Computers and Chemical Engineering 201 (2025) 109220

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

BioProcessNexus: An open-source platform for surrogate techno economic models

Tommaso De Santis [a,1], Matthias Medl [b,1,*], Peter Satzer [a], Gerald Striedner [a]

[a] Institute of Bioprocess Science and Engineering, BOKU University, Muthgasse 18, 1190 Vienna, Austria
[b] Institute of Statistics, BOKU University, Peter-Jordan-Straße 82/I, 1190 Vienna, Austria

ARTICLE INFO

Keywords:
Techno-economic analysis (TEA)
Surrogate-modeling
Enzymatic plastic recycling
Machine learning
Monte Carlo simulation
Open source

ABSTRACT

Techno-economic analysis (TEA) and life-cycle assessment (LCA) are essential tools for evaluating manufacturing processes, but the use of proprietary software creates barriers to accessibility and reproducibility. We present the BioProcessNexus, an open-source platform that democratizes access to process modeling through surrogate models trained on Monte Carlo data from proprietary TEA software. The platform facilitates model generation, analysis, and optimization while promoting standardization and collaboration across the scientific community. We demonstrate BioProcessNexus's capabilities through a comprehensive analysis of enzymatic PET recycling, comparing three surrogate modeling types: partial least squares, random forest, and Gaussian Process regression. Our analysis revealed that enzymatic PET recycling faces economic challenges, with a unit production cost of \$1.74 kg^{-1} TPA and an expected negative gross margin of -49.5 %. Sensitivity analysis identified feedstock cost and purification strategy as key areas for optimization. BioProcessNexus enables accessible, reproducible process modeling even when proprietary software was used for the initial model development. This approach advances the open innovation initiative and promotes transparent scientific collaboration while reducing barriers to advanced process modeling and optimization techniques. In this article, we will (i) introduce the BioProcessNexus platform and (ii) showcase a use case of an enzymatic PET recycling process.

1. Introduction

Conducting techno-economic and environmental modeling is crucial for newly developed and well-established manufacturing processes. In emerging processes, such modeling facilitates a deep understanding of the potential economic viability and environmental implications before bringing them into operation. This assessment aids in identifying areas for optimization, innovation, and cost reduction, guiding decision-makers toward sustainable and economically sound choices. For well-established manufacturing processes, ongoing modeling is equally crucial to address the evolving economic and environmental considerations landscape. It enables industries to adapt to changing market dynamics, regulatory requirements, and consumer preferences. For instance, in the context of the green transition, businesses must reduce their environmental footprint while remaining cost competitive (Maceno et al., 2018). This represents a major challenge that can be addressed via advanced techno-economic analyses (TEAs) and life-cycle analyses (LCAs) (Piekarski et al., 2013; Anastas and Lankey, 2000; Ögmundarson et al., 2020).

TEAs and LCAs pose challenges, including the high licensing cost of commonly used proprietary software, a lack of standardization hindering data transfer between platforms, and a steep learning curve (Cameron and Ingram, 2008). These tools often require significant financial investment and time to master. At the same time, the absence of standardized data analysis features within commonly used software complicates the extraction of meaningful insights from complex

Abbreviations: CAPEX, Capital Expenditure; EG, Ethylene Glycol; GP, Gaussian Process; GUI, Graphical User Interface; LCA, Life-Cycle Assessment; MSE, Mean Squared Error; NRMSE, Normalized Root Mean Squared Error; PBT, Polybutylene Terephthalate; PC-to-DFC, Purchase Cost to Direct Fixed Capital; PDO, 1,3-Propanediol; PEN, Polyethylene Naphthalate; PET, Polyethylene Terephthalate; PLS, Partial Least Squares; PMI, Process Mass Intensity; PTT, Polytrimethylene Terephthalate; RF, Random Forest; RMSE, Root Mean Squared Error; RSS, Residual Sum of Squares; SHAP, SHapley Additive exPlanations; TEA, Techno-Economic Analysis; TPA, Terephthalic Acid; VBA, Visual Basic for Applications.

* Corresponding author.
E-mail address: matthias.medl@boku.ac.at (M. Medl).
[1] Shared first author.

https://doi.org/10.1016/j.compchemeng.2025.109220
Received 10 March 2025; Received in revised form 19 May 2025; Accepted 1 June 2025
Available online 3 June 2025

datasets. On top of that, gathering data to build inventories requires a significant amount of time, effort, and resources. As the demand for sustainability and economic assessments grows among academics and businesses, overcoming these obstacles is vital for widespread adoption (Weule, 1993). Another significant challenge is the limited accessibility and shareability of TEA models developed using proprietary software. This limitation can hinder collaboration and the reproducibility of research findings. To address this, open-source initiatives such as BioSTEAM have emerged as promising alternatives (Cortes-Peña et al., 2020). BioSTEAM is a python-based open-source alternative for building and analyzing process models, particularly biorefineries. Such initiatives are integral for open and transparent research, as models generated with open-source software can be shared without limitation. However, despite their potential, open-source solutions like BioSTEAM come with their own set of challenges. For instance, they often require researchers to be proficient in programming, which can be a barrier for those without a coding background. Additionally, these platforms may lack the user-friendly interfaces and streamlined workflows that proprietary software typically offers, making them less accessible to some users. Furthermore, proprietary software often includes specialized unit operations and advanced features that are not yet available in open-source alternatives, limiting their applicability in certain contexts. For these reasons, many researchers still decide to use proprietary software.

In previous studies, surrogate modeling has been employed to reduce the computational time of TEA simulations or as an optimization strategy. (Romero et al., 2023; Taras and Woinaroschy, 2012; Yildiz and Sayar, 2020). In this study, we introduce the BioProcessNexus, which leverages surrogate modeling to facilitate the sharing of models originally developed using proprietary software. Specifically, Monte Carlo data generated from proprietary process modeling software is used to train surrogate models, which can then be freely shared. This approach allows researchers to efficiently create and share models using open-source software equipped with a user-friendly graphical user interface, ensuring accessibility and ease of use. Once the surrogate models are trained, the software facilitates advanced data analyses, enabling the identification of major cost drivers and the execution of comprehensive risk evaluations. The primary advantages of using surrogate models include their ease of sharing and the ability to streamline and extend analysis pipelines. The main goal of BioProcessNexus is to make techno-economic process models more accessible and to promote the publishing of surrogate models alongside scientific articles. BioProcessNexus will establish a comprehensive database of Monte Carlo data and corresponding surrogate models to achieve this. Academics and professionals from various fields will be encouraged to contribute their Monte Carlo data and surrogate models to this database, fostering a collaborative and extensive resource for the scientific community. This initiative aligns with the mandatory data publishing requirements that have become a prerequisite for scientific articles across various disciplines. By encouraging collaboration and leveraging the expertise of a broad user base, BioProcessNexus aims to enhance accessibility and applicability, ultimately contributing to the advancement of scientific research.

We use an enzymatic polyethylene terephthalate (PET) recycling process to demonstrate the capabilities and standard workflow of BioProcessNexus. Several enzymes have been discovered and optimized in recent years for the degradation of polyester polymers (Guo et al., 2024). In this study, we take the enzyme PHL7 as a reference, which was discovered and characterized by Sonnendecker et al. (Sonnendecker et al., 2022) and adopted by the EU consortium Enzycle (acib GmbH, 2020). Even though the enzymatic recycling technology is approaching the commercial stage, many uncertainties are currently present for key process parameters (Uekert et al., 2023).

2. Materials and methods

In this section, we will provide an overview of the techno-economic process model, describe the data generation pipeline for the Monte Carlo dataset associated with the techno-economic process model, outline the workflow for training the surrogate models, and detail the methods utilized for their analysis.

2.1. Techno-economic process model

A techno-economic analysis of the enzymatic PET recycling process was performed employing a process simulation framework based on fundamental mass and energy balance principles and economic evaluation; in this study, these principles were applied through the use of SuperPro Designer® v.12 (Intelligen, Inc.). SuperPro Designer is capable of simulating and optimizing various unit operations typically employed in bio-manufacturing. Detailed information about the equations utilized by SuperPro Designer for mass and energy balances, as well as economic calculations, can be found in the software manual (Intelligen Inc, 2021) . After constructing the flowchart that represents the enzymatic recycling process in batch mode (Fig. 1), the data provided by the Enzycle consortium were used as inputs for the process parameters (Table 1). The model assumes a greenfield scenario, which means that a new plant is built, and the relative depreciated capital costs are incorporated into the product unit cost. The databases of SuperPro Designer and other sources have been utilized to fill in any missing data.

The SuperPro Designer process model focused on three primary operations: raw materials storage, mixing units, and the depolymerization reaction, as illustrated in Fig. 1. The process was configured to operate in batch mode with 330 working days annually, without assuming any cycle slack time between batches. Moreover, the process was modelled at an industrial scale with the capability to recycle 57,791 MT of PET per year. This capacity was selected as it represents a commercially relevant scale that aligns with planned industrial installations, such as Carbioś first commercial enzymatic PET recycling plant (Mathe, 2022).

The model assumes the recycling of generic low crystallinity PET waste streams (such as PET trays) as feedstock. Due to the low crystallinity of these materials, which makes them readily susceptible to enzymatic degradation, no pretreatment step was required in the process design (Sonnendecker et al., 2022).

The purification section of the recycling process was simplified using a split operation to separate terephthalic acid (TPA, the main product), ethylene glycol (EG), and waste materials. This was necessary due to the lack of pilot-scale data and due to the availability of multiple purification methods. To manage these uncertainties, the purification section was treated as a variable unit cost in the economic model, adjustable by users. The cost range, set between $0.1 and $2 per kg of TPA (see Table 1), was used in a Monte Carlo analysis as agreed within the Enzycle consortium. To provide a reference, a techno-economic study by the BOTTLE Consortium estimated a purification cost of $0.42 per kg of TPA, incorporating activated carbon adsorption, crystallization, and distillation (Singh et al., 2021).

In this study, we selected seven key output metrics modeled with SuperPro Designer to ensure a comprehensive evaluation across economic, operational, and environmental categories. The economic indicators — gross margin [%], TPA unit cost excluding capital expenditure (CAPEX) [$ kg^{-1}], and TPA unit cost including CAPEX [$ kg^{-1}] — provide a clear overview of the manufacturing costs and the overall process profitability. The operational indicators — number of batches [batches $year^{-1}$] and total reactor volume [m^3] — offer insights into process scheduling and equipment sizing. Lastly, the environmental indicators — process mass intensity (PMI) with water [kg MT^{-1}] and PMI without water [kg MT^{-1}] — measure the resource intensity of the process.

Process mass intensity (PMI) is defined as the total mass of all input materials used per metric ton of product (Eq. (1)).

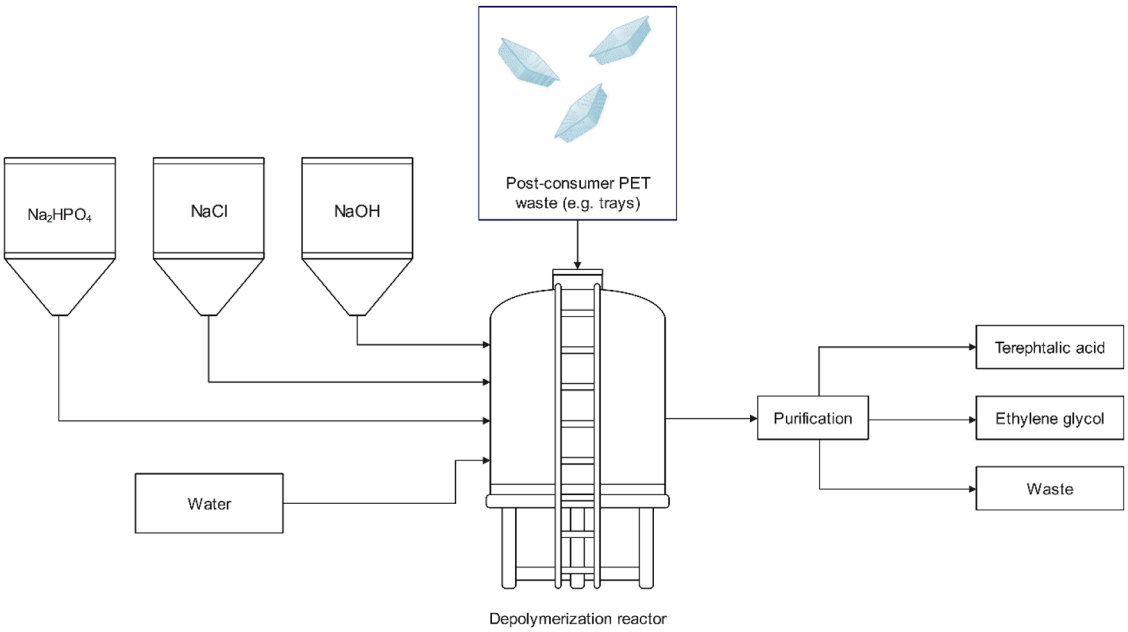

Fig. 1. Simplified process flow diagram of the enzymatic recycling process modelled in SuperPro Designer. Post-consumer PET waste is processed in a stirred depolymerization reactor along with process chemicals (Na_2HPO_4, NaCl, NaOH), the enzyme and water. The reaction products undergo purification to yield the main products terephthalic acid and ethylene glycol, while separating waste streams.

$$Total\ PMI = \frac{Total\ water,\ raw\ materials,\ consumables\ used\ in\ process\ (kg)}{Main\ product\ (TPA,\ kg)} \quad (1)$$

2.2. Monte Carlo dataset generation

A Monte Carlo dataset has been generated using the enzyme recycling model. This dataset was subsequently used to train surrogate models. The input parameters during the Monte Carlo sampling varied across wide parameter ranges to ensure that the surrogate models performed well under various parameter settings. The assumptions for the input parameter ranges can be found in Table 1 and were derived from a combination of literature review, market price databases, publicly available data, and, where public data was insufficient, from internal experimental results and supplier information obtained within the Enzycle consortium. To perform the Monte Carlo sampling, SuperPro Designer was integrated with Microsoft Excel through a component object model interface, a client service enabling communication between software applications such as Microsoft Excel to programmatically interact with SuperPro Designer for tasks like running simulations, retrieving data, and updating process parameters. This step was followed by defining the lower and upper limits for selected input variables and choosing the output metrics for the Monte Carlo dataset. Subsequently, a VBA macro was employed to execute the Monte Carlo sampling from uniform distributions. A summary of the chosen bounds of the uniform distributions can be found in Table 1. In total, 20,000 samples were generated this way. The generation of this 20,000-sample dataset required approximately 11.1 h of computational time on a standard desktop computer, averaging 2 s per simulation run. The scripts that were used for this purpose are consolidated within the Pro Monte Carlo Excel tool, which is accessible for download via https://github.com/mmedl94/bioprocessnexus.

2.3. Analysis of Monte Carlo data

All further analysis was performed using the BioProcessNexus graphical user interface (GUI) developed in Python 3.11 (Van Rossum and Drake, 2009). The source code of the graphical user interface (GUI), an executable launcher of it, and a comprehensive documentation of the software, including other case studies, can be found at https://github.com/mmedl94/bioprocessnexus. A brief description of previously published applied mathematical methods can be found in Appendix 1.

2.3.1. General mathematical notation

We refer to the Monte-Calo data set as $\mathscr{D} = [X,y]$, where $X \in \mathbb{R}^{n\times m}$ is an $n \times m$ feature matrix containing $i = 1, \ldots, n$ samples with $j = 1, \ldots, m$ features (variables) and $y \in \mathbb{R}^{n\times d}$ is the corresponding response matrix with $k = 1, \ldots, d$ features. x_i is the sample of X at index i, x_j the feature vector of X at index j and $x_{i,j}$ the value of X of feature j and sample i. For the surrogate models it is assumed that there is a functional relationship $f: \mathbb{R}^m \rightarrow \mathbb{R}^d$, mapping feature matrix X to response matrix y, with machine learning model $\hat{f}$ approximating f. $\hat{f}_k$ denotes the surrogate model estimating response vector y_k and $\hat{f}_k(X)$ being the model estimate for y_k based on feature matrix X.

2.3.2. Data preprocessing and generation of surrogate models

Initially, X and y have been standardized by subtracting the vector of column-wise means μ_j and μ_k and dividing by the vector of column-wise standard deviations σ_j and σ_k resulting in the standardized matrices X_{std} and y_{std} as shown in Eq. (2).

$$X_{j,\ std} = \frac{X_j - \mu_j}{\sigma_j}, y_{k,std} = \frac{y_k - \mu_k}{\sigma_k} \quad (2)$$

For the rest of the manuscript X_{std} and y_{std} are denoted as X and y unless mentioned otherwise to improve readability. Standardization has been performed as it is beneficial for generating PLS and GP regression models and searching optimal feature settings for minimizing or maximizing linear combinations of responses (see Appendix 2). Model predictions have been transformed to original scale prior to visualization, performance evaluation, and calculation of SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017).

All models shown in this study have been trained to predict individual response vectors. Thus, d separate models have been trained in case of d-dimensional response matrices. Three different model types have been generated: (a) partial least squares (PLS), (Wegelin, 2000) (b)

Table 1
Overview of the bounds of the uniform distributions that were used to sample the input parameters for the Monte Carlo simulation. All input parameters are varied independently, meaning changes in one do not affect the others. [1]TPA refers to terephthalic acid, the primary product from PET depolymerization. [2]EG stands for ethylene hlycol, a byproduct whose sales are considered a credit. Thus, its revenue per kg of TPA reduces the unit cost of TPA. [3]Purchase cost to direct fixed capital (PC-to-DFC), a user-specified factor used to estimate the capital expenditure (CAPEX) by multiplying it with the equipment purchase costs. [4]Refers to the depolymerization reaction. [5]Refer to the concentrations inside the depolymerization reactor where these ingredients are mixed. [6]Refers to purification of TPA.

Process parameters	Lower bound	Upper bound	Sources
Enzyme cost [$ kg^{-1}]	25	200	Enzycle consortium
Na_2HPO_4 cost [$ kg^{-1}]	0.4	0.7	echemi.com (ECHEMI Digital Technology Co. Ltd., 2025)
NaCl cost [$ kg^{-1}]	0.4	0.7	chemanalyst.com
NaOH cost [$ kg^{-1}]	0.4	0.7	echemi.com (ECHEMI Digital Technology Co. Ltd., 2025)
Water cost [$ kg^{-1}]	0.002	0.01	Water Price Index (Holindu GmbH, 2025)
PET feedstock cost [$ kg^{-1}]	0.2	0.7	Enzycle consortium, Singh et al. (Singh et al., 2021)
Selling price TPA [$ kg^{-1}][1]	0.5	2	chemanalyst.com (ChemAnalyst, 2025)
Selling price EG [$ kg^{-1}][2]	0.3	1	echemi.com (ECHEMI Digital Technology Co. Ltd., 2025)
Operator basic labor rate [$ hr^{-1}]	10	30	erieri.com (ERI Economic Research Institute Inc., 2025)
Depolymerization reaction time [hr]	1	96	Enzycle consortium
Reaction extent [0 to 1][4]	0.7	1	Enzycle consortium
PET loading [g L^{-1}][5]	100	300	Enzycle consortium
Enzyme loading [kg kg_{PET}^{-1}][5]	0.0003	0.006	Enzycle consortium
Na_2HPO_4 concentration [g L^{-1}][5]	0.1 (0.0007 M)	163.9 (1.15 M)	Enzycle consortium
NaCl concentration [g L^{-1}][5]	5.84 (0.1 M)	116.89 (2 M)	Enzycle consortium
NaOH mix-in mass ratio [0 to 1][5]	0.2	0.4	Enzycle consortium
Purification cost [$ kg^{-1}][6]	0.1	2	Singh et al. (Singh et al., 2021)
PC-to-DFC Factor (CAPEX)[2]	3	8	SuperPro Designer
Power cost [$ kWh^{-1}]	0.05	0.15	statistik.at (STATISTIK AUSTRIA, 2025)
PET fraction in feedstock [0 to 1]	0.6	0.97	Enzycle consortium, Singh et al. (Singh et al., 2021)
Depolymerization reaction temperature [°C]	25	90	Enzycle consortium
Batch throughput [kg $year^{-1}$]	10,000	180,000	Singh et al. (Singh et al., 2021)

random forest (RF) (Breiman, 2001) and (c) Gaussian process (GP) regression models (Rasmussen and Williams, 2008).

For all three model types, the scikit-learn implementations have been used (Pedregosa et al., 2011). The number of components of the PLS models was tuned as described below. The default hyperparameters were used for all RF models. For the GPR models, the final kernel consisted of adding white noise kernel with a constant kernel multiplied by a radial basis function kernel. The number of restarts of the log marginal likelihood optimizer was set to three.

To ensure that the models are not overfitting, cross-validation has been employed. Therefore, $\mathscr{D}$ has been split into two different subsets: the training and test set. The models have been trained with the training set, and the accuracy calculation and sensitivity analysis were performed on the test set. For data splitting, $\mathscr{D}$ has first been sorted so that y_k was in monotone increasing order. As a consequence of y being a d-dimensional matrix and each y_k resulting in a different sorting outcome, d different data splits were performed. One for each response. After sorting, every $5th$ element of X and y_k was shifted to the test set, while the remaining rows of X formed the training set. For PLS models, every $5th$ element of the training set was shifted into another subset, which was used to tune the number of latent components. The number of latent components that resulted in the smallest root mean squared error on this additional subset was then used to train the final PLS model (for that data split) on the full training set.

2.3.3. Performance evaluation of surrogate models

The predictive performance of each surrogate model $\widehat{f_k}$ was assessed using the mean squared error (MSE), root mean squared error (RMSE), and the normalized root mean squared error (NRMSE). All performance metrics presented in this study were calculated with test set data.

$$MSE_k = \frac{1}{n}\sum_{i=1}^{n}\left(y_{k,i} - \widehat{f}_k(x_i)\right)^2 \tag{3}$$

$$RMSE_k = \sqrt{MSE_k} \tag{4}$$

$$NRMSE_k = \frac{RMSE_k}{max(y_k) - min(y_k)} \tag{5}$$

In addition to assessing the general model performance, the performance of each model type with varying training set sizes was calculated. Therefore, only random subsets of varying sizes of the training set were used.

2.4. Software implementation

A comprehensive overview of the software can be found at https://github.com/mmedl94/bioprocessnexus, https://bioprocessnexus.readthedocs.io/en/latest/ and a brief overview can be found in Appendix 2. The GUI of the software was made with CustomTkinter, (Schimansky, 2024) the plotting was implemented with matplotlib, (Hunter, 2007) the data handling and processing was performed with numpy, (Harris et al., 2020) the surrogate models were generated with scikit-learn, (Pedregosa et al., 2011) probability distributions were fit with distfit, (Taskesen, 2020) the shapley values were computed with SHAP, (Shapley, 1953); (Lundberg and Lee, 2017) and parameter optimization was implemented with hyperopt. (Bergstra et al., 2011)

3. Results

To assess the validity of the surrogate models trained with the Monte Carlo data we computed their error on test set data. The test set errors of the surrogate models are summarized in Table 2. We can see that the GP model was the most accurate for all responses, except for the number of batches, for which the RF was the most accurate. The PLS model outperformed the RF model for the gross margin, the TPA unit cost excluding the CAPEX and including the CAPEX, while the RF model was better than the PLS model for the PMI with and without water, and the total reactor volume.

In Fig. 2 we can see a comparison of observations vs. predictions plots for the number of batches for all models. Ideally, the predictions agree precisely with the test set observations, which can be seen by the data points on the red diagonal line. The PLS model (Fig. 2A) was highly inaccurate, and it is apparent that the model could not capture the non-linearity of the relationship between the features and the number of batches. The GP model (Fig. 2C) showed a somewhat better performance but failed to fully capture the non-linearity, and the predictions were noisier. However, the RF model (Fig. 2B) predicted the number of batches accurately.

Table 2
The RMSE and NRMSE are shown for three model types and all seven responses. The performance metrics were calculated on test set data. The GP models resulted in the lowest RMSEs and NRMSEs for all responses except the "Number of batches" for which the Random Forst model was better.

Responses	Partial Least Square		Random Forest		Gaussian Process	
	RMSE	NRMSE [%]	RMSE	NRMSE [%]	RMSE	NRMSE [%]
Gross margin [%]	34.508	4.195	38.767	4.713	**12.843**	**1.561**
Number of batches [batches year^{-1}]	188.930	10.871	**0.288**	**0.017**	92.669	5.332
PMI with water [kg^{-1}]	1065.200	5.604	225.260	1.185	**115.170**	**0.606**
PMI without water [kg^{-1}]	245.410	4.046	121.350	2.001	**22.480**	**0.371**
Total reactor volume [m^3]	216.300	5.395	32.754	0.817	**17.794**	**0.444**
TPA unit cost excl. CAPEX [$ kg^{-1}]	0.233	3.607	0.302	4.662	**0.034**	**0.525**
TPA unit cost incl. CAPEX [$ kg^{-1}]	0.387	4.504	0.425	4.953	**0.148**	**1.722**

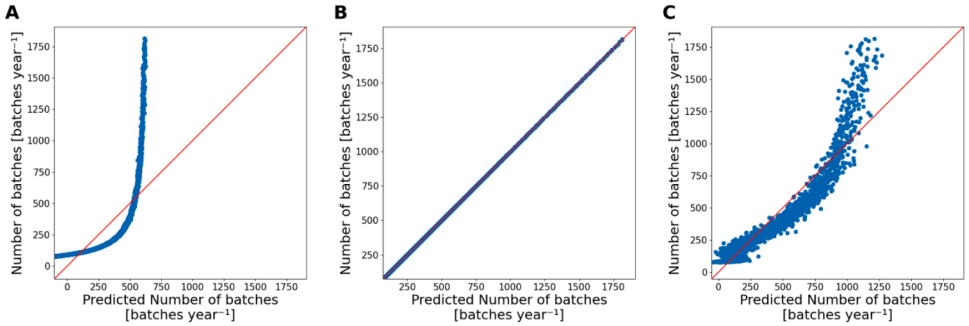

Fig. 2. Observations vs. predictions plots of the number of batches response for all three model types (A: PLS, B: RF and C: GP). The PLS model was unable to capture the non-linearity in the data, the GP model performed better, but still was inadequate, and the predictions of the RF model were highly accurate.

Fig. 3 shows the observation vs. prediction plots for the GP models for all other responses. It confirms that the GP models are capable of accurately predicting the responses (except the number of batches). The data points are very close to the diagonal line for the PMI with water (Fig. 3B), PMI without water (Fig. 3C), total reactor volume (Fig. 3D), and the TPA unit cost excluding CAPEX (Fig. 3E). For the gross margin we can see some heteroscedasticity with larger deviations towards larger negative values. The few data points off the diagonal can be neglected considering that there are 4000 data points within the test set. The same goes for the TPA unit cost including CAPEX (Fig. 3F); settings with larger values resulted in a larger variance in the residuals.

In the next step, we analyzed whether the number of observations in the Monte Carlo dataset was sufficient or if additional observations should be drawn. To do this, we artificially reduced the number of observations in the training set while keeping the test set constant. This approach allows us to observe trends in model performance scaling with

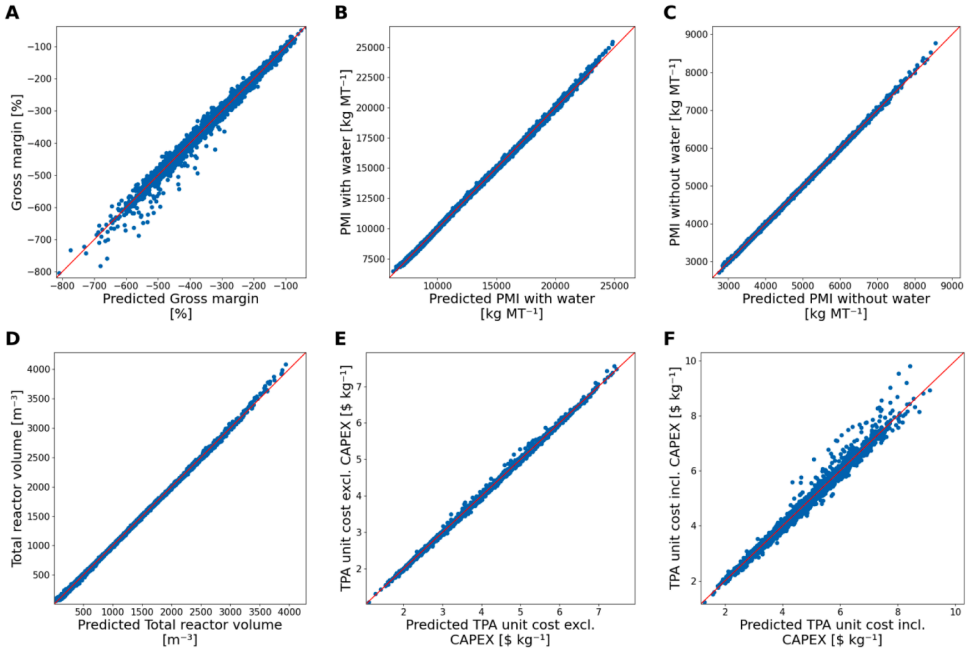

Fig. 3. Observation vs. prediction plots of the (A) gross margin (%), (B) PMI with water (kg^{-1}), (C) PMI without water (kg^{-1}), (D) total reactor volume (m^3), (E) TPA unit cost excluding CAPEX ($ kg^{-1}) and the (F) TPA unit cost including CAPEX ($ kg^{-1}). The model predictions were very accurate for most responses but showed some acceptable disagreement regarding the gross margin and the TPA cost including CAPEX.

an increasing number of observations. The results are visualized in Fig. 4.

The scaling performance plots reveal several insightful trends across different response variables for the models (Fig. 4). In general, both GP and RF models show noticeable performance improvements as the sample size increases, while the performance of the PLS models remains relatively similar, indicating limited responsiveness to additional data. This suggests that PLS may have a more restricted capacity to leverage extra observations effectively in complex scenarios. However, the stability of PLS performance at lower observation counts also means that it may be a more efficient choice in cases where data and computational resources are limited.

For GP and RF models, we observe that the performance improvements start to flatten with an increasing number of training observations, indicating that the dataset size used in this study is likely sufficient. If the RMSE continued to decrease noticeably with larger observation fractions, it would suggest a need to increase the Monte Carlo dataset size to achieve satisfactory performance. Thus, this analysis provides a basis for determining the optimal dataset size, ensuring efficiency and model accuracy without drawing excessive additional observations.

Since we have ensured that the surrogate models are of high quality, we can now investigate their properties in more detail. We have calculated the SHAP values for all models and generated beeswarm plots for all responses. We have selected beeswarm plots for two interesting process variables for further analysis: the gross margin (Fig. 5A) and the PMI without water (Fig. 5B).

The beeswarm plot for the gross margin (Fig. 5A) reveals that almost all process variables have some influence, with varying degrees of impact. The five most influential variables are the purification cost of the TPA, the cost of the PET feedstock, the depolymerization reaction time, the PET loading in the reactor, and the extent of the depolymerization reaction. The color coding of the individual data points in the beeswarm plot indicates whether a process variable of a high (red) or low (blue)

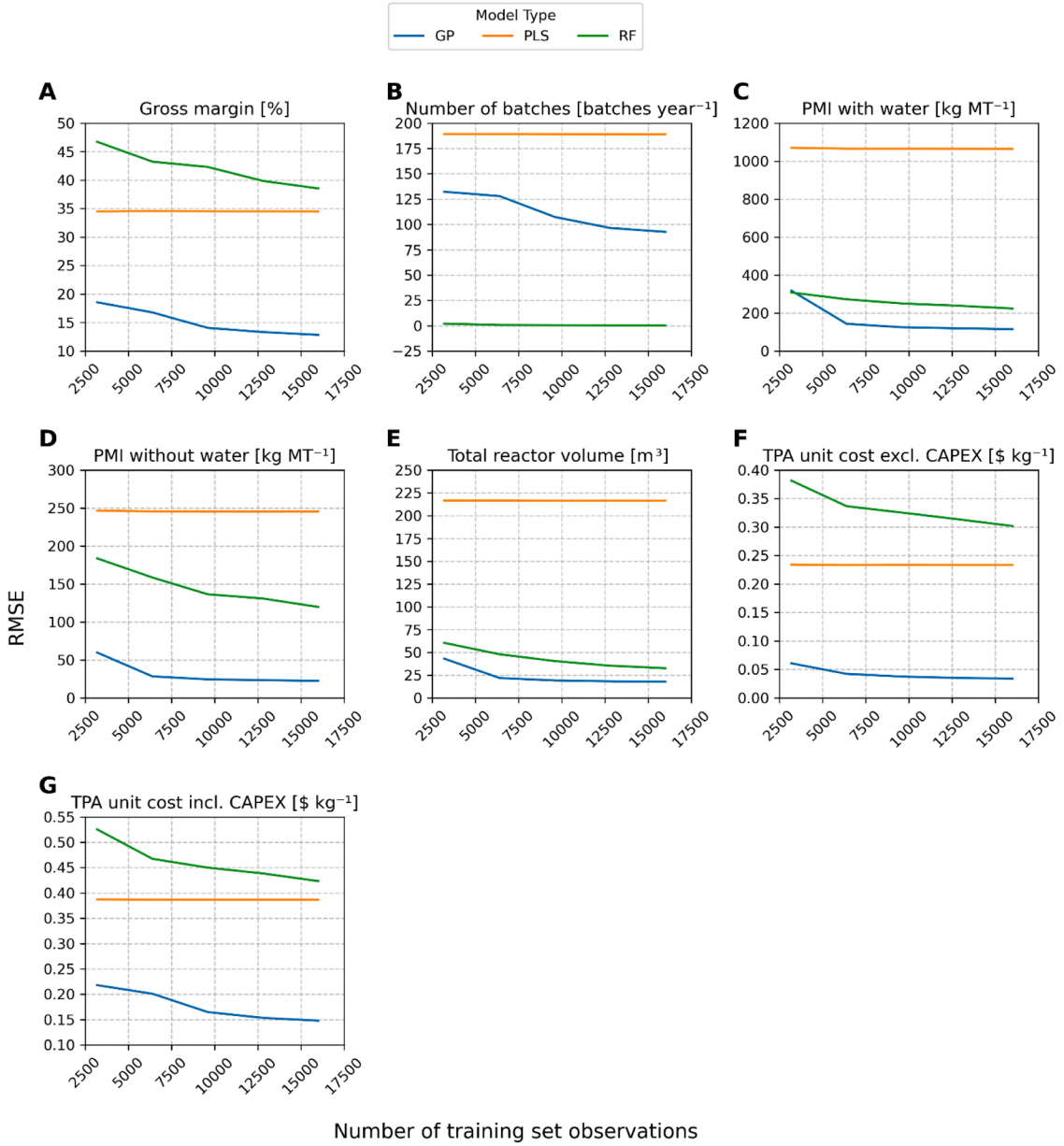

Fig. 4. Overview of the scaling behavior of GP, PLS and RF models with the number of training observations for all responses; (A) the gross margin, (B) the number of batches, (C) the PMI with water, (D) the PMI without water, (E) the total reactor volume, (F) the TPA unit cost excluding CAPEX and (G) the TPA unit cost including CAPEX.

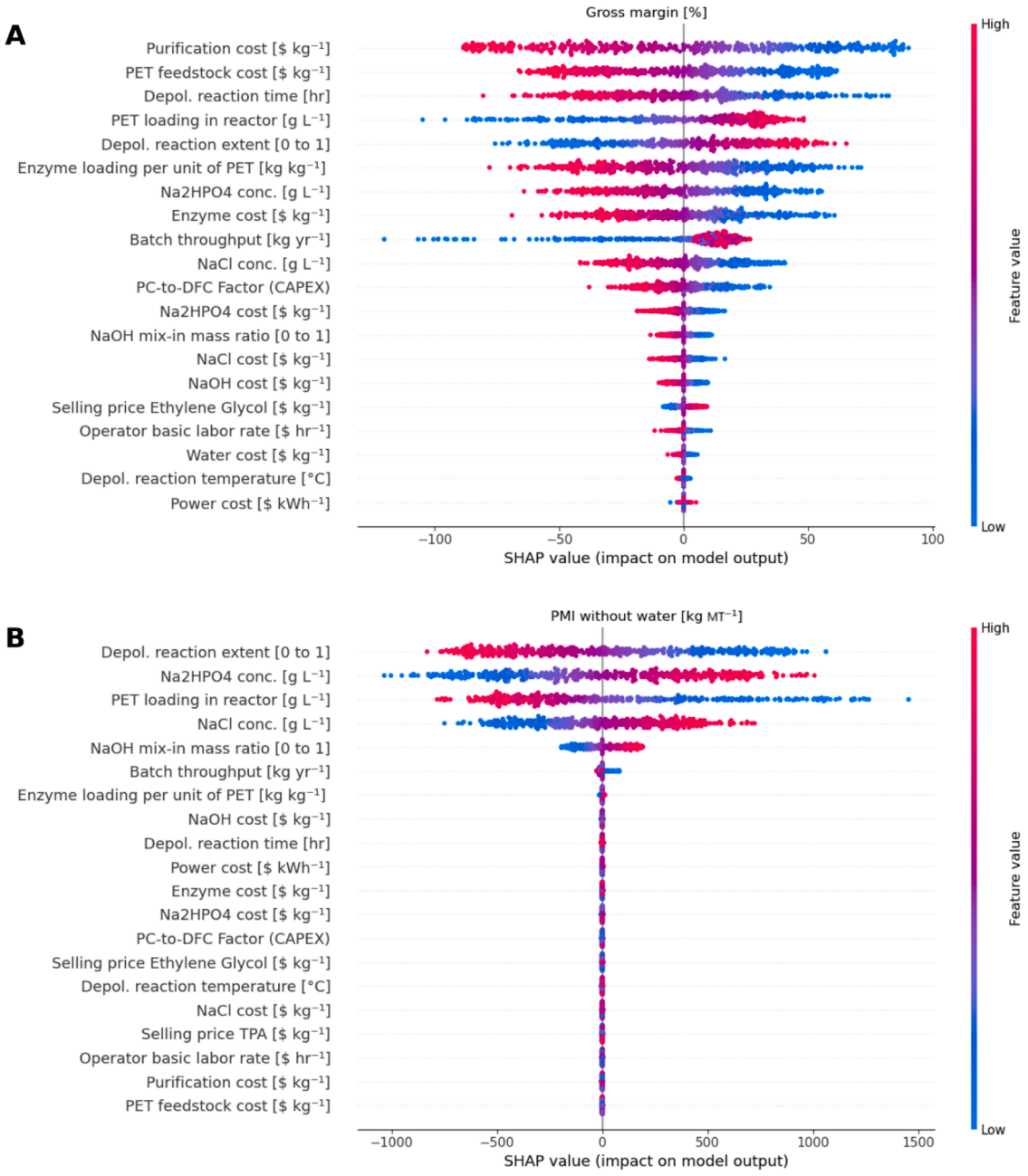

Fig. 5. Beeswarm plots of the SHAP values for (A) the gross margin and (B) the PMI without water, visualizing the influence of multiple process variables on the respective responses. Data points with high feature values are indicated in red, while those with low feature values are shaded in blue. The plot for the gross margin shows that almost all process variables have an influence, whereas the plot for the PMI without water indicates that only six variables have a noteworthy impact.

value resulted in that specific data point, while the position of the data point on the x-axis indicates the corresponding SHAP value.

For instance, the purification cost of the TPA shows negative SHAP values when the feature value is high, meaning that high purification costs result in a reduction of the gross margin by roughly 80 %. It is also interesting to observe the pattern for the PET loading in the reactor. High values of PET loading result in a higher gross margin, but at generally high values, it does not influence the gross margin much. However, extremely low values of PET loading result in a more pronounced decrease in the gross margin. This indicates that very low values must be avoided at all costs, while moderately low values arent as problematic. We can also see that costs for power, water, and labor cost have a relatively small impact on the gross margin. Here, it must be noted that the chosen lower and upper bounds of the individual process parameters greatly influence the importance of the parameters on the response. For example, the bounds of the labor rate of labor cost were assumed to be 10 and 30 $ hr^{-1}. If they were changed to 10 and 100 $ hr^{-1} the influence of the parameter would naturally increase as well.

In the beeswarm plot for the PMI without water, we can see that it is important to focus on high values for the extent of the depolymerization reaction and the PET loading, whereas low values are better for the Na2HPO4 and NaCl concentration. The other parameters had negligible effects on the PMI without water.

The training dataset was generated using uniform distributions with wide parameter ranges (Table 1), ensuring that the models can be applied to many scenarios without the need for extrapolation. However, these uniform distributions dont necessarily represent realistic feature distributions. Therefore, another Monte Carlo sampling has been

performed, with more realistic feature distributions, and the surrogate models were sampled instead of the SuperPro Designer model. Triangular distributions with more realistic, process-specific boundaries were used (Table 3). Notable refinements in selected process parameters include the narrowing of reaction conditions, with depolymerization temperature converging around 65 °C (range: 25–70 °C) compared to the training range of 25–90 °C. Similarly, reaction extent expectations were elevated and narrowed for the resampling (0.9–1.0) compared to the training range (0.7–1.0), reflecting more optimistic process performance assumptions. Economic parameters show significant focusing of ranges, particularly in the purchase cost to direct fixed capita factor (used to estimate CAPEX), which was narrowed from a uniform distribution between 3 and 8 to a triangular distribution centered at 5.8 (range: 3–6). The selling price of TPA was adjusted to have a most likely value of 1.07 $ kg^{-1}, compared to the uniform distribution between 0.5–2 $ kg^{-1} used for model training.

Table 3
Process parameter boundaries used for the surrogate model Monte Carlo simulation. Each parameter follows a triangular probability distribution defined by lower bound, most likely value, and upper bound. The boundaries represent a refined, more realistic operating space compared to the uniform distributions used in training and their values have been approved by the Enzycle consortium.

Process parameters	Lower bound	Mode	Upper bound	Sources
Enzyme cost [$ kg^{-1}]	25	100	200	Enzycle consortium
Na_2HPO_4 cost [$ kg^{-1}]	0.4	0.52	0.7	echemi.com (ECHEMI Digital Technology Co. Ltd., 2025)
NaCl cost [$ kg^{-1}]	0.4	0.52	0.7	chemanalyst.com
NaOH cost [$ kg^{-1}]	0.4	0.57	0.7	echemi.com (ECHEMI Digital Technology Co. Ltd., 2025)
Water cost [$ kg^{-1}]	0.002	0.005	0.01	Water Price Index (holidu.com) (Holindu GmbH, 2025)
PET feedstock cost [$ kg^{-1}]	0.2	0.2	0.5	Enzycle consortium, Singh et al. (Singh et al., 2021)
Selling price TPA [$ kg^{-1}]	1.07	1.07	2	chemanalyst.com (ChemAnalyst, 2025)
Selling price EG [$ kg^{-1}]	0.3	0.68	1	echemi.com (ECHEMI Digital Technology Co. Ltd., 2025)
Operator basic labor rate [$ hr^{-1}]	10	15	20	erieri.com (ERI Economic Research Institute Inc., 2025)
Depolymerization reaction time [hr]	12	24	96	Enzycle consortium
Reaction extent [0 to 1]	0.90	0.95	1	Enzycle consortium
PET loading [g L^{-1}]	100	200	250	Enzycle consortium
Enzyme loading [kg kg_{PET}^{-1}]	0.0003	0.0006	0.001	Enzycle consortium
Na_2HPO_4 concentration [g L^{-1}]	0.1 (0.0007 M)	24.52 (0.173 M)	70.98 (0.5 M)	Enzycle consortium
NaCl concentration [g L^{-1}]	5.84 (0.1 M)	54.29 (0.929 M)	54.29 (0.929 M)	Enzycle consortium
NaOH mix-in mass ratio [0 to 1]	0.2	0.2	0.4	Enzycle consortium
Purification cost [$ kg^{-1}]	0.1	0.42	1	Singh et al. (Singh et al., 2021)
PC-to-DFC Factor (CAPEX)	3	5.8	6	SuperPro Designer
Power cost [$ kWh^{-1}]	0.05	0.1	0.15	statistik.at (STATISTIK AUSTRIA, 2025)
PET fraction in feedstock [0 to 1]	0.6	0.75	0.97	Enzycle consortium, Singh et al. (Singh et al., 2021)
Depolymerization reaction temperature [°C]	25	65	70	Enzycle consortium
Batch throughput [kg $year^{-1}$]	80,000	93,354	180,000	Singh et al. (Singh et al., 2021)

The differences between the probability distributions of the features used for the initial Monte Carlo sampling and the resampling using the surrogate model are reflected in shifts in the response distributions (Fig. 6). The total reactor volume distribution undergoes a notable transformation between datasets. The training distribution shows a right-skewed pattern with a peak around 1000 m^3 and extends to 4000 m^3 (Fig. 6E), while the surrogate model generates a more concentrated distribution centered at 1040 m^3 (Fig. 6E). This concentration effect aligns with the narrowed batch throughput range in the surrogate model (triangular distribution between 80,000–180,000 kg $year^{-1}$ and a mode at 93,354 kg $year^{-1}$) compared to the training range (uniform distribution between 10,000–180,000 kg $year^{-1}$).

The number of batches per year reveals stark differences between the datasets (Fig. 6B). The training data exhibits a right-skewed distribution with a sharp peak at approximately 85 batches $year^{-1}$, while the resampled response peaks at 132 batches $year^{-1}$.

The Process Mass Intensity (PMI) metrics show systematic shifts that reflect the refined input parameters. The PMI with water shifts from a peak of 9225 kg^{-1} in the training data to 9304 kg^{-1} in the surrogate model (Fig. 6D), maintaining similar right-skewed characteristics but with a more pronounced tail in the surrogate model. The PMI without water shows a more substantial change, with the training distribution centered around 4266 kg^{-1} shifting to a more concentrated distribution around 3361 kg^{-1} in the surrogate model (Fig. 6C). This shift can be attributed to the optimization of Na2HPO4 and NaCl concentrations, which were given more focused ranges in the surrogate model (0.0007–0.5 M and 0.1–0.93 M, respectively) compared to the training ranges (0.0007–1.15 M and 0.1–2 M).

The economic parameters exhibit marked improvements in the surrogate model distributions. The TPA unit cost excluding CAPEX shows a shift from a distribution centered at 3.79 $ kg^{-1} to approximately 1.80 $ kg^{-1} in the surrogate model (Fig. 6F). Similarly, the TPA unit cost including CAPEX moves from 4.29 $ kg^{-1} to 2.11 $ kg^{-1} (Fig. 6G), with both surrogate distributions showing more symmetric and concentrated patterns. The gross margin distribution shifted from a mode of −301 % in the training dataset to −97 % in the surrogate model (Fig. 6A). This substantial improvement is largely a result of the more optimistic selling price assumptions (TPA selling price centered at 1.07 $ kg^{-1}) and narrowed operating cost ranges, particularly enzyme cost (triangular distribution centered at 100 $ kg^{-1}) and purification cost (centered at 0.42 $ kg^{-1}). Both distributions maintain approximately symmetric characteristics, but the surrogate model distribution shows a markedly reduced spread, suggesting more consistent economic performance under the refined operating conditions.

Due to the shifts in the response distributions depicted in Fig. 6, we reevaluated the surrogate models' errors. Specifically, we recalculated the responses using the original SuperPro Designer model and compared them to the surrogate models' predictions. The model errors are visualized in Fig. 7 and summarized in Table 4.

Fig. 7 shows that the respective models slightly underestimates the gross margin (Fig. 7A) and overestimates the TPA cost, including and excluding CAPEX (Fig. 7F and G). When comparing the RMSEs and NRMSEs calculated with the resampled dataset to those calculated with the stratified test set (Table 4), we observe that they are similar for all responses except for the TPA unit cost excluding CAPEX, where the RMSE is approximately four times larger. It is noteworthy that the models for the gross margin and TPA unit cost including CAPEX are slightly less accurate despite similar RMSE and NRMSE values. This is

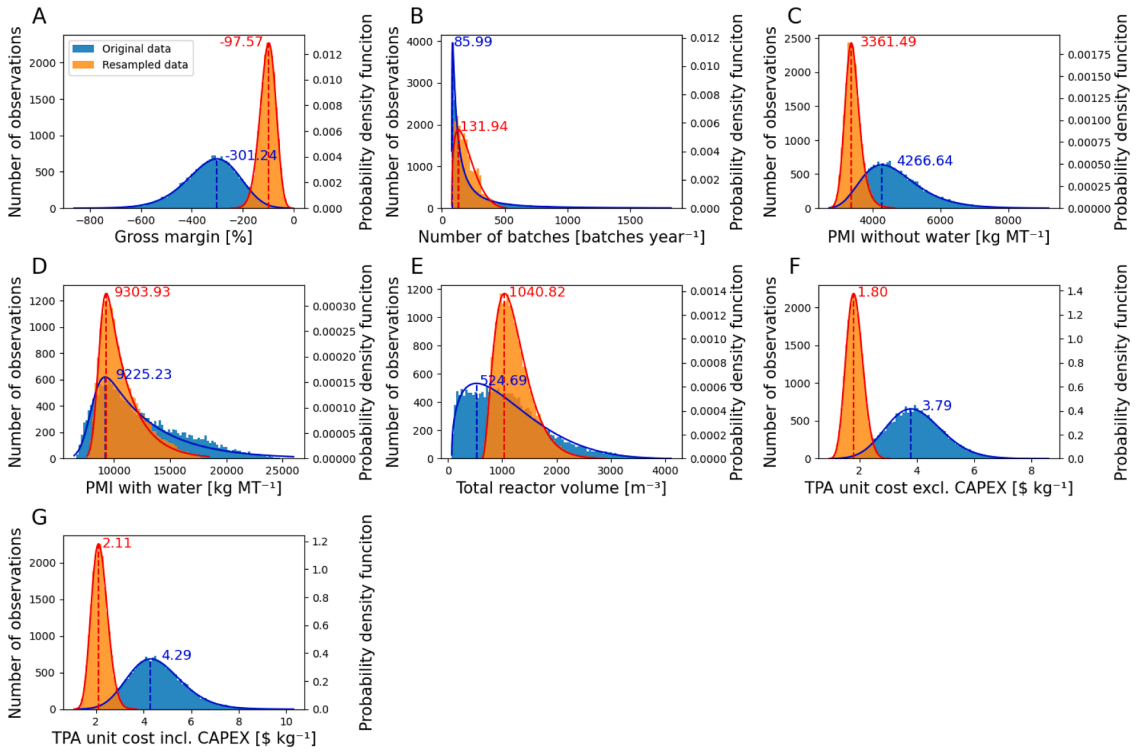

Fig. 6. Comparison of probability distributions of the training Monte Carlo dataset generated using uniform input distributions (blue) and the surrogate model responses generated using triangular input distributions with refined parameter ranges (orange). (A) Gross margin (%), (B) Number of batches (batches year^{-1}), (C) PMI without water [kg^{-1}], (D) PMI with water [kg^{-1}], (E) Total depolymerization reactor volume [m^3], (F) TPA unit cost excluding CAPEX [$ kg^{-1}], and (G) TPA unit cost including CAPEX [$ kg^{-1}]. The fitted probability density functions are indicated by solid lines and modes by dashed lines.

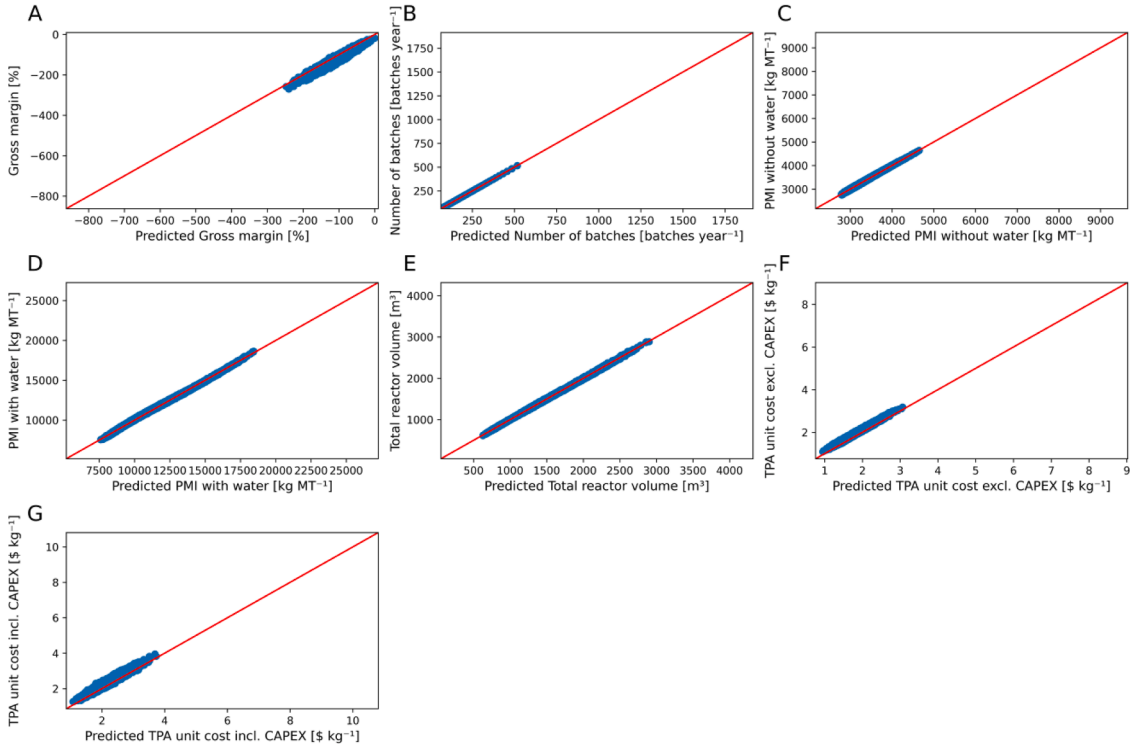

Fig. 7. Observations vs. predictions plots of the resampled dataset of the (A) gross margin (%), (B) number of batches per year, (C) PMI with water (kg^{-1}), (D) PMI without water (kg^{-1}), (E) total reactor volume (m^3), (F) TPA unit cost excluding CAPEX ($ kg^{-1}) and the (G) TPA unit cost including CAPEX ($ kg^{-1}). The model predictions correlate strongly with the observations indicating high model performance. The respective models slightly overestimate the gross margin and underestimate the TPA costs including and excluding CAPEX.

Table 4
Two performance metrics (RMSE and NRMSE) of the mixed model are shown for all seven responses. The performance metrics were calculated by comparing the responses of the resampled dataset with the corresponding responses computed with SuperPro Designer. The NRMSE is normalized with the original variable ranges of the responses.

Responses	Stratified test set		Resampled dataset	
	RMSE	NRMSE [%]	RMSE	NRMSE [%]
Gross margin [%]	12.843	1.561	15.691	1.820
Number of batches [batches year^{-1}]	0.288	0.017	0.209	0.012
PMI with water [kg^{-1}]	115.170	0.606	96.600	0.495
PMI without water [kg^{-1}]	22.480	0.371	17.172	0.265
Total reactor volume [m^3]	17.794	0.444	10.891	0.269
TPA unit cost excl. CAPEX [$ kg^{-1}]	0.034	0.525	0.143	1.872
TPA unit cost incl. CAPEX [$ kg^{-1}]	0.148	1.722	0.170	1.854

due to the response values of the resampled dataset being numerically smaller, leading to similar errors being larger in comparison. The reduced accuracy likely stems from the response distribution being at the edges of the response distributions of the original training dataset. To mitigate this effect, one could increase the number of observations in that region of the training dataset or broaden the parameter ranges of the features to align the response distributions of the resampled dataset closer to the center of the response distributions of the training dataset.

However, it is essential to note that the overall model performance remains high. Additionally, the predictions of the original SuperPro Designer model are rough estimations, and minor deviations from them are less influential in the broader context. For instance, in the case of the gross margin, it is more critical to understand which process features influence it and to what extent, and to recognize that it is generally negative, rather than having perfectly accurate predictions at each specific setting.

Another feature of BioProcessNexus is the interactive histogram. It consists of two major elements: a display that shows a histogram of the Monte Carlo data (Fig. 8, left side) and a set of sliders corresponding to the features, which allow for subsetting the data displayed in the histogram (Fig. 8, right side). In Fig. 8A we show a histogram of the gross margin. The sliders in Fig. 8A weren't moved, and therefore, the gross margin values of the whole dataset are plotted. In Fig. 8B, the slider of the depolymerization reaction time has been moved. To be more precise, the left part of the slider was pushed towards the center. The values beside the slider say "49.51, 96.10". This means that only datapoints of the Monte Carlo dataset where the depolymerization reaction time was between 49.51 hr and 96.10 hr are considered in the histogram.

Furthermore, we can see an estimation of the probability distribution (a beta distribution) of the gross margin of the whole dataset as a red line. This estimation was started from within the GUI. The probability distribution shown in Fig. 8B is also calculated with the whole dataset. Therefore, we can see in Fig. 8B, how changing the subsetting parameters for the depolymerization reaction time shifts the distribution of the gross margin to the left.

We can also see an area shaded in red in Fig. 8A. This results from integrating the fitted probability distribution between a gross margin of −50 and 0. The integration revealed a 3.4 % probability of the gross margin being in this range given the assumptions of the feature distributions. This analysis can easily be performed from within the GUI and is essential for performing the modeled process risk analysis.

In the remaining section, we will focus more on the techno-economic analysis itself and less on the BioProcessNexus and its features.

Overall, the techno-economic analysis revealed that the total capital investment required for the greenfield plant was estimated at $30.8 m, with direct fixed capital costs accounting for $26.7 m. Given the process design parameters, the plant can process PET feedstock at 57,791 MT yr^{-1}, producing 26,979 MT yr^{-1} of terephthalic acid (TPA) as the main product.

The annual operating cost analysis showed total expenses of $47.0 m, with raw materials representing the largest share at 63.3 % ($29.7 m) of the total operating costs (Fig. 9B). The main cost drivers among raw materials were PET feedstock (41.7 %), NaCl (20.8 %), and NaOH (16.6 %). The process required relevant quantities of enzymes, contributing 8.8 % to the raw material costs at $2.6 m annually. Facility-dependent costs and purification expenses were the next most influential operating cost categories, accounting for 10.7 % ($5.0 m) and 24.1 % ($11.3 m) of the total operating costs, respectively. Utilities consumption was relatively modest, with annual costs of $204,120, primarily driven by steam usage (83.6 % of utility costs) for maintaining the process temperature during depolymerization.

Overall, the economic evaluation indicated challenging profitability metrics. With a unit production cost of $1,74 kg^{-1} TPA (Fig. 9A) and a selling price of $1,07 kg^{-1} TPA, the process operated at a negative gross margin of −49.5 %. The annual revenue from TPA sales was projected at $28.9 m, with additional credits of $3.8 m from recovered ethylene glycol (valued at $0.68 kg^{-1}).

The process included very large equipment, including a stirred reactor with a volume of 866.07 m^3 and multiple silos ranging from 649 to 7562 m^3. An analysis of the economics of scale of the process equipment can be found in Appendix 3. The 24-hour depolymerization step occurring in the stirred reactor represents the primary bottleneck of the process, impacting the overall throughput and batch cycle time. The process mass intensity (PMI) metrics revealed substantial material requirements, with a total PMI of 9493.63 kg^{-1} of when including water consumption, and 3293.09 kg^{-1} when excluding it. When analyzing the material intensity per kg of main product we can see that 2295.08 kg PET feedstock were required per kg of product. The other process chemicals that were needed in large quantities were NaCl (436.16 kg^{-1}), NaOH (321.31 kg^{-1}), and Na_2HPO_4 (197.75 kg^{-1}). Water usage was particularly intensive at 6200.54 kg^{-1}. Cleaning-in-place (CIP) operations required additional material inputs through CIP-caustic (26.15 kg^{-1}) and CIP-acid (15.68 kg^{-1}) solutions, while enzyme usage was relatively low at 0.96 kg^{-1}.

4. Discussion

In this section, we will first discuss BioProcessNexus itself and then the enzymatic PET recycling process.

Techno-economic modeling is a cornerstone for assessing and optimizing biotechnological processes, enabling economic feasibility and environmental impact evaluations. However, the limited accessibility of TEA models generated with proprietary software is problematic. The high cost of licenses restricts access and hinders reproducibility and collaboration across the scientific community. Moreover, multiple proprietary software solutions, each offering different tools for analyzing techno-economic models, further complicates the comparison of analyses and results. These constraints create bottlenecks in scientific progress, often leading to resource-intensive production processes that unnecessarily burden the environment, consumers, and shareholders.

The BioProcessNexus project addresses these challenges by providing easy-to-use open-source software that facilitates the generation, analysis, and optimization of surrogate models, a database for sharing surrogate models and Monte Carlo datasets, and tutorials to guide users. The core idea of the BioProcessNexus is to perform Monte Carlo sampling with the original techno-economic model developed using proprietary software and then train a surrogate model based on the Monte Carlo data. These surrogate models can then be used freely and be analyzed through a standardized and extendable pipeline, allowing for fair comparisons across models stemming from different software platforms. This ensures that advanced modeling and optimization techniques are accessible to a broad audience without cost and

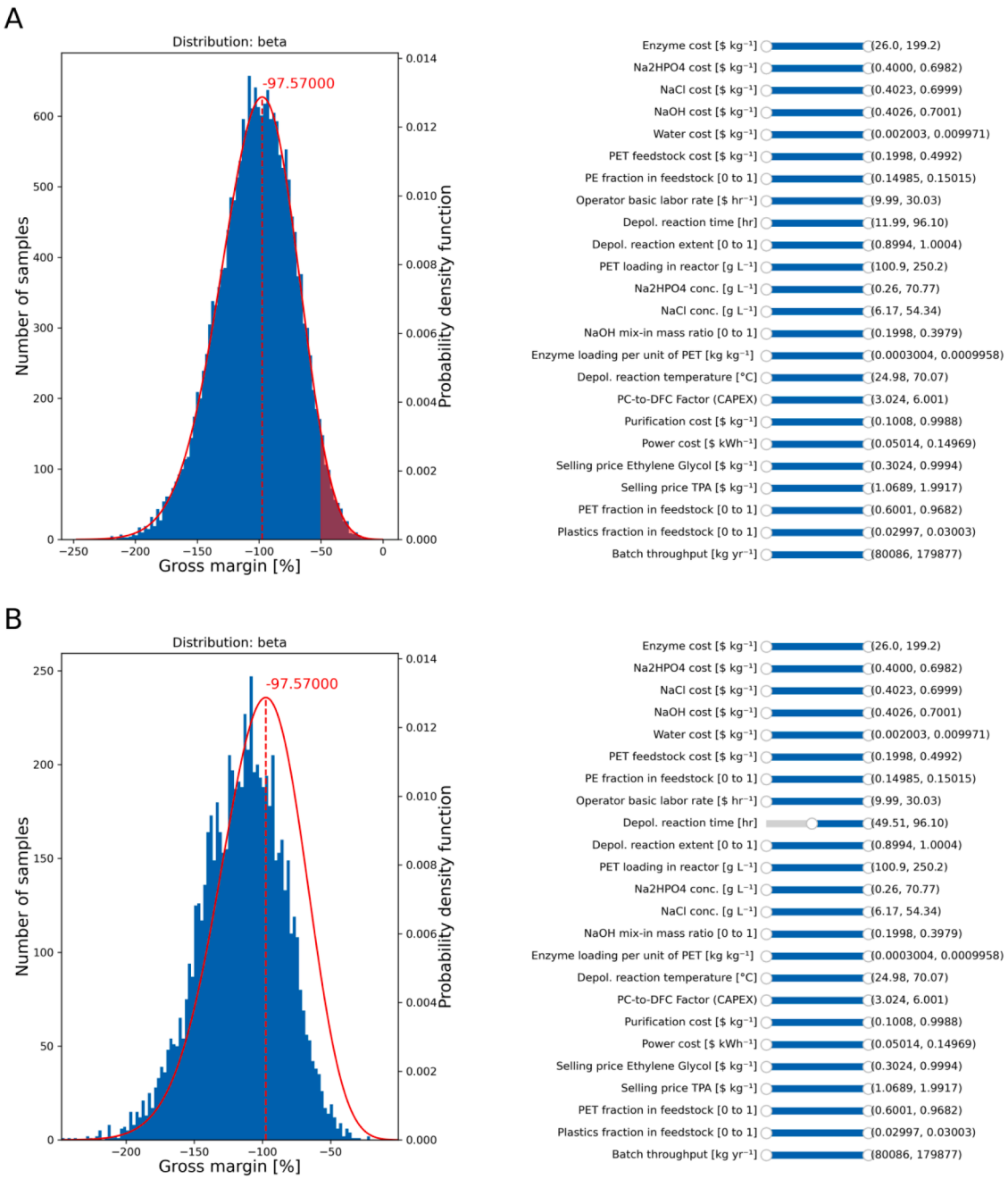

Fig. 8. The interactive histogram of BioProcessNexus comparing two different gross margin distributions (A and B) derived from the surrogate model Monte Carlo dataset. In panel A, the original distribution is shown, while panel B displays the effect of modifying the depolymerization reaction time parameter. The red lines represent the probability distribution fitted to the histogram without any subsetting (A), with red dashed lines indicating the mode of the distribution. The right panels display the features of the Monte Carlo dataset, with upper and lower bounds for various process parameters that can be adjusted using sliders. This visualization demonstrates how changing the depolymerization reaction time impacts the overall gross margin distribution.

programming knowledge barriers.

This project expands the open-source ecosystem and complements prior efforts in the field, such as BioSTEAM. While BioSTEAM is an excellent option for researchers with programming expertise and for modeling projects where all required unit operations are already implemented, the BioProcessNexus addresses scenarios where researchers prefer proprietary software. Proprietary tools are often more user-friendly and may include specialized unit operations not yet available in open-source software. Importantly, the analysis pipeline provided by the BioProcessNexus is not restricted to proprietary software. Monte Carlo sampling can also be performed using BioSTEAM models. Furthermore, this concept could be extended to enable the BioProcessNexus GUI to directly utilize original BioSTEAM models rather than relying on surrogate models. That would improve the accessibility of BioSTEAM models, as users would no longer need programming expertise to work with preexisting BioSTEAM models.

It is important to note that the BioProcessNexus does not allow users to extend or modify the original techno-economic models, such as by swapping or adding unit operations. Such modifications must be made directly using the software that generated the original model. This project aims to promote the accessibility and comparability of preexisting models and not to make TCA modelling software obsolete. On the contrary, we believe further development of TCA modelling software is necessary to facilitate scientific progress and that a lower barrier of entry

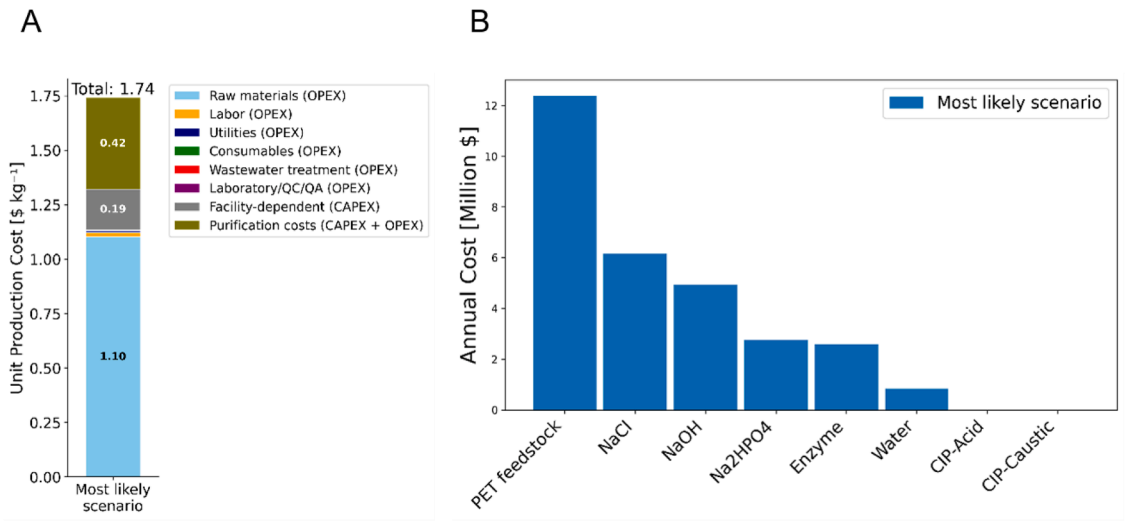

Fig. 9. Economic analysis of the enzymatic recycling process. (A) Breakdown of unit production costs showing the contribution of different cost categories including raw materials, labor, facility-dependent costs, and utilities. The total unit production cost was $1,74 kg TPA. (B) Distribution of material costs across major raw material inputs, with PET feedstock representing the highest cost component, followed by NaCl and NaOH. Material costs constitute 63.3 % of the total operating costs. The raw materials cost shown here only concerns the depolymerization reaction.

of preexisting models via the BioProcessNexus will result in a wider adoption of TCA modelling software. To this end, we strongly encourage scientists to share models of techno-economic models included in their publications. We believe that such sharing is essential for transparent, reproducible, and accessible science.

To demonstrate the value of accessible techno-economic modeling, we applied our approach to analyze the economic feasibility of enzymatic PET recycling, a promising but challenging technology for plastic waste valorization. The BOTTLE Consortium in the US released major previous work on this (Singh et al., 2021). In their study, the production cost for TPA was estimated to be 1.93 $ kg^{-1}. This value is close to the 1.74 $ kg^{-1} we achieved in our scenario adopting realistic process parameters. The main difference in the model of Singh et al. (Singh et al., 2021) is the use of cryo-grinding as a pre-treatment technique to amorphize the PET feedstock before the depolymerization reaction. This extra step accounts for extra costs that are avoided in our process since we assume the recycling of PET waste with a low crystallinity.

In another study, Uekert et al. (Uekert et al., 2023) modelled the repolymerization of recycled TPA obtained from enzymatic recycling, yielding a PET production cost of 4 $ kg^{-1} in their base case scenarios. The enzymatic recycling process was then compared with other recycling techniques such as mechanical recycling, glycolysis, and methanolysis. Although enzymatic recycling appeared to be the most expensive, it also showed benefits, such as higher tolerance for contaminants in the feedstock. This leads to the conclusion that enzymatic recycling could be attractive if price parity with other techniques is achieved.

Our analysis highlights the challenges that must be addressed to improve the economics of enzymatic recycling. Foremost among these is the feedstock cost, which represents 41.7 % of raw material costs and is the most significant expense in the process. PET, the most widely recycled plastic polymer, has seen increasing demand due to stricter regulations on plastic recyclate content in the European Union (Lee, 2019). Therefore, identifying low-cost PET waste streams neglected by mechanical recyclers is crucial. In this regard, the Enzycle consortium identified potential targets beyond PET trays, such as dark-colored PET flakes and PET lumps produced from defective machinery (East, 2023).

In this sense, the recently approved European Commission Packaging and Packaging Waste Regulation Directive, which mandates return systems for various packaging types, presents an opportunity (Packaging Europe, 2024). This directive should boost the availability of high-quality post-consumer PET waste. Additionally, advancements in municipal waste sorting technologies, such as AI-assisted robotic systems, are expected to increase outputs and further reduce PET waste costs (Tenore, 2023).

Future research must focus on optimizing the purification strategy for TPA, which appears to be the most impactful cost item in our analysis. The interaction kinetics between the enzyme and substrate are also fundamental, as our SHAP charts reveal that PET loading and enzyme loading substantially impact process economics and the PMI.

Given that TPA is a commoditized and inexpensive chemical (current market price of 1.07 $ kg^{-1}), even established mechanical recyclers struggle to compete with the prices of virgin TPA, leading some to bankruptcy (Laird, 2024). With our estimated unit production cost of 1.74 $ kg^{-1} and operating at a negative gross margin of −49.5 %, a strategic focus on premium applications, such as sports apparel, could be more profitable for emerging recycling technologies. Brands in this sector have higher profit margins and are willing to pay for premium sustainable materials, as evidenced by their collaborations with PET chemical recyclers (Patagonia Inc., 2023). Another strategy that is worth exploring is the targeting of other polyester polymers having a higher market value than PET, for example polybutylene terephthalate (PBT), whose monomer 1,4-butanediol commands a higher market price of 1.7–2.0 $ kg^{-1} (Business Analytiq Pte Ltd, 2024), and polytrimethylene terephthalate (PTT), which incorporates bio-based 1,3-propanediol (PDO) valued at around 2.4 $ kg^{-1} based on a glucose price of 400 $ t^{-1} and would be even higher if adjusted for inflation (Hermann and Patel, 2007). Polyethylene naphthalate (PEN), containing 2,6-naphthalene dicarboxylic acid, represents another high-value target, particularly in the growing flexible electronics sector (Scandurra et al., 2023). While these polymers have smaller market volumes than PET and a poor collection infrastructure, their established applications in high-performance sectors such as automotive, electronics, and specialty textiles, coupled with growing regulatory pressure for circularity in these industries, could provide more economically sustainable opportunities for emerging recycling technologies.

5. Future work

Future development of the BioProcessNexus may include expanding its library of available models, enhancing its user interface, and integrating emerging methods into its analysis pipeline. Another important

improvement is the speed of the surrogate models training, which can be enhanced by adopting parallelization strategies to exploit more effectively the computational power of the hosting computer. Another interesting feature would be to make it easier to calculate SHAP values based on specific scenarios (such as done with the interactive histogram) so the user could better evaluate which the most influential process features are in these specific scenarios. Furthermore, introducing the functionality of using BioSTEAM models with the BioProcessNexus GUI instead of surrogate models only, presents an interesting avenue. It must be mentioned that the community must drive the project's long-term success and growth. Researchers must adopt the tools presented in this article to make their work more accessible. Doing so will benefit the broader field by fostering transparency and collaboration enhancing the visibility and impact of their individual contributions.

6. Conclusion

We demonstrate that surrogate modeling can effectively improve access to complex process analysis while preserving the role of specialized software in process engineering. Gaussian Process surrogate models generated with the open-source software accompanying the BioProcessNexus achieved sub-2 % NRMSEs on the enzymatic recycling Monte Carlo dataset. This confirms the model's applicability for process optimization and analysis. While our case study revealed substantial economic challenges for enzymatic PET recycling, with a unit production cost of \$1.74 kg^{-1} TPA, it also identified clear optimization pathways through feedstock sourcing and purification strategies. BioProcessNexus represents a significant step toward standardized, collaborative process modeling in biotechnology. Its future impact will depend on community adoption and continued development of shared models and analytical tools to advance sustainable manufacturing processes.

CRediT authorship contribution statement

Tommaso De Santis: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Matthias Medl:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Peter Satzer:** Conceptualization. **Gerald Striedner:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Tommaso De Santis reports financial support and administrative support were provided by Enzycle consortium. Tommaso De Santis reports a relationship with Enzycle consortium that includes: funding grants, non-financial support, and travel reimbursement. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the DocSchool BioproEng for their support and the Enzycle consortium for providing the data for building the model for the enzymatic recycling process. This project has received funding from the Bio-Based Industries Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 887913.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.compchemeng.2025.109220.

Data availability

We have shared all software and data on a public repository.

References

acib GmbH, 2020. Microbial enzymes for treatment of non-recycled plastic fractions. Accessed: Nov. 27, 2024. [Online]. Available: https://www.enzycle.eu/.

Anastas, P.T., Lankey, R.L., 2000. Life cycle assessment and green chemistry: the yin and yang of industrial ecology. Green Chem 2 (6), 289–295. https://doi.org/10.1039/B005650M. Jan.

Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for Hyper-Parameter Optimization. Advances in Neural Information Processing Systems. Curran Associates, Inc.. Accessed: Nov. 27, 2024. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2011/hash/86e8f7ab32cfd12577bc2619bc635690-Abstract.html

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324. Oct.

Business Analytiq Pte Ltd, 2024. 1,3-Butanediol (1,3-BDO) Prices. businessanalytiq. Accessed: Nov. 27, 2024. [Online]. Available: https://businessanalytiq.com/procurementanalytics/index/13-butanediol-13-bdo-prices/.

Cameron, I.T., Ingram, G.D., 2008. A survey of industrial process modelling across the product and process lifecycle. Comput. Chem. Eng. 32 (3), 420–438. https://doi.org/10.1016/j.compchemeng.2007.02.015. Mar.

ChemAnalyst, 2025. Global Chemical and Petrochemicals, Specialty Chemicals, Elastomer and Rubber. Fertilizer and Feedstock - Latest Chemical Prices. News and Market Analysis. Accessed: Feb. 19, 2025. [Online] . Available: https://www.chemanalyst.com/.

Cortes-Peña, Y., Kumar, D., Singh, V., Guest, J.S., 2020. BioSTEAM: a fast and flexible platform for the design, simulation, and techno-economic analysis of biorefineries under uncertainty. ACS Sustain. Chem. Eng. 8 (8), 3302–3310. https://doi.org/10.1021/acssuschemeng.9b07040. Mar.

East, P., 2023. Recoup report - couloured PET. Accessed: Jan. 29, 2025. [Online]. Available: https://www.recoup.org/wp-content/uploads/2023/09/coloured-pet-document-final-v2-07oct20-1-1602070073.pdf.

ECHEMI Digital Technology Co. Ltd., 2025. Chemical Product Price Database: Compare and Download Prices Online. ECHEMI. Accessed: Feb. 19, 2025. [Online]. Available: https://www.echemi.com/price-database.html.

ERI Economic Research Institute Inc., 2025. Machine operator salary in Austria. Accessed: Feb. 19, 2025. [Online]. Available: https://www.erieri.com/salary/job/machine-operator/austria.

Guo, R.-T., et al., 2024. Natural and engineered enzymes for polyester degradation: a review. Environ. Chem. Lett. 22 (3), 1275–1296. https://doi.org/10.1007/s10311-024-01714-6. Jun.

Harris, C.R., Millman, K.J., van der Walt, et al., 2020. Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2.

Hermann, B.G., Patel, M., 2007. Today's and tomorrow's bio-based bulk chemicals from white biotechnology. Appl. Biochem. Biotechnol. 136 (3), 361–388. https://doi.org/10.1007/s12010-007-9031-9. Mar.

Holindu GmbH, 2025. The Water Price Index (EUR). Accessed: Feb. 19, 2025. [Online]. Available: https://www.holidu.com/magazine/water-price-index-intl.

Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. Python. https://doi.org/10.1109/MCSE.2007.55.

Intelligen Inc., 2021. "SuperPro_manualforprinting_v11.pdf." Accessed: Jan. 28, 2025. [Online]. Available: https://www.intelligen.com/wp-content/uploads/2020/05/SuperPro_ManualForPrinting_v11.pdf.

Laird, K., 2024. Umincorp teetering on brink of bankruptcy. Sustainable Plastics.. Accessed: May 31, 2024. [Online]. Available: https://www.sustainableplastics.com/news/umincorp-brink-bankruptcy

Lee, J., 2019. Recycled plastic is now more expensive than PET. That's Not Just an Economic Problem. ThePrint. Accessed: May 31, 2024. [Online]. Available: https://theprint.in/economy/recycled-plastic-is-now-more-expensive-than-pet-thats-not-just-an-economic-problem/302129/.

Lundberg, S.M., Lee, S.-I., 2017. A Unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. Curran Associates, Inc.. Accessed: Nov. 27, 2024. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Maceno, M.M.C., Pawlowsky, U., Machado, K.S., Seleme, R., 2018. Environmental performance evaluation – a proposed analytical tool for an industrial process application. J. Clean. Prod. 172, 1452–1464. https://doi.org/10.1016/j.jclepro.2017.10.289. Jan.

Mathe, A., 2022. Carbios to Build in France its First-Of-A-Kind Manufacturing Plant For Fully Bio-Recycled PET in Partnership With Indorama Ventures. Carbios. Accessed: Apr. 24, 2025. [Online]. Available: https://www.carbios.com/en/carbios-to-build-in-france-its-plant/.

Ögmundarson, Ó., Sukumara, S., Herrgård, M.J., Fantke, P., 2020. Combining environmental and economic performance for bioprocess optimization. Trends

Biotechnol. 38 (11), 1203–1214. https://doi.org/10.1016/j.tibtech.2020.04.011. Nov.
Packaging Europe, 2024. PPWR Vote Approves Collection Targets. PFAS Limits, and Fresh Produce Packaging Bans. Packaging Europe. Accessed: May 31, 2024. [Online]. Available: https://packagingeurope.com/news/collection-targets-pfas-limits-fresh-produce-packaging-bans-approved-in-ppwr-vote/11264.article.
Patagonia Inc., 2023. Jeplan. Accessed: May 31, 2024. [Online]. Available: https://www.patagonia.com/our-footprint/jeplan.html.
Pedregosa, F., et al., 2011. Scikit-learn. Accessed: Nov. 27, 2024. [Online]. Available: https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.
Piekarski, C.Moro, da Luz, L.Mendes, Zocche, L., de Francisco, A.C., 2013. Life cycle assessment as entrepreneurial tool for business management and green innovations. J. Technol. Manag.; Innov. 8 (1), 44–53. https://doi.org/10.4067/S0718-27242013000100005. Mar.
Rasmussen, C.E., Williams, C.K.I., 2008. *Gaussian Processes For Machine learning*, 3. Print. in Adaptive Computation and Machine Learning. MIT Press, Cambridge, Mass.
Romero, J.J., Jenkins, E.W., Husson, S.M., 2023. Surrogate-based optimization of capture chromatography platforms for the improvement of computational efficiency. Comput. Chem. Eng. 173, 108225. https://doi.org/10.1016/j.compchemeng.2023.108225. May.
Scandurra, G., Arena, A., Ciofi, C., 2023. A Brief review on flexible electronics for IoT: solutions for sustainability and new perspectives for designers. Sensors 23 (11). https://doi.org/10.3390/s23115264. Art. no. 11Jan.
Schimansky, T., 2024. TomSchimansky/CustomTkinter. Python. Nov. 27Accessed: Nov. 27, 2024. [Online]. Available: https://github.com/TomSchimansky/CustomTkinter.
Shapley, L.S., 1953. A value for n-person games. In: Contribution to the Theory of Games, 2. Accessed: Jan. 29, 2025. [Online]. Available: https://www.torrossa.com/gs/resourceProxy?an=5641636&publisher=FZO137#page=87.
Singh, A., et al., 2021. Techno-economic, life-cycle, and socioeconomic impact analysis of enzymatic recycling of poly(ethylene terephthalate). Joule 5 (9), 2479–2503. https://doi.org/10.1016/j.joule.2021.06.015. Sep.
Sonnendecker, C., et al., 2022. Low carbon footprint recycling of post-consumer PET Plastic with a metagenomic polyester hydrolase. ChemSusChem 15 (9), e202101062. https://doi.org/10.1002/cssc.202101062.
STATISTIK AUSTRIA, 2025. Energiepreise, -Steuern. Accessed: Feb. 19, 2025. [Online]. Available: https://www.statistik.at/statistiken/energie-und-umwelt/energie/energiepreise-steuern.
Taras, S., Woinaroschy, A., 2012. An interactive multi-objective optimization framework for sustainable design of bioprocesses. Comput. Chem. Eng. 43, 10–22. https://doi.org/10.1016/j.compchemeng.2012.04.011. Aug.
Taskesen, E., 2020. Distfit is a Python Library For Probability Density Fitting. Jupyter Notebook. Jan.Accessed: Nov. 27, 2024. [Online]. Available: https://erdogant.github.io/distfit.
Tenore, H., 2023. Sorting Out Recycling from Trash is the Perfect Example of a Job Few People Want to Do That AI is Better Than Humans at. Business Insider. Accessed: May 31, 2024. [Online]. Available: https://www.businessinsider.com/ai-robots-better-at-sorting-trash-recycling-waste-management-report-2023-11.
Uekert, T., et al., 2023. Technical, economic, and environmental comparison of closed-loop recycling technologies for common plastics. ACS Sustain. Chem. Eng. 11 (3), 965–978. https://doi.org/10.1021/acssuschemeng.2c05497. Jan.
Van Rossum, G., Drake, F.L., 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.
Wegelin, J.A., 2000. A Survey of Partial Least Squares (PLS) Methods, With Emphasis On the Two-Block Case. University of Washington, Tech. Rep.
Weule, H., 1993. Life-cycle analysis – a strategic element for future products and manufacturing technologies. CIRP Annals 42 (1), 181–184. https://doi.org/10.1016/S0007-8506(07)62420-2. Jan.
Yildiz, D.B., Sayar, N.A., 2020. Propagation of parametric uncertainty in a conceptually designed bioethanol production process," in *Computer Aided Chemical Engineering*. In: Pierucci, S., Manenti, F., Bozzano, G.L., Manca, D. (Eds.), 30 European Symposium on Computer Aided Process Engineering, 30 European Symposium on Computer Aided Process Engineering, 48. Elsevier, pp. 631–636. https://doi.org/10.1016/B978-0-12-823377-1.50106-3 vol. 48.