I COBBIND
I CONGRESSO BRASILEIRO DE BIOTECNOLOGIA INDUSTRIAL
XXIV SINAFERM ** XV SHEB ** XV ENZITEC FLORIANÓPOLIS ** 2024
Creating connections between biotechnology and industrial sustainability
August 25 to 28, 2024
Costão do Santinho Resort, Florianópolis, SC, Brazil
BIOPROCESS ENGINEERING

EFFECTS OF NONDIMENSIONALIZATION ON PHYSICS-INFORMED NEURAL NETWORK (PINN) APPLIED TO A LACTIC ACID PRODUCTION BIOREACTOR SIMULATION

Ivanildo J. Silva Jr[1]*, Amaro G. Barreto Jr[2] & Renan M. Frota[1]

[1] *Departamento de Engenharia Química, Universidade Federal do Ceará, Fortaleza, Ceará, Brazil*
[2] *Chemical and Biochemical Engineering Processes (EPQB), School of Chemistry— EQ/UFRJ, Universidade Federal do Rio de Janeiro (UFRJ), Cidade Universitária, Ilha do Fundão, Rio de Janeiro, Brazil*
** Corresponding author's email address: ivanildo@ufc.br*

ABSTRACT

Physics-Informed Neural Networks (PINN) is a framework capable of simulating physics-based systems with Neural Networks (NN). The present work explored the impact of nondimensionalization strategies (NdS) on the simulation of lactic acid production by *Lactobacillus casei* in a bioreactor using PINNs instead of traditional numerical methods. The optimal values of nondimensionalization factors for time, reactor volume, and concentrations of biomass, product, and substrate were determined. The scaling of nondimensionalized output variables between 0 and 10 produced significantly better results than nondimensionalization between 0 and 1, with Mean Absolute Deviation lower than 1.81. NdS were capable of reducing the error both on the training and extrapolation regions, improve the obedience to physical constraints and allowed more simple Neural Networks to simulate the reactor successfully.

Keywords: Bioreactor. PINN. Lactic Acid. Bioprocess.

1 INTRODUCTION

NN (Neural Network) is a class of highly flexible algorithms capable of working as universal approximators of any given function with inputs and outputs[1]. They can reproduce extremely complex phenomena. However, they often need a huge amount of data (big data) to be trained properly, and that is one of it's main disadvantages, alongside with possible huge computational costs. This can make the use of NNs very complex or expensive in some areas, like bioprocess engineering, since it requires many experiments to obtain enough data points that cost both time, technical expertise and material costs.

PINN (Physics-Informed Neural Networks) is a framework launched in 2018 that supports the optimization of NNs using a small amount of data (small data) [2]. One of the strongest points of PINNs is that they can be trained using equations and constraints based on physical, biological and chemical constraints. Thus, very little data despite the equation system that represents the phenomena studied is necessary. NNs produced using the PINN framework are trained using constraints based on the equations that model the system studied. They can be trained with virtually no experimental data, and also be used to predict data outside of the training range, allowing for extrapolation, prediction and even real-time process simulation. PINNs have been used successfully to simulate aerodynamics[3], adsorption processes[1], and cell diffusion[4]. While there are many Machine Learning (ML) studies applied to bioprocesses and bioreactors [5–8], PINNs are still needing further investigation in the bioprocess engineering field.

Bioprocesses are a challenge for PINNs because they frequently have variables that show multiple dependency. This makes the PINN solution to the equation system possibly more complex, requiring more time and technical knowledge to train, which may result in the use of PINNs being prohibitive. This work studies nondimensionalization strategies (NdS) applied to solve partial differential equations (PDEs) that represent the production of lactic acid by *L. casei* in a batch bioreactor. Lactic acid production was chosen due to abundant scientific literature data and high industrial and economical relevance [9].

2 MATERIAL & METHODS

A batch bioreactor for the production of lactic acid (LA) by *Lactobacillus casei* using lactose as a substrate was simulated using PINN with soft constraints on the programming language Python and the library DeepXDE [10]. The PINNs were trained on a computer with 16gb RAM and AMD Ryzen 5 4600G processor. Obtained PINN models have one input variable, t (time), and four output variables: X (biomass concentration), P (lactic acid concentration), S (lactose concentration) and V (reactor liquid volume). The error was calculated using Mean Absolute Deviation (MAD) of each output variable, comparing PINN and a reference numerical method (finite differences).

Equations 1 to 3 describe the simulated batch reactor and are are from another study [11]. The variables at t=0 were defined as: X= 1.15 $g.L^{-1}$, P= 6 $g.L^{-1}$, S=21.4 $g.L^{-1}$, V = 5 L. The parameters were obtained from the source [11] and are X_M=8 $g.L^{-1}$, P_M=90 $g.L^{-1}$, μ_{max}=0.265 h^{-1}, K_S = 0.72 $g.L^{-1}$, α=3.3, β=0.06 h^{-1}, Y_{PS}=0.682, m_S=0.03 h^{-1}, f=0.5, h=0.5. The PINN was trained with t between 0 and 10h, but simulated between 0 and 20h to evaluate extrapolation.

$$\frac{dX}{dt} = \frac{\mu_{max} S}{K_S + S}\left(1 - \frac{X}{X_m}\right)^f \left(1 - \frac{P}{P_m}\right)^h \quad (1)$$

$$\frac{dP}{dt} = \alpha \frac{dX}{dt} + \beta X \quad (2)$$

$$\frac{dS}{dt} = \frac{-1}{Y_{PS}} \frac{dP}{dt} - m_s X \quad (3)$$

The following assumptions were made: 1) the reactor is a closed system, 2) the reactor is perfectly mixed, and this mixing does not cause chemical, thermal or physical interference of any kind, 3) temperature, volume, pH and pressure are constant.

Each variable was nondimensionalized using the Equation 4. N represents dimensional variable, N_A is the nondimensional value of N and N_S the nondimensionalization factor of the given N variable. Additionally, outputs variable were tested using scale factors of x10 ($N_S \cdot 10$) and d10 ($N_S/10$). Table 1 describes the nondimensionalization factor applied to each variable.

$$N = N_A * N_S \quad (4)$$

Table 1 Nondimensionalization factors.

Variable	Meaning	Nondimensionalization factor	Value
t	Time	t_{sim} (Time of simulation)	20 h
X	Biomass concentration	X_M (maximum biomass concentration)	8 g . L^{-1}
P	Product concentration	P_M (maximum product concentration)	90 g . L^{-1}
S	Substrate Concentration	S_o (initial substrate concentration)	21.4 g . L^{-1}
V	Reactor liquid volume	V_{max} (maximum reactor volume)	5 L

3 RESULTS & DISCUSSION

Figure 1 shows the values of MAD versus NL (Neurons per layer) and HL (number of hidden layers) for each NdS applied to the simulated system. The color bar legend indicates the value of MAD. The common nondimensionalization produce average MAD of 12.54, while the dimensional version had a 2.92 MAD. The scaling x10 produced the highest error and the d10 scaling the lowest, with respective average MAD of 21.95 and 1.81. It can be concluded that the Nondimensional d10 version not only produced lower average MAD, but also allowed the central region of NL and HL to represent appropriately the reactor. This is significant because PINNs and NNs training can be time and knowledge consuming, and having a strategy that plainly increases the region of feasibility can allow more simple models do be trained, and optimal models to be found faster, easier and with lower computational cost. This can allow PINNs to be used more frequently in bioprocess simulation and also in realtime bioprocess simulation, which can increase both economical productivity and process security.

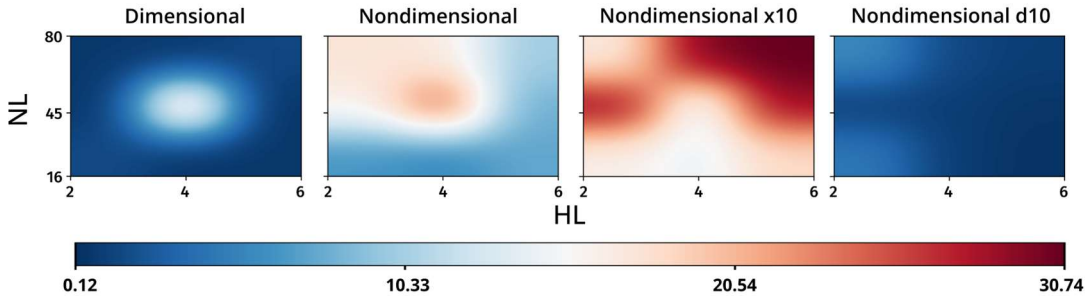

Figure 1 MAD vs NL (neurons per layer) and HL (number of hidden layers).

Figure 2 displays the output variables (X, P, S and V) for a PINN with NL=16, HL=6 without any NdS. PINN, Num (numerical method result) and Experimental (experimental data from source [11]) are displayed. The training region is of 10 h, only half of the simulation time. The model clearly violated physical constraints for the substrate (concentration of substrate is less than 0 after 12 h) and showed a significant error on volume prediction.

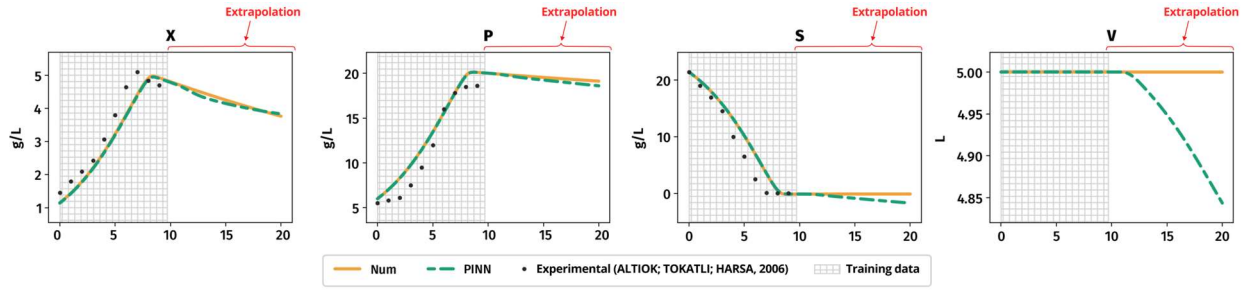

Figure 2 Output variables for a dimensional PINN with NL=16, HL=6.

Figure 3 displays the output variables (X, P, S and V) for a PINN with NL=16, HL=6 and Nondimensionalization d10. The model displayed good capabilities of extrapolation, respected physical and biological constraints and reproduced almost perfectly the data inside the training region. The only significant deviation was found on biomass concentration, after 15h, and was still relatively small.

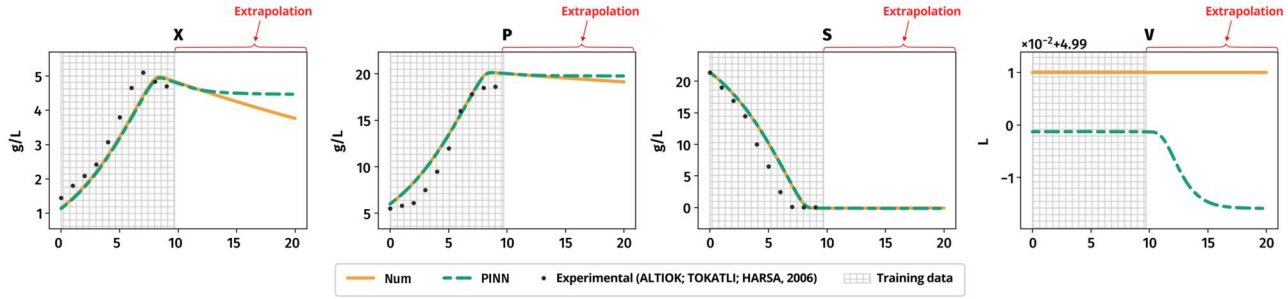

Figure 3 Output variables for a PINN with NL=16, HL=6 and Nondimensionalization d10.

4 CONCLUSION

NdS and PINNs were capable of simulating the production of lactic acid in a batch reactor. The nondimensionalization of output variables followed by scaling between 0 and 10 (d10) produced the best results, with an average MAD of 1.81. NdS were capable of improve the results obtained and allowed more simple (lower values of NL and HL) PINNs to produce simulations with lower errors in comparison to the dimensional PINNs.

REFERENCES

1. Santana, V. V. et al. A First Approach towards Adsorption-Oriented Physics-Informed Neural Networks: Monoclonal Antibody Adsorption Performance on an Ion-Exchange Column as a Case Study. ChemEngineering 6, 21 (2022).
2. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics 378, 686–707 (2019).
3. Mao, Z., Jagtap, A. D. & Karniadakis, G. E. Physics-informed neural networks for high-speed flows. Computer Methods in Applied Mechanics and Engineering 360, 112789 (2020).
4. Lagergren, J. H., Nardini, J. T., Baker, R. E., Simpson, M. J. & Flores, K. B. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. PLoS Comput Biol 16, e1008462 (2020).
5. Andrade Cruz, I. et al. Application of machine learning in anaerobic digestion: Perspectives and challenges. Bioresource Technology 345, 126433 (2022).
6. Bagherzadeh, F., Mehrani, M.-J., Basirifard, M. & Roostaei, J. Comparative study on total nitrogen prediction in wastewater treatment plant and effect of various feature selection methods on machine learning algorithms performance. Journal of Water Process Engineering 41, 102033 (2021).
7. Mateo Pérez, V., Mesa Fernández, J. M., Villanueva Balsera, J. & Alonso Álvarez, C. A Random Forest Model for the Prediction of FOG Content in Inlet Wastewater from Urban WWTPs. Water 13, 1237 (2021).
8. Mey, F., Clauwaert, J., Van Huffel, K., Waegeman, W. & De Mey, M. Improving the performance of machine learning models for biotechnology: The quest for deus ex machina. Biotechnology Advances 53, 107858 (2021).
9. López-Gómez, J. P., Alexandri, M., Schneider, R. & Venus, J. A review on the current developments in continuous lactic acid fermentations and case studies utilising inexpensive raw materials. Process Biochemistry 79, 1–10 (2019).
10. Lu, L., Meng, X., Mao, Z. & Karniadakis, G. E. DeepXDE: A Deep Learning Library for Solving Differential Equations. SIAM Rev. 63, 208–228 (2021).
11. Altıok, D., Tokatlı, F. & Harsa, Ş. Kinetic modelling of lactic acid production from whey by Lactobacillus casei (NRRL B-441). J. Chem. Technol. Biotechnol. 81, 1190–1197 (2006).

ACKNOWLEDGEMENTS

The authors thank CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support provided and an award of a scholarship.