Improving Value Chain Transparency in Bioprocess Life Cycle Assessment Using Surrogate Models

Michail Vousvoukis,[*1] Zhimian Hao,[*1] Polina Yaseneva,[1] Alexei A. Lapkin[1]
[*]*Equal contribution* [1]*Department of Chemical Engineering and Biotechnology, University of Cambridge*
Correspondence to: Polina Yaseneva py225@cam.ac.uk

1. Introduction

The rapid emergence of novel bioprocesses and bio-based products has increased the need for robust and fast Life Cycle Assessment (LCA) to determine whether these technologies deliver genuine environmental benefits [1]. However, most of these bioprocesses are in the early development stage and rely on confidential proprietary technologies [1], [2]. Both of these factors result in limited life cycle data availability, and ultimately missing life cycle impacts for bio-derived industrial chemicals. These data gaps propagate across value chains, hindering downstream industries from accurately quantifying environmental impacts via LCA [3], [4]. The main factor of this limitation is the risk of disclosing sensitive information related to upstream processing technologies, which restricts data sharing and reduces transparency [4]. To address this issue, we propose a surrogate modelling framework based on artificial neural networks (ANNs) trained on detailed first-principles process models calibrated with industrial data. This framework enables the prediction of environmental impacts via LCA without disclosing the underlying process, thus preserving the intellectual property.

2. Methodology and Results

The case study that was chosen to validate this framework was a biomass fractionation process developed and patented by Sonichem Ltd [5]. This process is based on the use of organic solvents (such as acids, alcohols, ketones, etc.) commonly known as "organosolv" [6] in combination with ultrasound technology. Based on the process flow diagrams and available data, a fully described process model was developed in Python and Aspen Plus. The process model was designed based on the typical biomass throughput per batch operation. The inputs to the model consisted of key design and operating parameters (e.g., oxalic acid concentration), reflecting the degrees of freedom available to the industrial partner. The model outputs were defined to be the life cycle inventory (LCI) flows such as the total electricity consumption. For the development of the surrogate model, an ANN was chosen since they are regarded as universal approximators and their capability to handle high dimensional input spaces [7]. A dataset comprising of 20,000 training simulations and 2,000 testing simulations was generated using the process model, within the operability limits specified by the company. The trained ANN was then used to predict the LCI flows directly from the process input parameters. Life Cycle Assessment (LCA) calculations for nine environmental impacts were then performed based on the predicted LCIs using a cradle-to-gate system boundary and applying the Environmental Footprint 3.0 methodology.
Through grid-search, the optimal ANN structure was identified as two hidden layers with 64 and 32 neurons, respectively. The trained model achieved an average coefficient of determination of $R^2 = 0.92$ across the predicted LCI flows. Consequently, the LCA results derived from surrogate-predicted LCIs exhibited a level of accuracy comparable to the process simulation calculations. The surrogate model successfully captured the nonlinear interaction between process units across different input combinations and identified the set of inputs associated with the lowest greenhouse gas emissions. Furthermore, both the mechanistic model and the surrogate-based LCA identified consistent environmental hotspots, demonstrating the reliability of the proposed framework. Overall, the proposed framework provides a robust methodology for predicting environmental impacts under data confidentiality constraints, supporting informed decision-making during early-stage process development.

Acknowledgments

The authors acknowledge funding of this work from UKRI (EPSRC) grants W031019/1. MV acknowledges funding of his PhD from the EPSRC Doctoral Landscape Award and Vice-Chancellor's Award scholarships. We acknowledge support from Sonichem Ltd with experimental data and technical expertise.

References

[1] M.-Y. Becker, N. Kohlheb, S. Hunger, S. Eschrich, R. Müller, and A. Aurich, 'Early-stage sustainability assessment of biotechnological processes: A case study of citric acid production', *Eng. Life Sci.*, vol. 20, no. 3–4, pp. 90–103, 2020, doi: 10.1002/elsc.201800198.

[2] K. Wowra, E. Hegel, A. Scharf, A. Grünberger, and K. Rosenthal, 'Estimating environmental impacts of early-stage bioprocesses', *Trends Biotechnol.*, vol. 41, no. 9, pp. 1199–1212, Sep. 2023, doi: 10.1016/j.tibtech.2023.03.011.

[3] M. Mba Wright *et al.*, 'Life Cycle Inventory Availability: Status and Prospects for Leveraging New Technologies', *ACS Sustain. Chem. Eng.*, vol. 12, no. 34, pp. 12708–12718, Aug. 2024, doi: 10.1021/acssuschemeng.4c02519.

[4] A. Stenzel and I. Waichman, 'Supply-chain data sharing for scope 3 emissions', *Npj Clim. Action*, vol. 2, no. 1, p. 7, Mar. 2023, doi: 10.1038/s44168-023-00032-x.

[5] 'Sonichem ultrasonic biorefinery technology', Sonichem. Accessed: Feb. 05, 2026. [Online]. Available: https://sonichem.com/

[6] D. Wei Kit Chin, S. Lim, Y. L. Pang, and M. K. Lam, 'Fundamental review of organosolv pretreatment and its challenges in emerging consolidated bioprocessing', *Biofuels Bioprod. Biorefining*, vol. 14, no. 4, pp. 808–829, 2020, doi: 10.1002/bbb.2096.

[7] Z. Hao, C. Zhang, and A. A. Lapkin, 'Efficient surrogates construction of chemical processes: Case studies on pressure swing adsorption and gas-to-liquids', *AIChE J.*, vol. 68, no. 6, p. e17616, 2022, doi: 10.1002/aic.17616.