Enhanced kinetic model parameters for xylitol bioproduction from *Candida mogii* ATCC 18364

Julio César Sánchez-Rendóna, Luis Gerónimo Matallanab, Ricardo Morales-Rodriguezc, Oscar Andrés Prado-Rubioa,d\*

aDepartamento de Ingeniería Química, Universidad Nacional de Colombia –Manizales 170003, Colombia

bGrupo de Investigación en Alimentos y Agroindustria, Universidad de Caldas, Calle 65, No. 26-10, C.P. Manizales 170002, Colombia

cDepartamento de Ingeniería Química, Universidad de Guanajuato, Noria Alta S/N; Guanajuato, Guanajuato, 36050, México.

dDepartment of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Lyngby DK-2800, Denmark

\*oaprador@kt.dtu.dk

Abstract

Xylitol is a common sweetener used in the dental and pharmaceutical industries. This molecule can be obtained from renewable sources and, is also a building block with the potential to produce a wide variety of chemical compounds. However, xylitol bioproduction is still economically unattractive and, there is a need for investigating optimal process design and operation. For that, a mathematical model that accurately predicts the xylitol fermentation is fundamental to reduce uncertainties during process production. In this study, a framework for structural and practical parameter identification is implemented to increase the number of identifiable kinetic parameters, using a mathematical model composed of 5 states and 11 parameters. The methodology uses well-established methods that unfortunately are not commonly used together such as a) experimental data processing, b) structural identifiability analysis, c) enhanced parameters identification using a nested optimization approach for self-tuning stochastic optimization and, d) validation. The employed methodology allows to determine that the full set of parameters is locally identifiable, and the number of identifiable parameters was increased from 4 to 10, also with reduced confidence intervals. The increased confidence in the model parameters brings interesting insights into this fermentation, which might permit peforming further and more certain analysis.

**Keywords:** xylitol bioproduction, parameter estimation, mathematical modeling, global optimization.

* 1. Introduction

The European sustainability agenda for 2,050 proposed to achieve up to 38 % of biobased products in the market to migrate to a more balanced future. One of the interesting platform components to be obtained from biomass is xylitol. Xylitol is a sweetener and additive used in food, pharma, dental, cosmetics, and others. Additionally, xylitol could be used to obtain xylaric and xylonic acids, propylene, and ethylene glycol, among others. Xylitol market size is expected to grow up to USD ~1.50 billion at a compound annual growth rate (CAGR) of 6 % (2023-2030) (Custom Market Insights, 2022). Xylitol bioproduction has been investigated, obtaining production cost from 1.59 USD/kg up to 5 USD/kg depending on the raw material and production scheme (Ruales-Salcedo et al., 2022). With a xylitol selling price of ~3.5 USD/kg, it can be said that the bioprocess is economically unattractive and, from the *in silico* process design, predictions could be still lacking of quality especially due to the fermentation model uncertainties. Recently, a xylitol kinetic model from *Candida mogii* ATCC 18364 was investigated, showing that only 4 out of 11 model parameters were identifiable using local optimization (Prado-Rubio et al., 2015). Therefore, there is a need for obtaining a more representative and interpretable xylitol kinetic model that could be reliably used for system understanding, experimental design, and optimal process design/control. This study aims to implement a robust framework for structural and practical parameter identification, which leads to an increase in the number of identifiable parameters. Thus, the quality and interpretability of the improved parameters allows for obtaining new system understanding, evidencing the time-variant nature of the repeated fed-batch fermentations and, the possible intracellular dynamic behavior of the biochemical system during the process.

* 1. Methodology

The methodology is shown in Fig. 1. As a starting point, both the mathematical model and the experimental data are needed. The data is treated to remove outliers (via signal filters and mobile median analysis) and random noise (via polynomial fitting with 4th order Savitzky-Golay polynomial). Considering the mathematical model and the characteristics of the experimental data (number and type of variables, type of experiments), a structural identifiability analysis is performed to determine if the parameters are identifiable globally/locally or not identifiable (GenSII tollbox). Whether the model is not structurally identifiable, either the experimental design or the mathematical model must be adjusted. When the model is structurally identifiable, an optimization problem is defined and used for optimizer tuning (via the IRACE package in R®). Hyperparameter tuning is an extremely important step that is unfortunately rarely reported, but it is important to mention that only through a specific configuration for a given optimization problem (incorporating experimental data), it possible to guarantee the correct behavior of the optimizer (Wolpert et al., 1997). The tuned hyperparameters are then used to solve the complete parameter estimation problem and calculate the optimal model parameters. As the mathematical model is complete, the practical identifiability (validation) test is performed within several categories as confidence intervals (in parameters and prediction), goodness-of-fit metrics (SSE, MRSE, GoF, R2, etc), and sensibility indexes (standardized regression coefficients method) (Sánchez-Rendón et al., 2020). If the validation process is not satisfactory, modifications in the model or the experimental design are required.

* + 1. Case of study: xylitol bioproduction

Xylitol is a C5 alcohol widely used as a sweetener and additive to dental and pharmaceutical products. Xylitol is also a platform chemical fundamental for other chemical compounds production (Ur-Rehman et al., 2015). Xylitol bioproduction experimental data used in this work was taken from Sirisansaneeyakul et al. (2013), who performed a sequential experiment of 22 fed-batch fermentations. The mathematical model of xylitol bioproduction is composed of 5 states (glucose, xylose, biomass, intra and extra xylitol) and 11 parameters that describe complex biological non-linear phenomena such as, inhibition of glucose and xylose uptake, intra- and extracellular transport and reaction stoichiometry (Tochampa et al., 2015, Hernández-Escoto et al., 2016).

A diagram of a scientific model

Description automatically generated with medium confidence

Figure 1. Methodology for robust parameter estimation in bioprocess mathematical models with inputs (Green block), methods (blue block) and results (orange block).

* 1. Results

After the data preprocessing, outliers and overall random noise were systematically removed. This is beneficial for the solution of the parameter estimation problem, because it reduces the number of possible trajectories that the model can describe. In terms of structural identifiability, the 11 parameters of the mathematical model were found to be locally identifiable as seen in Table 1. For practical identifiability, the nested optimization problem was solved with dataset 1, with the objective of hyperparameter tuning for the Particle Swarm Optimization (PSO) algorithm. This stochastic global optimizer has the desired property of guaranteed theoretical convergence, meaning that with a correct hyperparameter configuration, this algorithm of consistently find points within the neighborhood of the global optimum (Huang et al., 2023). IRACE can calculate the optimizer hyperparameters through a statistical process capable of increasing the reproducibility of the results while decreasing the computational effort. For this case, 96 candidate configurations were found during the tunning process. The best-calculated hyperparameters for the PSO algorithm are inertia range of [0.25, 0.50], MinNeighborsFraction of 0.36, SelfAdjustmentWeight of 1.58, SocialAdjustmentWeight of 1.44, and SwarmSize of 185. The value of the calculated hyperparameters contrasts with those of the default configuration i.e., inertia range of [0.1, 1.1], MinNeighborsFraction of 0.25, SelfAdjustmentWeight of 1.49, SocialAdjustmentWeight of 1.49, and SwarmSize of 100.

In terms of the PSO inner working, the calculated configuration favors local search with slow-moving particles that highly communicate with each other, while the global search is compensated with an increased number of particles. The previously described hyperparameter configuration and a new optimization problem defined with 9 datasets (datasets 1 to 9) led to the estimated parameter values shown in Table 1. However, as shown by the practical identifiability analysis, the estimated values can be classified into different categories in accordance with the confidence levels. First, the parameter  is practically non-identifiable (xylose uptake inhibition due to glucose). , , depict high uncertainty and , , , , have very low uncertainty (less than 1 %). These differences can be attributed to the type of experiments performed where there is only one experiment using diauxic growth (i.e., glucose and xylose). Consequently, there is insufficient information to estimate with high precision the parameters involving glucose behavior and the interaction of both carbon sources within the cell, reflected in the low accuracy of , , and the non-identifiability of .

Table 1. Structural identifiability and estimated parameter values.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter | Structural identifiability | Estimated value | Confidence interval | Hernández-Escoto et al., 2016 | Variation (%) |
|  | Local | 6.183x10-2 | 5.32x10-2 | - | - |
|  | Local | 7.5586 | 3.54x10-3 | - | - |
|  | Local | 2.1546 | 2.78x10-2 | - | - |
|  | Local | 9.747x10-2 | 0.798 | - | - |
|  | Local | 1.452x10-2 | 1.39x10-2 | - | - |
|  | Local | 4.458x10-3 | 2.26x10-3 | 0.189 | 4x103 % |
|  | Local | 5.909x10-3 | 4.34x10-3 | 11.761 | 198x103 % |
|  | Local | 5.116x10-5 | 1.99x10-5 | 16.06 | 31x106 % |
|  | Local | 8.284x10-2 | 3.47x10-4 | 0.342 | 312 % |
|  | Local | 19.318 | 0.28 | - | - |
|  | Local | 7.823x10-11 | 2.02x10-13 | - | - |

Interestingly, comparing these results with previous efforts in kinetic model tuning done by Hernández-Escoto et al. (2016), it can be seen that the parameter values highly depend on how the optimization problem is solved, as shown in Table 1. Specifically, the authors used the full data available with a local optimizer (gradient-based) and determined that only 4 parameters can be reliably estimated. In this work, 9 data sets and a tuned global optimizer were used, increasing the number of identifiable parameters to 10. The substantial differences between the estimated values highlight the importance and need for a carefully chosen setup and solution of the parameter estimation optimization problem.

The parity plot (Fig. 2, left) shows the descriptive quality of the mathematical model after parameter estimation, with a fair fitting in the 9 data sets used for calibration, and a Mean Absolute Percentage Error (MAPE) index of 226.32. Better accuracy can be observed in xylose and xylitol concentrations with most of their points lying between the 10% confidence interval lines in comparison with glucose and biomass. This can be attributed to a higher amount of information present in the data sets for these variables. However, after 150 hours of fermentation discrepancies between the model and the data arise (Fig. 3, top), possibly due to microorganism adaptation to increasing concentrations of xylose.

A graph of different colored lines

Description automatically generated with medium confidence

Figure 2. Parity plot for model output. Time progression is shown as darkening color shade. Left: descriptive quality (training datasets). Right: predictive quality (validation datasets).

A graph of different colored lines

Description automatically generated with medium confidence

Figure 3. Top: Model prediction for xylose (red line) with 95% confidence intervals (blue dashed line) and experimental data (magenta diamonds). Bottom: Example of standardized regression coefficients for xylose prediction. Each line corresponds to one model parameter.

The parity plot for the predictive quality of the model (Fig. 2, right) shows a lack of accuracy in the prediction of biomass and xylose concentration but a lower MAPE index of 149.31. Specifically, the model predicts a higher xylose uptake rate which increases in the last batches (Fig. 3, top). Then, to keep the cell growth, the model predicts xylitol consumption. These results confirm that the microorganism underwent metabolic adaptation that affect slowing down the growth, xylose uptake, and resulted in higher xylitol accumulation. This adds a complexity layer to the parameter estimation optimization problem given that the system is time-variant.

Finally, a sensitivity analysis was performed with the standardized regression coefficients method for the xylose state (Fig. 3, bottom). As expected, the relevance of the parameters changes dynamically during the fermentation process. The change in the preferred carbon source is shown as a change in the relative importance of parameters related to glucose and xylose. A high parameter interaction is presented during xylose consumption and xylitol production, which highlights the need for an alternative experimental design. In this particular case, the calibration of the mathematical model could be enhanced with more experiments of simultaneous feed of glucose and xylose.

* 1. Conclusions

Mathematical models comprise a wide range of useful tools for the understanding of complex systems and their practical applications, thus, assessing the quality of the model predictions and the interpretability of the model parameters becomes a fundamental task. In this sense, the present work makes use of previously established, but not combined, methods for experimental data cleaning, structural identifiability, global optimizer tuning, practical identifiability, and model validation with a distinction between the descriptive and predictive quality of model outputs. This methodology allows us to evaluate the adequacy of experimental design (structural identifiability), information quality and relevance (descriptive quality), and fair model prediction of the experimental data (predictive quality). As a result, 10 out of 11 model parameters were reliably estimated which allowed us to provide system insights.

**Acknowledgements**

This research was partially founded by Novo Nordisk Foundation grant NNF19SA0035474.

References

H. Hernández-Escoto, O. A. Prado-Rubio, R. Morales-Rodriguez, 2016. Model-based framework for enhanced and controlled operation of a fed-batch bioreactor: xylitol production. Computer Aided Chemical Engineering, 38, 301-306.

H. Huang, J. Qiu, K. Riedl, 2023, On the global convergence of particle swarm optimization methods, Applied Mathematics & Optimization, 88, 30.

O. A. Prado-Rubio, H. Hernández-Escoto, R. Rodriguez-Gomez, S. Sirisansaneeyakul, R. Morales-Rodriguez, 2015, Enhancing xylitol bio-production by an optimal feeding policy during fed-batch Operation, Computer-Aided Chemical Engineering, 37, 1757-1762.

A. V. Ruales-Salcedo, V. H. Grisales-Díaz, R. Morales-Rodríguez, J. Fontalvo, O. A. Prado-Rubio, 2022, Production of High-Added Value Compounds from Biomass. Book chapter in: Biofuels and Biorefineries.

J. C. Sánchez-Rendón, R. Morales-Rodriguez, L. G. Matallana-Pérez, O. A. Prado-Rubio. 2020. Assessing Parameter Relative Importance in Bioprocesses Mathematical Models through Dynamic Sensitivity Analysis. Computer Aided Chemical Engineering, 48, 1711-1716.

S. Sirisansaneeyakul, S. Wannawilai, Y. Chisti, 2013. Repeated fed-batch production of xylitol by Candida magnoliae TISTR 5663, Journal of Chemical Technology & Biotechnology, 88, 1121–1129.

W. Tochampa, S. Sirisansaneeyakul, W. Vanichsriratana, P. Srinophakun, H. H. C. Bakker, S. Wannawilai, Y. Chisti, 2015, Optimal Control of Feeding in Fed-Batch Production of Xylitol, Industrial & Engineering Chemistry Research, 54, 1992-2000.

S. Ur-Rehman, Z. Mushtaq, T. Zahoor, A. Jamil, M. A. Murtaza, M.A., 2015, Xylitol: A Review on Bioproduction, Application, Health Benefits, and Related Safety Issues, Critical Reviews in Food Science and Nutrition, 55, 1514-1528.

D. H .Wolpert, W. G. Macready, 1997, No free lunch theorems for optimization, IEEE Transactions on Evolutionary Computation, 1, 1, 67-82