Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

Development of a Predictive Model for Microbial Growth under Variable Conditions Using a Multilayer Perceptron Neural Network: Application *to Candida guilliermondii*

Jazmín Cortez-González[a], Juan Gabriel Segovia-Hernández[b], Salvador Hernández[b], Varinia López-Ramírez[a], Arturo Hernández-Aguirre[c], Rodolfo Murrieta-Dueñas[a*]

[a] Tecnológico Nacional de México / ITS Irapuato (ITESI), División de Ingeniería Química, ; Carretera Irapuato-Silao Km. 12.5, C.P:36821 Irapuato, Guanajuato, MEXICO
[b] Departamento de Ingeniería Química, División de Ciencias Naturales y Exactas, Universidad de Guanajuato, Noria Alta S/N, C.P. 36050, Guanajuato, Guanajuato, México.
[c] Centro de Investigación en Matemáticas (CIMAT) A.C., Departamento de Ciencias Computacionales, Callejón de Jalisco s/n, 36240, Mineral de Valenciana, Guanajuato, Gto., México.
* Corresponding Author: jazmin.cg @irapuato.tecnm.mx

ABSTRACT

In the field of biochemical process design, the accurate modeling of microbial growth is essential for the development and optimization of biological reactors used in the production of high-value compounds. Achieving this objective requires a detailed understanding of how environmental factors—such as pH and nutrient availability—influence microbial dynamics across the four distinct growth phases: lag, exponential, stationary, and death. Traditionally, reactor design relies heavily on the Monod model, which provides a simplified representation of microbial growth, focusing primarily on the exponential phase under constant operating conditions (1). However, this model presents substantial limitations when applied to dynamic environments where key parameters vary over time. To overcome these constraints, the present study proposes a data-driven modeling approach using a multilayer perceptron (MLP) artificial neural network for the prediction of microbial growth trajectories under varying pH conditions and substrate compositions. The yeast strain *Candida guilliermondii* was selected as the model microorganism due to its industrial relevance. Experimental growth data were collected through optical density measurements using a Multiskan™ FC Microplate Photometer (Thermo Scientific), covering a pH range from 6.0 to 8.5 and two substrate scenarios: pure xylose, and a 1:2 glucose–xylose mixture. The experimental data were used to train the MLP neural network, which generated predictive models capable of estimating growth behavior under the specified input conditions. This modeling approach enables the simulation of microbial growth curves at any given time point within the defined parameter space, providing a more flexible and comprehensive tool compared to classical models. The results of this study demonstrate that the proposed MLP-based model is a powerful computational tool for both the design and real-time control of bioprocesses. The integration of these tools into predictive control strategies is a key aspect of the bioprocess research. The model's versatility and its ability to predict microbial behavior make it a very important tool for bioreactor control.

Keywords: Artificial Intelligence, Modelling and Simulations, Machine Learning, Optimization, Biomass, microbial growth

INTRODUCTION

The design and operation of biological reactors critically depend on an accurate representation of microbial growth kinetics. These kinetics directly determine reactor productivity, stability, and control perfor-

mance. Among the most widely adopted kinetic formulations, the Monod model remains a foundational tool due to its simplicity and interpretability. Nevertheless, its validity is constrained to dilute systems, constant environmental conditions, and predominantly the exponential growth phase. Consequently, significant uncertainty arises when Monod-based models are employed for predicting microbial behavior under dynamic conditions, particularly during the lag and stationary phases.

In recent years, surrogate modeling techniques based on artificial intelligence and machine learning have emerged as promising alternatives for representing complex biological phenomena. Neural network–based models are capable of learning nonlinear relationships directly from experimental data without requiring explicit mechanistic assumptions. This makes them particularly suitable for systems involving multiple substrates, inhibitory effects, high substrate concentrations, and environmental variability.

Previous studies have explored surrogate models for microbial and microalgal growth under limited and mostly static conditions. For instance, Lee et al. (2015) reviewed kinetic models for microalgae cultivation, while Akkermans and Van Impe (2018) investigated the effect of pH inhibition using mechanistic approaches. More recent works have applied data-driven techniques to microbial systems; however, these efforts typically focus on one or two growth phases or assume fixed operating conditions. Notably, Murrieta-Dueñas et al. (2022) proposed a neural network approach to describe exponential growth under dynamic conditions, yet a comprehensive representation of the full growth trajectory remains largely unexplored.

To the best of the authors' knowledge, no previous study has demonstrated the prediction of multiple microbial growth phases under simultaneously varying pH and substrate conditions using a single surrogate model. Therefore, this work aims to fill this gap by developing an MLP-based surrogate model capable of predicting the complete microbial growth trajectory of *Candida guilliermondii* under dynamic environmental conditions.

METHODOLOGY

Figure 1 illustrates the overall methodology adopted in this study, which consists of two main stages: (i) experimental data acquisition and (ii) surrogate model development.

Experimental Design

Microbial growth experiments were conducted under controlled laboratory conditions to generate high-resolution kinetic data. The experimental matrix included pH values ranging from 6.5 to 8.5 and two substrate formulations: pure xylose and a glucose–xylose mixture (1:2). Cultivations were performed over a 24-hour period to capture the lag, exponential, and stationary growth phases. Optical density measurements were collected every 15 minutes in triplicate to ensure reproducibility and statistical robustness.

Surrogate Model Development

The second stage involved the construction of a surrogate model using a multilayer perceptron neural network. The MLP was trained to approximate the nonlinear mapping between the input variables—time and pH—and the output variable, optical density (absorbance). Different network architectures were evaluated by varying the number of neurons and activation functions. Model performance was assessed using standard statistical metrics, including AIC, MSE, RMSE, and the coefficient of determination (R^2).

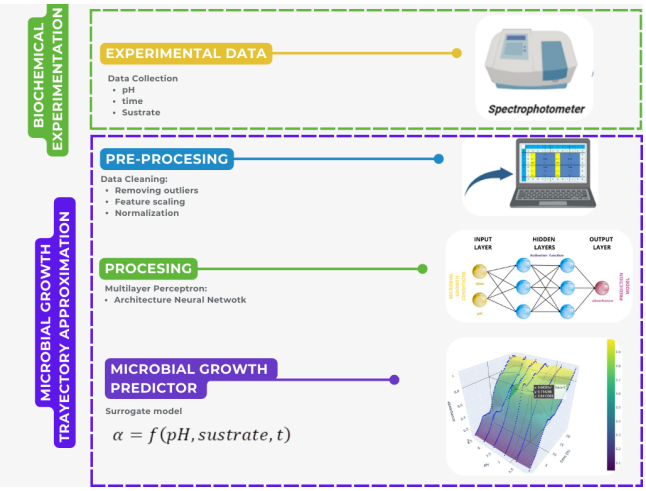

Figure 1. Methodology of Microbial growth trajectory approximation.

Biochemical Experimentation

Microbial growth is conventionally described through four distinct phases: lag, exponential, stationary, and death. The lag phase corresponds to the period immediately following inoculation, during which microorganisms adapt to the new environmental conditions and activate the metabolic pathways required for growth. Once adaptation is complete, the system enters the exponential phase, characterized by rapid cell division and substrate consumption at an approximately constant specific growth rate. As nutrients become depleted and inhibitory metabolites accumulate, the system transitions into the stationary phase, where cell growth and death rates are approximately balanced. Finally, the death phase is marked by a progressive decline in viable biomass, resulting in a significant reduction or cessation of microbial growth. These phases constitute a generalized kinetic framework that applies to a wide range of microorganisms, including bacteria,

fungi, yeasts, and microalgae.

In this study, the kinetic behavior of the yeast *Candida guilliermondii* was selected as the model system. This microorganism is widely distributed in natural environments and can be readily isolated from fruits and berries with high sugar content. In addition, it is commonly found as part of the native microflora of human skin and mucous membranes. Beyond its ecological prevalence, C. guilliermondii is of considerable biotechnological interest due to its ability to act as a biological control agent and its well-documented capacity to bioconvert xylose into xylitol, a high-value polyol with applications in the food and pharmaceutical industries. This metabolic capability relies on the efficient production of enzymes and cellular energy required for pentose assimilation.

Under optimal conditions, C. guilliermondii exhibits maximal growth at approximately pH 7 and a temperature of 28 °C. However, industrial bioprocesses often operate under suboptimal or fluctuating environmental conditions. For this reason, the present work explores a broader pH range, from 6.5 to 8.5, in order to capture the adaptive kinetic response of the microorganism. These controlled pH variations enable the observation of different growth regimes over a 24-hour cultivation period, thereby generating a comprehensive dataset that spans multiple growth phases.

Given the relevance of C. guilliermondii in xylose bioconversion, two substrate formulations were evaluated: a culture medium containing xylose as the sole carbon source and a mixed-substrate medium consisting of a glucose–xylose blend at a 1:2 ratio. The inclusion of glucose aims to assess its influence on reducing the lag phase and accelerating the onset of exponential growth, a strategy commonly employed to enhance process productivity in mixed-sugar fermentations.

All experiments were conducted in triplicate to ensure reproducibility and statistical reliability. Experimental analyses were performed at the Microbial Interaction Laboratory of the National Technological Institute of Mexico, Irapuato campus. Microbial growth was monitored via optical density measurements at 600 nm, which served as a proxy for biomass concentration. Measurements were recorded every 15 minutes over a 24-hour period at a constant temperature of 28 °C using a Thermo Scientific Multiskan™ FC Microplate Photometer. Cultivations were carried out in multiwell plates, allowing for high-throughput data acquisition and consistent experimental conditions across all runs.

Microbial Growth Trajectory Approximation

The second stage of the proposed methodology focuses on learning the underlying data patterns from the experimentally obtained microbial growth kinetics of C. guilliermondii. To this end, a data-driven surrogate model based on machine learning was developed using a multilayer perceptron (MLP) neural network. The well-established capabilities of modern artificial intelligence techniques, particularly neural networks, provide an efficient and flexible framework for predicting microbial growth under previously unobserved combinations of operating conditions.

The MLP architecture employed in this study consists of three main components: (i) an input layer, (ii) one or more hidden layers, and (iii) an output layer. Input signals are propagated forward through the network, where each neuron applies a nonlinear transformation to a weighted linear combination of its inputs. The final output of the network is a scalar value representing the predicted absorbance. In this work, the MLP is used as a nonlinear function approximator and predictive model to estimate the absorbance α as a function of time t and the pH of the culture medium.

From a theoretical standpoint, a multilayer perceptron is a universal approximator of continuous nonlinear functions, which makes it particularly suitable for modeling complex biological and biochemical processes. Formally, the MLP implements the functional mapping described by Equation (1), where the output is expressed as a composition of nonlinear activation functions parameterized by a set of weights and biases.

This modeling approach allows the MLP to implicitly capture phase transitions and nonlinear growth dynamics without explicitly segmenting the growth curve, thereby offering a significant advantage over classical kinetic models when dealing with dynamic and non-ideal operating conditions.

From a formal perspective, the MLP implements the function described by Equation 1.

$$y = f(x, \omega) \qquad \text{Eq. (1)}$$

Where f is a composition of nonlinear functions parameterized by weights and biases, ω.

The experimental setup is as follows: The number of neurons in the hidden layer was varied to determine the optimal topological architecture for our artificial neural network. Three activation functions—Sigmoid, ReLu, and Tanh—were analyzed. We implemented the Adam optimizer to automatically adjust the learning rate of each parameter, allowing for faster and more stable convergence. The number of neurons is as follows: 2, 4, 6, 8, and 10. We used a dataset of 540 feature vectors. We use the widely tested least squares function as the fitness function during the training stage. We validate our model using standard growth factors. We trained the machine learning model using three pH values (6.5, 7.5, 8, and 8.5) and substrate formulations in the culture medium (xylose at 100% and glucose-xylose at a ratio of 1:2). Our model predicts absorbance values for any combination of pH between 6.5 and 8.5, substrate (xy-

Table 1. Summary of the metrics that were obtained from the different architecture of the multilayer perceptron (MLP).

Configuration	Architecture		AIC			MSE			RMSE			R^2		
	Hidden layers	Neurons	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
1	2	2	-1822.761	-244.757	-518.618	0.008	0.007	0.007	0.088	0.082	0.083	0.910	0.937	0.926
2	2	6	-1825.331	-151.625	-451.464	0.006	0.006	0.005	0.076	0.075	0.070	0.932	0.948	0.947
3	4	2	-1583.158	-164.052	-408.564	0.014	0.019	0.015	0.117	0.138	0.122	0.842	0.822	0.839
4	4	6	-1775.455	4.179	-307.745	0.004	0.004	0.004	0.065	0.067	0.063	0.950	0.958	0.958
5	6	2	-855.842	-42.732	-183.566	0.086	0.111	0.093	0.294	0.334	0.305	0.000	-0.042	0.000
6	6	6	-463.842	349.268	208.434	0.086	0.111	0.093	0.294	0.334	0.305	0.000	-0.042	0.000
7	8	2	-831.842	-18.732	-159.566	0.086	0.111	0.093	0.294	0.334	0.305	0.000	-0.042	0.000
8	8	6	-1484.352	363.763	39.159	0.004	0.007	0.004	0.062	0.083	0.066	0.956	0.936	0.953
9	10	2	-807.842	5.269	-135.566	0.086	0.111	0.093	0.294	0.334	0.305	0.000	-0.042	0.000
10	10	6	-1146.487	522.436	243.249	0.006	0.006	0.006	0.077	0.076	0.077	0.931	0.946	0.935

lose or glucose-xylose), and timestamp (hours) between 0 and 24.

RESULTS

In this work, a machine-learning–based framework was developed for the estimation and prediction of microbial growth kinetics, as schematically illustrated in Fig. 6. The proposed framework comprises two main stages: model training and model testing. During the training stage, experimentally obtained kinetic data are used to fit the parameters of the neural network and to quantify the training error, expressed in terms of the mean squared error (MSE train). This stage results in the construction of the HC-MLP predictor model, which encapsulates the learned nonlinear relationship between the input variables and the microbial growth response.

To rigorously assess the generalization capability of the trained model, the testing stage employs an independent replicate of the experimental kinetic data that was not used during training. The prediction error obtained at this stage, quantified as MSE test, provides an unbiased evaluation of the model's predictive performance. Based on the results of the testing stage, the final surrogate model is obtained. This surrogate model generates a continuous response surface that represents the MLP approximation of the output variable—optical density (absorbance)—as a function of the input variables, time (t) and pH, for culture media containing xylose as the carbon source.

A case study was conducted to evaluate the predictive capability of the proposed MLP framework. The objective of this analysis was to assess the model's ability to accurately predict microbial growth for arbitrary combinations of pH and time within the studied domain. Specifically, the model was used to estimate absorbance values over a time horizon ranging from 0 to 24 h and pH values between 6.5 and 8.5, assuming xylose as the sole substrate. To identify the optimal network topology, the number of neurons in the hidden layer was systematically varied, allowing for an evaluation of the trade-off between model complexity and predictive accuracy.

To select the most appropriate neural network architecture, several complementary performance metrics were employed, including the Akaike information criterion (AIC), mean squared error (MSE), root mean squared error (RMSE), and the coefficient of determination (R^2). The AIC was used as the primary criterion for model selection, as it simultaneously accounts for goodness of fit and model parsimony by penalizing excessive complexity. Lower AIC values indicate a more favorable balance between predictive accuracy and information loss. The MSE quantifies the average squared deviation between predicted and observed values and was primarily used to compare model performance during the training phase. The RMSE, expressed in the same units as the experimental measurements, provides an intuitive measure of prediction error and was used for both validation and testing stages. Finally, the coefficient of determination (R^2) evaluates the proportion of variance in the experimental data explained by the model, with values approaching unity indicating excellent agreement between predictions and observations.

Table 1 summarizes the values of all performance metrics for the several MLP architectures evaluated in this study. Each metric is reported for the training, validation, and testing stages, providing a comprehensive basis for selecting the optimal neural network configuration.

As evidenced by the results summarized in Table 1, configuration 2 exhibits the most robust and consistent performance across all selected evaluation metrics. This configuration yields the lowest Akaike information criterion (AIC) value during the training phase (–1825.33), indicating an optimal trade-off between model fidelity and structural complexity. Comparable AIC behavior is observed in the validation and testing phases, which

further suggests that the selected architecture generalizes well beyond the training dataset.

In terms of predictive accuracy, configuration 2 achieves one of the lowest mean squared error (MSE) values, equal to 0.005842, demonstrating a high-quality fit to the experimental data. A similar trend is observed for the root mean squared error (RMSE), with values of 0.074527 and 0.069864 obtained for the validation and testing stages, respectively. These low RMSE values, expressed in the same units as the measured absorbance, confirm the model's ability to accurately reproduce microbial growth dynamics across the studied domain. Furthermore, the coefficient of determination (R^2) for this configuration approaches unity in all stages, indicating that the surrogate model explains nearly all of the variance observed in the experimental measurements.

Taken together, these results indicate that configuration 2 provides the most favorable balance between predictive performance, model parsimony, and computational efficiency. Consequently, the optimal MLP architecture corresponds to a network with two hidden layers comprising six neurons each, which minimizes the risk of overfitting while maintaining high predictive capability.

Figure 2 illustrates the three-dimensional response surface generated by the surrogate model corresponding to configuration 2 of the multilayer perceptron neural network. In this representation, the experimental data points are shown as blue markers, while the continuous surface depicts the MLP-based approximation of the microbial growth trajectory within the defined data domain. The smoothness and continuity of the predicted surface, together with its close agreement with the experimental observations, demonstrate the model's ability to capture the underlying nonlinear kinetics of *Candida guilliermondii* growth as a function of time and pH.

As illustrated in Figure 2, red neuronal perception can effectively predict the four phases of microbial growth: lag, exponential, stationary, and death. This observation is consistent across the pH range analyzed in the experimental setting (6.5–8.5). The resulting model is able to predict the trajectory of microbial kinetics from a point within the experimental data space. This trajectory can be verified by the gray line shown as an example in Figure 2. This trajectory can be verified at any point in the analyzed space and yields the corresponding coordinate values. Therefore, if we are aware of this condition, we can predict the microorganism's growth phase and implement the appropriate control measures to ensure the highest possible yield of the biological reaction.

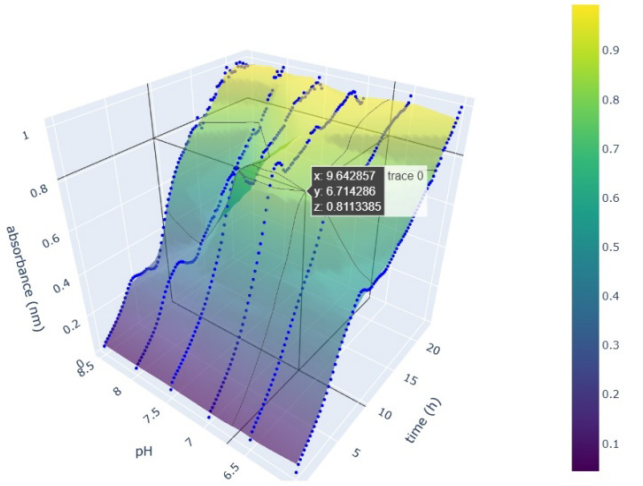

Figure 2. Best surrogate model of configuration 2, for predicted microbial growth kinetics trajectory of *Candida guillermondii*.

CONCLUSIONS

This study presents a robust data-driven framework for modeling microbial growth kinetics under dynamic environmental conditions using a multilayer perceptron neural network. Unlike classical kinetic models, the proposed surrogate model successfully captures nonlinear microbial behavior across multiple growth phases without relying on restrictive mechanistic assumptions.

The results demonstrate that the MLP-based model provides accurate predictions of microbial growth trajectories for *Candida guilliermondii* under varying pH and substrate conditions. This approach significantly enhances the flexibility and predictive capability of kinetic modeling tools, making it particularly suitable for applications in bioreactor design, optimization, and real-time control.

From an engineering perspective, the developed surrogate model enables rapid scenario analysis and supports the automation of experimental workflows, thereby reducing experimental costs and development time. Future work will focus on extending the model to include additional environmental variables, such as temperature and substrate inhibition, as well as integrating the surrogate model into advanced process control and digital twin frameworks for industrial bioprocesses.

AUTHOR IDENTIFIERS

Cortez-González J: 0000-0003-1613-0371
Segovia-Hernández J.G: 000-0003-4175-9809
Hernández S: 0000-0002-4598-3392
López-Ramírez V: N/A
Hernández-Aguirre A: 0000-0002-3744-9827
Murrieta-Dueñas R: 0000 0001 5759 2500

REFERENCES

1. Akkermans S, Van Impe JF. Mechanistic modelling of the inhibitory effect of ph on microbial growth. Food Microbiology 72:214-219 (2018). https://doi.org/10.1016/j.fm.2017.12.007
2. Lee E, Jalalizadeh M, Zhang Q. Growth kinetic models for microalgae cultivation: a review. Algal Research 12:497-512 (2015). https://doi.org/10.1016/j.algal.2015.10.004
3. Moreno?Paz S, Schmitz J, Martins dos Santos VAP, Suarez?Diez M. Enzyme?constrained models predict the dynamics of saccharomyces cerevisiae growth in continuous, batch and fed?batch bioreactors. Microbial Biotechnology 15:1434-1445 (2022). https://doi.org/10.1111/1751-7915.13995
4. Murrieta-Dueñas R, Serrano-Rubio JP, López-Ramírez V, Segovia-Dominguez I, Cortez-González J. Prediction of microbial growth via the hyperconic neural network approach. Chemical Engineering Research and Design 186:525-540 (2022). https://doi.org/10.1016/j.cherd.2022.08.021
5. Öksüz HB, Buzrul S. MONTE CARLO ANALYSIS FOR MICROBIAL GROWTH CURVES. J microb biotech food sci 10:418-423 (2020). https://doi.org/10.15414/jmbfs.2020.10.3.418-423

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See https://creativecommons.org/licenses/by-sa/4.0/