



Modeling the effect of temperature on ethanol fermentation using *Saccharomyces cerevisiae* CSI-1 by genetic programming

Cirilo Nolasco-Hipólito^a • Octavio Carvajal-Zarrabal^b • Jesús Carrillo-Ahumada^{a*} • Miguel Ángel Morales-Mora^c • Tan Yie-Hua^d • Fabiola Hernández-Sánchez^a • Miguel Ángel García-Muñoz^a • Mohammad Omar Abdullah^e

^aInstitute of Biotechnology. Universidad del Papaloapan. Tuxtepec, Oaxaca. Mex.

^bBiochemistry and Nutrition Chemistry Area, Universidad Veracruzana, Boca del Río, Veracruz, Mexico

^cColegio de Puebla. Dirección Académica. Puebla, México

^dDepartment of Chemical Engineering, Curtin University, Malaysia, Sarawak, Malaysia

^eDepartment of Chemical Engineering & Energy Sustainability, Faculty of Engineering, University Malaysia Sarawak (UNIMAS), Kota Samarahan, Sarawak, Malaysia

Received 06 02 2022; accepted 11 07 2022

Available 10 31 2023

Abstract: In this study, genetic programming (GP) was the systematic method used to identify the structure and parameters of a mathematical model for ethanol fermentation. The mathematical model simulated the effect of temperature on the kinetics of batch ethanol fermentation and helped to find out the optimum temperature for better performance of the process. *Saccharomyces cerevisiae* CSI-1 growing in cane molasses-based media was the microorganism used in all the experiments. Achieving the model's precision in describing the experimental observations involved the estimation of its structure (non-linear principally) and its constant parameters. The model found describes the fermentation kinetics and showed a fair prediction for dry cell weight (DCW), colony forming units/mL (CFU/mL), residual sucrose (RS), residual glucose (RG), and ethanol concentration (E). The model was used to optimize the operating conditions of the process. The predictions from the model in terms of mean square error (MSE) and sum squared error (SSE) fitted the experimental data well with fitness values in a range of $R^2 \geq 0.92$

Keywords: Ethanol fermentation, genetic programming, modeling, *Saccharomyces cerevisiae*, temperature effect, optimization

*Corresponding author.

E-mail address: jcarrillo@unpa.edu.mx (Jesús Carrillo-Ahumada).

Peer Review under the responsibility of Universidad Nacional Autónoma de México.

1. Introduction

For ethanol fermentation to yield maximum results, it is crucial to optimize the operating conditions of the process (Asaithambi et al., 2021; Ciesielski & Grzywacz, 2021; Chen et al., 2021; Estevão et al., 2021; Gao et al., 2022; Li et al., 2021; Rodman & Gerogiorgis, 2020; Shen et al., 2021; Urtubia et al., 2021). Among the key parameters, fermentation temperature plays a significant role in achieving high productivity. As ethanol production involves an exothermic reaction, it impacts the metabolism of the microorganism employed. While tropical regions naturally favor heat, controlling the fermentation temperature within the optimal range of 25-35 °C becomes challenging and economically unviable due to the substantial energy required to prevent heat-induced inactivation of yeast cells (Banat et al., 1992). Therefore, there is a quest for robust microorganisms that can efficiently ferment substrates in hot environments while tolerating high ethanol concentrations. The cost of controlling fermentation heat increases due to cooling requirements (Liu et al., 2019). It is worth noting that industrial ethanol fermentation relies on sugars derived from starchy materials, sugarcane juice, molasses (Lopes et al., 2016; McAloon et al., 2000; Samaniego-Sánchez et al., 2020), and specific substrates for distilled alcoholic beverages (Solís-García et al., 2017). Determining feasible products based on available raw materials and established technologies is essential in industrial zones (Zhao et al., 2020; González-Herrera et al., 2016). In ethanol production, sugarcane addresses the raw material availability concern, but temperature remains a critical factor to be studied (Rivera et al., 2017). Modeling has been employed to elucidate cell growth in relation to temperature (Abunde et al., 2019; Nor-Khaizura et al., 2019; Pereira et al., 2020). However, the outcome of simulation and optimization tools heavily relies on the quality of the mathematical model (Carrillo-Ahumada et al., 2020; Castillo-Santos et al., 2017; Darvishi et al., 2020; Díaz & Tost, 2018; Goelzer et al., 2009; Hebing et al., 2020; Jorayev et al., 2022; Meng et al., 2021; Müller et al., 2020; Rodríguez-Mariano et al., 2015; Salmi et al., 2021; Torralba-Morales et al., 2020; Vignesh & Chandraraj, 2021; Wu et al., 2015). Typically, fermentations rely on ideal laboratory conditions (e.g., synthetic media, stirring devices, heating modes), and it is preferable to consider industrial processes (real fermentation media, steady state, etc.) (de Andres-Toro et al., 1997). Modifying processes introduces new situations, necessitating extensive experimentation to generate data for constructing novel models. Consequently, developing a phenomenological process model becomes challenging due to limited understanding of the physicochemical phenomena and associated kinetic and transport mechanisms (Cheema et al., 2002). Moreover, the nonlinear dynamics of the process fur-

ther complicate modeling (Feil et al., 2004). Given these challenges, hybrid approaches have emerged, such as genetic programming (GP), an evolutionary artificial intelligence technique for developing mathematical models based on input-output data, in contrast to conventional regression and neural network modeling techniques (Babu & Karthik, 2007). The study and modeling of ethanol fermentation processes have proven effective in improving product quality, enhancing process control, and reducing costs (Fan et al., 2015). Genetic programming (GP) has successfully been utilized to model the glucose to gluconic acid bioprocess, resulting in increased overall productivity and improved interaction between dissolved oxygen and fungal mycelia (Babu & Karthik, 2007). In this research, GP was employed to develop a mathematical model that describes the impact of temperature on microorganism growth and ethanol yield. The approach followed the work of Madár et al. (2005), which focused on identifying the structure and parameters of a mathematical model using experimental data. Abonyi (2005) developed a MATLAB toolbox for this purpose, which was utilized in this research without modification. The algorithm selected for parameter and structure identification does not require a pre-defined experimental design and produces satisfactory results with limited data for correlation, as demonstrated in the work of Ramírez-Hernández et al. (2017). This methodology has found applications in various domains such as algorithms, biotechnology, computing, process control, data mining, and modeling (Banzhaf et al., 1998; Dorgo et al., 2021; Kumar et al., 2014). Germec et al. (2020) and Esfahanian et al. (2016) conducted parameter identification based on the Gompertz equation. However, this research focuses on establishing correlations using available data specific to the alcoholic fermentation process, rather than relying on a pre-defined mathematical structure.

1.1. Model identification for the fermentation process

Mathematical models play a crucial role in various scientific disciplines as they enable the description and prediction of system behavior under different conditions by utilizing variables. In this research, the symbolic optimization algorithm known as genetic programming (GP) was employed, following the approach proposed by Madár et al. (2005). This method facilitates the identification of both the model's structure and its parameters using experimental data. The key considerations in this approach include the following:

$$J(\theta)_{\min \theta \in \mathbb{R}^n} = [J_1(\theta), J_2(\theta), \dots, J_m(\theta)] \in \mathbb{R}^m \quad (1)$$

$$J_1(\theta) = \chi^2 = \frac{1}{N} \sum_{k=1}^N (y(k) - \hat{y}(k))^2 \quad (2)$$

$$\hat{y}(k) = \sum_{i=1}^M p_i F_i(\mathbf{x}(k)) \quad (3)$$

The target vector $J(\theta) \in \mathbb{R}^m$ represents the adjustment function based on the mean square error (MSE) between the calculated data and the measured output values. Here, $\theta \in \mathbb{R}^n$ denotes the decision vector. In "Eq. 2," the parameters include N , the number of samples used for model identification; $y(k)$, the experimental output; $\hat{y}(k)$, the calculated output; k , the sample index; and χ , the vector of regression variables. In "Eq. 3," the parameters consist of nonlinear functions F_1, \dots, F_M and model parameters p_1, \dots, p_M . genetic programming (GP) is a systematic method that employs natural evolution to automatically generate algorithms and expressions for the identification of mathematical models for specific problems (Koza & Poli, 2005). These expressions are encoded as a tree data structure, with functions as nodes and terminals as leaf nodes, allowing the generation of nonlinear input-output models. An orthogonal least squares algorithm is applied to estimate the contribution of tree branches and create a precise model (Brameier & Banzhaf, 2007). To develop the appropriate model for ethanol fermentation and similar systems, the GP MATLAB™ toolbox (available at <http://www.mathworks.com/matlabcentral/fileexchange/47197-genetic-programming-matlab-toolbox>) (Abonyi, 2005) was utilized (Datta et al., 2019; Grosman & Lewin, 2004; Kummer et al., 2019). The parameters of GP are presented in "Table 1."

Table 1. Parameters of GP.

Parameter	Value
Population size	100
Maximum number of tree levels	10
Symbols	+, *, /, ^
Type of selection and selection mode	Tournament selection
Generation gap	0.8
Probability of crossover	0.7
Probability of mutation	0.3
Mode of tree-recombination	1.0

In this research, the variables $\theta = [t, T]$ are considered as decision variables. In addition to the results obtained using "Eq. 1," the following adjustment metrics are utilized: the coefficient of determination R^2 ("Eq. 4"), which describes the fit between experimental and calculated data; the root mean squared error (RMSE) ("Eq. 5"), and the sum of squared error (SSE) ("Eq. 6"). The objective is to develop a model that can elucidate the impact of temperature on ethanol fermentation productivity, particularly how temperature influences the growth of microorganisms and, consequently, the system's ethanol concentration over time.

$$R^2 = 1 - \frac{SSE}{\sum_{k=1}^N (\hat{y}(k))^2} \tag{4}$$

$$RMSE = \frac{\sqrt{SSE}}{N} \tag{5}$$

$$SSE = \sum_{k=1}^N (y(k) - \hat{y}(k))^2 \tag{6}$$

The structure of the article is the following: Section 2 shows the experimental and computational methodologies; The results and discussion are shown in Section 3. Finally, some remarks and conclusions are exposed in Section 4.

1.2. Review of multi-objective optimization design procedures

A multi-objective optimization statement without loss of generality is defined as follows:

$$J(\theta)_{\min \theta \in \mathbb{R}^n} = [J_1(\theta), \dots, J_m(\theta)] \in \mathbb{R}^m \tag{7}$$

subject to: $g(\theta) \leq 0, h(\theta) \leq 0$ and $\underline{\theta}_i \leq \theta_i \leq \bar{\theta}_i$ with $i = [1, \dots, n]$. Where $\theta \in \mathbb{R}^n$ is defined as the decision vector, $J(\theta) \in \mathbb{R}^m$ as the objective vector and $g(\theta), h(\theta)$ as the inequality and equality constraint vectors, respectively; $\underline{\theta}_i, \bar{\theta}_i$ correspond to the lower and upper bounds in the decision space.

Since there is no single solution that is optimal for all objectives, a set of solutions called the Pareto set is defined. Each solution in the Pareto set represents an objective vector on the Pareto front. All solutions on the Pareto front are considered a set of Pareto-optimal and non-dominated solutions.

A design procedure employing multi-objective optimization techniques typically consists of three fundamental steps: 1) stating the multi-objective problem (MOP), 2) conducting the multi-objective optimization (MOO) process, and 3) performing the multi-criteria decision making (MCDM) stage (Meza et al., 2017).

- MOP statement

At this stage, the designer must make decisions regarding the design concept to address the problem, how to evaluate the performance of design alternatives, and which solutions are relevant, practical, or feasible. In the case of ethanol fermentation, the design concept refers to the operating conditions, while the design alternative pertains to specific time and temperature settings. Performance measurement requires the existence of a parametric model that establishes a correlation between the decision variables (which lead to specific design alternatives) and their performance

- MOO process

During this stage, the multi-objective optimization algorithm is implemented for the multi-objective problem (MOP). The algorithm can be ad-hoc or selected from a suitable pool of algorithms available. An algorithm is considered suitable for the problem at hand if it possesses desirable characteristics such as convergence, diversity, and relevance.

- Decision-making stage

Finally, with the approximate Pareto front, the designer will evaluate the trade-offs between conflicting design objectives and consider the design alternatives. The goal is to select a solution that strikes a preferable balance in performance for the specific problem. Procedures and visualization tools play a crucial role in assisting designers, particularly when dealing with four or more design objectives.

2. Methodology

2.1. Microorganism and culture conditions

Saccharomyces cerevisiae CSI-1 (abbreviated as CSI) yeast was used as the microorganism in this study. A vial of CSI stock cultures, stored at -10 °C, was thawed and used to refresh the cells in tubes containing 10 mL of culture medium prepared with 20 g/L of glucose and 5 g/L of yeast extract (YE). The medium, sterilized for 20 minutes at 121°C prior to inoculation, was then incubated at 37°C for 24 h in a Shel Lab SI incubator. Subcultures were performed monthly.

2.2. Inoculum preparation and fermentation medium

The refreshed culture served as the seed to prepare the inoculum using 200 mL of growth medium composed of 30 g/L of glucose and 5 g/L of YE. The inoculum was cultivated in a Shel Lab SI incubator at a temperature of 37°C for 24 h. The pre-culture was subsequently centrifuged at room temperature using a high-speed centrifuge (Kubota model CR21G) at a speed of 6000×g for 5 min to harvest the cells. The cell pellet was then resuspended in 100 mL of sterilized water and used as the inoculum. The fermentation medium consisted of 5 g/L of YE and 100 g/L of molasses (70 g/L sucrose and 30 g/L glucose). The molasses contained a low concentration of fructose and were not monitored. The media was autoclaved for 20 min at 121°C.

2.3. Batch fermentation

The harvested cells from the inoculum were transferred to the fermenter to initiate the fermentation process. Ethanol fermentations were conducted in a 1 L fermenter with a working volume of 800 mL. The experiment was initiated at temperatures of 30°C, 33°C, 35°C, and 37°C with an agitation speed set at 100 rpm. The optical density, temperature, and pH were monitored and recorded throughout the fermentation process. Samples were withdrawn every 3 h,

except at 12 h and 15 h, as these periods corresponded to the logarithmic phase where the monitored parameters were predictable, stable, and reproducible.

2.4. Analysis

2.4.1. Cell growth analysis

The growth of CSI yeast was monitored by measuring the optical density (OD) of cells at a wavelength of 575 nm using a UV-Vis Spectrophotometer and correlated to dry cell weight.

2.4.2. Colony forming units

Colony forming units (CFU/mL) were determined using the decimal serial dilutions method (100 - 900 µL) of 10^1 - 10^7 or 10^8 in 1.5 mL tubes. An aliquot of 100 µL was spread on a PDA plate medium.

2.4.3. Ethanol, glucose, and sucrose determination

Fermentable sugars (glucose and sucrose) and ethanol were analyzed using an enzymatic method with a model BF-5D (Oji scientific instruments Co., Ltd. Japan) analyzer.

2.4.4. Computational methodology

This section presents the computational methodology used in this research work. Firstly, the experimental dynamics of the fermentation process were observed using the original experimental data. Subsequently, based on the experimental data and utilizing GP (Banat et al., 1992), a nonlinear mathematical model was identified to represent the variable responses: dry cell weight (DCW), residual glucose (RG), residual sucrose (RS), ethanol (E), and colony forming units/mL (CFU/mL) as functions of time and temperature. Then, to validate the obtained model, correlation metrics between the model and experimental data were performed. Finally, multi-objective optimization was employed to determine if there is a certain correspondence between the previously obtained data and the results provided by optimization.

3. Results and discussion

3.1. Experimental dynamics of the fermentation process

The experimental dynamics of the fermentation process are presented using the experimental characteristics DCW, RG, RS, E, and CFU/mL (see "Tables 2-6" and "Figures 1-2").

"Table 2" displays the dynamics of DCW as a function of time and temperature. The parameters that are within a limited range are DCW and CFU/mL, which show a strong correlation. It is preferable to have low values of DCW and CFU/mL to achieve high specific productivity. However, DCW or its equivalent CFU/mL correlates with the volumetric productivity, represented by ethanol concentration (E) ("Figure 1"). Ethanol (E) can be considered as a reference for analyzing the coupling of the other parameters. This is the

optimal way to track the fermentation kinetics, considering that E is the final product of the substrate metabolism. Three intervals can be identified for E ("Figure 2"): a) $E > 40 \text{ g/L}$, b) $20 \text{ g/L} < E \leq 40 \text{ g/L}$ and c) $0 \text{ g/L} < E \leq 20 \text{ g/L}$. For the first interval a) $E > 40 \text{ g/L}$, high values of DCW , low values of RG , low values of RS , and high values of CFU are required and were obtained as expected. This trend is evident since higher CFU production leads to increased ethanol production.

Table 2. Experimental data for DCW (g/L).

c	Temperature (°C)				Mean	Std. dev.
	30	33	35	37		
0	1.95	2.2	2.25	2.28	2.17	0.15
3	2.2	2.51	2.46	3.03	2.55	0.35
6	2.6	2.82	3.48	5.64	3.64	1.39
12	6.03	6.36	6.93	8.25	6.89	0.98
18	10.1	9.1	11.1	8.55	9.71	1.13
21	10.5	9.3	11.1	8.65	9.89	1.11
24	9.75	9.55	10.08	8.7	9.52	0.59

Table 3. Experimental data for RG (g/L).

Time (h)	Temperature (°C)				Mean	Std. dev.
	30	33	35	37		
0	28.0	27.3	27.0	28.0	27.58	0.51
3	22.0	21.5	18.0	21.0	20.63	1.80
6	16.0	14.0	10.0	13.8	13.45	2.51
12	6.0	1.1	1.3	1.6	2.50	2.34
18	2.0	0.8	0.5	0.08	0.85	0.82
21	1.6	0.09	0.2	0.2	0.52	0.72
24	0.2	0.0	0.1	0.0	0.08	0.10

Table 4. Experimental data for RS (g/L).

Time (h)	Temperature (°C)				Mean	Std. dev.
	30	33	35	37		
0	70.0	70.0	72.0	72.0	71.0	1.15
3	67.5	67.0	68.0	67.0	67.38	0.48
6	65.0	64.0	64.0	65.0	64.5	0.58
12	65.0	61.3	64.0	62.4	63.18	1.65
18	5.6	5.3	5.7	5.2	5.45	0.24
21	5.0	4.0	5.0	4.5	4.63	0.48
24	4.7	4.6	4.3	2.1	3.93	1.23

Table 5. Experimental data for E (g/L).

Time (h)	Temperature (°C)				Mean	Std. dev.
	30	33	35	37		
0	0.4	0.6	0.0	0.4	0.35	0.25
3	3.0	2.9	0.8	1.7	2.10	1.05
6	4.2	4.8	4.9	6.0	4.98	0.75
12	20.0	18.0	24.0	26.0	22.0	3.65
18	37.4	40.1	42.4	38.9	39.70	2.11
21	41.4	43.2	45.3	44.4	43.58	1.69
24	44.8	48.0	49.5	49.6	47.98	2.24

Table 6. Experimental data for CFU/mL (Log_{10}).

Time (h)	Temperature (°C)				Mean	Std. dev.
	30	33	35	37		
0	6.54	6.82	6.70	6.48	6.64	0.15
3	6.88	7.98	7.54	7.96	7.59	0.51
6	7.15	8.32	8.88	8.64	8.25	0.77
12	8.20	8.85	9.08	8.95	8.77	0.39
18	8.93	8.95	9.12	8.66	8.92	0.19
21	8.90	9.0	9.01	8.68	8.90	0.15
24	8.91	8.98	9.04	8.52	8.86	0.23

For the second interval (b) $20 \text{ g/L} < E \leq 40 \text{ g/L}$, high and medium values of DCW , low values of RG , high values of RS , and high values of CFU are required. In contrast, for the third interval $0 \text{ g/L} < E \leq 20 \text{ g/L}$, low and medium values of DCW , high, medium, and low values of RG , high values of RS , and high, medium, and low values of CFU are observed to be required. This trend is also reasonable because lower CFU results in lower production of E . One solution to this situation is to allow the fermentation to continue for a longer time. However, this approach is unfavorable from an industrial perspective. This situation arises since the yeast acts as a biocatalyst, consuming the substrate according to the reaction rate. Glucose is the first substrate to be consumed due to the diauxic phenomenon, and then sucrose is metabolized once the initial glucose present in the molasses is depleted. At the end of fermentation, the desired product, E , is produced consequently correlated with the consumed substrate, the concentration of cells (DCW), and the operating temperature. "Table 3" shows the dynamics of RG as a function of time and temperature. The decrease in RG is inversely proportional to time, with the most drastic results at $t > 12 \text{ h}$. There are similarities in the RG values, specifically at: (a) $RG \cong 27 \text{ g/L}$, (b) $RG = 28 \text{ g/L}$, (c) $RG = 0.2 \text{ g/L}$ and (d) $RG = 0 \text{ g/L}$.

For (a) $RG \cong 27 \text{ g/L}$, the operating conditions correspond to: $T = 33 \text{ °C}$, $t = 0 \text{ h}$ and $T = 35 \text{ °C}$, $t = 0 \text{ h}$. For (b) $RG = 28 \text{ g/L}$, the operating conditions correspond to: $T = 30 \text{ °C}$, $t = 0 \text{ h}$ and $T = 37 \text{ °C}$, $t = 0 \text{ h}$. For (c) $RG = 0.2 \text{ g/L}$, the operating conditions correspond to: $T = 30 \text{ °C}$, $t = 24 \text{ h}$ and $T = 35 \text{ °C}$, $t = 21 \text{ h}$ and $T = 37 \text{ °C}$, $t = 21 \text{ h}$. Finally, for (d) $RG = 0 \text{ g/L}$ the operating conditions correspond to: $T = 33 \text{ °C}$, $t = 24 \text{ h}$ and $T = 37 \text{ °C}$, $t = 24 \text{ h}$. The decrease in glucose concentration is due to the cell's metabolism for cell reproduction, converting glucose into ethanol as a product. Glucose is preferred over sucrose as a substrate because it is thermodynamically favorable.

"Table 4" displays the dynamics of RS as a function of time and temperature. The decrease in RS varies with time, with the most significant changes occurring at $t > 12 \text{ h}$. This characteristic shows similar values for all operating conditions. However, the operating conditions $T = 33 \text{ °C}$, $t = 12 \text{ h}$, and $T = 37 \text{ °C}$, $t = 24 \text{ h}$ do not follow the same trend as the other operating conditions. As part of the normal yeast

metabolism of glucose and sucrose, the cell undergoes diauxic shift, where glucose is preferred over sucrose (Peng et al., 2015). Only when glucose is depleted does the yeast start to consume sucrose. It is known that *Saccharomyces cerevisiae* can utilize simple sugars but are unable to use monosaccharides such as xylose. Therefore, this model does not apply when the substrate comes from lignocellulosic fermentable sugars (Nawaz et al., 2020).

"Table 5" presents the dynamics of ethanol (E) as a function of time and temperature. The increase in E is directly proportional to time, with the most notable changes occurring at $t > 12$. For the time interval at $t > 12$ h, the production of E increases twice that of at $t > 12$ h. Some operating conditions that do not present the same trend as the others are the following: (a) $E = 0$ g/L, (b) $E = 0.8$ g/L and (c) $E = 6$ g/L with $T = 35$ °C, $t = 0$ h with; with $T = 35$ °C, $t = 3$ h and $T = 37$ °C, $t = 6$ h respectively. Again, this small difference is due to slight differences in the initial inoculum to the main fermentation.

"Table 6" displays CFU/mL as a function of time and temperature. The increase in CFU/mL is directly proportional to time, but there is a stationary phase observed at $t > 12$ h. This characteristic exhibits a similar trend across all operating conditions and is often observed in various microorganisms when important nutrients or cofactors are depleted. "Figures 1-2" illustrate the interrelationship among all the fermentation process characteristics in a combined manner.

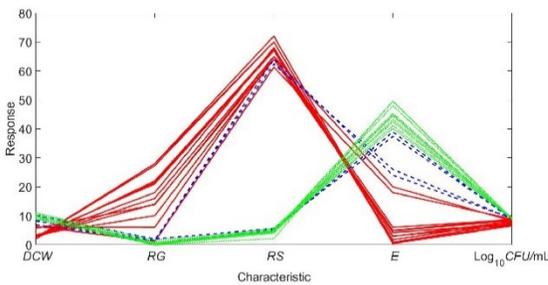


Figure 1. Coupling of the characteristics of the fermentation process. Green lines: $E > 40$ g/L; blue lines: $20 \text{ g/L} < E \leq 40$ g/L; red lines: $0 \text{ g/L} < E \leq 20$ g/L (experimental characteristics).

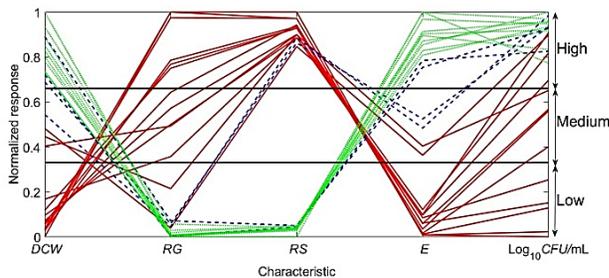


Figure 2. Coupling of the characteristics of the fermentation process. Green lines: $E > 40$ g/L; blue lines: $20 \text{ g/L} < E \leq 40$ g/L; red lines: $0 \text{ g/L} < E \leq 20$ g/L (normalized responses).

The characteristics that are in a limited range are DCW and CFU/mL , and their influence on the other characteristics is observed in the production of ethanol ("Figure 1"). Characteristic E (ethanol) can be considered as a reference to analyze the coupling of the other characteristics, for which there are three intervals ("Figure 2"): a) $E > 40$ g/L, b) $20 \text{ g/L} < E \leq 40$ g/L and c) $0 \text{ g/L} < E \leq 20$ g/L. For the first interval a) $E > 40$ g/L, high values of DCW , low values of RG , low values of RS and high values of CFU/mL are observed. For the second interval b) $20 \text{ g/L} < E \leq 40$ g/L, high and medium DCW values, low RG values, high RS values and high CFU/mL values are observed. Finally, for the third interval $0 \text{ g/L} < E \leq 20$ g/L, low and medium values of DCW , high, medium, and low values of RG , high values of RS and high, medium, and low values of CFU/mL are observed for the kinetics of the fermentation. The situation is due to the consumption of substrate by the yeast obeying the reaction rate. The first substrate consumed is glucose, then the diauxic phenomenon biases the reaction toward sucrose consumption. At the end of the fermentation, the ethanol produced as desired product is a consequence correlated with the substrate consumed and the temperature.

3.2. Model identification of the fermentation process

The experimental data were used in the GP MATLAB™ toolbox to obtain the mathematical model that describes the behavior of the fermentation process. The parameters of GP are described in "Table 1." The mathematical models are:

$$DCW = -0.001857t(T + t^2) + 0.059t^2 + 2.232 \quad (8)$$

$$RG = -\frac{28.130t^2}{(t^2 - t + T)} + 27.586 \quad (9)$$

$$RS = -61.36 \left(\frac{T}{T + \left(\frac{T-t}{t}\right)^T} \right) + 66.22 \quad (10)$$

$$E = 0.067tT \quad (11)$$

$$CFU = \begin{cases} -0.080t - \frac{305.94}{(t+T)} + 16.47 & \text{if } T \leq 33^\circ\text{C} \\ -0.224t - \frac{12.92T}{(T+2t)} + 19.56 & \text{if } T > 33^\circ\text{C} \end{cases} \quad (12)$$

where, the parameter models are: DCW is dry cell weight, CFU is colony forming units, RG is residual glucose, RS is residual sucrose, E is ethanol concentration, t is time and T is temperature considering the following restrictions: if $DCW < 0$ then $DCW = 0$; if $RG = 0$; if $RS < 0$ then $RS = 0$; if $E < 0$ then $E = 0$ and $CFU/mL < 0$ then $CFU/mL = 0$. The mathematical models are compared with the experimental data ("Figure 3"), and it can be observed that the mathematical model fitted the experimental data values fairly.

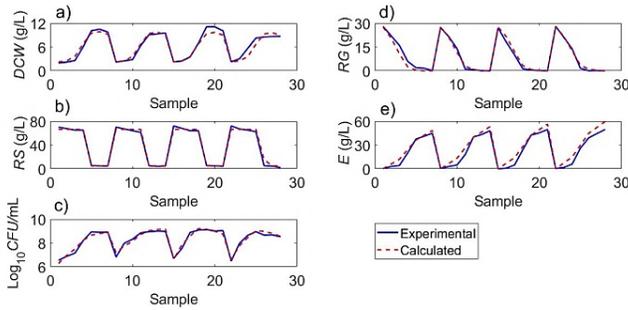


Figure 3. Validation of the mathematical model vs. experimental data: a) DCW (g/L); b) RS (g/L); c) $\log_{10}CFU/mL$; d) RG (g/L); e) E (g/L); experiments were performed by triplicate.

The fitness is higher than $R^2 \leq 0.92$ as reported in "Table 7." Therefore, "Eqs. 7-11" adequately represent the experimental data and are used to optimize the characteristics of the fermentation process.

Table 7. Fit Indices of experimental values.

Characteristic	R^2	RMSE	SSE
DCW	0.93	0.82	19.14
RG	0.97	1.74	84.78
RS	0.98	3.08	267.15
E	0.92	5.39	814.45
$\log_{10}CFU/mL$	0.94	0.19	1.11

3.3. Statement of optimization problem

In this research work, it is considered that the ethanol fermentation process presents the following characteristics: (DCW) g/L, (RG) g/L, (RS) g/L, (E) g/L and ($\log_{10}CFU/mL$). These characteristics are considered as dependent variables of the process. The independent variables of the process are time (t) and temperature (T). The relationship between the dependent and independent variables is shown in "Eqs 8-12". Some bibliographic references that consider a methodology like this work is Ramírez-Hernandez et al. (2017) where they consider:

a) set of data obtained experimentally from the dependent variables of independent, b) identification of the structure and parameters of the model and finally c) optimization in numerical simulation of the best operating conditions. For point c) optimization in numerical simulation of the best operating conditions in this research work, the following is considered. The multiobjective optimization problem can be posed by finding the values of X_1 and X_2 with $\mathbf{J}(\theta) \in \mathbb{R}^6 = [J_1, J_2, J_3, \frac{1}{J_3}, J_4, J_5]$ therefore:

$$\min \mathbf{J} \tag{13}$$

where θ is time (X_1) and temperature (X_2) subject to the decision variables are $0 < X_1 < 24$ and $30 < X_2 < 37$. The

definition of the decision variables allows the optimization algorithm to define the search space for potential solutions. In this work, the search space was determined with respect to a previous experimentation ("Tables 2-6"). To obtain potential solutions the MATLAB™ optimtool/gamultiobj (multiobjective optimization using genetic algorithm) toolbox was used.

3.4. Optimization

"Figures 4a-4b" present the operating conditions and corresponding characteristics obtained from the model and experimental data, showing a correspondence between them. In contrast, "Figures 4c-4d" display the set and Pareto front of operating conditions obtained through the optimization algorithm, along with the characteristics obtained through simulation. Additionally, the maximum values obtained from both experiments and simulation are also indicated.

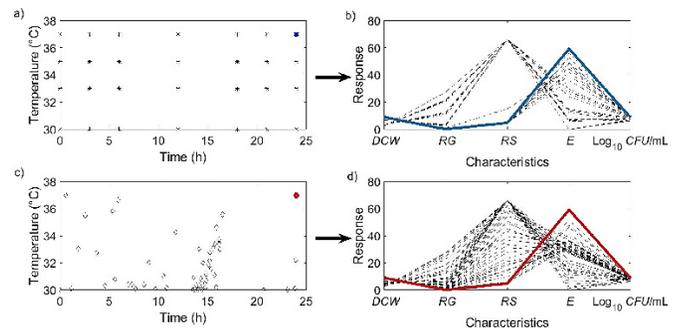


Figure 4. Set and Pareto front of the optimization process: a) Independent variables (t, T) proposed experimentally, b) dependent variables (DCW, RG, RG, E and $\log_{10}CFU/mL$) obtained with the mathematical model; c) independent variables (t, T) proposed by means of optimization, d) dependent variables (DCW, RG, RG, E and $\log_{10}CFU/mL$) obtained with the mathematical model and c). Blue lines indicate the maximum of E.

It is observed that there is a highly aggregated set in the region corresponding to a time interval of 10 to 15 h, but the maximum value is found with the operating conditions that meet $t > 37^\circ C$ and $t > 22$ h. One advantage of simulation is that it allows us to observe operating conditions that could lead to better performance of the process, which may include regions that were not explored in experimentation. The model was evaluated through experimentation and yielded satisfactory results.

These results precisely reveal the desired conditions for an exothermic process like ethanol fermentation. Operating ethanol fermentation above $37^\circ C$ is crucial for tropical countries due to its significant economic advantages. The advantage lies in the fact that if fermentation is conducted at temperatures higher than $37^\circ C$, it becomes easier to control using heat exchangers, especially considering that the cooling water temperature in tropical countries typically ranges from

30-35°C. Moreover, operating at temperatures above 37°C is also an economic advantage for processes with a duration of less than 24 h, as it reduces energy consumption required for cooling equipment. Operating at 37°C is a characteristic of the selected microorganism, which can produce ethanol even at temperatures as high as 45°C. However, operating at such elevated temperatures is not advisable as it negatively affects the viability of the microorganism. Maintaining viability is crucial to reuse the microorganism for multiple cycles of operation.

4. Conclusions

The values calculated with the mathematical model align well with the experimental data. Thus, the models can be effectively applied to predict the kinetics of the fermentation process, including cell growth and ethanol production from sugarcane molasses across a range of temperatures, demonstrating the accuracy of the proposed model. Finally, the optimization of ethanol fermentation characteristics highlights the identification of feasible operating conditions.

Conflict of interest

The authors do not have any type of conflict of interest to declare.

Acknowledgements

The authors would like to thank Ayaaki Ishizaki, Emeritus Professor of Kyushu University to allow the use of the strain *Saccharomyces cerevisiae* CSI-1 a property of his company Necfer corporation, Japan.

Funding

The authors received no specific funding for this work.

References

- Abonyi J. (2005). Genetic Programming MATLAB Toolbox, *MATLAB Central File Exchange*. Retrieved <https://www.mathworks.com/matlabcentral/fileexchange/47197-genetic-programming-matlab-toolbox>
- Abunde, N.F., Asiedu, N. Y., & Addo, A. (2019). Modeling, simulation and optimal control strategy for batch fermentation processes. *International Journal of Industrial Chemistry, 10*, 67-76. <https://doi.org/10.1007/s40090-019-0172-9>
- Asaithambi, N., Singh, S. K., & Singha, P. (2021). Current status of non-thermal processing of probiotic foods: A review. *Journal of Food Engineering, 303*, 110567. <https://doi.org/10.1016/j.jfoodeng.2021.110567>
- Babu, B. V., & Karthik, S. (2007). Genetic Programming for Symbolic Regression of Chemical Process Systems. *Eng. Lett., 14*(2), 42-55.
- Banat, I. M., Nigam, P., & Marchant, R. (1992). Isolation of thermotolerant, fermentative yeasts growing at 52 C and producing ethanol at 45 C and 50 C. *World Journal of Microbiology and Biotechnology, 8*, 259-263. <https://doi.org/10.1007/BF01201874>
- Banzhaf, W., Nordin, P., Keller, R. E., & Francone, F. D. (1998). *Genetic programming: an introduction: on the automatic evolution of computer programs and its applications*. Morgan Kaufmann Publishers Inc.
- Brameier, M. F., & Banzhaf, W. (2007). A comparison with tree-based genetic programming. *Linear Genetic Programming, 173-192*. https://doi.org/10.1007/978-0-387-31030-5_8
- Carrillo-Ahumada, J., Reynoso-Meza, G., Ruiz-López, I. I., & García-Alvarado, M. A. (2020). Analysis of open-loop and L2/D controlled closed-loop behavior of the Cholette's bioreactor under different operating conditions. *ISA transactions, 101*, 147-159. <https://doi.org/10.1016/j.isatra.2020.01.039>
- Castillo-Santos, K., Ruiz-López, I. I., Rodríguez-Jimenes, G. C., Carrillo-Ahumada, J., & García-Alvarado, M. A. (2017). Analysis of mass transfer equations during solid-liquid extraction and its application for vanilla extraction kinetics modeling. *Journal of Food Engineering, 192*, 36-44. <https://doi.org/10.1016/j.jfoodeng.2016.07.020>

- Cheema, J. J. S., Sankpal, N. V., Tambe, S. S., & Kulkarni, B. D. (2002). Genetic programming assisted stochastic optimization strategies for optimization of glucose to gluconic acid fermentation. *Biotechnology progress*, 18(6), 1356-1365.
<https://doi.org/10.1021/bp015509s>
- Chen, Z., Niu, Y., Chen, C., & Li, H. (2021). Optimization of bioethanol fermentation productivity in *Saccharomyces cerevisiae* by regulation of social behavior. *Chemical Engineering Science*, 246, 116980.
<https://doi.org/10.1016/j.ces.2021.116980>
- Ciesielski, A., & Grzywacz, R. (2021). Process maps with metabolic constraints for bioethanol production by continuous fermentation. *Chemical Engineering Science*, 229, 116134.
<https://doi.org/10.1016/j.ces.2020.116134>
- Darvishi, H., Farhudi, Z., & Behrooz-Khazaei, N. (2020). Multi-objective optimization of savory leaves drying in continuous infrared-hot air dryer by response surface methodology and desirability function. *Computers and electronics in agriculture*, 168, 105112.
<https://doi.org/10.1016/j.compag.2019.105112>
- Datta, S., Dev, V. A., & Eden, M. R. (2019). Developing non-linear rate constant QSPR using decision trees and multi-gene genetic programming. *Computers & Chemical Engineering*, 127, 150-157.
<https://doi.org/10.1016/j.compchemeng.2019.05.013>
- de Andres-Toro, B., Girón-Sierra, J. M., Lopez-Orozco, J. A., & Fernandez-Conde, C. (1997). Optimization of a batch fermentation process by genetic algorithms. *IFAC Proceedings Volumes*, 30(9), 179-184.
[https://doi.org/10.1016/S1474-6670\(17\)43157-0](https://doi.org/10.1016/S1474-6670(17)43157-0)
- Díaz, V. H. G., & Tost, G. O. (2018). Economic optimization of in situ extraction of inhibitors in acetone-ethanol-butanol (ABE) fermentation from lignocellulose. *Process biochemistry*, 70, 1-8.
<https://doi.org/10.1016/j.procbio.2018.04.014>
- Dorgo, G., Kulcsar, T., & Abonyi, J. (2021). Genetic programming-based symbolic regression for goal-oriented dimension reduction. *Chemical Engineering Science*, 244, 116769.
<https://doi.org/10.1016/j.ces.2021.116769>
- Esfahanian, M., Rad, A. S., Khoshhal, S., Najafpour, G., & Asghari, B. (2016). Mathematical modeling of continuous ethanol fermentation in a membrane bioreactor by pervaporation compared to conventional system: Genetic algorithm. *Bioresource technology*, 212, 62-71.
<https://doi.org/10.1016/j.biortech.2016.04.022>
- Estevão, S. T., e Silva, J. B. D. A., & Lourenço, F. R. (2021). Development and optimization of beer containing malted and non-malted substitutes using quality by design (QbD) approach. *Journal of Food Engineering*, 289, 110182.
<https://doi.org/10.1016/j.jfoodeng.2020.110182>
- Fan, S., Chen, S., Tang, X., Xiao, Z., Deng, Q., Yao, P., ... & Chen, C. (2015). Kinetic model of continuous ethanol fermentation in closed-circulating process with pervaporation membrane bioreactor by *Saccharomyces cerevisiae*. *Bioresource Technology*, 177, 169-175.
<https://doi.org/10.1016/j.biortech.2014.11.076>
- Feil, B., Abonyi, J., & Szeifert, F. (2004). Model order selection of nonlinear input-output models—a clustering based approach. *Journal of Process Control*, 14(6), 593-602.
<https://doi.org/10.1016/j.jprocont.2004.01.005>
- Gao, B., Wang, J., Wang, Y., Xu, Z., Li, B., Meng, X., ... & Zhu, J. (2022). Influence of fermentation by lactic acid bacteria and in vitro digestion on the biotransformations of blueberry juice phenolics. *Food Control*, 133, 108603.
<https://doi.org/10.1016/j.foodcont.2021.108603>
- Germec, M., Cheng, K. C., Karhan, M., Demirci, A., & Turhan, I. (2020). Application of mathematical models to ethanol fermentation in biofilm reactor with carob extract. *Biomass Conversion and Biorefinery*, 10, 237-252.
<https://doi.org/10.1007/s13399-019-00425-1>
- Goelzer, A., Charnomordic, B., Colombié, S., Fromion, V., & Sablayrolles, J. M. (2009). Simulation and optimization software for alcoholic fermentation in winemaking conditions. *Food Control*, 20(7), 635-642.
<https://doi.org/10.1016/j.foodcont.2008.09.016>
- González-Herrera, I. Y., Rabasa-Olazábal, G., Pérez-Martínez, A., González-Suarez, E., & Castro-Galiano, E. (2016). Herramienta para apoyar la toma de decisiones en el desarrollo de biorrefinerías. *Revista Mexicana de Ingeniería Química*, 15(3), 943-951.

- Grosman, B., & Lewin, D. R. (2004). Adaptive genetic programming for steady-state process modeling. *Computers & Chemical Engineering*, 28(12), 2779-2790. <https://doi.org/10.1016/j.compchemeng.2004.09.001>
- Hebing, L., Tran, F., Brandt, H., & Engell, S. (2020). Robust optimizing control of fermentation processes based on a set of structurally different process models. *Industrial & Engineering Chemistry Research*, 59(6), 2566-2580. <https://pubs.acs.org/doi/abs/10.1021/acs.iecr.9b05504>
- Jorayev, P., Russo, D., Tibbetts, J. D., Schweidtmann, A. M., Deutsch, P., Bull, S. D., & Lapkin, A. A. (2022). Multi-objective Bayesian optimisation of a two-step synthesis of p-cymene from crude sulphate turpentine. *Chemical Engineering Science*, 247, 116938. <https://doi.org/10.1016/j.ces.2021.116938>
- Koza, J., & Poli, R. (2005). In E. Burke & G. Kendall (Eds), *Search methodologies – Introductory tutorials in optimization and decision support techniques*. New York, USA: Springer Science and Business Media, Inc. [Chapter 5 – Genetic programming].
- Kumar, B., Jha, A., Deshpande, V., & Sreenivasulu, G. (2014). Regression model for sediment transport problems using multi-gene symbolic genetic programming. *Computers and electronics in agriculture*, 103, 82-90. <https://doi.org/10.1016/j.compag.2014.02.010>
- Kummer, A., Varga, T., & Abonyi, J. (2019). Genetic programming-based development of thermal runaway criteria. *Computers & Chemical Engineering*, 131, 106582. <https://doi.org/10.1016/j.compchemeng.2019.106582> [Get rights and content](#)
- Li, P., Tan, X., Fu, X., Dang, Y., & Li, S. (2021). Metabolomic analysis reveals *Kluyveromyces marxianus*'s stress responses during high-temperature ethanol fermentation. *Process Biochemistry*, 102, 386-392. <https://doi.org/10.1016/j.procbio.2021.01.024>
- Liu, C. G., Li, K., Wen, Y., Geng, B. Y., Liu, Q., & Lin, Y. H. (2019). Bioethanol: New opportunities for an ancient product. In *Advances in bioenergy* (Vol. 4, pp. 1-34). Elsevier. <https://doi.org/10.1016/bs.aibe.2018.12.002>
- Lopes, M. L., Paulillo, S. C. D. L., Godoy, A., Cherubin, R. A., Lorenzi, M. S., Giometti, F. H. C., ... & Amorim, H. V. D. (2016). Ethanol production in Brazil: a bridge between science and industry. *Brazilian journal of microbiology*, 47, 64-76.
- Madár, J., Abonyi, J., & Szeifert, F. (2005). Genetic programming for the identification of nonlinear input– output models. *Industrial & engineering chemistry research*, 44(9), 3178-3186.
- McAloon, A., Taylor, F., Yee, W., Ibsen, K., & Wooley, R. (2000). *Determining the cost of producing ethanol from corn starch and lignocellulosic feedstocks* (No. NREL/TP-580-28893). National Renewable Energy Lab.(NREL), Golden, CO (United States). <https://doi.org/10.2172/766198>
- Meng, Y., Yu, S., Qiu, Z., Zhang, J., Wu, J., Yao, T., & Qin, J. (2021). Modeling and optimization of sugarcane juice clarification process. *Journal of Food Engineering*, 291, 110223. <https://doi.org/10.1016/j.jfoodeng.2020.110223>
- Meza, G. R., Ferragud, X. B., Saez, J. S., & Durá, H. (2017). Controller tuning with evolutionary multiobjective optimization. *Switzerland: Springer*. <https://doi.org/10.1007/978-3-319-41301-3>
- Müller, J., Schenk, C., Keicher, R., Schmidt, D., Schulz, V., & Velten, K. (2020). Optimization of an externally mixed biogas plant using a robust CFD method. *Computers and electronics in agriculture*, 171, 105294. <https://doi.org/10.1016/j.compag.2020.105294>
- Nawaz, A., Ashfaq, A., Zaidi, S. M. A. M., Munir, M., Haq, I. U., Mukhtar, H., & Tahir, S. F. (2020). Comparison of fermentation and medical potentials of *Saccharomyces* with *Wickerhamomyces* genera. *Revista Mexicana de Ingeniería Química*, 19(1), 33-47.
- Nor-Khaizura, M. A. R., Flint, S. H., McCarthy, O. J., Palmer, J. S., & Golding, M. (2019). Modelling the effect of fermentation temperature and time on starter culture growth, acidification and firmness in made-in-transit yoghurt. *LWT*, 106, 113-121. <https://doi.org/10.1016/j.lwt.2019.02.027>
- Peng, B., Williams, T. C., Henry, M., Nielsen, L. K., & Vickers, C. E. (2015). Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. *Microbial cell factories*, 14, 1-11. <https://doi.org/10.1186/s12934-015-0278-5>
- Pereira, R. D., Badino, A. C., & Cruz, A. J. (2020). Framework based on artificial intelligence to increase industrial bioethanol production. *Energy & Fuels*, 34(4), 4670-4677. <https://doi.org/10.1021/acs.energyfuels.0c00033>

- Ramírez-Hernández, A., Aparicio-Saguilán, A., Reynoso-Meza, G., & Carrillo-Ahumada, J. (2017). Multi-objective optimization of process conditions in the manufacturing of banana (*Musa paradisiaca* L.) starch/natural rubber films. *Carbohydrate Polymers*, 157, 1125-1133.
<https://doi.org/10.1016/j.carbpol.2016.10.083>
- Rivera, E. C., Yamakawa, C. K., Saad, M. B., Atala, D. I., Ambrosio, W. B., Bonomi, A., ... & Rossell, C. E. (2017). Effect of temperature on sugarcane ethanol fermentation: Kinetic modeling and validation under very-high-gravity fermentation conditions. *Biochemical engineering journal*, 119, 42-51.
<https://doi.org/10.1016/j.bej.2016.12.002>
- Rodman, A. D., & Gerogjorgis, D. I. (2020). Parameter estimation and sensitivity analysis for dynamic modelling and simulation of beer fermentation. *Computers & Chemical Engineering*, 136, 106665.
<https://doi.org/10.1016/j.compchemeng.2019.106665>
- Rodríguez-Mariano, A., Reynoso-Meza, G., Páramo-Calderón, D. E., Chávez-Conde, E., García-Alvarado, M. A., & Carrillo-Ahumada, J. (2015). Análisis del desempeño de controladores lineales sintonizados en diferentes estados estacionarios del biorreactor de cholette mediante técnicas de decisión multi-criterio. *Revista mexicana de ingeniería química*, 14(1), 167-204.
- Salmi, T., Aguilera, A. F., Lindroos, P., & Kanerva, L. (2022). Mathematical modelling of oleic acid epoxidation via a chemo-enzymatic route—From reaction mechanisms to reactor model. *Chemical Engineering Science*, 247, 117047.
<https://doi.org/10.1016/j.ces.2021.117047>
- Samaniego-Sanchez, C., Marin-Garcia, G., & Quesada-Granados, J. J. (2020). A new fermented beverage from sugarcane (*Saccharum officinarum* L.) molasses: Analysis of physicochemical properties and antioxidant capacity, and comparison with other industrial alcohol products. *LWT*, 128, 109505.
<https://doi.org/10.1016/j.lwt.2020.109505>
- Shen, T., Wu, Q., & Xu, Y. (2021). Biodegradation of cyanide with *Saccharomyces cerevisiae* in Baijiu fermentation. *Food Control*, 127, 108107.
<https://doi.org/10.1016/j.foodcont.2021.108107>
- Solís-García, A., Rivas-García, P., Escamilla-Alvarado, C., Rico-Martínez, R., Bravo-Sánchez, M. G., & Botello-Álvarez, J. E. (2017). Methanol production kinetics during agave cooking for mezcal industry. *Revista Mexicana de Ingeniería Química*, 16(3), 827-834.
- Torralba-Morales, L. M., Reynoso-Meza, G., & Carrillo-Ahumada, J. (2020). Sintonización y comparación de conceptos de diseño aplicando la optimalidad de Pareto. Un caso de estudio del biorreactor de Cholette. *Revista Iberoamericana de Automática e Informática industrial*, 17(2), 190-201.
<https://doi.org/10.4995/riai.2019.11424>
- Urtubia, A., León, R., & Vargas, M. (2021). Identification of chemical markers to detect abnormal wine fermentation using support vector machines. *Computers & Chemical Engineering*, 145, 107158.
<https://doi.org/10.1016/j.compchemeng.2020.107158>
- Vignesh, N., & Chandraraj, K. (2021). Improved high solids loading enzymatic hydrolysis and fermentation of cotton microdust by surfactant addition and optimization of pretreatment. *Process Biochemistry*, 106, 60-69.
<https://doi.org/10.1016/j.procbio.2021.04.002>
- Wu, Z., Xu, E., Long, J., Wang, F., Xu, X., Jin, Z., & Jiao, A. (2015). Measurement of fermentation parameters of Chinese rice wine using Raman spectroscopy combined with linear and non-linear regression methods. *Food Control*, 56, 95-102.
<https://doi.org/10.1016/j.foodcont.2015.03.015>
- Zhao, G., Kuang, G., Wang, Y., Yao, Y., Zhang, J., & Pan, Z. H. (2020). Effect of steam explosion on physicochemical properties and fermentation characteristics of sorghum (*Sorghum bicolor* (L.) Moench). *LWT*, 129, 109579.
<https://doi.org/10.1016/j.lwt.2020.109579>