

A nonlinear approach for input signal design with persistent excitation applied to pH modelling in microalgae raceway reactors[★]

F. Campregher^{*,**} M. Caparroz^{***} J. L. Guzmán^{***}
A. Visioli^{*}

^{*} *Dipartimento di Ingegneria Meccanica e Industriale, University of Brescia, Brescia, Italy* {francesco.campregher, antonio.visioli}@unibs.it

^{**} *Dept. of Electrical and Information Engineering, Polytechnic of Bari, Bari, Italy* f.campregher@phd.poliba.it

^{***} *Department of Informatics, University of Almeria, CIESOL, ceiA3, 04120, Almería, Spain* {mcaparroz, joseluis.guzman}@ual.es

Abstract: The fast pace of urbanization, population growth, and fossil fuel dependency have brought environmental challenges such as global warming and water contamination, pressing for solutions in greenhouse gas reduction and wastewater treatment. Microalgae cultivation offers promising results by assimilating CO₂ and purifying wastewater. This study focuses on the identification of pH models in microalgae raceway reactors, essential for accurate control and optimization of growth conditions. A novel multisine-based persistent excitation approach combined with a range controller is proposed to enhance data quality and coverage across varying operating points, without violating output constraints. This method demonstrates improved operational stability and enriched dataset acquisition for model identification. Experimental results, conducted at the University of Almeria and IFAPA research center, confirm the method's effectiveness in generating reliable excitation data, supporting accurate pH model identification for industrial-scale applications.

Keywords: Input and excitation design, Closed loop identification, Nonlinear system identification, Microalgae, Open reactor

1. INTRODUCTION

The last few decades have come along with fast population growth, an intensive urbanization, and a significant increase in the use of fossil fuels. These aspects have brought great advances in the quality of life, such as the availability of electricity in every home, but also major problems that society must address as soon as possible. These include mainly water contamination, global warming, and climate changes.

In this sense, in recent years, microalgae industrial production has become very important in research centres, as it constitutes a solution to many of these issues. On the one hand, these microorganisms are able to remove 10 to 50 times more CO₂ than terrestrial plants (Vieira de Mendonça et al., 2021), resulting in a fixation of approximately 1.83 kg of CO₂ per kg of biomass, thanks to autotrophic growth (Chisti, 2007). This makes them a promising solution for greenhouse gases assimilation. On the other hand, in order to grow, they need nutrients such as phosphorus

and nitrogen, which happen to be the main contaminants of wastewater. Therefore, if microalgae are cultivated using effluents, they become a suitable solution for wastewater treatment, reducing the amount of pollutants in it and purifying it (Vieira de Mendonça et al., 2021).

Microalgae can be grown in closed or open reactors. The former are characterised by a physical barrier between the culture and the environment, thus preventing contamination of the culture with external agents. However, these types of reactors present high operation costs, elevated energy consumption and are difficult to scale up. Open reactors are large, shallow ponds, which contribute to better light penetration and therefore to an increase in productivity. The most extended ones among open systems are raceway reactors, which are object of this paper. They cover 90% of the total production worldwide, require a low initial investment, have a low energy consumption and are easy to scale up (Guzmán et al., 2020). One of the most important variable in this process is pH, given that it directly affects the microalgae growth, determining the solubility and availability of carbon and essential nutrients (Juneja et al., 2013; Nordio et al., 2023). It presents a highly varying dynamic and a great interaction with the rest of variables of the process, resulting in one of the most complex ones to model and control. It is mainly controlled with CO₂ injection, usually with on/off controllers with hysteresis that do not take into account the dynamics

[★] Project co-funded by the European Union – Next Generation Eu - under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 1 Investment 4.1 - Call for tender No. 2333 (22nd December 2023) of Italian Ministry of University and Research; Concession Decree No. 118 (2nd March 2023) adopted by the Italian Ministry of University and Research, Project code D93C23000450005, within the Italian National Program PhD Programme DAuSy; Spanish Ministry of Science: grant PID2023-150739OB-I00.

or the presence of disturbances, given the complexity of developing an accurate but simple model.

In the literature there exist first principle models derived from the theoretical knowledge. As shown in (Nordio et al., 2024; Solimeno et al., 2019; Fernández et al., 2017), they present the capability to provide an accurate description of the system’s internal behavior. They can be adopted as models for the so-called software sensors (García-Mañas et al., 2019), for high-level optimization, and for simulation. However, these types of models require a large amount of informative data to be calibrated due to their high amount of unknown parameters. They need more elaborated techniques to be handled compared with simpler data-driven models. Therefore, these types of models are not suitable for the design of low-level controllers.

Data-driven linear model representations (Rodríguez-Miranda et al., 2020; Carreño-Zagarra et al., 2019) are also present in the literature. They are usually identified through on-off relay experiments around an operating point. These kinds of models require less data and are easier to identify than first principle models, but, at the same time, they provide a less accurate description of the system. Linear models are usually utilized for tuning controllers such as Proportional-Integrative-Derivative (PID) (Carreño-Zagarra et al., 2019) and predictive controllers (Pawlowski et al., 2015). There are also nonlinear data-driven models in the literature, as shown in (Paladino et al., 2022; Caparroz et al., 2024; Pawlowski et al., 2019) which utilize different approaches to model the system such as nonlinear autoregressive models with exogenous input (NARX), linear models combined with regression trees and Wiener models. In most cases, the identification of these models is carried out through the use of on-off relay or open-loop experiments. The first type of experiment lacks the concept of persistent excitation due to the poor frequency content, instead, in the case of open-loop experiments they have to be more conservative due to the integrator-like dynamics present in the system, as shown in (Banerjee et al., 2024). Moreover, since the system is nonlinear, it is important to obtain data from different operating points to be able to capture the nonlinear behavior.

Therefore, in this paper, we aim to introduce an approach that can provide a persistent excitation while ensuring the maximization of the operating points and avoiding violation of the output constraint. To achieve that, we will design a suitable multisine signal in a closed-loop scheme utilizing a range controller. Experimental results are presented to show the capabilities of the proposed idea. The paper is organized as follows: section 2 describes the material and the methods adopted, section 3 presents the application of the described method and the obtained results, and section 4 draws the conclusions and outlines future works.

2. MATERIALS AND METHODS

2.1 Raceway reactor

The proposed algorithm was implemented in a raceway reactor located at the IFAPA research center, under agreement with the University of Almería (Almería, Spain). It consists of two 40 m long channels, through which the culture flows at 0.2 m/s, joined by a 1 m wide U-shaped

curve at their end (see Fig. 1). It presents a depth of 30 cm, although it is operated at 15 cm since it has been proven to be the optimal height (González Hernández et al., 2022). The reactor has a paddlewheel that impulses and mixes the culture, making sure all the microorganisms receive sunlight through the channel and, located right after, a 2 m depth sump where the CO₂ and air injection takes place for pH and dissolved oxygen (DO) control, respectively. The system is equipped with multiple sensors registering variables related with both the reactor and climatic conditions. In particular, this study focuses on the identification of the relation between solar irradiance, CO₂ injection and the pH, the sensor utilized to measure the pH level is localized 60 m after the CO₂ injection.

The produced strain is *Scenedesmus*, suitable for outdoor production under the environmental conditions presented in Southern Spain.



Fig. 1. Raceway reactor located at IFAPA research center

2.2 Experiment design

The experiment aims to provide an input signal to the system that can persistently excite the plant while maintaining the output inside the constraints. To achieve this objective, we have developed the scheme shown in Figure 2, which includes a multisine generator, a range controller, and the plant. The multisine generator provides the excitation while the range controller modulates the signal in amplitude multiplying it for a modulation factor based on the output signal of the plant.

Multisine generator Different types of input signals exist, as described in (Ljung, 1999), such as random binary signals, pseudo-random binary signals, and multisine. In this study, we adopted a multisine signal (scaled between 0 and 1) since it presents an analytical definition of the power spectrum and it contains smoother changes in it, avoiding the slew rate phenomena of the actuator. To design the input, the guidelines proposed by (McFarlane and Rivera, 1992) are followed. These guidelines utilize an approximate knowledge of the system dominant time

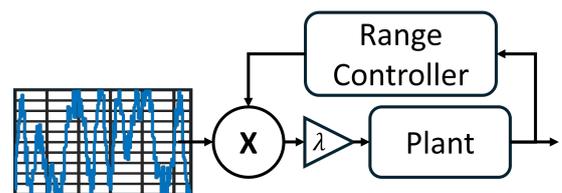


Fig. 2. Scheme of the experiment set up where on the left there is the multisine generator, on the right there are the range controller and the plant

constant to define the frequency content of the input signal, by applying the following formula:

$$\frac{1}{b \tau_{dom}^{high}} \leq \omega_{oi} \leq \frac{a}{\tau_{dom}^{low}}, \quad (1)$$

where a and b are two user-defined coefficients that determine the enlargement of the frequency content to take into account possible uncertainties in the dominant time constants first guess. Usually, they have to be kept higher or equal to 2. Then, τ_{dom}^{high} and τ_{dom}^{low} are the highest and lowest system dominant time constants, respectively. These two values are part of the *a priori* knowledge which can come from previous experiments or preliminary analysis of the system. Finally, ω_{oi} is the frequency content of the designed signal.

Another aspect to take into account while designing an input signal is the concept of persistent excitation which, according to (Ljung, 1999), relates the number of frequencies in the input signal $2n$ (signal of order n) and the number of coefficients present in the model to identify (n_b and n_f , numerator and denominator coefficients for a discrete-time model): it has to be $n = n_b + n_f - 1$.

Range controller The range controller maintains the system output within the feasibility constraints and maximizes the operating points in the experiment data in the case of nonlinear systems. This controller receives as input the plant output and computes the modulation factor. The base algorithm can be summarized as follows:

Algorithm 1 Range controller

```

if  $y_P > th_{sup}$  then
     $\mu$  increases of  $\Delta\mu$ 
else if  $y_P < th_{inf}$  then
     $\mu$  decreases of  $\Delta\mu$ 
end if
sat( $\mu$ , [0 1])

```

where y_p is the measured output of the plant, th_{sup} and th_{inf} are the superior and inferior thresholds of the controller which trigger the controller to act, and μ and $\Delta\mu$ are the modulation factor and the increment of it, respectively. The last statement of the algorithm saturates the modulation factor between 0 and 1. This is done for the sake of generality, as in this way the modulated input is forced to be always between 0 and 1. Then, to apply this method to a system it has to match the gain of the system using a further gain λ , which will be explained in more detail in section 3. This algorithm is evaluated at a fixed sampling period T_s .

The behavior of this controller is determined by the set of parameters $\Theta = \{T_s, th_{sup}, th_{inf}, \Delta\mu\}$ previously summarized. The first tuning of this controller is carried out through simulation results of a simple approximated plant model, which can be the one utilized for the input design phase. Finally, the controller requires fine-tuning through a trial-and-error procedure. The initial minimization procedure utilizes historical data of the measured disturbances to simulate different scenarios and the initial output of the model is set to a suitable value. Formally, the following optimization problem is solved:

$$\min_{\Theta} J, \quad (2)$$

subject to

$$th_{sup} - th_{inf} \geq hys_{min}, \quad (3)$$

where

$$J = w_1 \frac{N}{T_{\mu_{max}}} + w_2 \frac{N}{var} + w_3 \frac{out}{N} + w_4 \frac{1}{th_{sup} - th_{inf}}. \quad (4)$$

The cost function can be divided into four different sub-objectives: the first ($N/T_{\mu_{max}}$) is minimum when the time ($T_{\mu_{max}}$) for which the modulation factor reaches the maximum value in a simulation is maximum, where N is a normalization factor, which is the total number of samples utilized in the optimization procedure. This factor has the objective of maximizing the time interval when the modulation factor is high during a simulation. The second term (N/var) is minimum when the simulated output's variance (var) is maximum. This term aims to maximize the output signal variance, maximizing the number of operating points. The third term (out/N) minimizes the number of output instants that violate the constraints (out). The fourth term ($1/(th_{sup} - th_{inf})$) aims to simplify the minimization procedure by taking advantage of the knowledge that the variance maximization is directly proportional to the difference between the two thresholds ($th_{sup} - th_{inf}$). Each minimization term is weighted through user-defined weights w_i , $i = 1, \dots, 4$. Some guidelines for the decision on the weights are as follows. Regarding w_1 its value can be set as the average feasible value of $T_{\mu_{max}}$, w_2 can be set as the average feasible value of var , w_3 usually is set as 10^3 , and w_4 usually is set to a feasible value of the difference ($th_{sup} - th_{inf}$). The minimization is subject to the constraint (3) to ensure a minimum level of hysteresis through the use of the user-defined hys_{min} parameter. This minimization is a multi-objective non-convex problem due to its nature. In this study, we utilized a genetic algorithm to find a suitable solution to this minimization problem.

3. EXPERIMENTAL RESULTS

3.1 Experiment design

To acquire the *a priori* knowledge, we identify a First-Order-Plus-Dead-Time (FOPDT) continuous-time Multi-Inputs-Single-Output (MISO) model with an on-off relay with hysteresis experiment. The model structure is defined as follows:

$$pH(s) = \frac{K_{CO_2}}{Tp_{CO_2}s + 1} e^{-300s} CO_2(s) + \frac{K_I}{Tp_I s + 1} I(s) \quad (5)$$

where $CO_2(s)$ and $I(s)$ are the two main variables that influence the $pH(s)$ dynamics that are CO_2 and solar irradiance. The first input is manipulated while the second one is a measured disturbance. K_{CO_2} and K_I are the system gains. According to the *a priori* knowledge, the first gain has to be negative because the injection of carbon dioxide in the water produces carbonic acid which decreases the pH. Instead, the solar irradiance increases microalgae biomass production, which utilizes the carbon present in the medium, such as carbonic acid and inorganic carbon, producing oxygen increasing the pH. Then, Tp_{CO_2} and Tp_I are the system time constants, and the delay of

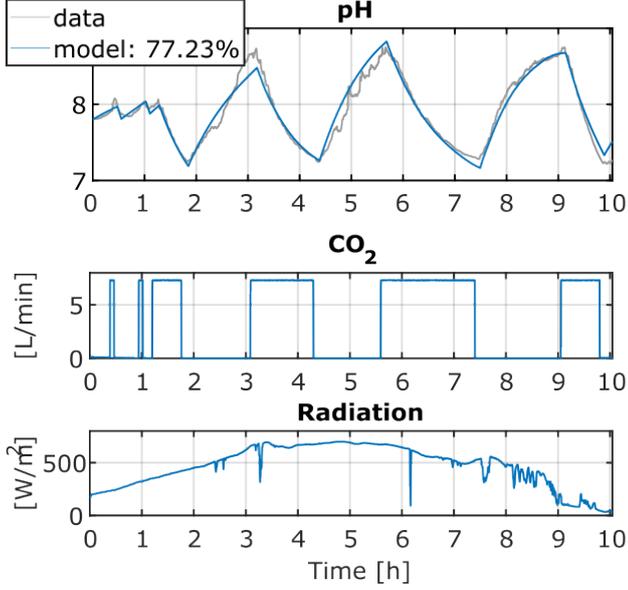


Fig. 3. On-off relay with hysteresis experiment and estimated model output. The experiment is conducted for around 10 hours.

CO_2 is due to the plant structure ($60\text{ m}/0.2\frac{\text{m}}{\text{s}} = 300\text{ s}$ as stated in subsection 2.1).

The on-off relay with hysteresis experiment is designed in order to capture an approximate dynamics of the system over the operating points of interest, which is determined by the pH constraints (7 - 9). Therefore, the hysteresis thresholds are selected to be 7.3 and 8.7, for the sake of safety. The CO_2 injection is chosen to be 7.3 L/min which is a value that ensures the pH to always decrease in any irradiance conditions. The identified linear model, through the use of the `procest` Matlab function, is the following one:

$$pH(s) = \frac{-0.32476}{3346s + 1} e^{-300s} CO_2(s) + \frac{2.253 \cdot 10^{-3}}{8183s + 1} I(s). \quad (6)$$

The dataset utilized and the result of the identification are illustrated in Figure 3, with a fit percentage of 77.23%, demonstrating the model ability to describe the data.

In the input design procedure, a multisine signal with crest factor minimization is utilized according to the guidelines outlined in subsection 2.2. In this case, the dominant time constant, according to the model presented in (6), is 3346s for the controlled input. From the simple linear model fit shown in Figure 3, we deduced that the plant can be described with a low-order model. Therefore, we selected the input signal with an order 8 of persistent excitation and we set $a = 9$ and $b = 2$. It implies that $\omega \in [1.45, 27.56] \cdot 10^{-4} \text{ rad/s}$ and, therefore, the minimum sampling frequencies, according to (Ljung, 1999), should be equal to $55.12 \cdot 10^{-4} \text{ rad/s}$, which implies a maximum sampling period of 1140s. However, we decided to use a sampling period of 60s to be more descriptive and also to be able to represent the known delay of 300s in an integer number of time instants. The resulting signal has a period of 12h and it is shown in Figure 4.

The range controller was tuned following the procedure

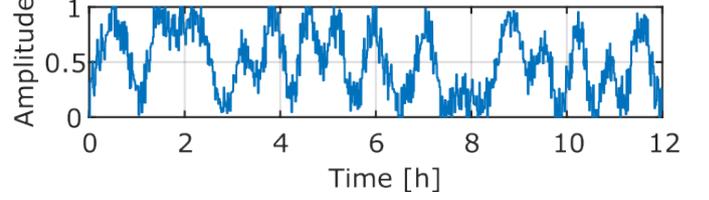


Fig. 4. Multisine signal designed with the following characteristics $T_s = 60\text{ s}$, $T = 12\text{ h}$, $w \in [1.45, 27.56] \cdot 10^{-4} \text{ rad/s}$

Table 1. Genetic algorithm parameters bounds

	T_s	th_{sup}	th_{inf}	$\Delta\mu$	λ
Lower bound	1000 s	7.2 pH	8.0 pH	0.2	10 L/min
Upper bound	4000 s	8.0 pH	8.7 pH	1.0	14 L/min

described in subsection 2.2 using the linear model (6) to simulate the plant and the data from September, October, and November 2022 of solar irradiance for a total of 30 days. Applying this procedure, we introduced the scaling factor λ as tuning parameter, which is responsible for scaling the input signal from amplitude 1 to a suitable value. It is expressed in L/min. Therefore, now the tuning parameter are $\bar{\Theta} = \{T_s, th_{sup}, th_{inf}, \Delta\mu, \lambda\}$. For the sake of implementation, the modulation factor was set to zero during the night, that is, from when the solar radiation is lower than 20 W/m^2 until it exceeds 50 W/m^2 . This last adjustment was done because the pH does not increase during the night due to the lack of solar irradiance. The weighting factors of the cost function were selected as $[w_1, w_2, w_3, w_4] = [21.6 \cdot 10^3, 1, 5.55 \cdot 10^{-3}, 10]$. The used parameters bounds are shown in Table 1 and the minimum hysteresis factor (hys_{min}) was set to be equal to 0.4 based on *a priori* knowledge. The tuning parameter determined by the genetic algorithm are $\bar{\Theta} = \{1550\text{ s}, 7.21, 8.68, 0.864, 11.666\text{ L/min}\}$. The tuning parameters were then finetuned through a trial and error procedure applying them to the plant due to the model mismatch of the real plant and the linear model. The lower and upper thresholds were reduced from 7.21 - 8.68 to 7.5 - 8.5 for the sake of avoiding undesired constraint violations.

The range controller ran for four days consecutively to collect enough data for the identification procedure since only 8 hours out of 24 hours in a day are available to excite the CO_2 injection signal without violating the lower constraint of 7. Furthermore, the usual aim of the identified model for control purposes is to describe the process dynamics during biomass production, which is performed thanks to daylight. Instead, during the night time microalgae do not perform photosynthesis so that the pH of the medium remains constant. The data from a day of the four days experiment are shown in Figure 5. From the plot of pH it is possible to notice that the controller ensures the output to not violate the constraints determined by the two black dashed lines while ensuring a good variety of operating points present in the collected data. Furthermore, from the CO_2 injection plot it is possible to observe the multisine signal modulated according to the control algorithm.

3.2 Identification results

A Wiener model was chosen to demonstrate the effectiveness of the excitation method. The Wiener model is

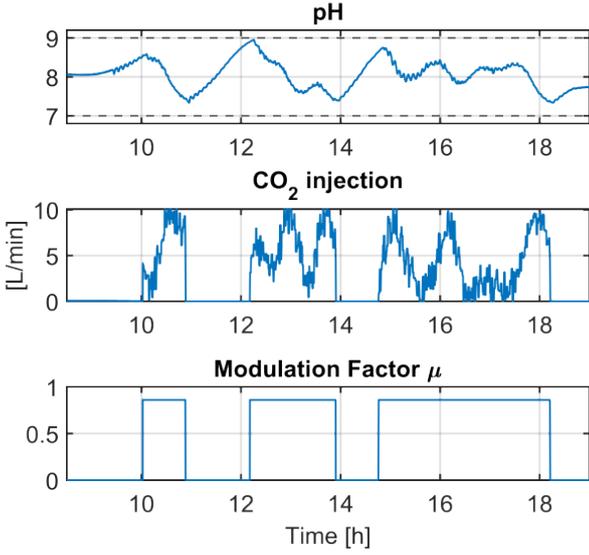


Fig. 5. Data collected in a day of October 2024. The first panel shows the pH (blue) and the output constraints (black dashed), the second panel CO_2 injection, and the third panel modulation factor

defined as a block-oriented model that is composed of linear dynamics and static nonlinearities (Ljung, 1999). The linear model can be of any kind such as PEM or state-space and the nonlinearities usually are sigmoidal, polynomial, or piecewise affine linear functions.

To perform the identification we utilized MATLAB Identification Toolbox. We divided the dataset into two parts: three days for the identification procedure and one for the validation.

In this study, we adopted as linear model a discrete-time MISO Output Error (OE) model, which belongs to the PEM family (Ljung, 1999), and it is defined as follows:

$$y(t) = M(q)U(t) + e(t), \quad (7)$$

$$M(q) = \begin{bmatrix} \frac{B_1(q)}{F_1(q)} & \dots & \frac{B_{n_u}(q)}{F_{n_u}(q)} \end{bmatrix} \quad (8)$$

$$U(t) = [u_1(t - n_{k,1}T) \dots u_{n_u}(t - n_{k,n_u}T)]^T \quad (9)$$

$$B_i(q) = b_{i,1} + b_{i,2}q^{-1} + \dots + b_{i,n_{b,i}}q^{-n_{b,i}+1}, \quad (10)$$

$$F_i(q) = 1 + f_{i,1}q^{-1} + \dots + f_{i,n_{f,i}}q^{-n_{f,i}}. \quad (11)$$

where q is the forward shift operator, $i = 1, \dots, n_u$ indicates the considered input where n_u is the number of inputs, $n_{b,i}$, $n_{f,i}$, and $n_{k,i}$ are called the model order because define numerator and denominator order of the system model ($n_{b,i}$, $n_{f,i}$) and the number of delay instant ($n_{k,i}$), $y(t)$ is the output of the system, $u_i(t - n_{k,i}T)$ is the delayed input of $n_{k,i}$ instants of time length T , and $e(t)$ is considered to be a zero mean white noise. This type of modelization permits to identify the system model $\tilde{p}(q)$ and the error model $\tilde{p}_e(q)$ as follows:

$$\tilde{p}(q) = \begin{bmatrix} \frac{B_1(q)}{F_1(q)} & \dots & \frac{B_{n_u}(q)}{F_{n_u}(q)} \end{bmatrix}, \quad \tilde{p}_e(q) = 1. \quad (12)$$

A sampling period of $T = 60$ s was considered. The OE order is determined by considering: *a priori* knowledge, the analysis of the model residual, and parsimony, which implies that the number of parameters to identify should

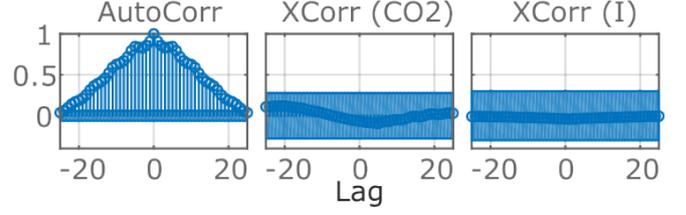


Fig. 6. Residual auto-correlation and cross-correlation evaluation for order selection of OE model from the consider experiment design, the dots are the data, instead the blue area is the confidence bounds

Table 2. OE order

	n_b	n_f	n_k
CO_2	2	4	5
I	2	2	1

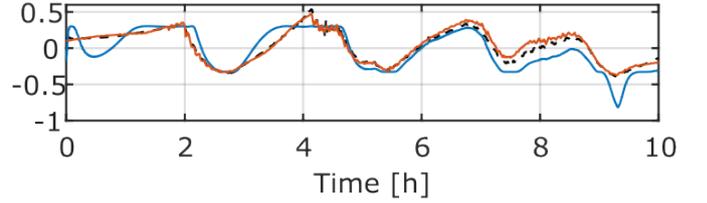


Fig. 7. Validation of the Wiener model identified with range controller data (in orange, fit percentage: 80.20%), relay data (in blue, fit percentage: 28.67%), and the validation data (in black dashed line). The data presented are subtracted by 8 and standardized

be as low as possible (in fact, by increasing them, the model can better fit the data, but it can be because of overfitting; furthermore, the accuracy of each parameter decreases). Thus, the n_k parameters of the order for the two inputs are selected based on the *a priori* knowledge of the delay, which is 300 s in the case of the CO_2 and no delay in the case of the solar irradiance, therefore $n_k = [5 \ 1]$. From the residual shown in Figure 6, it is possible to observe that the model of the system $\tilde{p}(q)$ well describes the system itself, as the cross-correlation between the input signal and the output error is approximable to zero. It means that all the linear correlation is described by the model.

The resulting nonlinearity is a cubic spline defined as follows:

$$f(r) = c_0 + c_1 r + c_2 |k_1 - r|^3 + \dots + c_{n+1} |k_n - r|^3, \quad (13)$$

where c_i $i = 0, \dots, n+1$ are the coefficients that multiply the different terms, k_i $i = 1, \dots, n$ are the nodes of the cubic spline where each term $|k_i - r|^3$ is equal to zero, and n is the number of nodes which is called the order of the cubic spline. The values of the order selected for the considered dataset are presented in Table 2 and the validation results are presented in Figure 7.

It turns out that the identification procedure reaches a satisfactory fit percentage on the validation data. Finally, for the sake of comparison, we identified the same model on on-off relay experiments to confirm the efficacy of the approach and the improvement introduced. The experiment characteristics are the same as the ones presented in Figure 3. We collected data for three days and validated

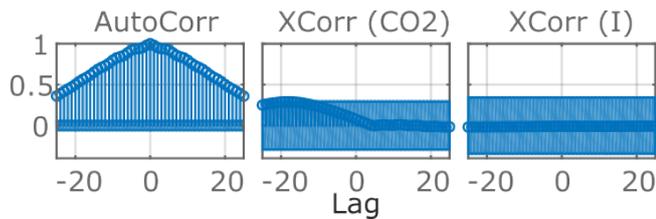


Fig. 8. Residual auto-correlation and cross-correlation evaluation for order selection of OE model from the on-off relay experiment, the dots are the data, instead the blue area is the confidence bounds.

the model on the same validation data previously used for comparison.

The procedure utilized to identify the Wiener model is the same as presented in the previous subsection. The OE order utilized is shown in Table 2. The validation results are shown in Figure 7, and from these results it can be observed that the obtained fitting percentage is lower than the one obtained with the range controller data.

4. CONCLUSIONS

In conclusion, we can state that the experiment design procedure achieves the desired objectives: the constraints on the output are ensured thanks to the range controller which also captures a wide variety of operating points in the dataset, the excitation of the model has a frequency focus on the dominant dynamics and it presents the persistent excitation property. Furthermore, the identification results, including the comparison with the standard approach, confirm the efficacy of this approach. Future improvements will include the extension of this work to non-linear MIMO systems and further generalization of this approach regarding robustness concerns.

REFERENCES

- Banerjee, S., Otálora, P., El Mistiri, M., Khan, O., Guzmán, J.L., and Rivera, D.E. (2024). Control-relevant input signal design for integrating processes: Application to a microalgae raceway reactor. *IFAC-PapersOnLine*, 58(15), 360–365. 20th IFAC Symposium on System Identification SYSID 2024.
- Caparroz, M., Guzmán, J.L., Berenguel, M., and Ación, F. (2024). A novel data-driven model for prediction and adaptive control of pH in raceway reactor for microalgae cultivation. *New Biotechnology*, 82, 1–13.
- Carreño-Zagarra, J., Guzmán, J.L., Moreno, J.C., and Villamizar, R. (2019). Linear active disturbance rejection control for a raceway photobioreactor. *Control Engineering Practice*, 85, 271–279.
- Chisti, Y. (2007). Biodiesel from microalgae. *Biotechnology Advances*, 25(3), 294–306.
- Fernández, I., Guzmán, J.L., Berenguel, M., and Ación, F.G. (2017). *Dynamic Modeling of Microalgal Production in Photobioreactors*, 49–87. Springer Singapore, Singapore.
- García-Mañas, F., Guzmán, J.L., Berenguel, M., and Ación, F. (2019). Biomass estimation of an industrial raceway photobioreactor using an extended kalman filter and a dynamic model for microalgae production. *Algal Research*, 37, 103–114.
- González Hernández, J., Rodríguez Miranda, E., Guzmán Sánchez, J.L., Ación Fernández, F.G., and Visioli, A. (2022). Temperature optimization in microalgae raceway reactors by depth regulation. *Revista Iberoamericana de Automática e Informática Industrial*, 19(2), 164–173.
- Guzmán, J.L., Ación, F.G., and Berenguel, M. (2020). Modelling and control of microalgae production in industrial photobioreactors. *Revista Iberoamericana de Automática e Informática industrial*, 18(1), 1–18.
- Juneja, A., Ceballos, R.M., and Murthy, G.S. (2013). Effects of environmental factors and nutrient availability on the biochemical composition of algae for biofuels production: A review. *Energies*, 6(9), 4607–4638.
- Ljung, L. (1999). *System Identification Theory For the User*. PTR Prentice Hall, New Jersey, second edition.
- McFarlane, R. and Rivera, D. (1992). Identification of distillation systems. *Practical Distillation Control*, 96–139.
- Nordio, R., Rodríguez-Miranda, E., Casagli, F., Sánchez-Zurano, A., Guzmán, J.L., and Ación, G. (2024). Abaco-2: a comprehensive model for microalgae-bacteria consortia validated outdoor at pilot-scale. *Water Research*, 248, 120837.
- Nordio, R., Viviano, E., Sánchez-Zurano, A., Hernández, J.G., Rodríguez-Miranda, E., Guzmán, J.L., and Ación, G. (2023). Influence of pH and dissolved oxygen control strategies on the performance of pilot-scale microalgae raceways using fertilizer or wastewater as the nutrient source. *Journal of Environmental Management*, 345, 118899.
- Paladino, O., Neviani, M., Ciancio, D., and De Francesco, M. (2022). Prediction of pH and microalgae growth in mixotrophic conditions by nonlinear black-box models for control purposes. *Biomass Conversion and Biorefinery*, 1–21.
- Pawlowski, A., Guzmán, J.L., Berenguel, M., and Ación, F. (2019). Control system for pH in raceway photobioreactors based on wiener models. *IFAC-PapersOnLine*, 52(1), 928–933. 12th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems DYCOPS 2019.
- Pawlowski, A., Mendoza, J.L., Guzmán, J.L., Berenguel, M., Ación, F.G., and Dormido, S. (2015). Selective pH and dissolved oxygen control strategy for a raceway reactor within an event-based approach. *Control Engineering Practice*, 44, 209–218.
- Rodríguez-Miranda, E., Guzmán, J.L., Berenguel, M., Ación, F., and Visioli, A. (2020). Diurnal and nocturnal pH control in microalgae raceway reactors by combining classical and event-based control approaches. *Water Science and Technology*, 82(6), 1155–1165.
- Solimeno, A., Gómez-Serrano, C., and Ación, F.G. (2019). Bio_algae 2: improved model of microalgae and bacteria consortia for wastewater treatment. *Environmental Science and Pollution Research*, 26, 25855–25868.
- Vieira de Mendonça, H., Assemany, P., Abreu, M., Couto, E., Maciel, A.M., Duarte, R.L., Barbosa dos Santos, M.G., and Reis, A. (2021). Microalgae in a global world: New solutions for old problems? *Renewable Energy*, 165, 842–862.