mainly because of the level of accuracy of the available analytical methods, generally located in the same range.

Table 1
Selected samples related to cases 1–2.

Experiments	Selected data from cases 1–2
2	a–b
3	b–e
4	a–b

Table 2
Vertices of the multistart parameter polytope.

Parameters	Lower bound	Upper bound
k_{X1}	0.1	1
k_{X2}	0.1	1
k_{X3}	0.01	1
k_{A2}	0.1	1
q_{Smax}	0.1	1
q_{Omax}	0.1	1
K_S	0.01	1
K_{iA}	1	10
K_A	0.01	1

(ii) When the inhibition is negligible, i.e. the concentration of acetate is much lower than the inhibition constant, expression (3b) is reduced to a classical Monod law. In this case, the identifiability of K_{iA} is, for the same reason as in (i), very poor. Since the values of K_{iA} range from 5 to 15 g/L, a large acetate concentration is required in order to be able to detect the inhibition.

Taking all the above-mentioned points into consideration, the experimental strategy will consist in fed-batch cultures using specific feed rate profiles triggering the metabolic switch.

For instance, an exponential feed rate F_{in} is imposed in the first experiment (see Fig. 1) and calculated as follows:

$$F_{in} = \frac{\mu_{set}}{k_{XS}} \frac{V_0 X_0 e^{\mu_{set}(t-t_0)}}{S_{in}} \tag{23}$$

where μ_{set} is the foreboded specific growth rate, k_{XS} is the yield coefficient defined as grams of produced biomass per grams of consumed substrate, V_0 and X_0 are the initial volume and biomass concentrations, respectively, t_0 is the initial feeding time and S_{in} is the substrate concentration in the inflow, i.e. 500 g/L of glucose. Expression (23) assumes that the fed-batch culture can be operated under glucose limitation, i.e. the remaining glucose concentration in the bioreactor is below the critical level. The tested μ_{set} values are in the range of 0.15–0.5 h^{-1}.

4.4. Experiment results

As above-mentioned, 4 fed-batch cultures are performed. In experiment 1 (Fig. 1), the bioreactor is inoculated with 150 mL of seed culture, the initial OD is 0.3, the glucose concentration measured at the beginning of the batch phase is 5.7 g/L, and the Initial culture volume is 3.15 L. During the batch phase, biomass reaches a concentration of 1.42 g/L (OD = 5.55). The on-line flags for the ending of batch phase, resulting from glucose depletion, are the sudden increase of pH from its set-point, requiring acid addition, the decrease of CO_{2out} (implying that O_{2out} increases) and the decrease of RPM since cells demand less oxygenation (red dashed circles). The fed-batch phase starts at 6.5 h of culture time. An exponential feed flow rate with $\mu_{set} = 0.2\ h^{-1}$ is applied in the time period 6.5–14 h (Fig. 1 between arrows *a* and *b*). During this period, CO_2 and RPM increase due to the important glucose oxidation. The pH decreases and base is added. The biomass follows an exponential growth and the glucose concentration remains between 0.04 and 0.15 g/l, so that no significant acetate production is observed. Then, the feeding is turned-off for one hour (Fig. 1b and c) before resuming it according to the initial exponential trajectory (Fig. 1c and d). After this exponential feeding, the feed rate is manipulated in a stepwise manner. The cells are evolving either in respirative regime or respiro-fermentative regime, close to the metabolic switch.

Table 3
Identified parameters for the first two scenarios.

Parameter	Value	Units
k_{X1}	0.1849 ± 0.0096	g/g
k_{X2}	0.2899 ± 0.0061	g/g
k_{A2}	0.4324 ± 0.0172	g/g
q_{Smax}	3.2818 ± 0.036	h^{-1}
q_{Omax}	1.4032 ± 0.0342	h^{-1}
K_S	0.0502 ± 0.0093	g/l
K_{iA}	2.0410 ± 0.2393	g/l

In experiment 2, the optical density and glucose concentration at the beginning of the batch phase are 0.2 and 5.2 g/l, respectively. The batch phase ends around 5 h when glucose is almost completely depleted (the remaining glucose concentration is lower than 0.1 g/l, see Fig. 2a). During the fed-batch phase, three exponential feed flow rate trajectories are applied, interrupted by intermediate step changes. During this experiment, cells switch several times between the metabolic regimes. In the culture period a–b (Fig. 2), an exponential feeding with $\mu_{set} = 0.30\ h^{-1}$ is applied, driving the cells into the respiro-fermentative metabolic pathway. This dataset therefore corresponds to cases 1 and 2. In *b*, the feed rate is suddenly set to zero and, as expected, acetate is rapidly re-consumed. A new exponential trajectory with $\mu_{set} = 0.15\ h^{-1}$ starts one hour later. Cells are following the respirative pathway, first facing case 4 (when the feed-rate is zero and glucose depleted) and 3 until the culture ends.

In experiment 3, the exponential feeding trajectory is divided in two parts with $\mu_{set} = 0.4$ and $0.5\ h^{-1}$. The two periods are separated by a zero-feed interval of two hours. Acetate is produced during the whole culture, even when the feed-rate drops to zero, reaching a concentration of 8.7 g/l; see Fig. 3. Cells are in respiro-fermentative regime, corresponding to case 2.

In experiment 4, the fed-batch phase is achieved with two exponential feed rates, first $\mu_{set} = 0.4\ h^{-1}$ and $\mu_{set} = 0.13\ h^{-1}$ towards the end of the culture. Acetate reaches a maximum concentration of 16.3 g/L, severely inhibiting biomass growth during the last hours while glucose is kept at a very low concentration in the culture medium (0.08–0.2 g/L). It is therefore assumed that cells grow following the fermentative regime (Fig. 4).

5. Data exploitation

The *E. coli* model (4) includes nonlinear differential equations for the biomass (X), glucose (S) and acetate (A) concentration, and an algebraic equation for the *OTR* evolution (quasi steady-state assumption on dissolved oxygen). The data will be partitioned according to its relation to a differential or algebraic state, and according to the metabolic switch, i.e. cases 1–2 or cases 3–4.

The parameter sets to be identified using nonlinear least squares are therefore:

$$\theta_{NL12} = [k_{X1} k_{X2} k_{A2} q_{Smax} q_{Omax} K_S K_{iA}] \tag{24}$$

and

$$\theta_{NL34} = [k_{X3} K_A (k_{OS}/k_{OA})] \tag{25}$$

while the parameter sets handled by the linear least squares are:

$$\theta_{L12} = [k_{O1} k_{O2}] \tag{26}$$

and

$$\theta_{L34} = [k_{O3}] \tag{27}$$

The details of the data partitioning are summarized in Table 1 for cases 1–2. Experiments 2 to 4 are selected for identification and

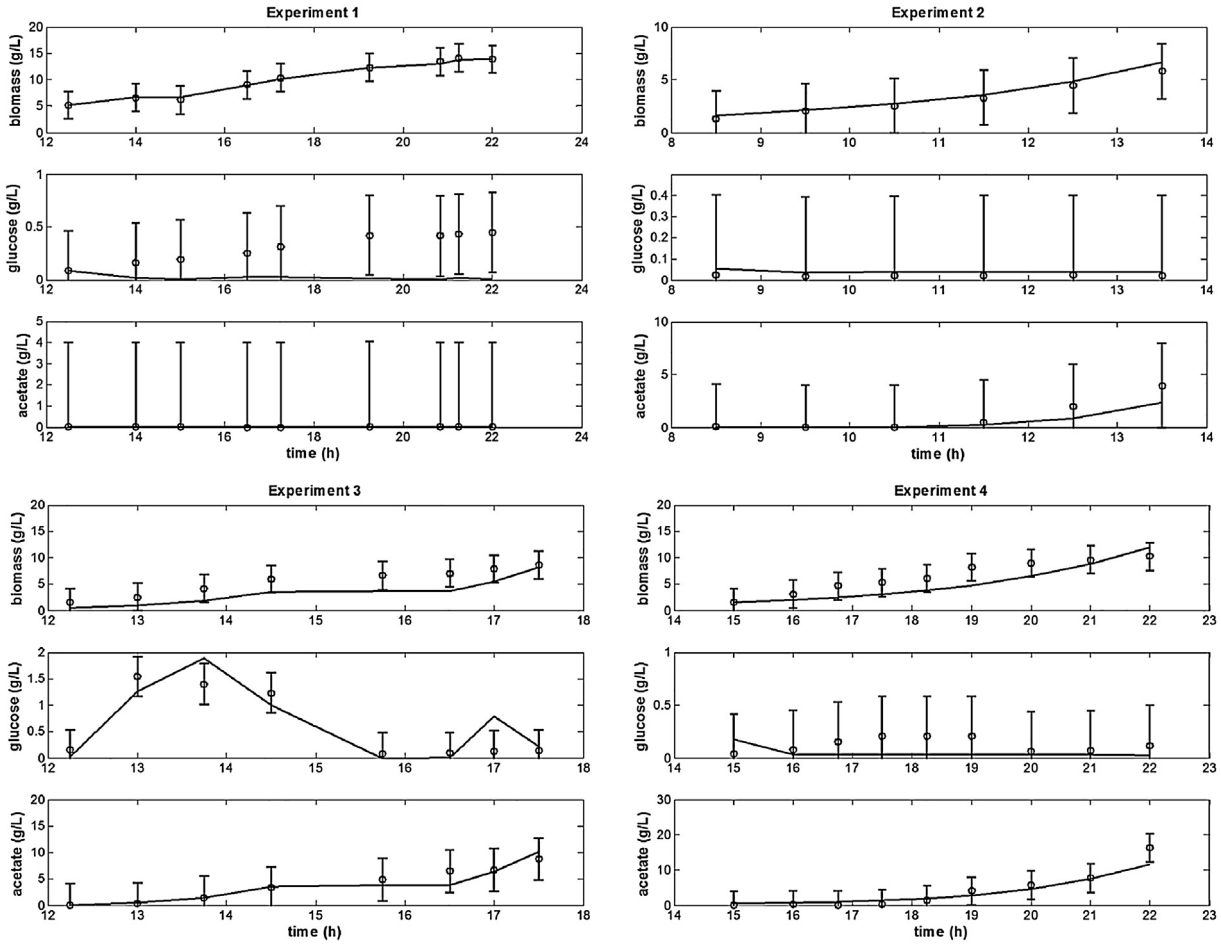

Fig. 5. Direct and cross validations applying nonlinear weighted least squares method to scenarios 1 and 2. -: simulated values, o: discrete measured values with 95% confidence intervals.

direct validation while Experiment 1 is dedicated to model cross-validation.

The ratio k_{OS}/k_{OA} can be considered as a single parameter close to 1 ([5] and [8]), which is also assumed here.

5.1. Parameter identifiability

In order to support data partitioning, parameter identifiability is assessed based on the FIM (17).

From the data related to cases 1–2, the following FIM is obtained:

$$FIM_{12} = 10^4 \begin{pmatrix} 1.0369 & 1.6993 & -0.0170 & 0.3649 & -0.0221 & -1.2345 & -0.0053 \\ 1.6993 & 3.4503 & 0.0786 & 0.7076 & -0.1388 & -2.5008 & -0.0214 \\ -0.0170 & 0.0786 & 0.1280 & 0.0211 & -0.0395 & -0.0785 & -0.0102 \\ 0.3649 & 0.7076 & 0.0211 & 0.1614 & -0.0311 & -0.5500 & -0.0044 \\ -0.0221 & -0.1388 & -0.0395 & -0.0311 & 0.0310 & 0.1083 & 0.0044 \\ -1.2345 & -2.5008 & -0.0785 & -0.5500 & 0.1083 & 1.9799 & 0.0159 \\ -0.0053 & -0.0214 & -0.0102 & -0.0044 & 0.0044 & 0.0159 & 0.0010 \end{pmatrix} \quad (28)$$

Since (28) is full-rank with a condition number of $6.0802 \cdot 10^4$, parameters can be considered as practically identifiable.

Regarding the data related to cases 3–4, the following FIM is obtained:

$$FIM_{34} = 10^8 \begin{pmatrix} 3.6490 & 0.1116 \\ 0.1116 & 0.0038 \end{pmatrix} \quad (29)$$

which is also full-rank with a condition number of $9.0119 \cdot 10^3$, confirming practical identifiability.

6. Sequential parameter identification

As parameter identification is always a difficult problem, a divide and conquer approach, where the initial problem is cut into smaller subproblems, is advisable. In this approach, the results of one identification subproblem feed the next one. Each subproblem is therefore dedicated to a small parameter subset, the other parameters being fixed from previous identification steps/results. There is a risk of parameter bias in such approach (as a parameter subset is fixed, it influence the search of the free parameter space). To alleviate this, it is recommended to complete the sweep through the several subproblems by a global nonlinear identification of all the parameters (starting from initial conditions which should not be too far away, even if some bias exists), and to use a multistart strategy for the full procedure. The sequential procedure is inspired from our successful results in another study [27], with the major difference that identification subproblems were guided by parameter sensitivities, whereas in the present study, partition is suggested by differential/agebraic states and metabolic switch.

The sequence can be outlined as follows:

1. nonlinear sub-model identification of the assumed respiro- fermentative scenarios 1 and 2 and corresponding biomass, glucose and acetate data partition.

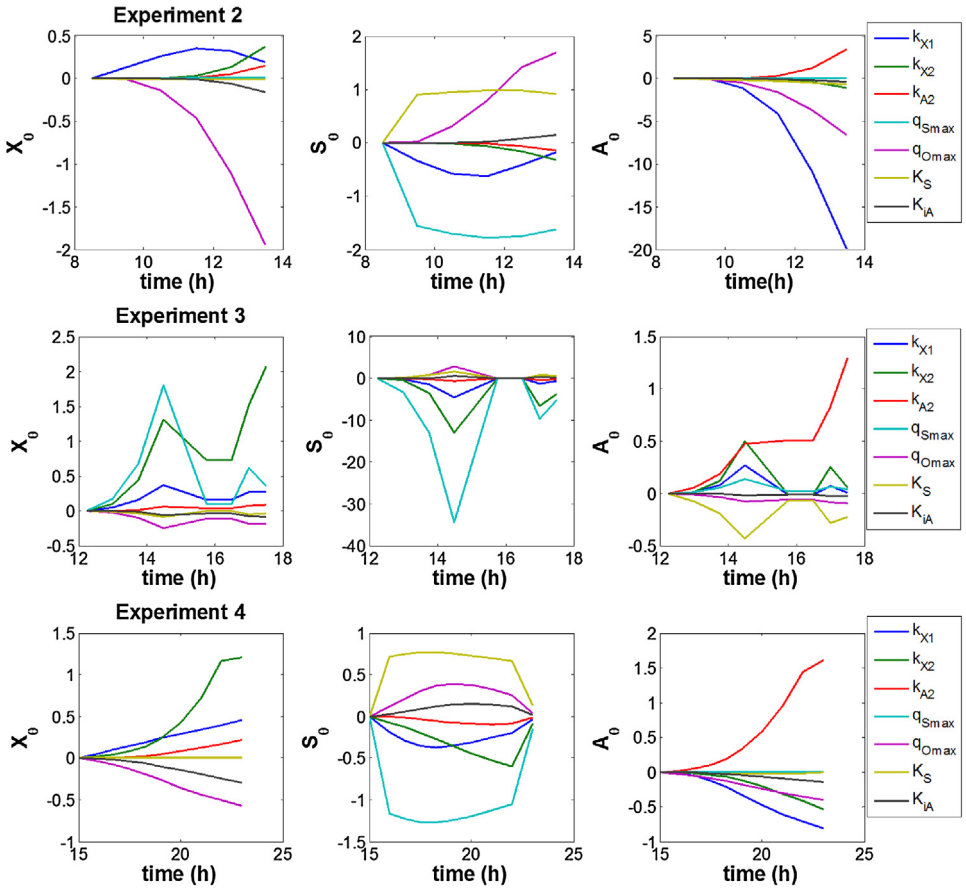

Fig. 6. Parametric sensitivities related to scenarios 1 and 2.

2. nonlinear sub-model identification of the assumed respirative scenarios 3 and 4 and corresponding biomass, glucose and acetate data partition. Identified parameters from the first step are assumed to be known and not re-identified.
3. linear sub-model identification of the assumed respiro- fermentative scenarios 1 and 2 and corresponding *OTR* and *CTR* data partition. Identified parameters from the first two steps are assumed to be known and not re-identified. Biomass, glucose and acetate concentrations are predicted by the first nonlinear sub-model.
4. linear sub-model identification of the assumed respirative scenarios 3 and 4 and corresponding *OTR* and *CTR* data partition. Identified parameters from the first three steps are assumed to be known and not re-identified. Biomass, glucose and acetate concentrations are predicted by the first two nonlinear sub-models.
5. full nonlinear model identification (considering both metabolic pathways) using the previously identified parameter sets as initial guesses and corresponding biomass, glucose and acetate data partitions.
6. full linear model identification (considering both metabolic pathways) using parameters identified during the fifth step and corresponding *OTR* and *CTR* data partitions. Biomass, glucose and acetate concentrations are predicted by the corresponding full nonlinear model.

6.1. Nonlinear sub-models

Parameter identification related to *X*, *S* and *A* data is achieved through nonlinear weighted least squares calling the MATLAB optimizer "*fmincon*" and the ODE solver "*ode*15*s*" several times (typically three times, new parameter initial guesses being the result of the previous optimizer call) using a covariance matrix Σ playing the role of a diagonal scaling matrix with maximum output values as in (13). *Fmincon* is a Successive Quadratic Programming (SQP) optimizer using a gradient-based method that is designed to work on constrained problems.

A multistart procedure is used to explore the parameter space and possibly detect the global minimum. Initial guesses for the parameters to be identified belong to a polytope whose vertices correspond to lower and upper bounds (which are used in "*fmincon*" under the form of optimization constraints), and are listed in Table 2. The number of multistart iterations is set to 30. In order to select the best result, three criteria, presented in descending order of importance, are suggested: fitting between model prediction and experimental data during direct and cross validations (qualitative assessment), the cost function residual (estimation precision) and estimated parameter discrepancies quantified by confidence intervals (estimation accuracy).

6.1.1. First sub-model

The resulting first sub-model validation (best result from the 30 runs related to scenarios 1 and 2) is shown in Fig. 5 and the corresponding parameter values are summarized in Table 3.

As expected from (28), confidence intervals are quite tight. The most important uncertainty concerns K_{iA} (12% of the estimated value). Regarding direct validation, the model fits well with the experiment (2, 3 and 4) data and cross-validation (experiment 1) also shows satisfactory results even if glucose seems slightly underestimated. Relatively large error bounds on glucose and acetate measurements are observed when the state values are almost zero, indicating the level of precision (and information on the concerned states) that can be expected from the model for such experiments

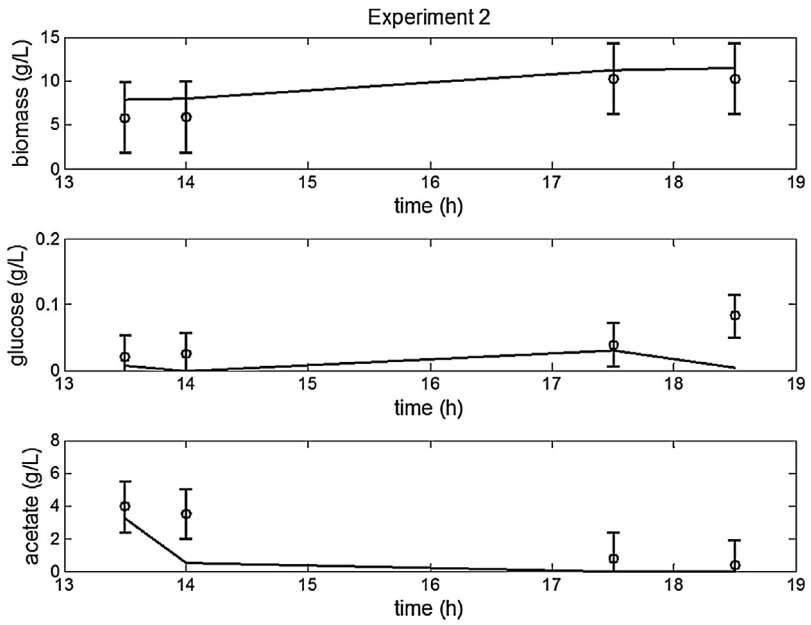

Fig. 7. Direct and cross validations applying nonlinear weighted least squares method to scenarios 3 and 4. -: simulated values, o: discrete measured values with 95% confidence intervals.

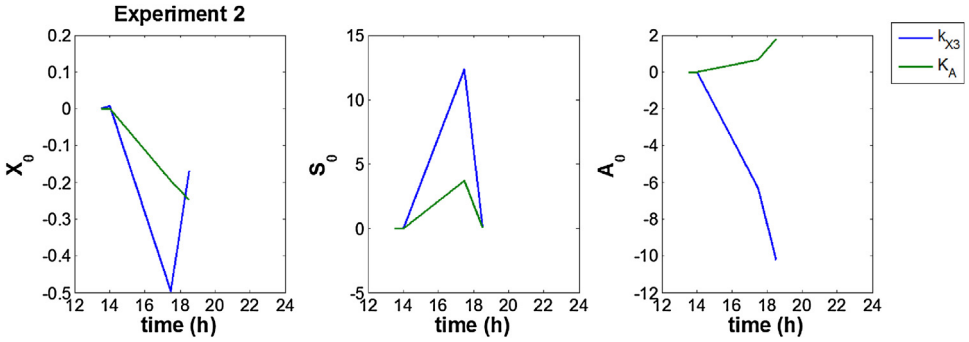

Fig. 8. Parametric sensitivities for the identification related to scenarios 3 and 4.

Table 4
Identified parameter set related to scenarios 3 and 4.

Parameter	Values	Units
k_{X3}	0.0412 ± 0.0001	g/g
K_A	0.3928 ± 0.002	g/l

Table 5
Parameter associated to the gas dynamics.

Parameter	Value	Units
k_{O1}	1.2622 ± 0.1794	g/g
k_{O2}	0.2519 ± 0.0704	g/g
k_{O3}	0.5658 ± 0.1342	g/g
k_{C1}	0.9436 ± 0.1794	g/g
k_{C2}	0.1223 ± 0.0704	g/g
k_{C3}	0.4218 ± 0.1342	g/g

reaching the sensitivity level of the analytical methods (located around 0.1 g/L). Accurate estimation becomes impossible and the concentrations should be considered as zero.

Fig. 6 shows the time evolution of the sensitivity functions of biomass, glucose and acetate with respect to parameters during experiments used for direct validation, i.e. experiments 2, 3 and 4.

From the sensitivities related to the three experiments, k_{X1} seems to influence significantly the three states while, in experiment 2, the influence of q_{Omax} and q_{Smax} is quite important on glucose concentration and to less extent on biomass and acetate concentrations. Acetate is influenced mainly by k_{X1} and k_{A2}. The remaining parameters (i.e. k_{X2}, K_S and K_{iA}) have moderate influence. Sensitivities from experiment 4 are comparable to experiment 2. However, concerning experiment 3, biomass concentration is influenced mostly by q_{Smax}, q_{Omax} and k_{X2}. Glucose is largely influenced by q_{Smax}, followed by k_{X2}. The major influence on acetate concentration variations is provided by k_{A2}. On the other hand, parameters associated to the glucose uptake rate are influencing acetate concentration, as well. Then, almost all the parameters have some degree of influence on biomass, glucose and acetate. Hence, this experiment is by far the most informative for parameter identification. Considering the three experiments as a whole, K_{iA} has a much more moderate influence, significantly on biomass concentration, influenced by acetate inhibition. This could explain the higher uncertainty level compared to other parameters.

6.1.2. Second sub-model

The previously identified parameters from scenarios 1 and 2 are now assumed to be known and set to the identified values of Table 3. The resulting second sub-model validation (best result from the 30 runs related to scenarios 3 and 4) is shown in Fig. 7 and the corresponding parameter values are shown in Table 4.

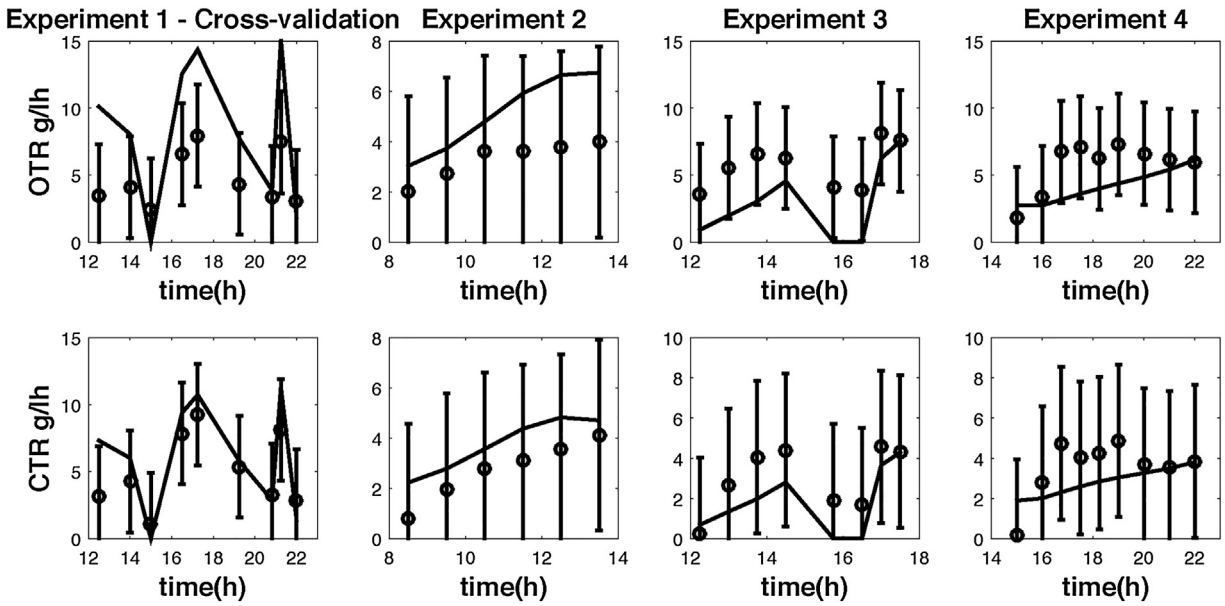

Fig. 9. Model validation for parameter associated to the gas dynamics for scenarios 1 an 2: *OTR* and *CTR*. Experiment 4: cross validation. -: simulated values, o: discrete measured values.

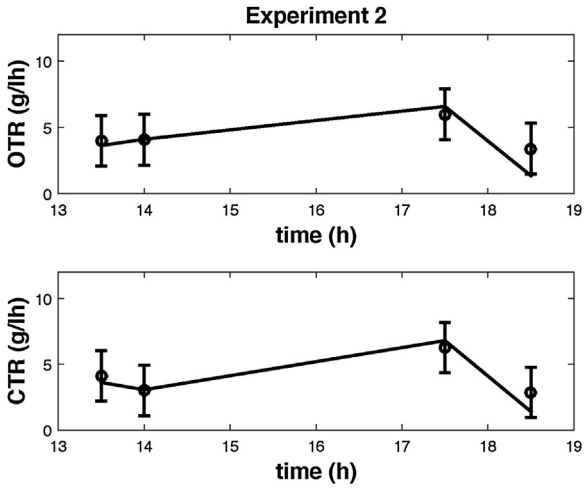

Fig. 10. Model validation for parameter associated to the gas dynamics for scenarios 3 an 4: *OTR* and *CTR*. -: simulated values, o: discrete measured values.

Confidence intervals are still quite satisfactory and parametric sensitivities are shown in Fig. 8. As observed, the model fits moderately well the measurements. Regarding sensitivities, parameters k_{X3} and K_A have a large influence on glucose and acetate concentrations. Conversely, parameter influence on biomass is almost negligible.

6.2. Linear sub-model

The next step consists in identifying the parameters associated to the gas dynamics using the linear least squares method. As shown in (7) oxygen is assumed to be in steady state so that the oxygen transfer rate (*OTR*) is almost equal to the oxygen uptake, leading to a linear relation $OTR = (k_{O1}\mu_1 X + k_{O2}\mu_2 X + k_{O3}\mu_3 X)X$ where the explicative variables μ_i ($i = 1, 2, 3$), are now supposed to be perfectly known, that is, considering parameter values from the nonlinear sub-models and the corresponding biomass, glucose and acetate predictions.

Table 6
Full model.

Parameters	Values	Units
k_{X1}	0.1849 ± 0.0096	g/g
k_{X2}	0.2899 ± 0.0061	g/g
k_{X3}	0.0412 ± 0.0001	g/g
k_{A2}	0.4324 ± 0.0172	g/g
q_{Smax}	3.2818 ± 0.036	h^{-1}
q_{Omax}	1.4032 ± 0.0342	h^{-1}
K_S	0.0502 ± 0.0093	g/l
K_{iA}	2.0410 ± 0.2393	g/l
K_A	0.3928 ± 0.002	g/l
k_{OS}/k_{OA}	1	
k_{O1}	0.7374 ± 0.0698	g/g
k_{O2}	0.3194 ± 0.0667	g/g
k_{O3}	1.3418 ± 0.4614	g/g
k_{C1}	0.7608 ± 0.0698	g/g
k_{C2}	0.1055 ± 0.0667	g/g
k_{C3}	0.8461 ± 0.4614	g/g

Rearranging (7) to get a linear expression, we obtain:

$$\varphi^T = [\mu_1 X \ \mu_2 X \ \mu_3 X] \quad \theta = [k_{O1} \ k_{O2} \ k_{O3}] \tag{30}$$

As explained in [8], the condition of dissolved carbon dioxide in the medium is more complex since carbon dioxide is by far more soluble than oxygen and, rigorously, $CTR \approx CER$ (where *CER* is the carbon transfer evolution rate) is not true. However, since no dissolved carbon dioxide measurement from the experimental plant is currently available, this rough assumption is made.

Applying the quasi-steady state assumption and neglecting dilution, k_{C1}, k_{C2} and k_{C3} are also estimated as in (30).

The identified parameters associated to the gas dynamics are shown in Table 5. Confidence intervals are relatively small even for k_{O3} and k_{C3} where the data set is though limited to 12 samples.

The fitting between model prediction and experimental data of *OTR* and *CTR*, shown in Figs. 9 and 10 in direct and cross validations, are still quite satisfactory.

Interestingly, although the determination of *CTR* is based on a steady-state assumption which is a priori not always valid, model validation shows that the model correctly fits the measurements.

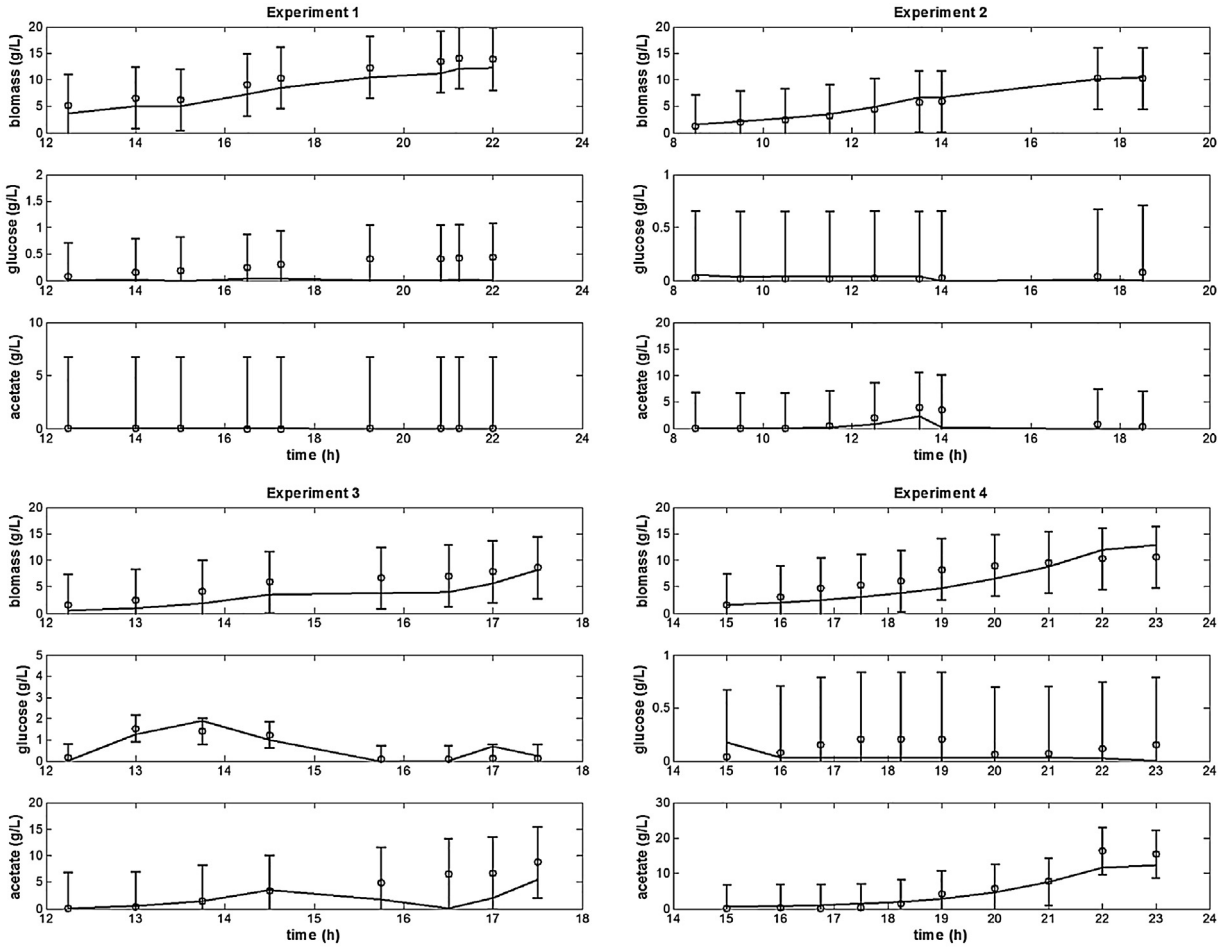

Fig. 11. Direct validation of the nonlinear full-model using parameter values from Tables 3 and 4. -: simulated values, o: discrete measured values.

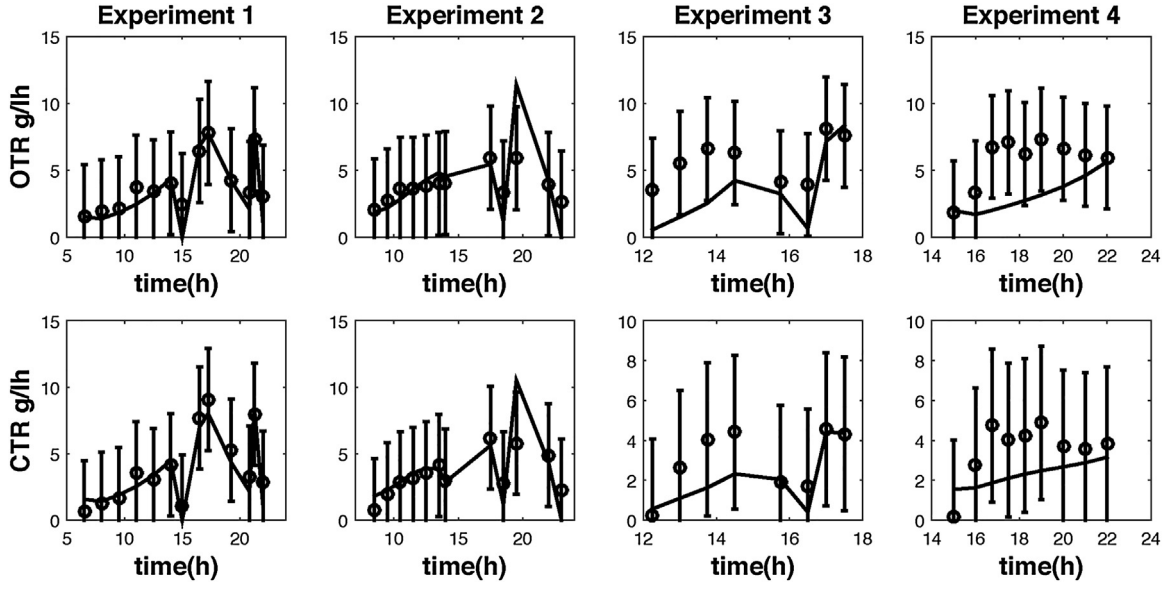

Fig. 12. Direct validation of the linear full-model using parameter values from Tables 3 and 4. -: simulated values, o: discrete measured values.

6.3. *The full model*

The full model is now validated using the four experiments and previous sequentially identified parameter values are used as new initial guesses for the nonlinear least square method. As a matter of fact, taking into account model fitting with the data, the solver does not lead to any improvement, suggesting that the sequential identification method leads to the best results shown in Fig. 11.

The model predicts satisfactorily data general tendencies (see Fig. 11) even if glucose concentration predictions still seem

less accurate when reaching low levels. However, biomass and acetate concentrations, which are important indicators of *E. coli* metabolism, are, in overall, quite well predicted. Parameter values are shown in Table 6.

Conversely to the nonlinear part, the full linear model fits better the experimental data as shown in Fig. 12. However, parametric uncertainty is unfortunately not improved as shown in Table 6.

The global lack of improvement supports the idea that the proposed sequential method performs well and allows to approach quickly very good parameter estimates while decreasing model complexity (i.e. the number of parameters and therefore the search space dimension) and, consequently, computational effort. By way of example, one sub-model identification run using three times the "fmincon" solver is achieved within 30 seconds to 2 min by a $core - i7 - 6560U$ computer 2.20–2.21 GHz with 8 Go RAM, while one equivalent full model identification run is achieved within 10 minutes to half an hour using the same computational device.

7. Conclusion

In this study, parameter identification of a BL21(DE3) *E. coli* strain is carried out using 4 fed-batch experiments. The experimental strategy aims at designing informative cultures in which cells follow two possible metabolic pathways. The suggested mechanistic model structure indeed assumes the existence of overflow metabolism leading to either respirative or respiro-fermentative regime and accurate feed flow rate trajectories are carefully set to meet both of them.

Achieving a specific partitioning of the data, parts of the stoichiometry and kinetics are successively identified applying sequentially nonlinear and linear least squares methods. A multistart procedure is also applied in order to explore the parameter space and possibly detect the global minimum. The selection of the best set is carried out taking into account, in descending order of importance, the fitting during direct and cross-validations, the cost function values and the uncertainty on the identified parameters. Parametric sensitivities are used to check the informative quality of the experiments as well as to characterize parametric uncertainty.

Validation of the full model using all the available data from the 4 experiments is quite satisfactory: even if the model shows moderate estimation accuracy for very low glucose concentrations, model predictions reproduce the right tendencies of experimental data.

Acknowledgements

The authors gratefully acknowledge the support of the Walloon Region in the framework of the research project HYPRO2COM. 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.

References

[1] H. Crabtree, Observations on the carbohydrate metabolism of tumors, Biochem. J. 23 (1929) 536–545.
[2] J.-N. Phue, S. Noronha, R. Hattacharyya, A. Wolfe, J. Shiloach, Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and northern blot analyses, Biotechnol. Bioeng. 90 (7) (2005) 805–820.
[3] G. Vemuri, E. Altman, D. Sangurdekar, A. Khodursky, M. Eiteman, Overflow metabolism in *Escherichia coli* during steady-state growth: transcriptional regulation and effect of the redox ratio, Appl. Environ. Microbiol. 72 (5) (2006) 3653–3661.
[4] G. Luli, W. Strohl, Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations, Appl. Environ. Microbiol. 56 (4) (1990) 1011–1104.
[5] B. Xu, M. Jahic, G. Blomsten, S. Enfors, Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*, Appl. Microbiol. Biotechnol. 51 (1999) 564–571.
[6] M. Alba, E. Calvo, Characterization of bioreaction processes: aerobic *Escherichia coli* cultures, J. Biotechnol. 84 (2000) 107–118.
[7] M. Eiteman, E. Altman, Overcoming acetate in *Escherichia coli* recombinant protein fermentations, Trends Biotechnol. 24 (11) (2006) 530–536.
[8] I. Rocha, Model-Based Strategies for Computer-aided Operation of a Recombinant *E. coli* Fermentation, Ph. D. Thesis, Universidade do Minho, 2003.
[9] Y. Usuda, Y. Nishio, S. Iwatani, S. Van Dien, A. Imaizumi, K. Shimbo, N. Kageyama, D. Iwahata, H. Miyano, K. Matsui, Dynamic modeling of *Escherichia coli* metabolic and regulatory systems for amino-acid production, J. Biotechnol. 147 (2010) 17–30.
[10] K. Valgepea, K. Adamberg, R. Nahku, P. Lahtvee, L. Arike, R. Vilu, Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase, BMC Syst. Biol. 4 (166) (2010) 13.
[11] B. Sonnleitner, O. Kappeli, Growth of *S. cerevisiae* is controlled by its limited respiratory capacity: formulation and verification of a hypothesis, Biotechnol. Bioeng. 28 (1986) 927–937.
[12] V. Galvanauskas, R. Simutis, N. Volk, A. Lubbert, Model based design of a biochemical cultivation process, Bioprocess Eng. 18 (1998) 227–234.
[13] A. Cockshott, I. Bogle, Modelling the effects of glucose feeding on a recombinant, Bioprocess Eng. 20 (1998) 83–90.
[14] M. Akesson, E. Karlsson, P. Hagander, J. Axelsson, A. Tocaj, Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* w3110, Biotechnol. Bioeng. 64 (5) (1999) 590–597.
[15] L. Dewasme, G. Goffaux, A.-L. Hantson, A. Vande Wouwer, Experimental validation of an extended Kalman filter estimating acetate concentration in *E. coli* cultures, J. Process Control 23 (2013) 148–157.
[16] M. Cruz-Bournazou, T. Barz, D. Nickel, F. Glauche, A. Knepper, P. Neubauer, Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities, Biotechnol. Bioeng. 114 (2017) 610–619.
[17] E. Anane, D. López Cárdenas, P. Neubauer, M. Cruz-Bournazou, Modelling overflow metabolism in *Escherichia coli* by acetate cycling, Biochem. Eng. J. 125 (2017) 23–30.
[18] P. Neubauer, L. Haggstrom, S. Enfors, Influence of substrate oscillations on acetate formation and growth yield in *Escherichia coli* glucose limited fed-batch cultivations, Biotechnol. Bioeng. 47 (1995) 139–146.
[19] T. Paalme, R. Elken, A. Kahru, K. Vanatalu, R. Vilu, The growth rate control in *Escherichia coli* at near to maximum growth rates: the a-stat approach, Antonie Van Leeuwenhoek 71 (1997) 217–230.
[20] A. Lara, L. Caspeta, G. Gosset, F. Bolivar, O. Ramirez, Utility of an *Escherichia coli* strain engineered in the substrate uptake system for improved culture performance at high glucose and cell concentrations: an alternative to fed-batch cultures, Biotechnol. Bioeng. 99 (2008) 893–901.
[21] A. Lara, H. Taymaz-Nikerel, M. Mashego, W. van Gulik, J. Heijnen, O. Ramirez, W. van Winden, Fast dynamic response of the fermentative metabolism of *Escherichia coli* to aerobic and anaerobic glucose pulses, Biotechnol. Bioeng. 104 (2009) 1153–1161.
[22] M. Basan, S. Hui, H. Okano, Z. Zhang, Y. Shen, J. Williamson, T. Hwa, Overflow metabolism in *Escherichia coli* results from efficient proteome allocation, Nature 528 (2015) 99–104.
[23] L. Dewasme, B. Srinivasan, M. Perrier, A. Vande Wouwer, Extremum-seeking algorithm design for fed-batch cultures of microorganisms with overflow metabolism, J. Process Control 21 (2011) 1092–1104.
[24] L. Santos, L. Dewasme, D. Coutinho, A. Vande Wouwer, Nonlinear model predictive control of fed-batch cultures of micro-organisms exhibiting overflow metabolism: assessment and robustness, Comput. Chem. Eng. 39 (2012) 143–151.
[25] S. Tebbani, D. Dumur, G. Hafidi, A. Vande Wouwer, Nonlinear predictive control of fed-batch cultures of *Escherichia coli*, Chem. Eng. Technol. 33 (7) (2010) 1112–1124.
[26] L. Dewasme, S. Fernandes, Z. Amribt, L. Santos, P. Bogaerts, A. Vande Wouwer, State estimation and predictive control of fed-batch cultures of hybridoma cells, J. Process Control 30 (2015) 50–57.
[27] I. De Jesus Saraiva, A. Vande Wouwer, A.-L. Hantson, Parameter identification of a dynamic model of CHO cell cultures, Bioprocess Biosyst. Eng. 38 (11) (2015) 2231–2248.
[28] Y. Wang, S.-L. Wu, W. Hancock, R. Trala, M. Kessler, A. Taylor, P. Patel, J. Aon, Proteomic profiling of *Escherichia coli* proteins under high cell density fed-batch cultivation with overexpression of phosphogluconolactonase, Biotechnol. Prog. 21 (2005) 1401–1411.
[29] A. Matin, The molecular basis of carbon-starvation-induced general resistance in *Escherichia coli*, Mol. Microbiol. 5 (1991) 3–10.
[30] D. Siegele, R. Kolter, Life after log, J. Bacteriol. 174 (1992) 345–348.
[31] E. Walter, L. Pronzato, Identification of Parametric Models From Experimental Data, Springer, COUNTRY, 1997.