



# Biomass estimation of an industrial raceway photobioreactor using an extended Kalman filter and a dynamic model for microalgae production

F. García-Mañas<sup>a</sup>, J.L. Guzmán<sup>a,\*</sup>, M. Berenguel<sup>a</sup>, F.G. Ación<sup>b</sup>

<sup>a</sup> Department of Informatics, University of Almería, CIESOL, ceiA3, E04120 Almería, Spain

<sup>b</sup> Department of Chemical Engineering, University of Almería, CIESOL, ceiA3, E04120 Almería, Spain

## ARTICLE INFO

### Keywords:

Extended Kalman filter  
Microalgae  
Biomass concentration  
On-line estimation

## ABSTRACT

Production of microalgae is one of the emerging biotechnological processes due to its potential applications to produce high value-added compounds. In photobioreactors for microalgae production, the biomass concentration is a desirable variable to be measured on-line to optimize the yield of the systems. However, biomass concentration can hardly be monitored in real time. There are few expensive commercial sensors that in fact provide uncertain measurements. State estimators, also known as software sensors, are algorithms designed to estimate unmeasured (or non-easily measurable) variables of a process. In this work, a state estimator using the extended Kalman filter algorithm is developed to estimate biomass concentration for an outdoor industrial raceway photobioreactor. The state estimator is based on a dynamic model for microalgae production specifically designed for this type of photobioreactor. Results demonstrate that, despite the complex non-linear dynamics that characterise this kind of bioprocess, a state estimator can provide a relatively accurate estimation of the biomass concentration. Furthermore, a state estimator could be used to optimize the operation of industrial photobioreactors by utilizing the estimated biomass concentration for automatic control of the process.

## 1. Introduction

Biotechnological processes require optimization and advanced control techniques to guarantee continuous and reliable production. Bioprocess control, as performed in photobioreactors, is fundamentally based on dissolved oxygen, pH and temperature measurements which can be obtained by on-line monitoring of the culture. For improving the quality of the automatic control of production step, it is necessary to have real time measurements of biomass concentration. To that end, complex and expensive instruments are applied [1] (although recently cheaper sensors are being developed [2]). Moreover, most commercial sensors have limitations to provide reliable measurements of this and other variables of bioprocesses [3].

Production of microalgae is one of the most relevant emerging biotechnological processes because of its potential applications to produce high value-added compounds, such as antioxidants, polyunsaturated fatty acids, cosmetics, pharmaceuticals, etc. [4]. In microalgae photobioreactors the biomass concentration is a desirable variable to be monitored, despite it is very problematic to be on-line measured in real time with sensors [5], as for example turbidimeters. Low sensibility at high biomass concentrations, so high turbidity, is a major challenge. Biofouling (defined as adhesion of microalgae,

bacteria and organic compounds to sensors surface and photobioreactor walls), is also a relevant existing issue with *in situ* biomass concentration measurement, because it causes a distortion of turbidity measurements [6,7]. Conventional methods of monitoring biomass concentration based on light absorption are not confinable by the changes on pigments content of biomass according to changes in the culture conditions, thus volumetric determinations and gravimetric measurements being the most extended method. However, dry weight measurements are tedious and can introduce errors if don't performed adequately, for that large volumes of culture being required which are not always available in small systems, additionally disturbing biological conditions of microalgae [8].

Because of the described limitations, state estimators, also named software sensors, are of special interest to be applied in microalgae based processes. State estimators (based on statistical algorithms and mathematical models of the process) can predict or estimate unmeasured (or non-easily measurable) variables using other measured variables [9]. There is a wide diversity of techniques for the development of state estimators. The most employed method in literature is the Kalman filter [10]. Basic structures of Kalman filtering are commonly used for dynamical processes that can be described by linear mathematical models. However, the majority of bioprocesses are

\* Corresponding author at: Department of Informatics, University of Almería, Carretera Sacramento s/n, E04120 Almería, Spain.

E-mail addresses: [fgm754@inlumine.ual.es](mailto:fgm754@inlumine.ual.es) (F. García-Mañas), [jose Luis.guzman@ual.es](mailto:jose Luis.guzman@ual.es) (J.L. Guzmán), [beren@ual.es](mailto:beren@ual.es) (M. Berenguel), [facien@ual.es](mailto:facien@ual.es) (F.G. Ación).

characterized by complex nonlinear dynamics, thus the extended Kalman filter (EKF) and the unscented Kalman filter (UKF) are encouraged choices [11]. Some of these model-based estimators have been developed to track culture states in lab-scale microalgal photobioreactors achieving promising results [12–14]. Nevertheless, it should be emphasized that mathematical models for bioprocesses often presents a certain degree of uncertainty which can result in a difficulty for the operation of state estimators [15]. For this reason, parameter estimation methods are also applied [16,17]. Currently, hybrid state estimators are being developed to combine diverse techniques and to improve convergence velocity while minimizing estimation errors. Artificial intelligence is also employed, such as fuzzy logic and neuronal networks [18].

In this work, a state estimator based on EKF algorithm is designed to estimate the biomass concentration for an outdoor industrial raceway photobioreactor. The state estimator has been developed using a dynamic model for microalgae production previously published in [19]. The model has been readapted and recalibrated with experimental data from one year. Results demonstrate that the state estimator can provide a more accurate estimation of biomass concentration compared to the dynamic model itself. Additionally, the state estimator is presented as a solution to measure the biomass concentration in real time and could be used for automatic control to focus on optimizing production.

## 2. Materials and methods

### 2.1. Description of the raceway photobioreactor

The raceway photobioreactor is located at the Research Centre “Las Palmerillas” (Almería, Spain), property of Cajamar Foundation. The photobioreactor, as showed in Fig. 1a, consists of two channels of 50 m length each and connected by U-shaped bends (180°), with a sump of 0.59 m<sup>3</sup> (0.65 m long × 0.90 m wide × 1 m deep) where air or CO<sub>2</sub> are automatically injected to control the dissolved oxygen and pH of the culture [19–21]. The culture is circulated by a paddle-wheel of 1.2 m diameter. Fig. 1b shows the real experimental facility made of 3 mm thick fiberglass.

For the experiments described in this paper, dissolved oxygen and pH measurements were obtained using sensors (5083T and 5120, Crison, Barcelona, Spain) located in each representative part of the photobioreactor (paddle-wheel, sump and channel) for on-line monitoring of the culture. In Fig. 1a, dissolved oxygen sensors are represented by blue dots, while pH sensors are represented by red dots. The spatial distribution of these variables is estimated using the model in [19]. Those sensors also measured the temperature of the culture while other instruments registered solar radiation, ambient temperature and humidity to know surrounding conditions of the photobioreactor. Considering the growth of microalgae in the photobioreactor as a dynamical process, the inputs to control the optimal culture conditions (air and CO<sub>2</sub>) were measured by mass flow meters (PFM 725S-F01-F, SMC, Tokyo, Japan). The raceway was operated at a constant depth of 0.16 m.

Regarding biomass concentration (slow state variable), previous works of some of the authors ([19] and references therein) analysed that spatial variations are negligible (and thus perfect mixing assumption applies) in the fast time scale describing pH and dissolved oxygen evolution, so that single measurements are enough and can be done anywhere. Biomass concentration was off-line measured by filtering 100 mL samples of the culture at different hours of the day. The samples were heated to 80 °C during 24 h. This significant challenge to measure one of the most important process variables underlines the urgent necessity of evaluating the inclusion of a state estimator to on-line measure biomass concentration.

### 2.2. Microorganism and culture medium

The microalgae strain *Scenedesmus almeriensis* (CCAP 276/24) was used. Inoculum for the raceway reactor was produced in a 3.0 m<sup>3</sup> tubular photobioreactor under controlled conditions: at pH = 8 and at a temperature ranging from 18 to 22 °C using freshwater and Mann & Myers medium prepared using fertilizers (0.14 g·L<sup>-1</sup> K(PO<sub>4</sub>)<sub>2</sub>, 0.18 g·L<sup>-1</sup> Mg(SO<sub>4</sub>)<sub>2</sub>, 0.9 g·L<sup>-1</sup> NaNO<sub>3</sub>, 0.02 mL·L<sup>-1</sup> Welgro, and 0.02 g·L<sup>-1</sup> Kalentol). The raceway reactor was inoculated and operated in batch mode for one week, after which it was operated in semi continuous mode at 0.2 day<sup>-1</sup> dilution rate using the same culture medium as before; in this case, the pH was controlled (at pH = 8) by injecting pure CO<sub>2</sub> whereas dissolved oxygen was maintained below 250%Sat. by supplying air.

### 2.3. Software and hardware

Simulations and tests with real experimental data were performed with *MATLAB*® (Mathworks, Massachusetts, USA), version 9.1, R2016b. The development of the state estimator has been executed with specific functions for EKF available in *Control System Toolbox*™ (version 10.1).

State estimator simulations were carried out with the following computer hardware specifications: Intel® Core™ i5-6500 processor (up to 3.60 GHz), 8GB DDR4 (2133 MHz) RAM memory, and 500GB solid-state drive. Simulations performance was evaluated with *MATLAB*® Profiler, verifying the execution time and optimizing the code for the model and estimator.

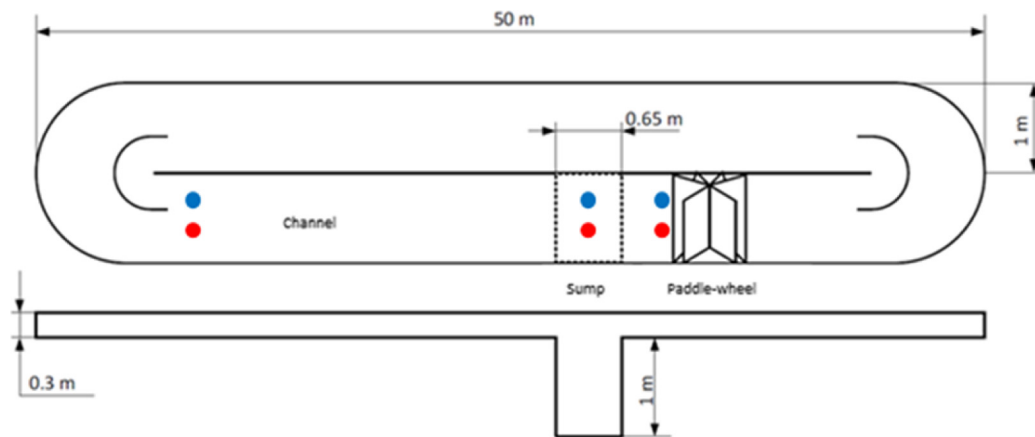
### 2.4. Dynamic model and state vector

The dynamic model consists of a set of equations that reproduce thermodynamic relationships, transport phenomena, mass balances and biological phenomena that occur in microalgae cultures on raceway photobioreactors. Therefore, the model can simulate evolution in time and space of the main variables of the process [19]. These variables are stored in a state vector that retains the characteristic dynamics of the process in the three representative parts (paddle-wheel, sump and channel). As the culture circulates along the channel the value of the variables (Table 1) changes due to the photosynthetic activity of cells. In order to account for this evolution, the model considers that the 100 m channel can be treated as 100 sections of 1 m each, providing an adequate trade-off between performance (accuracy) and computational effort. In consequence, the state vector presents the structure showed in Table 2, with a total of 512 states.

Variables, constants and characteristic parameters shown in Table 1 are used into different sub-models: solar radiation model, biological model, and engineering model of the reactor. For better understanding of the effect of these variables and constants, some equations of the model [19] are presented in Appendix A. The majority of the constants exposed in this work are related to the biological model. Eq. (A.1) calculates the average irradiance of the reactor ( $I_{av}$ ) as a function of the biomass concentration and parameters  $K_a$ . In Eq. (A.2), photosynthesis rate ( $P_{O_2}$ ) is determined as a function of dissolved oxygen concentration, pH and average irradiance, and it is affected by  $K_i$ ,  $K_{O_2}$ ,  $m$ ,  $n$ ,  $P_{O_2,max}$ ,  $R_{O_2}$ , and  $z$ . The rest of the constants are related with mass balances. Eq. (A.3) is applied to calculate the biomass concentration in the channel and is affected by parameter  $Y_{b/O_2}$ . Dissolved oxygen concentration in the channel is calculated with Eq. (A.4). Similar expressions were implemented to determine state variables for the paddle-wheel and sump as shown in Eqs. (A.5) to (A.10). The detailed version of the model can be found in [19].

### 2.5. Genetic algorithm for calibration of the model

The values of the parameters of the dynamic model were originally calibrated and validated with experimental data from the year 2014



(a)



(b)

Fig. 1. Raceway photobioreactor. (a) Photobioreactor scheme (top and side view) and (b) Real experimental facility at “Las Palmerillas”. Retrieved from [20].

Table 1  
Variables and constants of the model. Extracted from [19].

Variable	Definition	Units
<b>Dynamic variables</b>		
$C_b$	Biomass concentration	$\text{kg m}^{-3}$
$[CO_2]$	Carbon dioxide concentration in the liquid phase	$\text{mol m}^{-3}$
$[C_T]$	Total inorganic carbon concentration	$\text{mol m}^{-3}$
$[H]$	Hydrogen ions concentration	$\text{mol m}^{-3}$
$[O_2]$	Dissolved oxygen concentration	$\text{mol m}^{-3}$
$YO_2$	$O_2$ to $N_2$ molar ratio in gas phase	$\text{mol } O_2 (\text{mol } N_2)^{-1}$
$YCO_2$	$CO_2$ to $N_2$ molar ratio in gas phase	$\text{mol } CO_2 (\text{mol } N_2)^{-1}$
<b>Constants and characteristic parameters</b>		
$K_a$	Extinction coefficient	$\text{m}^2 \text{kg}^{-1}$
$K_i$	Form parameter	$\mu\text{E m}^{-2} \text{s}^{-1}$
$K_{O_2}$	Oxygen inhibition constant	$\text{mol m}^{-3}$
$m$	Form parameter	
$n$	Form exponent	
$P_{O_2, \text{max}}$	Maximum photosynthesis rate	$\text{kg } O_2 \text{ kg}^{-1} \text{ s}^{-1}$
$R_{CO_2}$	Respiration coefficient for carbon dioxide	$\text{kg } CO_2 \text{ kg}^{-1} \text{ s}^{-1}$
$R_{O_2}$	Respiration	$\text{kg } O_2 \text{ kg}^{-1} \text{ s}^{-1}$
$Y_{b/O_2}$	Biomass yield coefficient	$\text{kg}$
$z$	Form parameter	

under different operational conditions for the photobioreactor compared to those from year 2016. Some of the model parameters are directly related with the temperature. However, other parameter values depend on the status of the culture and if sudden metabolic changes occur, the only way to re-calculate those parameter values is by performing lab-scale experiments [22]. To avoid that tedious procedure, global identification techniques are a common alternative. Thus, a calibration process was accomplished to retune characteristic parameters of the model equations to achieve proper concordance between simulated dynamics and real dynamics of the process from year 2016.

For that purpose, an optimization problem was resolved to minimize root sum square error (RSS) between simulated outputs and real measured outputs from the photobioreactor (dissolved oxygen and pH). Outputs vectors were composed of signals with different units: dissolved oxygen was measured in %Sat. and pH corresponds to hydrogen ions concentration. To obtain a total error as an index of calibration, both dissolved oxygen and pH outputs were previously normalized and scaled from 0 to 1 as a result of dividing them by their respective maximum values. For this reason, the RSS error is expressed in non-dimensional units.

Taking into account the mathematical complexity of the model and

**Table 2**  
Model internal state vector structure.

Section	Channel					Paddle-wheel					Sump							
	Variable	[O <sub>2</sub> ]	C <sub>b</sub>	[C <sub>T</sub> ]	[H]	[CO <sub>2</sub> ]	[O <sub>2</sub> ]	C <sub>b</sub>	[C <sub>T</sub> ]	[H]	[CO <sub>2</sub> ]	[O <sub>2</sub> ]	C <sub>b</sub>	[C <sub>T</sub> ]	Y <sub>O<sub>2</sub></sub>	Y <sub>CO<sub>2</sub></sub>	[H]	[CO <sub>2</sub> ]
Dimension	100	100	100	100	100	1	1	1	1	1	1	1	1	1	1	1	1	1

the great number of parameters that are susceptible to be calibrated, a genetic algorithm [23] was employed to execute the calibration procedure. Genetic algorithms can efficiently explore in less computational time a region of solutions (minimization of RSS error) until founding certain areas of the region with more probabilities to contain the solution of the problem. These areas are intensively explored by different techniques of the genetic algorithm, such as mutation, crossing and individuals selecting (sets of values for parameters). The algorithm can be stopped either when the error is below a predefined threshold or when a maximum execution time is reached (probably obtaining a suboptimal solution due to local minima).

2.6. Extended Kalman filter for discrete-time systems

The dynamic model has been restructured to be represented in discrete time (where  $k$  represents the actual time instant) according to the following set of equations:

$$\begin{aligned} x_k &= f(x_{k-1}, u_{s_k}) + w_{k-1} \\ y_k &= h(x_k, u_{m_k}) + v_k \end{aligned} \tag{1}$$

where  $x$  is the defined state vector for the model,  $f$  is the non-linear state transition function (that contains the rest of functions with the equations of the model) and  $h$  is a function that relates the state vector with the measurable outputs of the model. In measurement function,  $h$ , the calculation (simulation) of dissolved oxygen and pH for each part of the photobioreactor is performed. Functions  $f$  and  $h$  have input arguments denoted as  $u_s$  and  $u_m$  respectively. These arguments can be the process inputs for function  $f$ , for example. Process noise,  $w$ , and measurement noise,  $v$ , are assumed to be zero-mean white noises, with no correlation and with covariance matrices  $Q$  and  $R$ , respectively:

$$\begin{aligned} w_k &\sim (0, Q) \\ v_k &\sim (0, R) \end{aligned} \tag{2}$$

The recursive Kalman filter algorithm consists of two steps: correction (*a priori* estimation) and prediction (*a posteriori* estimation) [24]. At prediction step, estimation for the states is generated based on the previous values of the state vector. At correction step, the predicted state vector is updated with the available information from the measurable variables of the process.

The designed state estimator initially performs (in each cycle) the correction step because measurable variables are always available from the beginning. In the first instance, the state estimator is initialized based on a first value for the state vector,  $x_0$ . In Eq. (3), predicted state vector and state estimation covariance are denoted as  $\hat{x}$  an  $P$ , respectively.

$$\begin{aligned} \hat{x}_{0|0} &= E(x_0) \\ P_{0|0} &= E[(x_0 - \hat{x}_{0|0})(x_0 - \hat{x}_{0|0})^T] \end{aligned} \tag{3}$$

For discrete-time instants  $k = 0, 1, 2, 3, \dots$ , recursive cycle of correction-prediction is applied. At correction step, corresponding to Eqs. (4) and (5), Jacobians  $C$  and  $S$  are calculated for measurement function and Kalman gain,  $K$ , is computed. The predicted state vector is corrected using real output measurement data,  $y$ , and state estimation covariance is updated.

$$\begin{aligned} C_k &= \left. \frac{\partial h}{\partial x} \right|_{\hat{x}_{k|k-1}} \\ S_k &= \left. \frac{\partial h}{\partial v} \right|_{\hat{x}_{k|k-1}} \end{aligned} \tag{4}$$

$$\begin{aligned} K_k &= P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + S_k R S_k^T)^{-1} \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k (y_k - h(\hat{x}_{k|k-1}, u_{m_k})) \\ P_{k|k} &= P_{k|k-1} - K_k C_k P_{k|k-1} \end{aligned} \tag{5}$$

At prediction step, Jacobians  $A$  and  $G$  are calculated for state transition function and future state vector,  $\hat{x}_{k+1|k}$ , and state estimation covariance,  $P_{k+1|k}$ , are predicted according to Eqs. (6) and (7).

$$\begin{aligned} A_k &= \left. \frac{\partial f}{\partial x} \right|_{\hat{x}_{k|k}} \\ G_k &= \left. \frac{\partial f}{\partial w} \right|_{\hat{x}_{k|k}} \end{aligned} \tag{6}$$

$$\begin{aligned} P_{k+1|k} &= A_k P_{k|k} A_k^T + G_k Q G_k^T \\ \hat{x}_{k+1|k} &= f(\hat{x}_{k|k}, u_{s_k}) \end{aligned} \tag{7}$$

As can be noticed in the shown algorithm, the EKF attempts to approximately linearize model functions  $f$  and  $h$  every instant. A relevant aspect that negatively affects to the computational time of the algorithm is that Jacobians matrices  $A$  and  $C$  must be numerically computed since some equations of the dynamic model have static balances. On the contrary, Jacobians  $S$  and  $G$  are calculated as identity matrices because the process expression in Eq. (1) assumes additive noise.

Properly setting the values for matrices  $Q$  and  $R$  is also an important aspect for the performance of the state estimator. Matrix  $Q$  describes process noise covariance and represents the inaccuracy of the model to reproduce real dynamics of the process. Matrix  $R$  defines measurement noise covariance associated to real process outputs, so it can be interpreted as the noise covariance of sensors. The election of both matrices has an impact on the execution of the state estimator, as can be observed in Eqs. (5) and (7). Inappropriate tuning of these matrices can lead to bad calculations of Kalman gain and state estimation covariance, resulting in wrong estimations and even causing the instability of the state estimation procedure.

Tuning of matrix  $Q$  is extremely difficult to achieve when using non-linear models with uncertainty in characteristic parameters and constants. In contrast, matrix  $R$  can be adjusted based on standard deviations of sensors that measure the outputs of the process. Nevertheless, both matrices are usually tuned by trial and error procedures and evaluating the quality of the estimations provided by the filter [25].

Matrices  $Q$  and  $R$  contain their covariance values in the main diagonal and the rest of elements are null since there is no correlation between process and measurement noises. Both are square matrices with dimensions given by the number of states and the number of measurable outputs, respectively.

3. Results and discussions

3.1. Adaptation of the model

Adaptation of the dynamic model [19] was a previous stage to

modify the programmed code for certain functions in order to simulate and perform off-line trials with real experimental data registered in the photobioreactor from year 2016. The most important modification was related with operational conditions and process control characteristics. Injected gas for pH control from year 2016 was 100% pure CO<sub>2</sub>, while in year 2014 the injected gas had only 10% CO<sub>2</sub> composition because it was obtained from a diesel-fuelled heating boiler. Consequently, if CO<sub>2</sub> gas is pure, less gas flow is required to achieve pH control of the culture, so the programmed code was modified to consider new values for gas flow rate and its composition.

Multiple simulations were executed with real experimental data to evaluate computing demands, such as elapsed time per simulation with *MATLAB*®. Several enhancements were included in the structure of the programmed functions for the dynamic model and simulation times were reduced. For instance, a simulation with 24 h of registered experimental data initially lasted 3 min of computing and after the improvements the same simulation lasted 1 min. The developed state estimator also benefited with the enhancements because it employees the programmed function from the dynamic model.

In this work, adequate computational performance is also pursued since the state estimator should be implemented on a hardware device for on-line operation with the rest of the sensors in the photobioreactor.

### 3.2. Calibration of the model

Dynamics of bioprocesses are highly complex because they deal with living organisms. Dynamic models need to be re-calibrated due to metabolic changes in the culture or by the dependence of parameters with temperature and, therefore, with seasons of the year [26]. The equations of model [19] do not account for the effect of environmental temperature on the dynamics. Therefore, re-calibration of the model is required so that model uncertainty is accounted by modifications in several parameters. The model [19] was originally calibrated for fall in year 2014. Hence, the recalibration of parameters was executed with experimental data from November 2016. The selected data were free of inconsistencies and showed the representative dynamics of the photobioreactor.

Before performing a calibration procedure, and especially when involving a large number of parameters, selection of best parameters to be calibrated is crucial. A sensibility analysis was performed in a series of simulations in which  $\pm 10\%$  changes in identified parameters were performed, so that the error between simulated and real outputs was studied and it was concluded that parameters which mostly affect the performance of the model were:  $P_{O_2,max}$ ,  $K_a$ ,  $n$ ,  $K_b$ ,  $m$ ,  $Y_{b/O_2}$ ,  $R_{O_2}$ ,  $R_{CO_2}$ ,  $K_{O_2}$  and  $z$  (see Table 1).

The calibration was performed with a genetic algorithm [23] and constrained values for parameters. The limits for values were selected as conservative ranges since the parameters must retain the physico-chemical and biological meaning inside the model as well as they should allow some freedom for variability and hence expanding the solution space. All simulations for calibration procedure were executed with experimental data corresponding to daylight period with a short span of nocturnal data as initialization. The calibration time per simulation was reduced because less experimental data were employed by focusing on daylight dynamics and the period of time when real biomass samples were recorded.

Experimental data for calibration were selected from November 3, 2016. The limits and values applied to parameters in this calibration are presented in Table 3. Results of a simulation with calibrated parameters are shown in Fig. 2. The total RSS error (dissolved oxygen and pH) is 25.71 units.

Validation for calibrated parameters was executed with selected experimental data from November 17, 2016. Results for this day are presented in Fig. 3, with a total RSS error of 30.26 units.

Similar results were obtained in simulations with different days. It was also determined that the calibration provides acceptable results for

**Table 3**  
Limits and values for parameters in calibration.

Parameter	Initial value (units)	Limits	Calibrated value
$K_a$	79.9980 (m <sup>2</sup> kg <sup>-1</sup> )	[50, 150]	74.1757
$K_i$	173.9944 (μE m <sup>-2</sup> s <sup>-1</sup> )	[100, 300]	292.3446
$K_{O_2}$	0.8373 (mol m <sup>-3</sup> )	[0.282, 1]	0.9112
$m$	0.0021	[0.001, 0.01]	0.0074
$n$	1.0450	[1,3]	1.6363
$P_{O_2,max}$	2.0641·10 <sup>-5</sup>	[1.03205·10 <sup>-5</sup> , 4.1282·10 <sup>-5</sup> ]	2.7111·10 <sup>-5</sup>
$R_{CO_2}$	4.2752·10 <sup>-6</sup>	[0.85504·10 <sup>-6</sup> , 1.2826·10 <sup>-5</sup> ]	1.2182·10 <sup>-5</sup>
$R_{O_2}$	9.5752·10 <sup>-7</sup>	[1.91504·10 <sup>-7</sup> , 2.87257·10 <sup>-6</sup> ]	6.9481·10 <sup>-7</sup>
$Y_{b/O_2}$	0.2 (kg)	[0.05, 0.4]	0.3611
$z$	5.4356	[1,6]	5.4809

around one month of operation (out of this period higher RSS errors -over 50 units- are achieved). Hence, it was concluded that the model should be calibrated once or twice each season of the year or it may be modified to include the variation of temperature [27] and its effect on the dynamics of the process.

RSS errors are expressed in units because it is necessary for the resolution of error minimization problem in calibration with genetic algorithm (as explained in Section 2.5). However, it is also interesting to analyse relative errors presented in Table 4. These relative errors are remarkably similar to original obtained errors for the model in year 2014 [19]. Therefore, it can be confirmed that the model has been properly recalibrated for experimental data and operational conditions in year 2016. In view of the results in Figs. 2 and 3, it can be deduced that the photobioreactor was operating in strongly changing conditions and performance of the model is acceptable due to the preponderant uncertainty of the process.

Thus, it has been highlighted the existence of important limitations in the operation of industrial raceway photobioreactors, related with changing ambient conditions and the difficulty of repeating similar tests on different days of a year. This issue is related with the variation of environmental conditions in real outdoor cultures, both daily and along the year, and the effect that microalgae adaptation has in the variation of the main characteristic parameters of the biological system. Thus, the biological behaviour is slightly different when continuous sunny days take place that when cloudy and sunny days alternate, this makes more difficult to find reliable models and imposes the necessity to manage large sets of data to accurately calibrate them.

### 3.3. Development of the state estimator

Fig. 4 shows the schematic concept for the developed state estimator, combining an EKF and the dynamic model of the photobioreactor. The state estimator is designed to predict the complete state vector of the model, but it has been optimized to principally estimate biomass concentration by tuning  $Q$  and  $R$  matrices.

Matrix  $R$  was adjusted with standard deviation values of installed sensors on real photobioreactor for outputs measurement. Standard deviation for dissolved oxygen is of 5% in saturation, and for pH is consider of 0.2 (in pH scale). Eq. (8) presents matrix  $R$  squared values to express variance.  $Q$  matrix was tuned by analysing the range of variation for state vector of the model and reasoning how much variation might be necessary to make the state vector values more similar to real states on the photobioreactor. Due to its great dimensions (512 files and 512 rows),  $Q$  matrix is shown in a reduced manner in Table 5, with the values for process noise covariance allocated in the main diagonal.

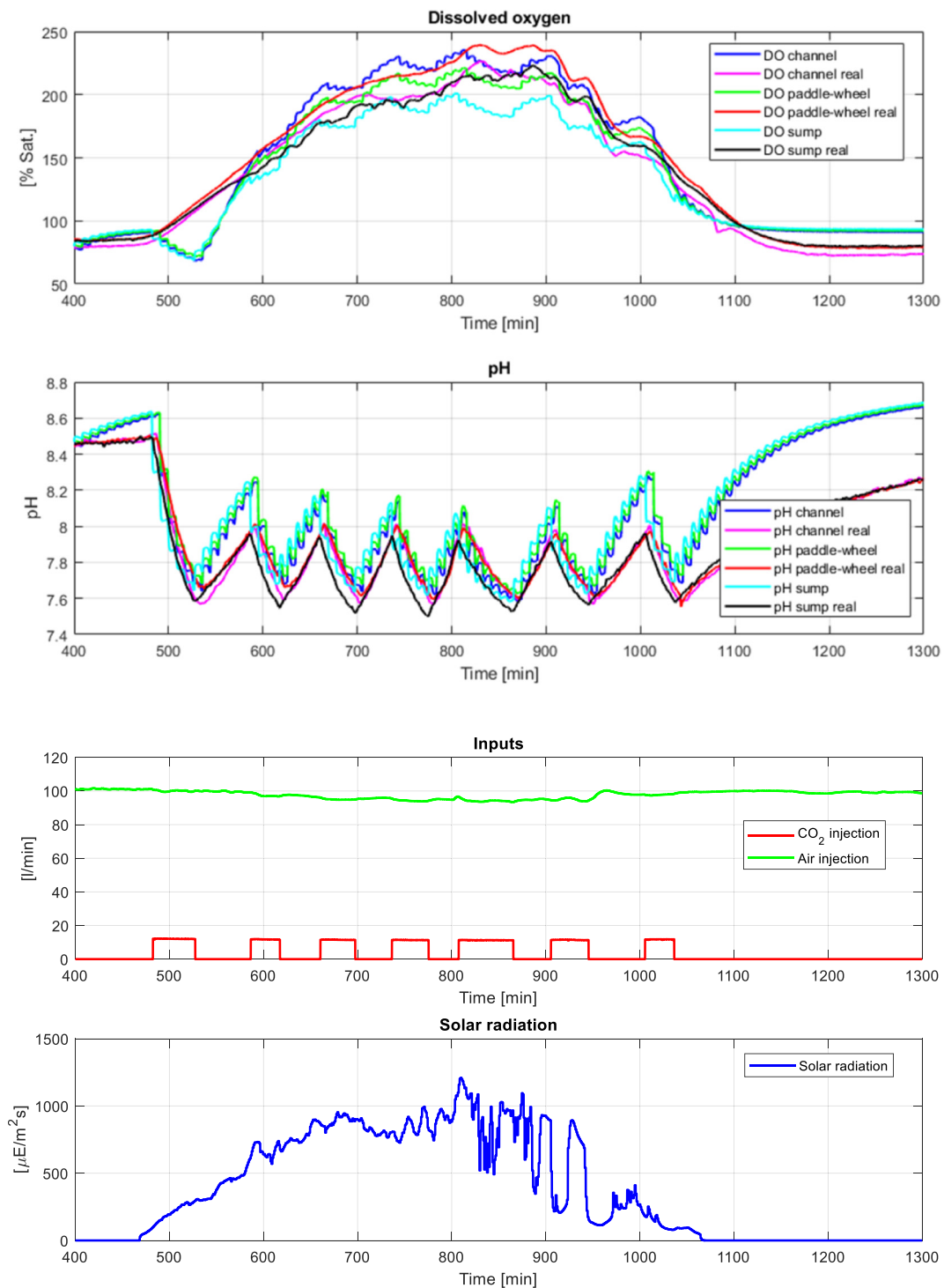


Fig. 2. Simulation results with calibrated parameters for November 3, 2016. Time is shown in minutes from 00:00 a.m. (local hour).

$$R = \begin{pmatrix} 25 & 0 & 0 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.04 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.04 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.04 \end{pmatrix} \quad (8)$$

Fig. 5 shows the results for an off-line test of the estimator using experimental data from November 3, 2016. To evaluate the behaviour of the estimator, Fig. 5e should be initially analysed. In this figure, the quality of the estimator adjustment to real data can be highly perceptible compared with the rigidity of the biomass concentration simulated

by the model. The estimated biomass concentration is closer to real experimental data than the biomass concentration simulated by the model and evolves with a decreasing tendency to the nocturnal period, while this behaviour cannot be reproduced by the dynamical model. This type of response is consistent with microalgae metabolism: cellular growth takes place during daylight hours due to photosynthesis, and biomass loss occurs because of dark respiration at night-time. Unfortunately, biomass concentration measurements were not performed during night-time due to limitations in the experimental station access and, therefore, the behaviour of the state estimator is not totally tested.

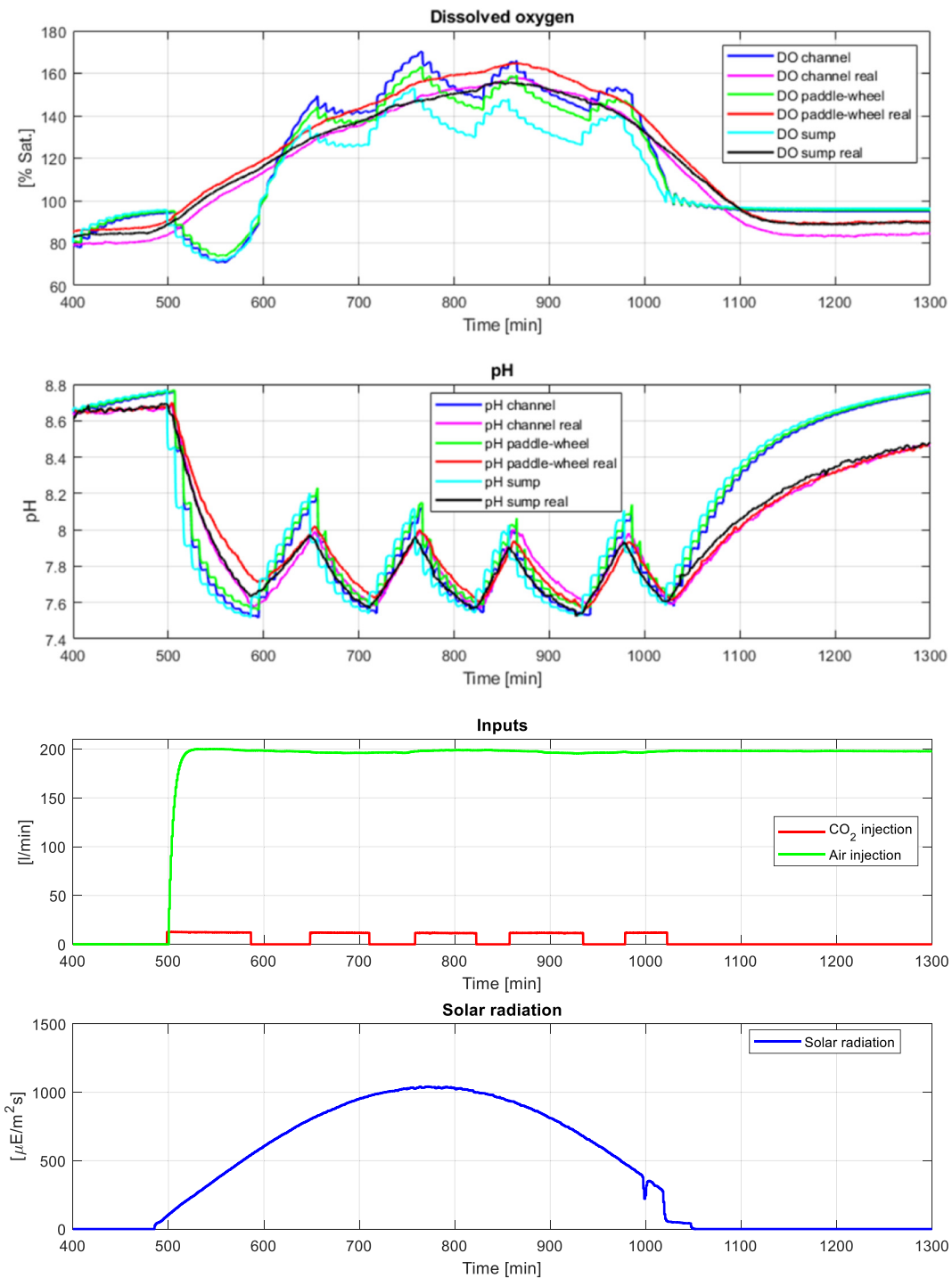


Fig. 3. Validation for calibrated parameter simulating with data from November 17, 2016.

Table 4

Relative errors for calibration and validation results.

Date	Relative error (%) without calibrated parameters						Relative error (%) with calibrated parameters					
	DO channel	DO paddle	DO sum	pH channel	pH paddle	pH sump	DO channel	DO paddle	DO sump	pH channel	pH paddle	pH sump
11/03	26.95	31.17	29.39	3.72	3.61	3.56	10.30	8.37	8.87	2.45	2.58	2.85
11/17	18.32	19.81	20.03	4.17	4.05	4.15	8.73	7.77	9.29	1.73	1.83	1.92

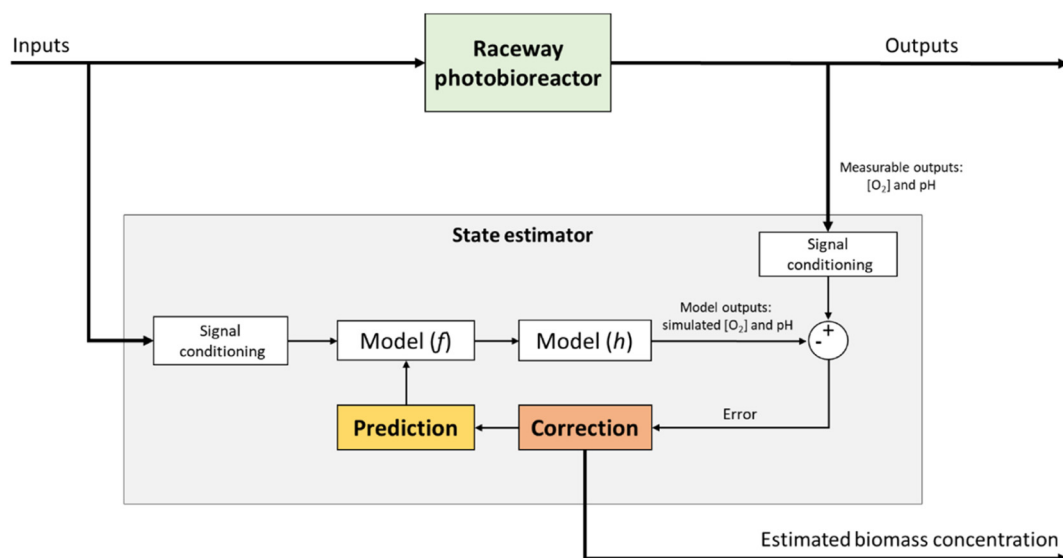


Fig. 4. Schematic concept for the developed state estimator, designed to work on-line with the photobioreactor. Model (f) and Model (h) blocks refer to f and h functions as explained in Section 2.6.

Table 5  
Process noise covariance for matrix Q.

Section	State	Covariance
Channel	[O <sub>2</sub> ]	0.1
	C <sub>b</sub>	0.01
	[C <sub>T</sub> ]	2
	[H]	5·10 <sup>-7</sup>
	[CO <sub>2</sub> ]	0.5
Paddle-wheel	[O <sub>2</sub> ]	0.05
	C <sub>b</sub>	0.005
	[C <sub>T</sub> ]	0.1
	[H]	1·10 <sup>-6</sup>
	[CO <sub>2</sub> ]	0.001
Sump	[O <sub>2</sub> ]	0.05
	C <sub>b</sub>	0.005
	[C <sub>T</sub> ]	0.1
	Y <sub>O<sub>2</sub></sub>	2·10 <sup>12</sup>
	Y <sub>CO<sub>2</sub></sub>	1
	[H]	1·10 <sup>-6</sup>
	[CO <sub>2</sub> ]	0.1

Fig. 5a corroborates that correction step is properly performed, as the simulated outputs adequately correspond with real outputs, so the Kalman gain from Eq. (5) is correctly calculated by the estimator. On the other side, Fig. 5b shows the predictions obtained by simulating with corrected states from correction step. In terms of errors, Fig. 5c evidences that substantial errors appear in prediction step for dissolved oxygen in channel and paddle-wheel, which may be indicative that a readjustment for Q matrix is needed. However, it should be emphasized that any change in Q matrix can cause an improvement but also a worsening in biomass estimation. For this reason, if biomass concentration is adequately estimated in relation with real samples, it is advisable not to change values of matrix Q.

In Fig. 5a and b some sudden peak deviations can be observed for pH state in paddle-wheel. Values of Q and R matrices, and numeric calculation for Jacobians in Eqs. (4) and (6) may cause that some states turn into negative or null values, which is not possible from a physicochemical perspective. In order to ensure that the estimated state vector does not contain wrong values, a restriction function was developed. In each discrete time instant, this function checks the values for pH and dissolved CO<sub>2</sub> states, and negative or null values are replaced with a logic value from some previous instants. Thanks to this function, although some divergences may appear, estimated biomass

concentration is not affected. The programmed restriction function is a preliminary solution to perform experimental simulations and, if the state estimator is incorporated to real process measurement system, it should be mathematically modified to impose constraints to states in an integrated manner with the estimation algorithm. Some examples of state constraints for Kalman filters were proposed in [28].

Similar results were obtained for other experimental data in days of October and November. For instance, Fig. 6 shows the results for October 26, 2016. In this test, a reduced simulation was executed using experimental data from time interval in which real biomass samples are congregated. Fig. 6a and b do not exhibit any type of peak deviation and errors for correction step in Fig. 6c are nearly zero, so this confirms a good performance for the estimator in different ambient conditions compared to Fig. 5d.

In Fig. 6e, estimated biomass concentration can be evaluated. As observed, estimated biomass begins to decrease gradually at 600 min (10:00 a.m. local hour), sooner than expected by real biomass samples, until it reaches the lowest value of 0.3 kg m<sup>-3</sup>. Subsequently, estimated biomass starts an increasing trend due to the high leap in solar radiation at 750 min (12:30 p.m. local hour). When solar radiation rises, photosynthetic activity is also augmented, causing a greater growth of culture cells and thus it leads to increase biomass concentration. The estimator response clearly fits to different varieties of changes that occur in the real photobioreactor.

In multiple off-line test with the estimator, computational times were surprisingly high (2–5 h, depending on the quantity of experimental data). This is caused by numeric resolution for Jacobians, that consumes the 93% of the time for each off-line test because the computer performs partial derivatives for 512 states by numeric approximation of non-linear functions. Since the objective of implementing a state estimator for biomass concentration is to use it as a software sensor in the real installation, it is necessary to know if the computation time of the developed estimator is compatible with the measurement frequency of the existing sensors in the photobioreactor. To that end, the total computation time was divided by the number of estimated samples in each test, obtaining a constant ratio of 1 s of computation per estimated sample. As measurement from the sensors in the real photobioreactor is performed every minute, the developed estimator can work online and in real time on the process.



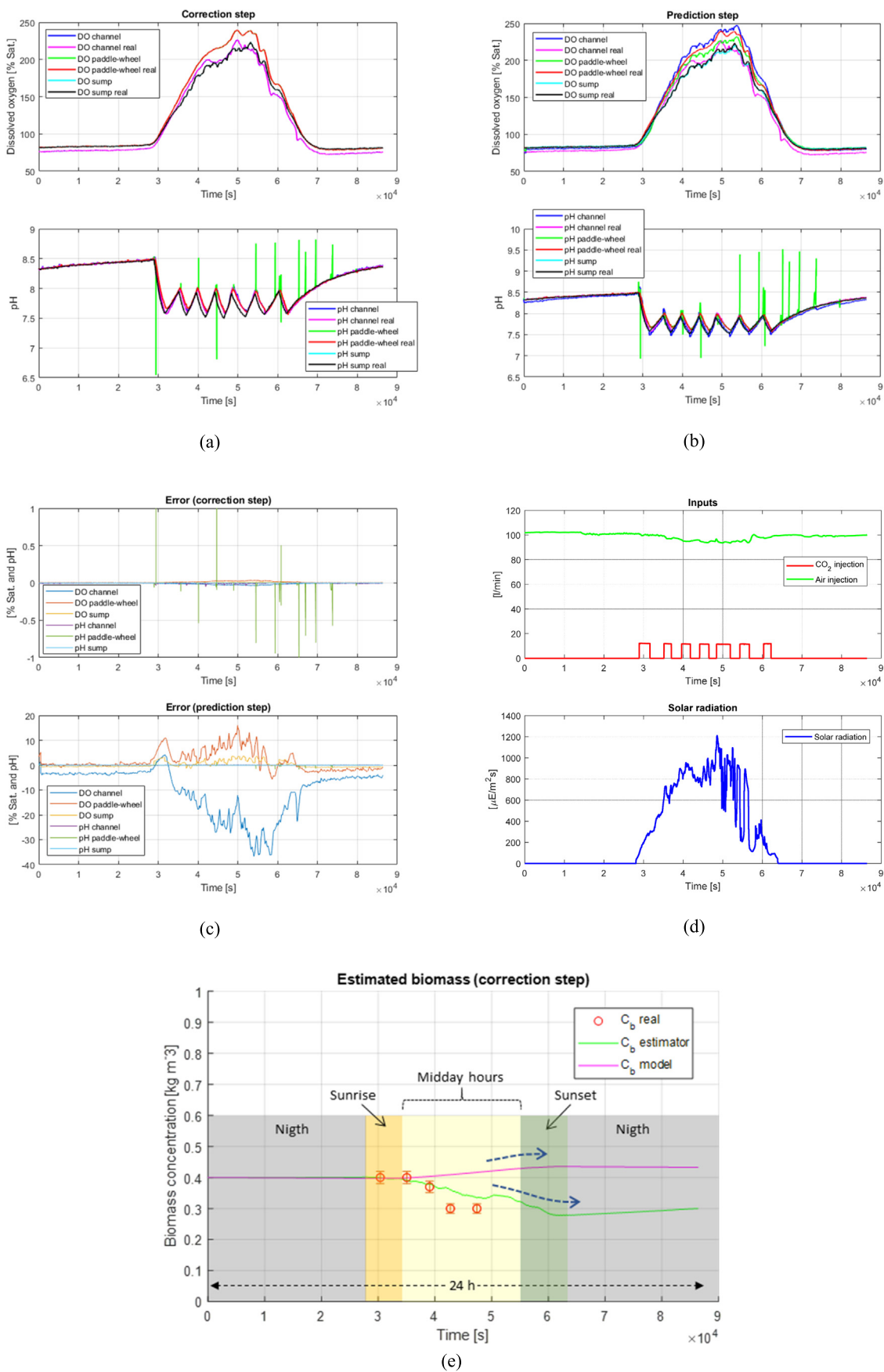


Fig. 5. Results for the state estimator with data from November 3, 2016.

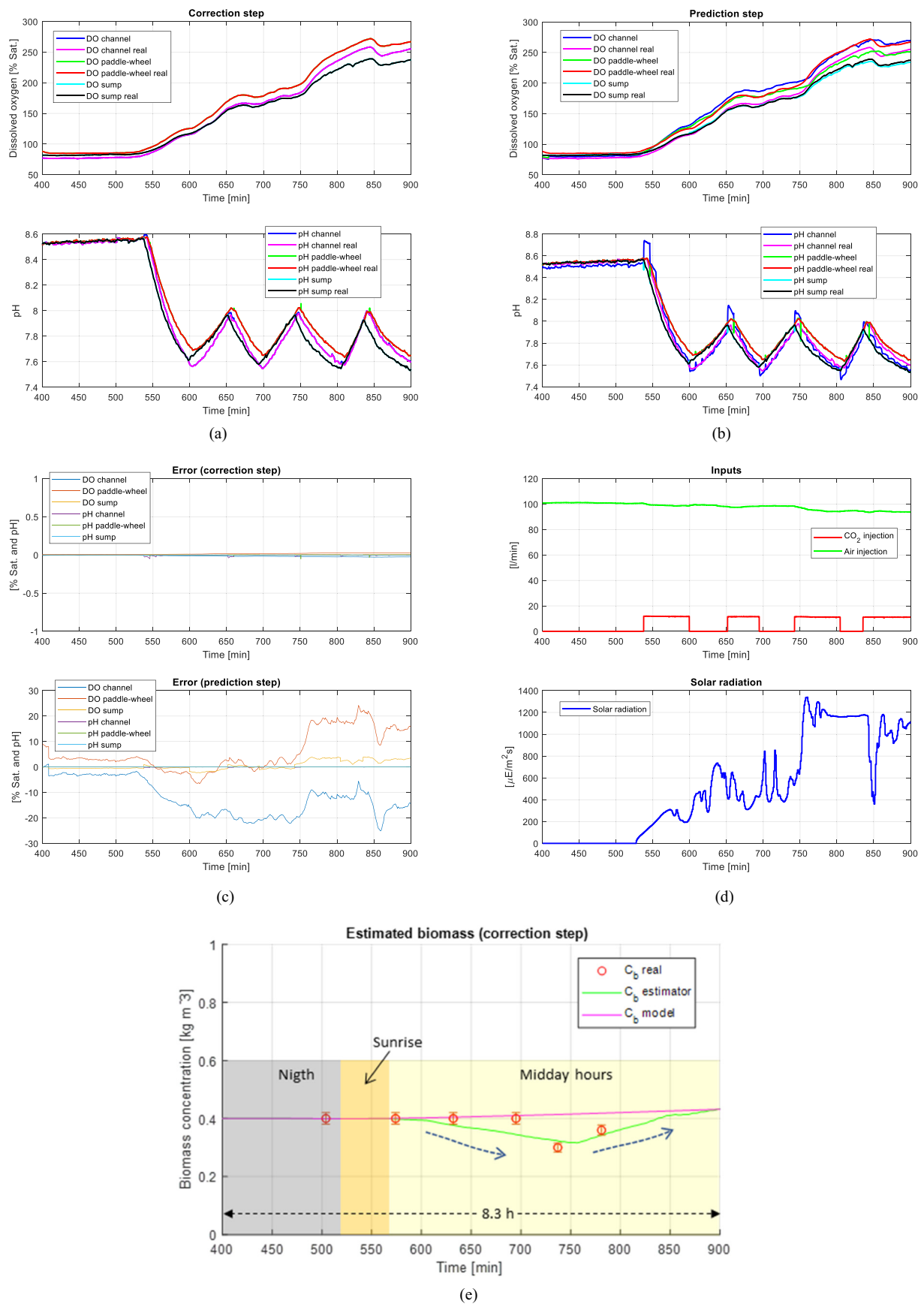


Fig. 6. Results for the state estimator with data from October 26, 2016.

#### 4. Conclusions

This work tries to respond to one of the biggest problems affecting the monitoring of microalgae production processes, which is that biomass concentration cannot be accurately measured in real time on the process. An estimator using the extended Kalman filter has been developed to estimate the biomass concentration using the available experimental measurements of dissolved oxygen, pH, gas injections and solar radiation. The developed estimator presents an acceptable efficiency, but it requires multiple “trial and error” tests with high computation times to determine the appropriate values of  $Q$  and  $R$  matrices, its tuning parameters. Proposed values for  $Q$  and  $R$  were a result of a series of simulations that could not be representative of the true distribution that these covariance matrices may have for all the experimental data of a complete year. Despite this, the results show a clear potential in the performance of the estimator, although more improvements have to be performed in the future. For instance, in future works, modern estimation algorithms will be tested to compare their performance against the extended Kalman filter.

In conclusion, the developed state estimator offers a successful performance, in terms of trends, compared to the rigidity of the dynamic model. This encourages to consider on-line use of the estimator on the process to generate advanced control actions for the photobioreactor, focused on optimizing biomass productivity and, by extension, to generate greater industrial benefits with high value-added compounds that can be obtained from microalgae cultures.

#### Acknowledgments

This work has been partially funded by the following projects: DPI2014-55932-C2-1-R and DPI2017-84259-C2-1-R (Spanish Ministry

#### Appendix A

An overview of the equations of the model (extracted from [19]) is presented in this appendix.

##### Biological model

$$I_{av}(t, x) = \frac{I_0(t)}{K_a C_b(t, x) h} (1 - \exp(-K_a C_b(t, x) h)) \tag{A.1}$$

$$P_{O_2}(t, x) = (1 - \alpha_s) \frac{P_{O_2, max} I_{av}(t, x)^n}{K_i \exp(I_{av}(t, x) m) + I_{av}(t, x)^n} \left( 1 - \left( \frac{[O_2](t, x)}{K_{O_2}} \right)^2 \right) \left( B_1 \exp\left( \frac{-C_1}{pH(t, x)} \right) - B_2 \exp\left( \frac{-C_2}{pH(t, x)} \right) \right) - \alpha_s R_{O_2} \tag{A.2}$$

##### Engineering model

- Mass balances in the liquid phase

1. Channel:

$$wh \frac{\partial C_b(t, x)}{\partial t} = -whv \frac{\partial C_b(t, x)}{\partial x} + whP_{O_2}(t, x) C_b(t, x) Y_{b/O_2} \tag{A.3}$$

$$wh \frac{\partial [O_2](t, x)}{\partial t} = -whv \frac{\partial [O_2](t, x)}{\partial x} + wh \frac{P_{O_2}(t, x) C_b(t, x)}{M_{O_2}} + whK_{laO_2} ([O_2^*](t, x) - [O_2](t, x)) \tag{A.4}$$

2. Paddle-wheel:

$$\frac{dCb_{out}(t)}{dt} = -\frac{Q_{liq}}{V_p} (Cb_{out}(t) - Cb_{in}(t)) + P_{O_2}(t) Cb_{in}(t) Y_{b/O_2} \tag{A.5}$$

$$\frac{d[O_2]_{out}(t)}{dt} = -\frac{Q_{liq}}{V_p} ([O_2]_{out}(t) - [O_2]_{in}(t)) + \frac{P_{O_2}(t) Cb_{out}(t)}{M_{O_2}} + K_{laO_2p} ([O_2^*](t) - [O_2](t))_{lm} \tag{A.6}$$

of Economy, Industry and Competitiveness and FEDER funds), and the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 727874 SABANA, and supported by Cajamar Foundation.

#### Author contributions statement

F. García-Mañas was the responsible of the work, the estimator development, and the writing of the manuscript. J. L. Guzmán Sánchez was the responsible of data processing, model calibration and contribute to the elaboration of the manuscript. M. Berenguel was responsible of estimator choice and its tuning parameters by using genetic algorithm. He also contributed to text and coordination of the work. F. G. Ación Fernández was the responsible of the nonlinear model adaptation and the elaboration of the manuscript, mainly participating in the discussion of the results and finalization of the manuscript.

#### Conflict of interest statement

The authors declare any potential financial or other interests that could be perceived to influence the outcomes of the research.

#### Statement of informed consent, human/animal rights

“No conflicts, informed consent, human or animal rights applicable”.

#### Declaration of authors

All the participants are authorship of this work and agree to submit the manuscript for peer review to Algal Research.

## 3. Sump:

$$\frac{dCb_{out}(t)}{dt} = -\frac{Q_{liq}}{V_s(1-\varepsilon_s(t))}(Cb_{out}(t) - Cb_{in}(t)) + P_{O_2}(t)Cb_{in}(t)Y_{b/O_2} - \frac{Q_m}{V_s(1-\varepsilon_s(t))}Cb_{out}(t) \quad (A.7)$$

$$\frac{d[O_2]_{out}(t)}{dt} = -\frac{Q_{liq}}{V_s(1-\varepsilon_s(t))}([O_2]_{out}(t) - [O_2]_{in}(t)) + \frac{P_{O_2}(t)Cb_{out}(t)}{M_{O_2}} + K_{laO_2s}([O_2^*](t) - [O_2](t))_{lm} + \frac{Q_m}{V_s(1-\varepsilon_s(t))}([O_2]_{lm} - [O_2]_{out}(t)) \quad (A.8)$$

## • Mass balances to the gas phase (sump)

$$\frac{dYO_{2,out}(t)}{dt} = -\frac{Q_{gas}}{V_s(1-\varepsilon_s(t))}(YO_{2,out}(t) - YO_{2,in}(t)) - K_{laO_2s} \frac{V_{mol}(1-\varepsilon_s(t))}{y_{N_2} \varepsilon_s(t)}([O_2^*](t) - [O_2](t))_{lm} \quad (A.9)$$

$$\frac{dYCO_{2,out}(t)}{dt} = -\frac{Q_{gas}}{V_s(1-\varepsilon_s(t))}(YCO_{2,out}(t) - YCO_{2,in}(t)) - K_{laCO_2s} \frac{V_{mol}(1-\varepsilon_s(t))}{y_{N_2} \varepsilon_s(t)}([CO_2^*](t) - [CO_2](t))_{lm} \quad (A.10)$$

## References

- [1] M. Benavides, D. Coutinho, A.L. Hantson, J. Van Impe, A. Vande Wouwer, Robust Luenberger observers for microalgal cultures, *J. Process Control* 36 (2015) 55–63, <https://doi.org/10.1016/j.jprocont.2015.09.005>.
- [2] B.T. Nguyen, B.E. Rittmann, Low-cost optical sensor to automatically monitor and control biomass concentration in microalgal cultivation, *Algal Res.* 32 (2018) 101–106, <https://doi.org/10.1016/j.algal.2018.03.013>.
- [3] A.J. De Assis, R. Maciel Filho, Soft sensors development for on-line bioreactor state estimation, *Comput. Chem. Eng.* 24 (2000) 1099–1103, [https://doi.org/10.1016/S0098-1354\(00\)00489-0](https://doi.org/10.1016/S0098-1354(00)00489-0).
- [4] E. Ibañez, A. Cifuentes, Benefits of using algae as natural sources of functional ingredients, *J. Sci. Food Agric.* 93 (2013) 703–709, <https://doi.org/10.1002/jsfa.6023>.
- [5] F. Ortega, O. Pérez, E. López, Comparación del desempeño de estimadores de estado no lineales para determinar la concentración de biomasa y sustrato en un bioproceso, *Inf. Tecnol.* 26 (2015) 35–44.
- [6] L. Delauney, C. Compare, M. Lehaitre, Biofouling protection for marine environmental sensors, *Ocean Sci.* 6 (2010) 503–511, <https://doi.org/10.5194/os-6-503-2010>.
- [7] O. Zerrouh, J.V. Reinoso-Moreno, L. López-Rosales, M. del C. Cerón-García, A. Sánchez-Mirón, F. García-Camacho, E. Molina-Grima, Biofouling in photobioreactors for marine microalgae, *Crit. Rev. Biotechnol.* 37 (2017) 1006–1023, <https://doi.org/10.1080/07388551.2017.1299681>.
- [8] J. Albiol, J. Robusté, C. Casa, M. Poch, Biomass estimation in plant cell cultures using an extended Kalman filter, *Biotechnol. Prog.* 9 (1993) 174–178.
- [9] C. Komives, R.S. Parker, Bioreactor state estimation and control, *Curr. Opin. Biotechnol.* 14 (2003) 468–474, <https://doi.org/10.1016/j.copbio.2003.09.001>.
- [10] R.E. Kalman, A new approach to linear filtering and prediction problems, *ASME J. Basic Eng.* 82 (1960) 35–45, <https://doi.org/10.1115/1.3662552>.
- [11] R. Simutis, V. Galvanuskas, D. Levisauskas, J. Repsyte, V. Grincas, State estimation of a biotechnological process using extended Kalman filter and particle filter, *World Acad. Sci. Eng. Technol. Int. J. Biol. Biomol. Agric. Food Biotechnol. Eng.* 8 (2014) 937–941.
- [12] J. Li, N.S. Xu, W.W. Su, Online estimation of stirred-tank microalgal photobioreactor cultures based on dissolved oxygen measurement, *Biochem. Eng. J.* 14 (2003) 51–65, [https://doi.org/10.1016/S1369-703X\(02\)00135-3](https://doi.org/10.1016/S1369-703X(02)00135-3).
- [13] S.J. Yoo, D.H. Jung, J.H. Kim, J.M. Lee, A comparative study of soft sensor design for lipid estimation of microalgal photobioreactor system with experimental validation, *Bioresour. Technol.* 179 (2015) 275–283, <https://doi.org/10.1016/j.biortech.2014.12.029>.
- [14] S. Tebbani, M. Titica, S. Caraman, L. Boillereaux, Estimation of *Chlamydomonas reinhardtii* growth in a torus photobioreactor, *IFAC Proc. Vol.*, IFAC, 2013, pp. 72–77, <https://doi.org/10.3182/20131216-3-IN-2044.00053>.
- [15] D. Dochain, State and parameter estimation in chemical and biochemical processes: a tutorial, *J. Process Control* 13 (2003) 801–818, [https://doi.org/10.1016/S0959-1524\(03\)00026-X](https://doi.org/10.1016/S0959-1524(03)00026-X).
- [16] G. Marafioti, S. Tebbani, D. Beauvois, G. Becerra, A. Isambert, M. Hovd, Unscented Kalman Filter State and Parameter Estimation in a Photobioreactor for Microalgae Production, *IFAC*, 2009, <https://doi.org/10.3182/20090712-4-TR-2008.00131>.
- [17] W.W. Su, J. Li, N.S. Xu, State and parameter estimation of microalgal photobioreactor cultures based on local irradiance measurement, *J. Biotechnol.* 105 (2003) 165–178, [https://doi.org/10.1016/S0168-1656\(03\)00188-3](https://doi.org/10.1016/S0168-1656(03)00188-3).
- [18] J. Mohd Ali, N. Ha Hoang, M.A. Hussain, D. Dochain, Review and classification of recent observers applied in chemical process systems, *Comput. Chem. Eng.* 76 (2015) 27–41, <https://doi.org/10.1016/j.compchemeng.2015.01.019>.
- [19] I. Fernández, F.G. Ación, J.L. Guzmán, M. Berenguel, J.L. Mendoza, Dynamic model of an industrial raceway reactor for microalgae production, *Algal Res.* 17 (2016) 67–78, <https://doi.org/10.1016/j.algal.2016.04.021>.
- [20] A. Pawlowski, J.L. Mendoza, J.L. Guzmán, M. Berenguel, F.G. Ación, S. Dormido, Effective utilization of flue gases in raceway reactor with event-based pH control for microalgae culture, *Bioresour. Technol.* 170 (2014) 1–9, <https://doi.org/10.1016/j.biortech.2014.07.088>.
- [21] A. Pawlowski, J.L. Guzmán, M. Berenguel, F.G. Ación, S. Dormido, Application of predictive feedforward compensator to microalgae production in a raceway reactor: a simulation study, *Energies* 11 (2018), <https://doi.org/10.3390/en11010123>.
- [22] R. Muñoz-Tamayo, P. Martinon, G. Bougaran, F. Mairet, O. Bernard, Getting the most out of it: optimal experiments for parameter estimation of microalgae growth models, *J. Process Control* 24 (2014) 991–1001, <https://doi.org/10.1016/j.jprocont.2014.04.021>.
- [23] C.R. Houck, J. Joines, M.G. Kay, A genetic algorithm for function optimization: a Matlab implementation, *Ncsu-le Tr.* (1995) 95.
- [24] D. Simon, *Optimal State Estimation*, Wiley-Interscience, Hoboken, 2006.
- [25] J. Valappil, C. Georgakis, Systematic estimation of state noise statistics for extended Kalman filters, *AIChE J.* 46 (2000) 292–308, <https://doi.org/10.1002/aic.690460209>.
- [26] D. Dochain, *Bioprocess Control*, ISTE, London, 2008.
- [27] Q. Béchet, N. Coulombier, C. Vasseur, T. Lasserre, L. Le Dean, O. Bernard, Full-scale validation of an algal productivity model including nitrogen limitation, *Algal Res.* 31 (2018) 377–386, <https://doi.org/10.1016/j.algal.2018.02.010>.
- [28] D. Simon, Kalman filtering with state constraints: a survey of linear and nonlinear algorithms, *IET Control Theory Appl.* 4 (2010) 1303–1318, <https://doi.org/10.1049/iet-cta.2009.0032>.