



Reusable surrogate models for distillation columns

Martin Bubel¹*, Tobias Seidel¹, Michael Bortz¹

Department of Optimization, Fraunhofer Institute for Industrial Mathematics, Fraunhofer-Platz 1, Kaiserslautern, D-67663, Rheinland-Pfalz, Germany

ARTICLE INFO

Keywords:

Surrogate modeling
Fluid modeling
Machine learning
Property prediction methods
Process optimization
Entrainer distillation

ABSTRACT

Surrogate modeling is a powerful methodology in chemical process engineering, frequently employed to accelerate optimization tasks. Despite their popularity, most surrogate models are trained for a narrow range of fixed chemical systems and operating conditions, which limits their reusability. This work introduces a paradigm shift towards reusable surrogates by developing a single model for distillation columns that generalizes across a vast design space. The key enabler is a novel ML-fueled modelfluid representation which allows for the generation of datasets of more than 1000000 samples. This allows the surrogate to generalize not only over column specifications but also over the entire chemical space of homogeneous ternary vapor-liquid mixtures. We validate the model's accuracy and demonstrate its practical utility in a case study on entrainer distillation, where it successfully screens and ranks candidate entrainers, significantly reducing the computational effort compared to rigorous optimization.

1. Introduction

High-fidelity simulation is a cornerstone of modern chemical process engineering, enabling the detailed analysis of complex phenomena, from molecular-level interactions to full-scale plant operations (Forrester et al., 2008). Distillation columns, in particular, represent a class of unit operations that are critical to the chemical industry but their simulation is known to be computationally expensive. Rigorous models often involve solving large systems of nonlinear MESH (Mass, Equilibrium, Summation, Heat) equations (Biegler et al., 1997). Solving these systems usually comes with a high computational effort, which limits the use of high-fidelity models in tasks that require many repeated evaluations, such as large-scale optimization, comprehensive design studies, and rapid exploration of what-if scenarios.

To bridge this gap, surrogate models have become a popular and powerful methodology in the field (McBride and Sundmacher, 2019). A surrogate model acts as an efficient, optimization-friendly prediction model of a physical system or its high-fidelity simulation. It is trained on data from experiments or simulations to mimic the system's behavior, providing near-instantaneous predictions. A crucial aspect of surrogate models is their formulation as an **explicit map**, $f : \mathcal{X} \rightarrow \mathcal{Y}$, where the model output $y \in \mathcal{Y}$ is computed in a single forward pass given the model inputs $x \in \mathcal{X}$. This stands in contrast to rigorous models that require solving an implicit system of equations. The benefits of this

explicit formulation for optimization are manifold: the computational cost is significantly reduced, analytical derivatives are readily available for efficient Newton-type optimizers, and the models are not prone to the convergence failures that can plague rigorous simulators.

The fundamental concepts of surrogate modeling are well-established across engineering disciplines (Forrester et al., 2008) and their importance is continuously growing in fields from materials science (Peivaste et al., 2025) to the development of digital twins (Bárkányi et al., 2021). Within chemical engineering, a variety of methods are employed. In small data regimes, Bayesian methods like Gaussian Processes are popular, often coupled with adaptive sampling for Bayesian Optimization to find local optima without requiring global accuracy (Caballero and Grossmann, 2008; Winz et al., 2025). In contrast, the rise of large-scale data generation has fueled the use of Artificial Neural Networks (ANNs) to create *global* surrogate models that are accurate across vast feature spaces, which is the focus of this work.

Furthermore, surrogates are applied at various levels of abstraction, each with a trade-off between ease of use and reusability:

- **Flowsheet-level surrogates:** These models are built for an entire process but are brittle to changes in flowsheet topology (Palmer and Realf, 2002a,b). Furthermore, generating training data can

* Corresponding author.

E-mail address: martin.bubel@itwm.fraunhofer.de (M. Bubel).

List of operators, variables, and abbreviations

Operator	Meaning
\hat{a}	Estimate of a
a^*	Optimal value of a according to some objective function
a_i	Index i of a
a^{phase}	Property a in given phase
a^{type}	Property a of given type
Δa	Difference in a (e.g., error)
Variable	Meaning
ℓ_i	Molar fraction of component i in the liquid phase
v_i	Molar fraction of component i in the vapor phase
ℓ, v	Vector of molar fractions in liquid/vapor phase
T	Temperature
p	Pressure
γ_i	Activity coefficient of component i
$\gamma_i _j$	Activity coefficient of component i at infinite dilution in component j
RR	Reflux ratio
s	Bottoms-to-feed split ratio
N_S	Number of equilibrium stages
N_S^{AF}, N_S^{BF}	Stages above/below feed
\dot{Q}^{Reb}	Reboiler heat duty
\dot{Q}^{Con}	Condenser heat duty
h_i^{AV}	Enthalpy of vaporization of component i
h	Molar enthalpy
\dot{Q}	General heat duty \dot{Q}
x, \mathcal{X}	Feature vector, feature space
y, \mathcal{Y}	Target/output, target space
θ	Model parameter
x^{MF}	Modelfluid feature vector
θ^{MF}	Modelfluid parameter vector
f	Surrogate model function
\mathcal{L}	Loss function \mathcal{L}
J	Jacobian matrix J
CI	Confidence interval (CI)
$\ell_{B,i}$	Bottom stream molar fractions
$\ell_{D,i}$	Distillate stream molar fractions

All variables are in SI units unless stated otherwise.

Abbreviation	Meaning
ANN	Artificial Neural Network
ML	Machine Learning
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MINLP	Mixed-Integer Nonlinear Programming
CAPEX	Capital Expenditure
OPEX	Operational Expenditure
PC-SAFT	Perturbed Chain - Statistical Associating Fluid Theory
MESH	Mass, Equilibrium, Summation, Heat
VLE	Vapor-Liquid Equilibrium
NRTL	Non-Random Two-Liquid (model)
DIPPR	Design Institute for Physical Properties (database)
SMILES	Simplified Molecular Input Line Entry System
CI	Confidence Interval
AI	Artificial Intelligence

EA	Extended Antoine (model)
SA	Simplified Antoine (model)
MF	Modelfluid
E	Entrainer
Rig	Rigorous (model)
SM	Surrogate Model
Ac	Acetone
Az	Azeotrope
Reb	Reboiler
Con	Condenser
B	Bottom
D	Distillate
F	Feed
AF	Above Feed
BF	Below Feed
S	Stage

be paradoxical, as it requires numerous successful simulations of the entire flowsheet, the very task that is often difficult and prone to convergence difficulties.

- **Unit operation-level surrogates:** A more robust, modular approach where surrogates for individual units can be interconnected, improving flexibility and reusability (Freund and Sundmacher, 2008; Caballero and Grossmann, 2008; Bubel et al., 2021). This modularity is a core principle of our work.
- **Property-level surrogates:** Models for specific, expensive calculations within a unit model, such as phase equilibria (Nentwich and Engell, 2019; Winz et al., 2021).

Despite the sophistication of the available methods, a common and fundamental limitation pervades the literature: the validity of a trained surrogate is almost always restricted to a fixed chemical system. Whether at the flowsheet, unit, or property level, a model trained for a specific mixture (e.g., water/ethanol) must be completely regenerated if the chemical components change (Sethi et al., 2025; Abranches et al., 2023), hindering the scalability and industrial applicability of the surrogate modeling paradigm. This system-specific nature has been a barrier to the transition of surrogate models from academic prototypes to widely applicable, reusable tools.

Recent work by Sun et al. represents a first significant step towards overcoming this limitation (Sun et al., 2023, 2025). They developed surrogate models for binary vapor-liquid equilibria (VLE) that generalize across different chemical pairs by using pure component properties and molecular fingerprints as inputs. While this demonstrated the potential for generalization at the property level, the challenge of creating a reusable, general-purpose surrogate for a complete *unit operation* model — with its complex interplay of mass balances, energy balances, and equipment parameters — remains an open and critical challenge.

This work introduces a shift towards **reusable surrogate models** for distillation columns, fundamentally addressing this challenge of limited generalizability. We demonstrate that a single surrogate model can be trained to have a vast range of validity, not only across different column specifications — such as flow rates, reflux ratio, pressure, number of stages, and feed location — but also, most significantly, across the entire chemical space of homogeneous vapor-liquid ternary mixtures. The key enabler for this is the use of a **novel modelfluid representation**, which we introduced in a previous publication (Bubel et al., 2025). This representation, combined with a machine learning-fueled dataset generation strategy, allows the surrogate model to learn the underlying relationships between fluid properties and column behavior, rather than simply memorizing the behavior of one specific system. A schematic overview of this entire workflow is presented in Fig. 1. The main contributions and goals of this paper are to:

- Introduce a new class of surrogate models for distillation columns that generalizes over the entire space of homogeneous ternary vapor–liquid systems, making it truly reusable across different flowsheets, operating conditions, and chemical systems.
- Detail a large-scale, ML-fueled data generation strategy, creating a dataset of over 1,000,000 unique data points for well-beyond 100,000 individual mixtures using modern property prediction methods and established pure component data (Wilding et al., 1998).
- Demonstrate the practical utility of the trained surrogate model in a challenging entrainer selection case study, where a single model is used to rapidly screen candidates and guide process design.
- Highlight the potential of such general-purpose surrogates as powerful tools for comprehensive design space exploration, large-scale optimization, and as a method to provide high-quality initial guesses (warm-starts) for rigorous process simulators, thereby improving their robustness and convergence speed.

The result is a powerful and reusable tool that significantly lowers the barrier for comprehensive process optimization and design space exploration. By interconnecting it with other unit operation models to flowsheets, the surrogate model can be used in a variety of case studies for any ternary vapor–liquid system.

While the main focus is on the ternary distillation column, surrogate models for binary and ternary VLE as well as a binary distillation column are provided in the Supporting Information as a reference.

The remainder of this paper is organized as follows.

Section 2 details the development of the surrogate model, from the feature and target selection to the ML-fueled dataset generation and model training.

Section 3 presents a comprehensive case study on entrainer distillation to validate the model's performance and demonstrate its practical utility.

Finally, we conclude in Section 4 with a summary of our findings and an outlook on future research directions.

2. Development of the reusable surrogate model

This section details the development of the reusable surrogate model for distillation columns.

We first define the features and targets of the model, including the crucial modelfluid representation that enables generalization across chemical systems (Section 2.1).

Next, we describe our novel ML-fueled dataset generation strategy, which allows us to create a large and diverse training dataset far exceeding the scale of previous works (Section 2.2).

Finally, we outline the training procedure for the surrogate model, quantify its prediction uncertainty, and evaluate the achieved predictive performance (Section 2.3).

We further note that the proposed surrogate model is reusable within the space of ternary mixtures obeying homogeneous vapor–liquid phase behavior.

While the modelfluid representation generally extends to multi-component systems, such an extension is beyond the scope of this work. We revisit this potential in Section 4, where we discuss future directions.

2.1. Feature and target selection

The predictive power and generalizability of the surrogate model are fundamentally determined by the choice of its input features and output targets. We designed these to capture both the thermodynamic behavior of the chemical mixture and the physical configuration of the unit operation.

The modelfluid representation. The key enabler for generalization across the ternary mixture space is the modelfluid representation introduced in Bubel et al. (2025). For the sake of a self-contained manuscript, we summarize its core principles here. Instead of using discrete component identities (e.g., SMILES strings), the framework uses a vector of continuous, physically meaningful descriptors that characterize the thermodynamic behavior of a binary fluid system. This allows the surrogate model to learn the underlying relationships between fluid properties and column performance, rather than memorizing the behavior of a few specific systems.

For a ternary system, the modelfluid is defined by the following features for each of the three components and their binary pairs:

- **Pure Component Properties:** The saturated vapor temperatures (T_i^{SV}) at a given system pressure and the corresponding vaporization enthalpies (h_i^{AV}). These anchor the system's boiling behavior and energy requirements.
- **Mixture Interaction Properties:** The activity coefficients at infinite dilution ($\gamma_i|_j$) and the derivative of the vapor mole fraction with respect to the liquid mole fraction at infinite dilution ($\partial v_i / \partial \ell_i|_j$). The former is a direct measure of liquid-phase non-ideality and is predictable via methods like matrix completion (Jirasek et al., 2020; Damay et al., 2021). The latter is a powerful, derivative-based feature that directly quantifies the relative volatility at infinite dilution, a key indicator of separation difficulty. For a detailed thermodynamic treatment of this feature, the reader is referred to Bubel et al. (2025).

We define the modelfluid feature vector as follows:

$$x^{\text{MF}} = \left[p, T_1^{SV}, T_2^{SV}, T_3^{SV}, h_1^{AV}(T_1^{SV}), h_2^{AV}(T_2^{SV}), h_3^{AV}(T_3^{SV}), \gamma_1|_2, \gamma_2|_1, \gamma_1|_3, \gamma_3|_1, \gamma_2|_3, \gamma_3|_2, \frac{\partial v_1}{\partial \ell_1}|_3, \frac{\partial v_2}{\partial \ell_2}|_1, \frac{\partial v_3}{\partial \ell_3}|_1 \right]. \quad (1)$$

Note that in (1), we omit $\frac{\partial v_1}{\partial \ell_1}|_2$, $\frac{\partial v_2}{\partial \ell_2}|_3$, and $\frac{\partial v_3}{\partial \ell_3}|_2$. This is due to the requirement of unique vapor pressure models for each component in a ternary system. This physical requirement, which is derived in detail in Bubel et al. (2025), reduces the number of independent features from 19 to 16, as in (1). While we could also have chosen another subset of features for (1), the choice of features in this work advocates the incorporation of prediction methods for the activity coefficient at infinite dilution.

In our modeling (1), the activity coefficients at infinite dilution directly parameterize a Margules activity coefficient model. The remaining features are used to explicitly calculate the parameters of a simplified two-parameter Antoine equation for vapor pressure and a composition-weighted model for the enthalpy of vaporization. This feature-to-parameter mapping is deterministic and explicit, and its full set of equations is featured in Bubel et al. (2025).

Furthermore, the features in (1) generally occupy a bounded space; temperatures, pressures, and enthalpies are physically constrained, while the activity coefficients and their derivatives are observed to be empirically limited within the datasets considered. A detailed proposal on modelfluid feature bounds is provided in Bubel et al. (2025).

Unit operation features. To ensure the surrogate model is also generalizable across different column designs and operating conditions, the following features of the unit operation are included as inputs:

- The number of equilibrium stages above and below the feed stage (N_S^{BF}, N_S^{AF}).
- The ratio of the bottom flow rate to the feed flow rate (s).
- The reflux ratio (RR).

The respective feature vector is defined as

$$x^{\text{Co}} = [s, RR, N_S^{BF}, N_S^{AF}], \quad (2)$$

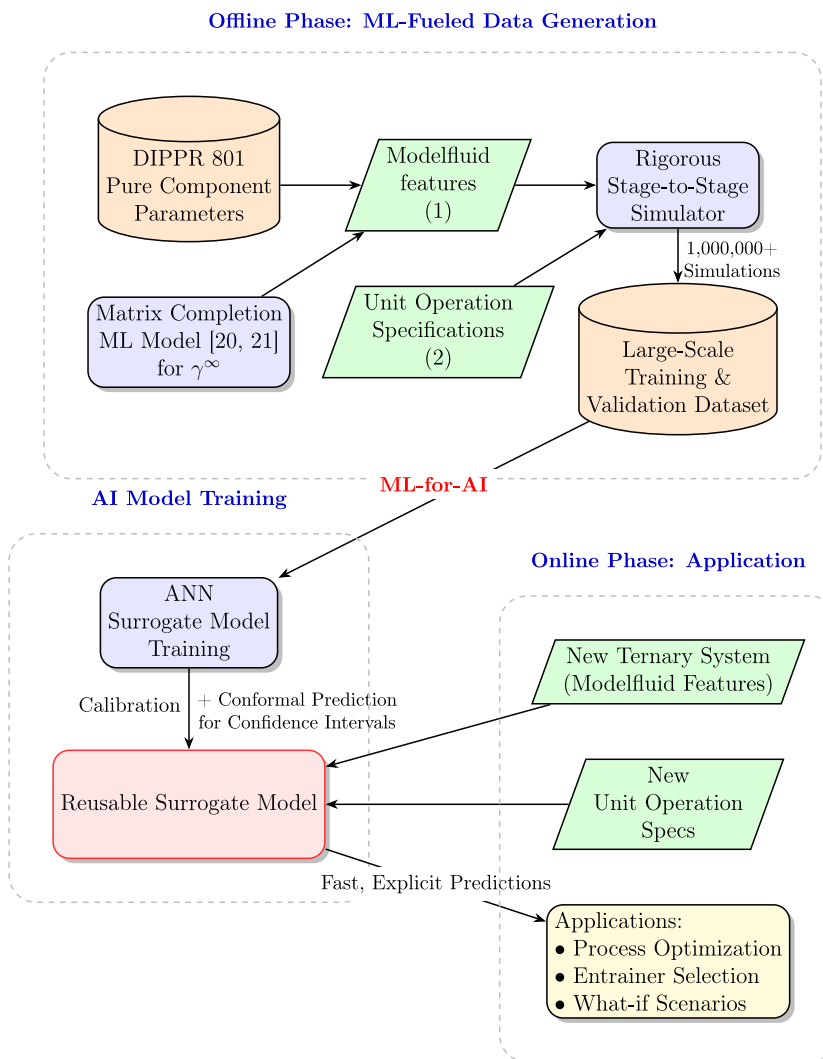


Fig. 1. Schematic overview of the proposed workflow. A machine learning model (Matrix Completion) is used to generate a large dataset (the “ML” part), which then enables the training of a generalizable Artificial Neural Network surrogate (the “AI” part). The final, calibrated model serves as a reusable Digital Twin for rapid process optimization and design space exploration.

while the pressure p is already included in the modelfluid features x^{MF} . The final combined feature vector serving as input for the surrogate model is thus defined as:

$$x = [x^{\text{MF}}, x^{Co}]. \quad (3)$$

Rigorous distillation column simulation. All training, validation, and test data points – as described in the following section – are generated using a rigorous stage-to-stage distillation column model. Given a feature vector (3), we can simulate a distillation column based MESH equation-based modeling (we refer to this as *rigorous*) presented in Bubel et al. (2025). The features from (3) that describe the modelfluid (1) are mapped to thermodynamic model parameters using the explicit feature-to-parameter map, as discussed above. The features that describe the column specifications can straightforwardly be used in the rigorous column modeling. Furthermore, to use the modelfluid-specific variant of MESH equation-based distillation column simulation presented in Bubel et al. (2025), we need to make the following assumptions:

1. Streams outside columns are at liquid-boiling state.
2. Columns operate at a constant, uniform pressure.
3. Columns are adiabatic, with heat exchange only at the condenser and reboiler.

4. Total condensation and evaporation are assumed for the condenser and reboiler.
5. The enthalpy of mixing is neglected.

A sketch of the column, as considered in this work, is shown in Fig. 2.

Model targets. The surrogate model is trained as an explicit map to predict key performance indicators and separation outcomes. Given the full feature vector (3), the model outputs the following targets:

- The reboiler heat duty (\dot{Q}^{Reb}) (condenser duty can be added but is not considered in this work).
- The bottom stream molar fractions ($\ell_{B,i} \quad \forall i = 1, \dots, N_C - 1$).
- The distillate stream molar fractions ($\ell_{D,i} \quad \forall i = 1, \dots, N_C - 1$).

We only predict the molar fractions of $N_C - 1$ components, as the last component’s fraction is determined by the summation constraint

$$\ell_{N_C} = 1 - \sum_{i=1}^{N_C-1} \ell_i.$$

While the prediction of both bottom and distillate compositions is redundant, we choose to predict both in order to analyze and compare the model’s predictive accuracy on each stream.

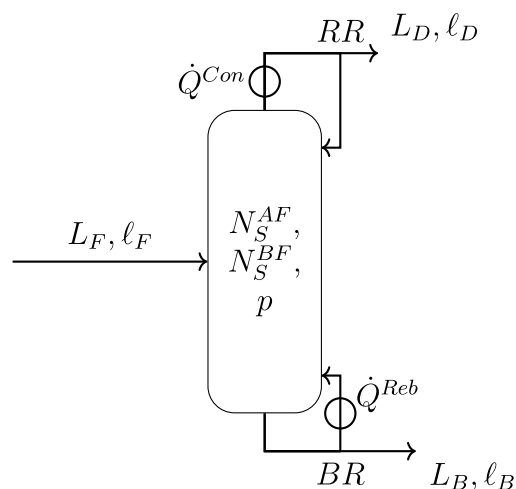


Fig. 2. Schematic of a distillation column used in this work. The column processes feed (F) into distillate (D) and bottom (B) streams, operating at pressure p . Its size is defined by the number of stages above (N_S^{AF}) and below (N_S^{BF}) the feed stage. Heat duties at the reboiler and condenser are represented by \dot{Q}^{Reb} and \dot{Q}^{Con} , respectively. The column's operation is characterized by the boilup ratio (BR) and reflux ratio (RR), respectively, which control how much of the bottom and distillate output is recycled back to the column. For simplicity, internal stages are not shown.

2.2. Dataset generation

The generation of datasets for the development of a distillation column surrogate model includes both, the sampling of modelfluid features and the rigorous simulation of the distillation column for those modelfluid features at various operating conditions – as described in Section 2.1. While the final surrogate model is trained to operate solely on the modelfluid features (1), the explicit map of modelfluid features to thermodynamic model parameters is the essential bridge that enables the use of rigorous column simulation for creating the large-scale dataset.

For the generation of training-, validation-, and test datasets in this work, we do not randomly sample modelfluid features. Instead, we use both, information available from pure component property- and mixture property databases and combine them with ML-predicted mixture properties. The below paragraphs, describe how the modelfluid features are obtained for the different datasets used to train, validate and test the surrogate model.

The generation of all datasets includes distillation column simulation for the modelfluid feature samples. To prevent simulations at thermodynamically inconsistent modelfluid feature samples, which is generally possible, we perform a series of VLE-based consistency checks, as detailed in Appendix A. If any of those checks fails for a modelfluid feature vector, it is discarded from the dataset.

For each modelfluid feature vector, we perform 100 rigorous distillation column simulations at random column specifications, to sample the column features (2). The column specifications are uniformly sampled from the following bounds:

- $\ell_F \in [0, 1]^3$ s.t. $\sum x_i = 1$
- $RR \in [0.1, 40]$
- Bottom-to-Feed flow rate split: $s \in [0.001, 0.999]$
- Number of stages above feed: $N_S^{AF} \in [2, 30]$
- Number of stages below feed: $N_S^{BF} \in [2, 30]$
- Pressure: $p \in [0.5, 10]$ bar

ML-fueled training and validation dataset. A cornerstone of this work is the generation of a large-scale dataset sufficient for training a global surrogate model in a high-dimensional feature space. Our approach represents a story of **ML-for-AI**: we leverage an existing machine learning method to generate a large dataset, which in turn enables the successful training of our artificial intelligence-based surrogate model. What this means is that the data for training and validation is generated synthetically. The modelfluid representation for each ternary mixture is constructed by combining pure component data with ML-predicted mixture properties. Specifically, pure component Antoine equation parameters and the pure component vaporization enthalpies model parameters are sourced from the commercial DIPPR 801 database (Wilding et al., 1998). Using those, we can compute the saturated vapor temperatures T_i^{SV} and the pure component vaporization enthalpies at the saturation temperature h_i^{SV} . The crucial mixture interaction information is obtained by predicting the activity coefficients at infinite dilution (γ^∞). While many powerful prediction methods have recently been developed (Rittig et al., 2023; Sanchez Medina et al., 2023; Caprio et al., 2023; Specht et al., 2024; Damay et al., 2021), we use the Matrix Completion method from Jirasek et al. (2020), as it has been successfully applied with this modelfluid representation in Babel et al. (2025). These $\gamma_i|_j$ values parameterize the Margules model of the mixture. Using the vapor pressure Antoine equations and the Margules activity coefficient model, we compute the $\frac{\partial \gamma_i}{\partial \ell_j}|_j$ at a chosen pressure p .

This yields all features of (1), which means that we can use the rigorous modelfluid-based distillation column simulation as discussed in Section 2.1.

The power of this setup comes from the usage of the ML-based activity coefficient prediction. We can compute the modelfluid feature vector for each ternary combination of components found in our pure component property database (which is vast), resulting in 167,221 unique and consistent modelfluid feature vectors (excluding samples that fail the VLE checks Appendix A), i.e. 167,221 unique ternary mixtures. As a reference, the works of Sun et al. (2023, 2025) were based on 210 binary mixtures.

Performing distillation column simulation, as described above, we obtain a final dataset of 2,465,106 feature and target vectors.

Test dataset. For final testing, a separate, higher-fidelity dataset was constructed using curated NRTL parameters from the Aspen Plus V10 mixture property database (Aspen Technology Inc., 2017). This means we combine the vapor pressure- and enthalpy models from the DIPPR 801 database (Wilding et al., 1998) with the NRTL parameters from Aspen (Aspen Technology Inc., 2017) and compute the vapor-liquid equilibria at some pressure p . From the obtained VLE, we determine the modelfluid features for 420 unique ternary mixtures, which form the basis of the test dataset. For each of these mixtures, we generate a range of distillation column specifications by sampling the relevant column features as described above. This results in a comprehensive test dataset comprising 13,494 distinct feature-target pairs, each representing a unique combination of mixture and column operating conditions.

This dataset, while smaller, is considered a suitable test of the surrogate model's predictive performance for real-world industrially-relevant systems, after being trained on synthetically generated data using ML-based predictions of limiting activity coefficients.

Discussion on data distribution. As illustrated in Fig. 3, the sampling of the modelfluid feature space in our training data is not perfectly uniform. This is a consequence of the natural distribution of the predicted activity coefficients and the filtering effect of our VLE consistency checks. This non-uniformity appears to be beneficial compared to a uniform sampling of modelfluid features, as it better reflects the distribution of real-world chemical systems found in our test set. This may also be due to undetected artifacts in the generated data, when sampling uniformly in the modelfluid feature space, which remains a challenge and motivates future work on more robust consistency

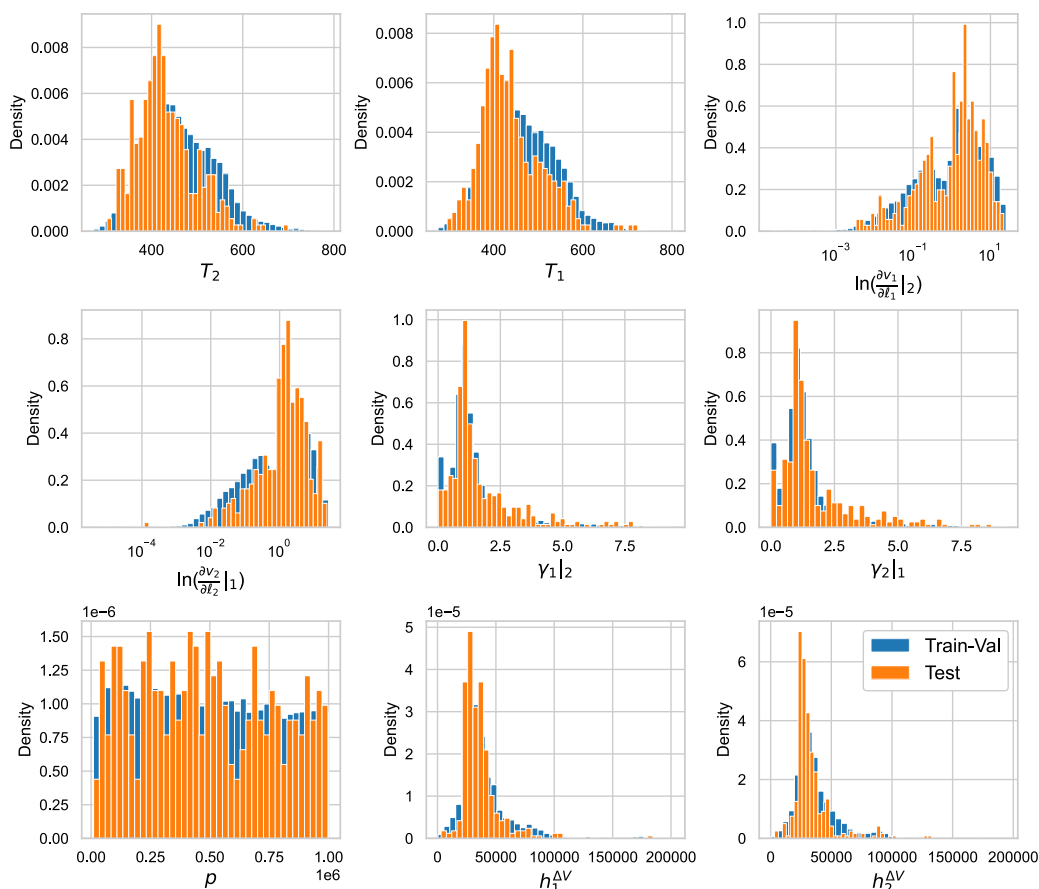


Fig. 3. Distribution of selected modelfluid features in the training and validation dataset (blue) and the test dataset (orange), showing a reasonable overlap despite different data sources. The term *density* on the vertical axes means that the vertical axis values (frequency) have been normalized such that the total area under the distribution curve sums to one.

checks for large-scale thermodynamic data generation. Additional visualizations of the modelfluid feature space samples of the training-, validation- and test dataset, including binary projection scatter plots, are provided in the Supporting Information.

Note on the use of ML-based prediction models for modelfluid features.

For the generation of the surrogate model training dataset, ML-based prediction methods for $\gamma_i|_j$ are used. While using such methods significantly increases the size training dataset, it also means that the predictions obtained usually are not perfectly accurate. In this work, we use the Matrix Completion Method as proposed in Jirasek et al. (2020), who report an absolute prediction error of $\ln(\gamma_i|_j)$ of less than 0.3 for about 80% of their data, at an experimental uncertainty of $\ln(\gamma_i|_j)$ of about 0.1 to 0.2 (Jirasek et al., 2020). However, since the goal of the training dataset is not to incorporate explicit feature vectors of popular solvent-solute couples, but rather to sample the feature space, this prediction error does not introduce any disadvantage for the quality of the training dataset or the surrogate model trained on it. As noted for the test dataset, which is used for model validation, we do not use ML-based predictions for the modelfluid feature vectors.

In Section 3, we also use pure component models from the DIPPR 801 database (Wilding et al., 1998) in combination with NRTL parameters from Aspen (Aspen Technology Inc., 2017) to obtain the modelfluid feature vectors of the entrainer candidates.

2.3. Model training and analysis

With the features, targets, and datasets defined, the final step is the training and calibration of the surrogate model. The model is an

Artificial Neural Network (ANN), chosen for its ability to approximate highly complex, nonlinear functions.

We train a total of four individual ANN-based surrogate models in this work: a VLE prediction model for binary and ternary systems, and a distillation column surrogate model for binary and ternary systems. In this section, we only discuss the distillation column surrogate model for ternary systems, as it is the most difficult to train and generalize. Detailed descriptions of the model architectures, training hyperparameters, and performance evaluations for the remainder of the models are provided in the Supporting Information

The ML-generated dataset was split randomly into training (80%) and validation (20%) sets.

Even though overfitting was not observed to be an issue in this work, as confirmed by the learning curves shown in Fig. 4, we manually excluded all ternary mixtures used in the entrainer selection case study Section 3 from the training dataset, to provide a rigorous out-of-sample validation.

We use a feed-forward neural network with linear layers and rectified linear unit (ReLU) activation functions on all but the final layer. The model uses one input layer, four hidden layers, and one output layer, with (22, 1024), (1024, 512), (512, 256), (256, 128), (128, 64), (64, 5) neurons in each layer, where the first number is the number of input features and the second number is the number of output features of each layer, respectively. We used a batch size of 64 and the model was trained for 830 epochs on a learning rate of 0.0001 and another 100 epochs on a learning rate of $5e-5$, before training was stopped due to slow progress. Reducing the learning rate after 830 iterations results in a sudden drop of the training loss, as shown in Fig. 4. The Adam optimizer (Kingma and Ba, 2014) was

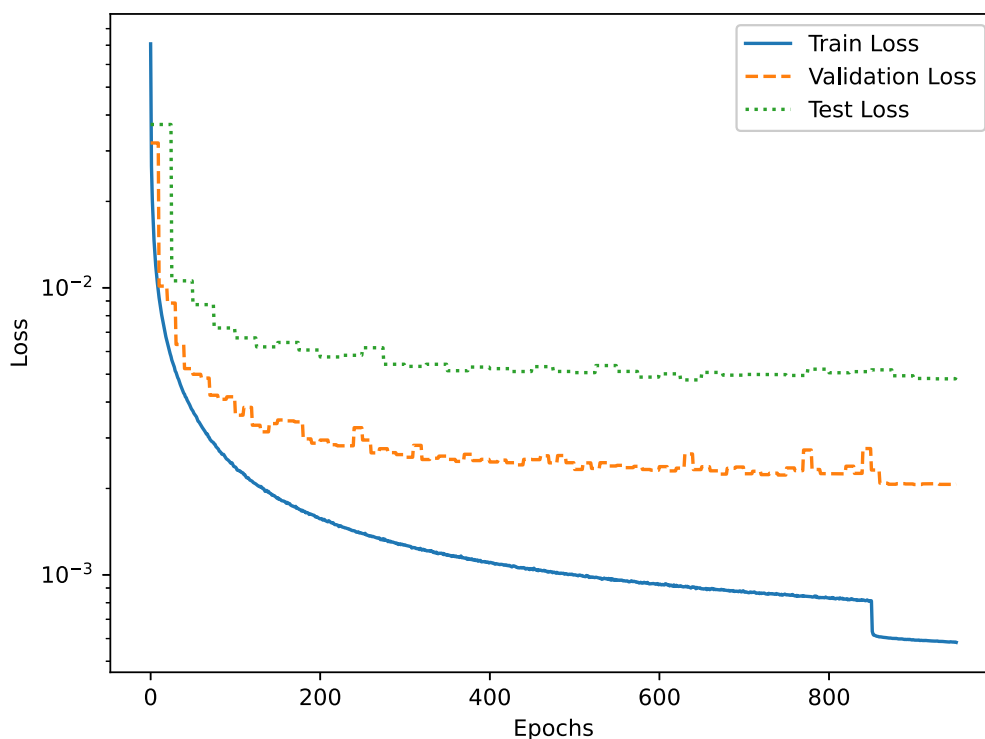


Fig. 4. Evolution of the Mean Squared Error (MSE) loss on the training (blue solid line), validation (orange dashed line), and test (green dotted line) sets over the training epochs.

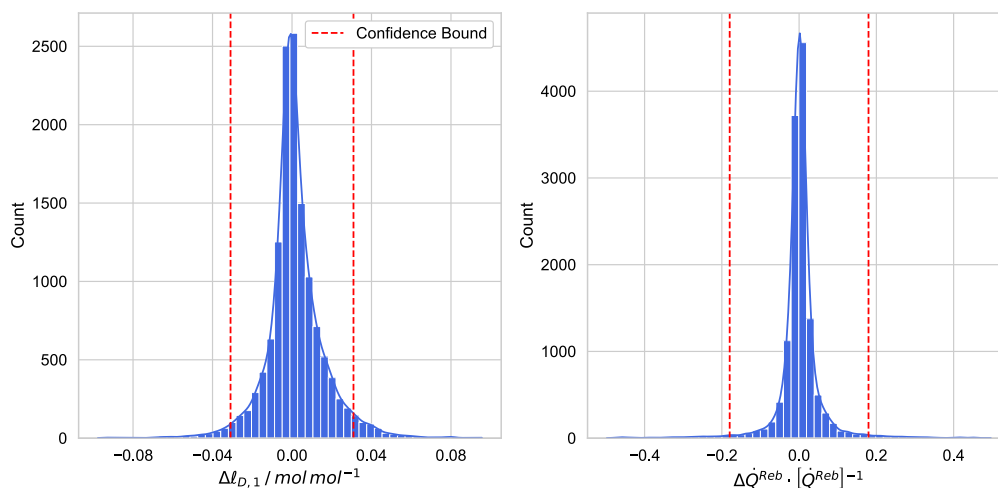


Fig. 5. Histogram of prediction errors where $\Delta\ell_{D,1} = \ell_{D,1}^{Rig} - \ell_{D,1}^{SM}$ and $\Delta\dot{Q}^{Reb} = (\dot{Q}^{Reb,Rig} - \dot{Q}^{Reb,SM}) / \dot{Q}^{Reb,Rig}$. The property with superscript *Rig* denotes the result of rigorous distillation column simulation and the property with superscript *SM* denotes the result of the surrogate model prediction. The red-dashed lines show the confidence regions obtained from calibrating the conformal prediction on the validation dataset. There are a couple (less than 20 overall) of outliers with $\Delta\ell_{D,1}$ errors >0.1 , which we removed from the figure for the sake of axis scaling.

employed to update the model parameters during training, owing to its efficiency and widespread adoption for neural network optimization. Those settings were chosen based on previous experience, as well as a small number of experimental runs. The training, validation, and test dataset consist of a total of 1,972,085, 493,041, and 13,494 data points, respectively.

Performance evaluation. The predictive accuracy of the trained surrogate model is evaluated by comparing its predictions against the ground truth values from the test dataset. The distribution of the prediction errors, shown in Fig. 5, indicates that the model achieves high accuracy for the vast majority of cases.

The results shown in Figs. 5 and 6 demonstrate that the trained surrogate model successfully generalizes across both the space of distillation column specifications and the space of ternary systems with homogeneous vapor–liquid phase behavior. This makes the trained surrogate model **reusable**, as it can be used for various systems and operating conditions, without the necessity of re-calibration or fine-tuning. For brevity, error histograms and measured vs. predicted diagrams for all molar fraction targets are not presented, as their distributions are very similar. The Root Mean Square Error (RMSE) for all targets, summarized in Table 1, confirms this observation and quantifies the overall model performance.

According to Fig. 6, the model predicts the reboiler duty with higher accuracy than the bottom and distillate compositions. This is because

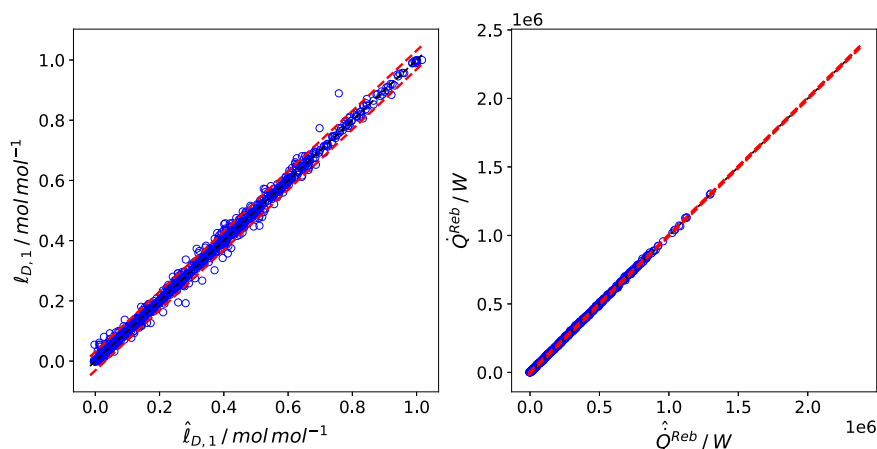


Fig. 6. Measured versus predicted values for the ternary distillation column surrogate model on the test dataset. Surrogate model predictions (horizontal axis, denoted by a $\hat{\cdot}$ symbol) are compared to ground truth values from rigorous simulation (vertical axis). The red dashed lines indicate the 95% prediction intervals, calibrated via conformal prediction on the validation set. The close clustering along the diagonal demonstrates the model's high accuracy and generalizability across diverse ternary systems and column configurations. The figure displays a randomly selected subset of the test dataset, as visualizing the entire dataset is impractical due to its large size.

Table 1

Root Mean Square Error (RMSE) values for the surrogate model predictions.

Property	RMSE
$\ell_{B,1}$	0.0257 mol mol ⁻¹
$\ell_{D,1}$	0.0271 mol mol ⁻¹
$\ell_{B,2}$	0.0189 mol mol ⁻¹
$\ell_{D,2}$	0.0193 mol mol ⁻¹
\hat{Q}^{Reb}	4216.271 W

the reboiler duty's dependency on the surrogate model input features is largely monotonic. For example, a lower bottom-to-feed split results in a higher reboiler duty. This behavior is easier for the model to learn and generalize compared to the output concentration values, which depend on the complex, often non-monotonic azeotropic behavior of the system.

An additional study was performed to analyze whether the prediction error of the trained surrogate model shows a correlation to the distance of the test dataset features to the training dataset. This is based on the idea that higher prediction error is expected for regions less densely covered in the training dataset. However, for the surrogate model and the datasets of this work, no such trend could be identified. This likely is the case since the ML-fueled training dataset generation, covering more than 100,000 mixture systems at various process specifications, fully covers the test dataset, which follows a similar feature distribution, c.f. Fig. 3. For the sake of brevity, the description and results of this analysis are provided in the supporting information (S6) to this work.

Confidence interval calibration via conformal prediction. To provide a reliable measure of the model's prediction uncertainty, we employ the method of conformal prediction (Angelopoulos and Bates, 2023). This post-processing technique is particularly well-suited for this application as it uses the validation dataset to calibrate prediction intervals with a statistically rigorous guarantee on marginal coverage. This means that, for a chosen confidence level of 95%, we can expect 95% of the true values to fall within their respective predicted intervals over the long run, adding a crucial layer of trust and reliability to the surrogate model's outputs.

The method yields a symmetric 95% confidence interval (CI) for each prediction \hat{y} of the form:

$$CI = [\hat{y} - \Delta, \hat{y} + \Delta] \quad (4)$$

Table 2

Confidence interval radii for surrogate model predictions.

Property	Value
$\Delta \ell_{B,C_1}$	0.0355 mol mol ⁻¹
$\Delta \ell_{B,C_2}$	0.0355 mol mol ⁻¹
$\Delta \ell_{D,C_1}$	0.031 mol mol ⁻¹
$\Delta \ell_{D,C_2}$	0.0355 mol mol ⁻¹
$\Delta \hat{Q}^{Reb}$	8.500 kW

where Δ is the confidence interval radius calibrated on the validation set. The calibrated radii for each of the model's output targets are presented in Table 2.

These radii, which correspond to the red-dashed lines shown in the error histogram in Fig. 5, quantify the model's expected predictive uncertainty. For instance, the model predicts molar fractions with a typical uncertainty of approximately ± 0.035 mol mol⁻¹.

Eventually, it is the availability of an uncertainty quantification of the trained surrogate model that qualifies its use in practical applications, e.g. optimization case studies, where it is mandatory to estimate the uncertainty contained in the surrogate model-based predictions. Now, the trained surrogate model can be interconnected to flowsheets and used in process simulation and optimization to support rigorous optimization and decision-making processes. In the following section, we demonstrate this by applying the surrogate model to the problem of entrainer selection for an extractive distillation.

3. Case study: Surrogate model-based entrainer selection

In this section, we demonstrate the practical application of the generalizable surrogate model by applying it to the problem of entrainer selection for extractive distillation. The goal is to efficiently screen a pool of candidate entrainers for the separation of a maximum-boiling azeotropic mixture of Acetone and Chloroform. This case study was also analyzed using a rigorous modeling approach in Bubel et al. (2025), providing a direct benchmark against which we can compare the performance and computational advantages of our surrogate-based methodology.

The entrainer distillation process and optimization problem. Entrainer distillation is a process that enhances the separation of azeotropic mixtures by introducing a third component, the entrainer, which favorably alters the vapor-liquid equilibrium (Düssel and Stichlmair, 1995). A

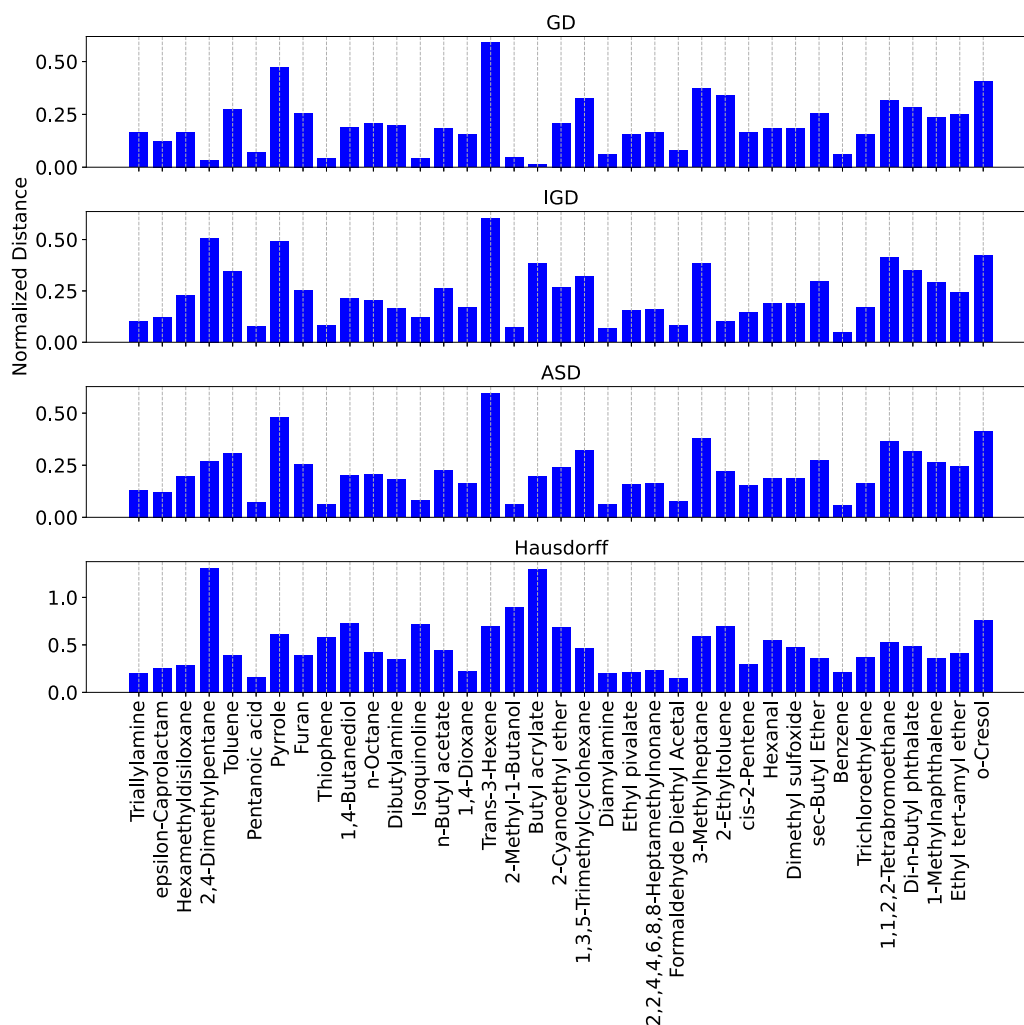


Fig. 8. Histogram of distance metrics obtained from comparing the surrogate model-based NQ curves to those obtained from rigorous optimization. The component names on the horizontal axis correspond to the entrainer candidates used in the case study. The distance metrics correspond to the *generational distance* (C.1), *inverted generational distance* (C.2), *average symmetric distance* (C.3), and *Hausdorff distance* (C.4) and are defined in Appendix C.

measures between Pareto frontiers is non-trivial (Ishibuchi et al., 2015), we use multiple metrics to better quantify the results. We use the generational distance (GD), inverted generational distance (IGD), average symmetric distance (ASD), and Hausdorff distance measures for the comparison in this section. While motivated and explained in more detail in Appendix C, a brief explanation of these metrics is:

- GD: Measures the average distance of the surrogate model-based Pareto points to their closest neighbor on the rigorous Pareto frontier.
- IGD: Measures the average distance of the rigorous Pareto points to their closest neighbor on the surrogate model-based frontier.
- ASD: The mean value of GD and IGD.
- Hausdorff: Measures the maximum distance of a point on one frontier to its closest neighbor on the other frontier.

For all of these distance measures, a lower value indicates a better approximation of the true Pareto frontier. The overall observation from Fig. 8 is that the computed distance measures vary for the different entrainer candidates considered in this comparison. The results shown in Fig. 8 show the surrogate model's ability to generalize across different operating conditions and systems. While there are candidates with particularly small values in all of the distance measures, there are some outlier candidates that exhibit above-average values in some of the distance measures. For example, the distance measures displayed

in Fig. 8 exhibit significant variation between the GD value and the Hausdorff distance for 2,4-Dimethylpentane. Such a result indicates the existence of at least one point on the rigorous frontier that is far away from the surrogate model-based frontier. Using the NQ curves for all entrainer candidates, provided in the Supporting Information to this work, this observation is confirmed. While the distance measures shown in Fig. 8 provide an overview on the distribution of NQ curve approximation quality of surrogate model-based optimizations for the different entrainer candidates considered in this work, they are not a generic measure of how useful this approximation is for a certain design study. Therefore, the suitability of the surrogate model-based NQ curves is demonstrated on the task of ranking of candidate entrainers by their efficiency, as presented later in this section.

Furthermore, the goal of surrogate modeling usually is not to replace the rigorous modeling. Instead, the results obtained from surrogate model-based optimization can be used to support the rigorous optimization. As shown in Fig. 9, using the optimal solutions from the surrogate-based optimization as warm-starts, the rigorous (MESH-equation based) optimization results in an approximation of the NQ curve, including points at a lower total number of column stages than what was found using 100 random multi-starts alone. Also, in addition to the rigorous NQ curves and the surrogate model-based NQ curve, we computed a surrogate model-based NQ curve representing a worst- and best-case scenario, considering the confidence intervals Table 2 of the trained distillation column surrogate model. Best case means, that

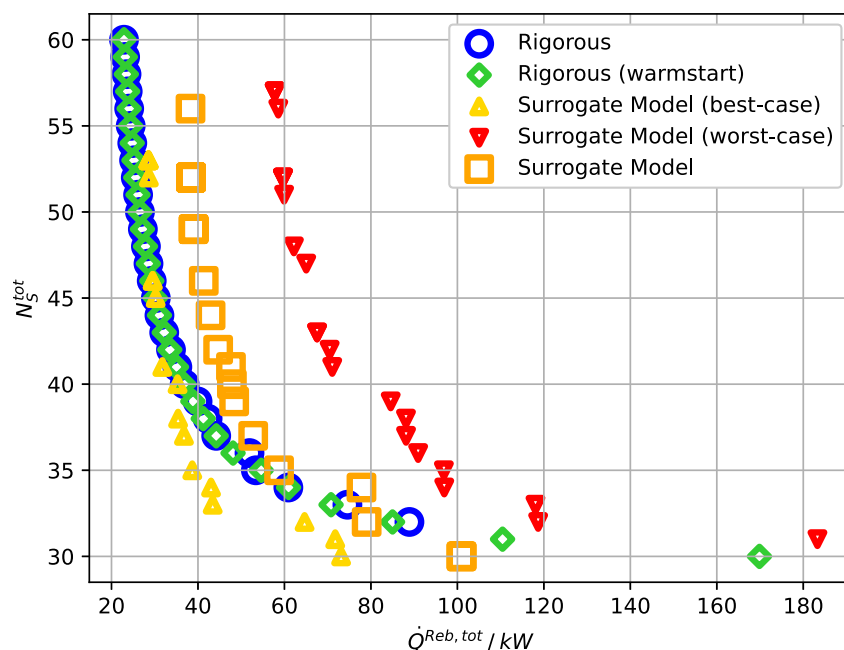


Fig. 9. NQ curves for Pentanoic acid (CAS 109-52-4) as entrainer in the system Acetone+Pentanoic acid+Chloroform. The NQ curves were obtained using the Rigorous (MESH-equation based modeling) for the optimization, using 100 random start locations (blue circles) and the solutions of the surrogate model-based optimizations (green diamonds). The NQ curves obtained using the ternary reusable surrogate model for the distillation columns are shown by the orange squares. The best- and worst-case results using the surrogate model in the optimizations are shown by the yellow upside-triangles and the red downside-triangles, respectively, while best- and worst-case means adding $\pm 0.5 \cdot \Delta$ of the confidence regions Table 2 to the surrogate model predictions.

we add $0.5 \cdot \Delta \mathcal{L}_{ProC_{O_i}}$ and subtract $0.5 \cdot \Delta \dot{Q}_{C_{O_i}}^{Reb}$ of the surrogate model predictions for each column i in the flowsheet during optimization, c.f. (Table 2). This represents the best possible situation under the calibrated surrogate model confidence regions, where the product purity is overestimated and the reboiler duty is underestimated. The worst case is computed vice versa. For the system displayed in Fig. 9, the surrogate model-based NQ curve is too pessimistic and the best case curve is closer to the rigorous NQ curve. The NQ curves for all entrainer candidates are featured in the Supporting Information to this work.

3.2. Discussion: Ranking accuracy and model limitations

The ultimate goal of the case study is to rank the entrainer candidates based on their predicted performance, thereby identifying the most promising options for further, more rigorous study. To achieve this, a dominance criterion is established to compare the NQ curves. An entrainer candidate A is considered superior to candidate B if its Pareto frontier represents a more favorable CAPEX-OPEX trade-off, i.e., achieving a lower total heat duty for a given total number of stages for the majority of points along the frontier. Using this pairwise comparison, the full set of candidates is sorted to produce the final ranking. In Fig. 10, we show the error of that ranking, compared to the results from ranking the NQ curves obtained from fully-rigorous optimization. Note that the results obtained for the NQ curves using the surrogate model-based and rigorous optimizations are only representative for the non-optimizable specifications chosen in this work. That is, the feed flowrate is chosen as 1 kmol h^{-1} and the columns operate at a pressure of $p = 1 \text{ bar}$. While changing these conditions requires recomputation of the NQ curves, and the ranking may change, the surrogate model remains usable as it was trained for various pressures.

The ranking derived from our surrogate model successfully identifies the top-performing entrainers and correctly isolates the low-performing ones. In particular the most promising candidates (top 10 category in Fig. 10) are successfully ranked by the surrogate model. This shows that the trained surrogate model has successfully been used in a reusable manner: The same surrogate model was deployed

in a flowsheet consisting of three distillation columns, all operating at different conditions, performing different separations. The surrogate models were interconnected to form the entire flowsheet and optimization on this interconnected system yields a reasonably close approximation to the true ranking for most of the entrainer candidates. Still, there are some entrainer candidates that are ranked less precisely than others. These typically are systems whose modelfluid features lie in a sparse region of the surrogate's training data space. Also, in the optimization case study used to demonstrate the reusable surrogate model, even small inaccuracies in the predicted product compositions can have a considerable impact on the final NQ curve. This is because there are regions in the design space where the total reboiler duty is highly sensitive to small changes in process parameters.

The computational effort required for the optimization based on the pre-trained reusable surrogate model is considerably cheaper than that of a full rigorous optimization. The optimization based on the surrogate model achieves a speedup of approximately two orders of magnitude compared to optimization that requires solving the MESH equations directly.

In Appendix B, we compare the ranking accuracy achieved in this work — using optimization based on the reusable surrogate model — with the results reported in Bubel et al. (2025). As discussed in more detail in Appendix B, the authors of Bubel et al. (2025) employ a different approach, relying on a local approximation of the rigorous entrainer distillation flowsheet derived from the solution of a process fluid optimization problem.

4. Conclusion and outlook

This work addresses a foundational limitation of surrogate modeling in chemical process engineering: the system-specific nature of state-of-the-art models, which restricts their reusability. We have demonstrated the development of a single, global surrogate model for distillation columns capable of generalizing not only across a wide range of operating and design parameters but, for the first time, across the entire chemical space of homogeneous ternary vapor-liquid mixtures. This

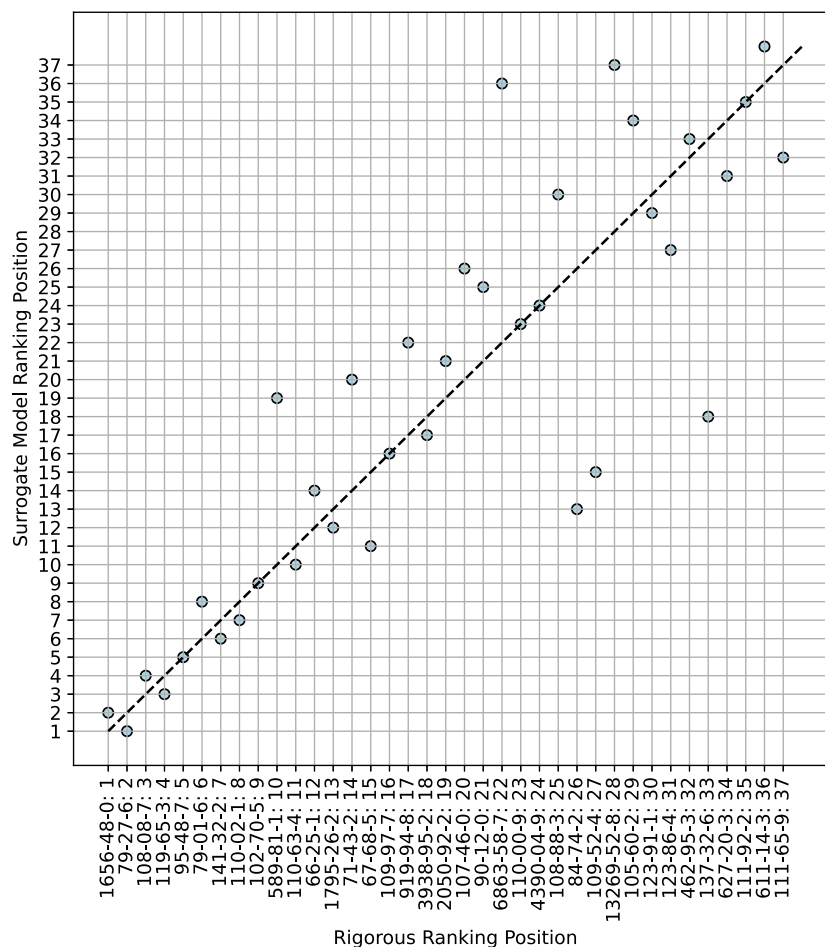


Fig. 10. Validation of the surrogate model-based ranking of entrainer candidates. The results are visualized as a parity plot. On the horizontal axis, the validation candidates are ordered from best (top, rank 1) to worst (bottom, rank 37) according to the true (rigorous optimization-based) ranking. The surrogate model-based ranking positions are presented on the vertical axis. The closer the scatter points are to the ideal diagonal line (black dashed line), the more accurate the surrogate model-based ranking.

surrogate model can be used in a variety of flowsheets and processes, supporting simulation studies and optimization problems.

This advancement in surrogate model reusability is based on two core contributions. First, we apply a modelfluid representation in a novel way that enables the surrogate to learn the fundamental relationships between thermodynamic behavior and process performance across the entire mixture space, a scope of generalization not previously achieved for distillation column surrogate models. Second, and most critically, we employ a powerful ML-for-AI strategy to overcome the data scarcity that has traditionally constrained such ambitious modeling efforts. This approach leverages an existing machine learning model for property prediction to generate a large and physically representative dataset, which in turn provides the necessary foundation to train our more complex, general-purpose AI-based surrogate. The outcome is a surrogate model of a distillation column that can be used for any ternary mixture obeying homogeneous vapor–liquid phase behavior.

In a case study on entrainer distillation, we demonstrated the practical benefit of this reusable surrogate. By providing an explicit mapping from inputs to outputs, the model enables the rapid and robust generation of Pareto-optimal frontiers for an entrainer candidate pool, a screening task that would be computationally expensive with rigorous models alone. The resulting entrainer ranking showed a strong correlation with rankings derived from rigorous optimization, proving the surrogate’s utility as a powerful tool to support and accelerate

process synthesis. Furthermore, the integration of conformal prediction provides statistically rigorous confidence intervals, offering a crucial measure of reliability for engineering applications. The ability to quantify prediction uncertainty is a key enabler for surrogate models in practical applications.

This work is intended to showcase what is possible for a new generation of surrogate models and to outline a path towards true reusable prediction models for unit operations. The goal is not to deliver a final, perfect model, but rather to establish a new methodology and a proof-of-concept for what can be achieved when ML-based data generation is combined with AI-based surrogate modeling.

While this work focuses on ternary mixtures, the modelfluid representation generally extends to multi-component systems. However, the cost of training data generation and model training increases significantly with the number of components, which poses a challenge for practical implementation. Also, in order to prevent training separate models for each number of components, an architectural approach to modeling multi-component mixtures would be of high value on the road to broader reusability. Another crucial next step would be to enhance the ML-for-AI workflow by developing more sophisticated and robust thermodynamic consistency checks for large-scale, synthetically generated fluid systems. Improving the quality and reliability of the training data at its source is the most direct path to increasing the accuracy and robustness of the final surrogate model. Furthermore,

to manage the risk of the surrogate predicting physically infeasible operating points, future work should explore the development of a complementary feasibility classifier, as demonstrated in Höller et al. (2024). Integrating such a classifier would ensure that surrogate-based optimizations converge to solutions that are not only mathematically optimal but also physically realizable.

Ultimately, this work presents the potential of pre-trained, globally valid surrogate models that can serve as off-the-shelf digital assets for the process industries.

CRedit authorship contribution statement

Martin Bubel: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tobias Seidel:** Writing – review & editing, Supervision, Investigation, Formal analysis, Conceptualization. **Michael Bortz:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

All authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors gratefully acknowledge Hans Hasse and Fabian Jirasek for the fruitful and valuable discussions.

Appendix A. VLE consistency checks

This section details the set of rules used to classify the vapor–liquid equilibrium (VLE) behavior of synthetically generated binary mixtures as thermodynamically consistent or inconsistent. These checks are a critical step in our data generation workflow, ensuring the quality and physical realism of the dataset used to train the reusable surrogate model.

The necessity for these checks arises from the empirical nature of the thermodynamic property models commonly used in process simulation, including those in this work. Models for vapor pressure (e.g., Antoine equation) and liquid-phase activity coefficients (e.g., Margules model) are not derived from first principles and can be parameterized to describe behavior that is not physically attainable.

When these models are parameterized using ML-predicted data, as done in our large-scale dataset generation (Section 2.2), the risk of producing such thermodynamic artifacts increases. Therefore, a robust filtering mechanism is required to discard these unphysical systems.

Our workflow involves generating a vast number of potential ternary systems by combining pure component data from the DIPPR 801 database (Wilding et al., 1998) with ML-predicted activity coefficients at infinite dilution (Jirasek et al., 2020; Damay et al., 2021). Before a ternary system is accepted, we analyze the VLE of its three constituent binary sub-systems. If any of the binary pairs fail one or more of the checks described below, the entire ternary system is removed from the dataset. While the checks are discussed for isobaric conditions, many of their principles apply to isothermal systems as well.

The set of checks presented in this section reflects the current state of our research and provides a robust filter for the large-scale dataset generation. It is important to note, however, that this collection of rules is based on a combination of thermodynamic principles and empirical heuristics, and does not constitute a formal proof of thermodynamic consistency. Consequently, it is possible that some systems passing these checks may still exhibit VLE profiles that appear atypical or thermodynamically questionable, even if they do not violate a specific criterion defined below.

Figures illustrating systems that violate these checks are provided in the Supporting Information to this work.

Note that the checks introduced in this section are supposed to filter x_{MF} values that are considered unphysical.

These checks are not considered in Section 3, where the x_{MF} are fixed.

Also, the feature vectors x_{MF} used in Section 3 have all been tested to satisfy the VLE consistency checks covered in this section. If the model fluid features (or part of them) were to optimize too, the checks in this section can be added to the optimization problem as constraints. For further phase stability- and azeotropy constraints for process fluid optimization, see Bubel et al. (2025).

A.1. Endpoint convergence check

This check ensures that the VLE curves correctly converge to the pure component endpoints. It is a fundamental, physically motivated test to prevent artifacts at the composition boundaries. The conditions checked at the limits of the mole fraction ℓ_i are:

- $v_i(\ell_i \rightarrow 0) = 0$
- $v_i(\ell_i \rightarrow 1) = 1$

Given the nature of the isobaric VLE calculation, the bubble and dew temperatures are guaranteed to match the pure component boiling points at these limits.

A.2. Continuity check

This check is motivated by the empirical observation that some parameterizations lead to unphysical discontinuities, or jumps, in the VLE curve. It identifies systems where an infinitesimally small change in liquid composition causes a large, finite change in the corresponding vapor composition. Since such behavior is not observed in real fluid systems, we implement this check with a configurable tolerance to filter out these artifacts. This is a purely heuristic check and the threshold needs to be set by the user. In this work, we use a threshold of $|v_i^{(j+1)} - v_i^{(j)}| \leq 0.2 \text{ mol mol}^{-1}$, where superscripts $(j+1)$ and (j) denote two consecutive points on the VLE curve.

A.3. Monotonicity check

This check enforces that the bubble and dew curves are monotonic in regions where physical principles require them to be. For a simple binary system, the vapor mole fraction v_i must be a monotonically increasing function of the liquid mole fraction ℓ_i . Violations often appear as non-physical hooks or reversals in the VLE diagram and frequently co-occur with other artifacts.

A.4. Phase envelope check

This check enforces a direct consequence of the second law of thermodynamics: at any given composition, the dew point temperature must be greater than or equal (singular VLE points) to the bubble point temperature. A direct comparison can be complicated because the standard VLE calculation yields bubble temperatures at discrete liquid compositions and dew temperatures at discrete vapor compositions, while the liquid and vapor compositions usually differ. To perform the check, we interpolate the bubble and dew temperatures onto a common composition axis and ensure that $T^{\text{dew}} \geq T^{\text{bubble}}$ across the entire range. A small tolerance is permitted to account for numerical noise, especially near azeotropes. In this work, we tolerate violations of this criterion in the range of $\pm 0.05 \text{ mol mol}^{-1}$ of an azeotropic point.

In fact, there exists the possibility to create VLEs satisfying extended Raoult's Law where for all concentrations visited, the bubble temperature is larger than the dew temperature. Using the model fluid features of a binary system and replacing the saturated vapor temperatures of the pure components generates such a system.

A.5. Azeotrope and extrema consistency check

This check combines several criteria to ensure the geometric shape of the VLE diagram is physically consistent, particularly for systems with one or more azeotropes. It validates the relationship between azeotropes and the extrema of the temperature curves. The check proceeds as follows:

- **Azeotrope Counting:** The number of azeotropes is determined by counting the sign changes of $(v_i(\ell_i) - \ell_i)$ of a series solutions of extended Raoult's Law for equidistantly monotonically increasing values of ℓ_i .
- **Extrema Matching:** The number of extrema (maxima or minima) in the bubble and dew temperature curves is counted and must match the number of azeotropes.
- **Polyazeotropy Nature:** For systems with multiple azeotropes, it enforces rules on their sequence to filter out unphysical combinations (e.g., see Anjum et al. (2024)).

A violation of any of these conditions indicates an inconsistent VLE shape.

A.6. Zeotropic curvature check

This check is designed specifically for zeotropic systems to detect unphysical S-shapes in the temperature-composition curves. Such shapes, characterized by an inflection point in the central composition range, are often indicators of (incipient) liquid-liquid immiscibility, which is not captured by the VLE model being used. The check computes the second derivative of the bubble and dew curves and flags any system with a sign change (inflection point) in the core region.

In this work, we use a tolerance of 0.1 mol mol^{-1} of the pure component locations as many systems seemed to fail this check near the pure component endpoints, while generally exhibiting a smooth and consistently looking VLE.

A.7. Numerical validity check

This final, straightforward check ensures data integrity by scanning all computed VLE data points for non-numerical values, such as NaN or Inf. The presence of such values indicates a failure in the numerical VLE solver and results in the system being discarded.

Appendix B. Comparison of ranking accuracy to related work

In addition to the discussion of the ranking error using surrogate model-based

optimization in Section 3.1, we compare the entrainer ranking results achieved in this work, by the use of the reusable surrogate model, to those presented in Bubel et al. (2025), who use the same case study, in Fig. B.11.

In Bubel et al. (2025), the authors used a first-order approximation of the objective function ($\dot{Q}^{Reb,tot}$), at the solution to their process fluid optimization problem. In that work, they did not only optimize the process specifications of the columns in the flowsheet but also the modelfluid features that describe the entrainer component or its interactions with the solutes. The results achieved in this work are not only more accurate but come with a significantly lower computational effort compared to the rigorous process fluid optimization conducted in Bubel et al. (2025). In Fig. B.11, we show the differences (error) in the entrainer candidate ranking from the rigorous approximation used in Bubel et al. (2025) and the surrogate model-based ranking presented in this work. It is found that the majority of systems are ranked in close vicinity while the surrogate model-based ranking outperforms the results from Bubel et al. (2025) in identifying the best-performing entrainers. For the mid- and low-performing candidates, the ranking

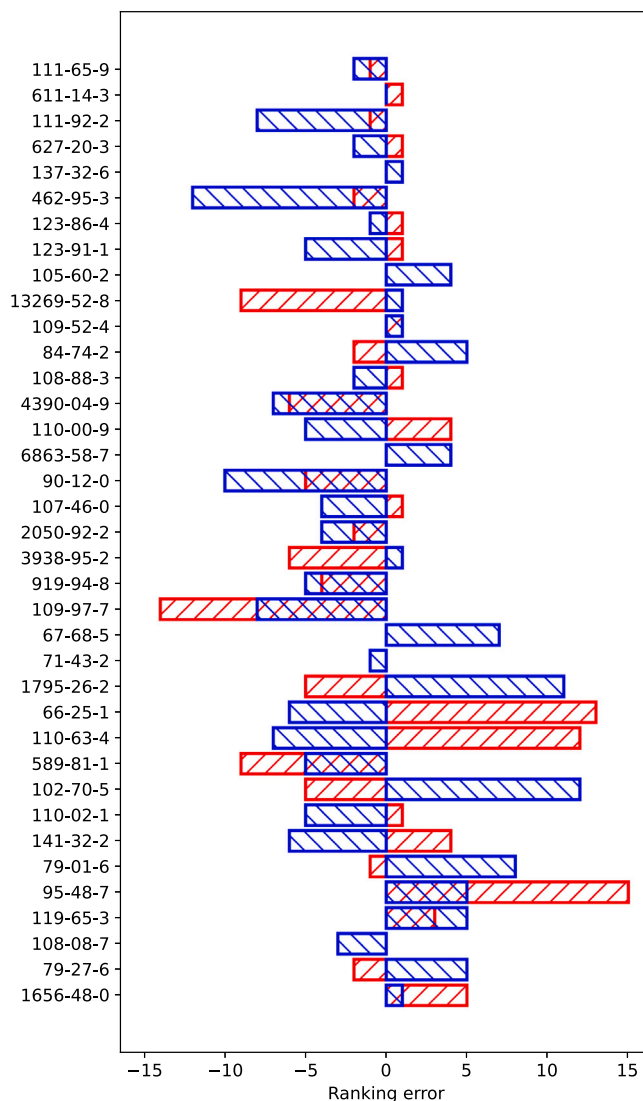


Fig. B.11. Comparison of the error using surrogate model-based ranking of NQ curves from this work to the error of the approximate ranking presented from Bubel et al. (2025). Candidates are ordered from best (top, rank 1) to worst (bottom, rank 37) according to the rigorous optimization-based ranking. Red bars (“//”) show the ranking error of this work (surrogate modeling); blue bars (“\\”) show the ranking error of Bubel et al. (2025). The horizontal bars indicate the magnitude of the ranking error, defined as the absolute difference between the true rank obtained from the rigorous NQ curves and the approximate ranking obtained using the surrogate model-based ranking or the one from Bubel et al. (2025).

error varies between the two approaches. The mean value over the absolute ranking error for all entrainer candidates is 3.73 for the surrogate model-based ranking and 4.84 for the ranking from Bubel et al. (2025).

A further important consideration arises from the differing domains of the rigorous and surrogate models. Rigorous simulators can fail to converge for certain input specifications, effectively creating holes or infeasible regions in the design space. The surrogate model, by its nature as an explicit function, will always return a prediction for any given input. This is favorable for an optimization algorithm, avoiding convergence failures that can prematurely terminate rigorous optimization runs. However, this robustness also introduces a critical challenge. The optimizer may identify an optimal point in a region where the rigorous model would not have a feasible solution, as the surrogate is

trained only on data from successful simulations and has no inherent knowledge of this feasibility boundary. While this behavior was not observed to be a dominant issue in this work, it remains a crucial aspect to manage when applying surrogate models in practice. A robust strategy to mitigate this risk is to complement the regression surrogate with a classification model trained to predict process feasibility, an approach that is discussed further in the outlook. Also, the optimization problem using the surrogate model may be different to that based on the rigorous modeling. An immediate effect of this may be that convex NQ curves using the rigorous modeling can become non-convex using the surrogate models. This can be observed by the NQ curves for some of the entrainer candidates presented in the Supporting Information to this work.

Appendix C. Distance metrics for Pareto frontier comparison

In this section we describe in distance measures used to compare the Pareto frontiers obtained from the surrogate model-based optimizations to the reference frontier using the rigorous modeling of the distillation columns in the flowsheet.

We use the generational distance (GD) and inverted generational distance (IGD) measures, as described in Ishibuchi et al. (2015), to assess the proximity of the surrogate model-based frontiers to that of the rigorous optimization and vice versa. We also use the average symmetric distance (ASD), which is the mean value of GD and IGD, and the Hausdorff distance to provide a more comprehensive comparison. For all of these distance measures, a lower value indicates a better approximation of the true Pareto frontier. The GD metric (C.1), which measures the average distance of the Pareto points of the surrogate model-based optimization to their respective closest-neighbor on the rigorous Pareto frontier, may not reveal if the surrogate model-based frontier does not cover the entire frontier from the rigorous optimization. On the other hand, the IGD metric (C.2), which measures the average distance of the Pareto points of the rigorous optimization to their respective closest-neighbor on the surrogate model-based frontier, may not reveal if the surrogate model-based frontier contains any points that are far away from the rigorous Pareto frontier. This is why we also use the average symmetric distance (ASD) (C.3), which yields the average of both the GD and IGD metrics. In addition, we use the Hausdorff distance metric (C.4), which measures the maximum distance of a point on one frontier to its closest-neighbor on the other frontier. This metric can be used as a warning flag, indicating when there are outliers on any of the curves in the comparison, both for the surrogate model-based and the rigorous optimization.

Since, we are only interested in the geometric distance between the approximate frontiers and the reference (rigorous) frontiers, we refrain from using any metrics that consider dominance of frontiers, such as hypervolume or IGD+ (Ishibuchi et al., 2015).

For the distance measures described in the following, we let $P_{\text{true}} = \{t_1, t_2, \dots, t_{|P_{\text{true}}|}\}$ be the set of points representing the true Pareto front, and let $P_{\text{approx}} = \{a_1, a_2, \dots, a_{|P_{\text{approx}}|}\}$ be the set of points representing an approximate front.

When using the distances in Section 3, we consider the NQ curves obtained from the optimization on the rigorous distillation column model as the true front, and that NQ curves from the surrogate model-based optimizations as the approximate front. The distance between any two points u and v in the objective space is given by the Euclidean distance $d(u, v)$. The below formulas and definitions are obtained from Ishibuchi et al. (2015) and the references therein.

Generational Distance (GD). The GD metric measures the average distance from each point in the approximate front to its nearest neighbor in the true front. It quantifies how close the found solutions are to the true solutions. A lower value is better.

$$GD(P_{\text{approx}}, P_{\text{true}}) = \frac{1}{|P_{\text{approx}}|} \sum_{a \in P_{\text{approx}}} \min_{t \in P_{\text{true}}} d(a, t) \quad (\text{C.1})$$

Inverted Generational Distance (IGD). The IGD metric measures the average distance from each point in the true front to its nearest neighbor in the approximate front. It is a comprehensive metric that evaluates both the convergence and the diversity (coverage) of the approximate front. A lower value is better.

$$IGD(P_{\text{approx}}, P_{\text{true}}) = \frac{1}{|P_{\text{true}}|} \sum_{t \in P_{\text{true}}} \min_{a \in P_{\text{approx}}} d(t, a) \quad (\text{C.2})$$

Average Symmetric Distance (ASD). The ASD, provides a balanced, symmetric measure of the average geometric distance between the two fronts. It is simply the mean of the GD and IGD metrics. A lower value indicates a better overall geometric match.

$$ASD(P_{\text{approx}}, P_{\text{true}}) = \frac{GD(P_{\text{approx}}, P_{\text{true}}) + IGD(P_{\text{approx}}, P_{\text{true}})}{2} \quad (\text{C.3})$$

Hausdorff distance. The Hausdorff distance (d_H) measures the maximum worst-case distance between the two fronts. It is defined as the maximum of two directed distances, where each directed distance finds the point in one set that is farthest from any point in the other set. It is highly sensitive to outliers. A lower value is better.

$$d_H(P_{\text{approx}}, P_{\text{true}}) = \max \left(\max_{a \in P_{\text{approx}}} \left\{ \min_{t \in P_{\text{true}}} d(a, t) \right\}, \max_{t \in P_{\text{true}}} \left\{ \min_{a \in P_{\text{approx}}} d(t, a) \right\} \right) \quad (\text{C.4})$$

Appendix D. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.compchemeng.2025.109523>.

Data availability

The authors do not have permission to share data.

References

- Abranches, D.O., Maginn, E.J., Colón, Y.J., 2023. Activity coefficient acquisition with thermodynamics-informed active learning for phase diagram construction. *AIChE J.* 69 (8), e18141. <http://dx.doi.org/10.1002/aic.18141>, <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.18141>.
- Angelopoulos, A.N., Bates, S., 2023. Conformal prediction: A gentle introduction. *Found. Trends Mach. Learn.* 16 (4), 494–591. <http://dx.doi.org/10.1561/2200000101>.
- Anjum, N., Oba, S., Hiaki, T., Hsieh, C.J., Chen, C.C., 2024. Thermodynamic modeling of double azeotropy systems. *Fluid Phase Equilib.* 578, 113993. <http://dx.doi.org/10.1016/j.fluid.2023.113993>, <https://www.sciencedirect.com/science/article/pii/S037838122300273X>.
- Aspen Technology Inc., 2017. Aspen plus, version 10.
- Bárkányi, Á., Chován, T., Németh, S., Abonyi, J., 2021. Modelling for digital twins—Potential role of surrogate models. *Processes* 9 (3), 476. <http://dx.doi.org/10.3390/pr9030476>, <https://www.mdpi.com/2227-9717/9/3/476>.
- Biegler, L.T., Grossmann, I.E., Westerberg, A.W., 1997. *Systematic Methods of Chemical Process Design*. Prentice Hall PTR, ISBN: 0134924223.
- Bubel, M., Ludl, P.O., Seidel, T., Aspiron, N., Bortz, M., 2021. A modular approach for surrogate modeling of flowsheets. *Chem. Eng. Tech.* 93 (12), 1987–1997. <http://dx.doi.org/10.1002/cite.202100077>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/cite.202100077>.
- Bubel, M., Seidel, T., Bortz, M., 2025. A machine learning-fueled modelfluid for flowsheet optimization. <https://arxiv.org/abs/2509.02242>. [arXiv:2509.02242](https://arxiv.org/abs/2509.02242) [cs.CE].
- Caballero, J.A., Grossmann, I.E., 2008. An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE J.* 54 (10), 2633–2650. <http://dx.doi.org/10.1002/aic.11579>, <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.11579>.
- Caprio, U.Di., Degréve, J., Hellinckx, P., Waldherr, S., Leblebici, M.E., 2023. HybridGamma: A thermodynamically consistent framework for hybrid modelling of activity coefficients. *Chem. Eng. J.* 475, 146104. <http://dx.doi.org/10.1016/j.cej.2023.146104>, <https://www.sciencedirect.com/science/article/pii/S1385894723048350>.
- Damay, J., Jirasek, F., Kloft, M., Bortz, M., Hasse, H., 2021. Predicting activity coefficients at infinite dilution for varying temperatures by matrix completion. *Ind. Eng. Chem. Res.* 60 (40), 14564–14578. <http://dx.doi.org/10.1021/acs.iecr.1c02039>.

- Düssel, R., Stichlmair, J., 1995. Separation of azeotropic mixtures by batch distillation using an entrainer. *Comput. Chem. Eng.* 19, 113–118. [http://dx.doi.org/10.1016/0098-1354\(95\)87024-5](http://dx.doi.org/10.1016/0098-1354(95)87024-5), <https://www.sciencedirect.com/science/article/pii/S0098135495870245>.
- Forrester, A.I.J., Söbester, A., Keane, A.J., 2008. *Engineering Design Via Surrogate Modelling: A Practical Guide*. John Wiley & Sons, Ltd, <http://dx.doi.org/10.1002/9780470770801>.
- Freund, H., Sundmacher, K., 2008. Towards a methodology for the systematic analysis and design of efficient chemical processes: Part 1. From unit operations to elementary process functions. *Chem. Eng. Process.: Process. Intensif.* 47 (12), 2051–2060. <http://dx.doi.org/10.1016/j.cep.2008.07.011>, <https://www.sciencedirect.com/science/article/pii/S0255270108001669>.
- Höller, J., Bubel, M., Heese, R., Ludl, P.O., Schwartz, P., Schwientek, J., Aspiron, N., Wlotzka, M., Bortz, M., 2024. Adaptively exploring the feature space of flowsheets. *AIChE J.* 70 (7), e18404. <http://dx.doi.org/10.1002/aic.18404>, <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.18404>.
- Ishibuchi, H., Masuda, H., Tanigaki, Y., Nojima, Y., 2015. Modified distance calculation in generational distance and inverted generational distance. In: Gaspar-Cunha, A., Henggeler Antunes, C., Coello, C.C. (Eds.), *Evolutionary Multi-Criterion Optimization*. Springer, Cham, pp. 110–125. http://dx.doi.org/10.1007/978-3-319-15892-1_8.
- Jirasek, F., Alves, R.A.S., Damay, J., Vandermeulen, R.A., Bamler, R., Bortz, M., Mandt, S., Kloft, M., Hasse, H., 2020. Machine learning in thermodynamics: Prediction of activity coefficients by matrix completion. *J. Phys. Chem. Lett.* 11 (3), 981–985. <http://dx.doi.org/10.1021/acs.jpcclett.9b03657>.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980). <https://arxiv.org/abs/1412.6980>.
- McBride, K., Sundmacher, K., 2019. Overview of surrogate modeling in chemical process engineering. *Chem. Ing. Tech.* 91 (3), 228–239. <http://dx.doi.org/10.1002/cite.201800091>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/cite.201800091>.
- Nentwich, C., Engell, S., 2019. Surrogate modeling of phase equilibrium calculations using adaptive sampling. *Comput. Chem. Eng.* 126, 204–217. <http://dx.doi.org/10.1016/j.compchemeng.2019.04.006>, <https://www.sciencedirect.com/science/article/pii/S0098135418312626>.
- Palmer, K., Realf, M., 2002a. Metamodeling approach to optimization of steady-state flowsheet simulations: Model generation. *Chem. Eng. Res. Des.* 80 (7), 760–772. <http://dx.doi.org/10.1205/026387602320776830>, <https://www.sciencedirect.com/science/article/pii/S0263876020272248X>.
- Palmer, K., Realf, M., 2002b. Optimization and validation of steady-state flowsheet simulation metamodels. *Chem. Eng. Res. Des.* 80 (7), 773–782. <http://dx.doi.org/10.1205/026387602320776849>, <https://www.sciencedirect.com/science/article/pii/S02638760202722491>.
- Peivaste, I., Belouettar, S., Mercuri, F., Fantuzzi, N., Dehghani, H., Izadi, R., Ibrahim, H., Lengiewicz, J., Belouettar-Mathis, M., Bendine, K., Makradi, A., Horsch, M., Klein, P., El Hachemi, M., Preisig, H.A., Rezgui, Y., Konchakova, N., Daouadi, A., 2025. Artificial intelligence in materials science and engineering: Current landscape, key challenges, and future trajectories. *Compos. Struct.* 372, 119419. <http://dx.doi.org/10.1016/j.compstruct.2025.119419>, <https://www.sciencedirect.com/science/article/pii/S0263822325005847>.
- Rittig, J.G., Felton, K.C., Lapkin, A.A., Mitsos, A., 2023. Gibbs-duhem-informed neural networks for binary activity coefficient prediction. *Digit. Discov.* 2 (6), 1752–1767. <http://dx.doi.org/10.1039/D3DD00103B>.
- Sanchez Medina, E.I., Linke, S., Stoll, M., Sundmacher, K., 2023. Gibbs–Helmholtz graph neural network: capturing the temperature dependency of activity coefficients at infinite dilution. *Digit. Discov.* 2 (3), 781–798. <http://dx.doi.org/10.1039/D2DD00142J>.
- Sethi, S., Zhang, X., Sundmacher, K., 2025. Process-driven solvent screening for efficient extractive distillation using interpolative rational functions. *Chem. Eng. Sci.* 301, 120675. <http://dx.doi.org/10.1016/j.ces.2024.120675>, <https://www.sciencedirect.com/science/article/pii/S0009250924009758>.
- Specht, T., Nagda, M., Fellenz, S., Mandt, S., Hasse, H., Jirasek, F., 2024. HANNA: Hard-constraint neural network for consistent activity coefficient prediction. *Chem. Sci.* 15 (47), 19777–19786. <http://dx.doi.org/10.1039/D4SC05115G>.
- Sun, J., Xue, J., Yang, G., Li, J., Zhang, W., 2025. Vapor–liquid phase equilibrium prediction for mixtures of binary systems using graph neural networks. *AIChE J.* 71 (2), e18637. <http://dx.doi.org/10.1002/aic.18637>, <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.18637>.
- Sun, G., Zhao, Z., Sun, S., Ma, Y., Li, H., Gao, X., 2023. Vapor–liquid phase equilibria behavior prediction of binary mixtures using machine learning. *Chem. Eng. Sci.* 282, 119358. <http://dx.doi.org/10.1016/j.ces.2023.119358>, <https://www.sciencedirect.com/science/article/pii/S0009250923009144>.
- Wilding, W.V., Rowley, R.L., Oscarson, J.L., 1998. DIPPR project 801 evaluated process design data. *Fluid Phase Equilib.* 150–151, 413–420. [http://dx.doi.org/10.1016/S0378-3812\(98\)00341-0](http://dx.doi.org/10.1016/S0378-3812(98)00341-0), <https://www.sciencedirect.com/science/article/pii/S0378381298003410>.
- Winz, J., Fromme, F., Engell, S., 2025. Bayesian optimization of gray-box process models using a modified upper confidence bound acquisition function. *Comput. Chem. Eng.* 194, 108976. <http://dx.doi.org/10.1016/j.compchemeng.2024.108976>, <https://www.sciencedirect.com/science/article/pii/S0098135424003946>.
- Winz, J., Nentwich, C., Engell, S., 2021. Surrogate modeling of thermodynamic equilibria: Applications. *Sampl. Optim. Chem. Ing. Tech.* 93 (12), 1898–1906. <http://dx.doi.org/10.1002/cite.202100092>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/cite.202100092>.