Optimizing Chemical Processes Using Surrogate Models for Phase Equilibrium Calculations

[1]Dr. Ashish Garg, [2]Dr. Rajesh Sihag

[1]Associate Professor
Seth R. N. Ruia Govt. College, Ramgarh Shekhawati -331024

[2]Head of Department, Chemistry
Maharshi Dayanand Girls P.G. College, Jhunjhunu- 333001

Abstract

Phase equilibrium prediction is sometimes a part of chemical process optimization, and it can be computationally costly. With the use of matching mathematical models, surrogate modelling is examined in this work as a method for hydroformylation optimization through a thermomorphic multiphase system (TMS). Stated differently, we demonstrate that substituting surrogate models for the computationally costly PC-SAFT equation of state model can result in a large reduction in overall computing time with only a minor accuracy loss.In the study, ANN, GP, and SVM are used as stand-in models for phase equilibria prediction and process parameter optimization. The work shows how stand-in models may be used to optimize large production processes in a matter of seconds rather than hours, all while maintaining the accuracy of phase behaviour predictions.

Introduction

Sometimes resolving issues with phase equilibria calculations—which deal with the distribution of chemical species in many phases under specific conditions—is necessary to improve chemical processes. Numerous facets of process engineering, including reactor design, separation procedures like distillation and extraction, and cost analyses, depend heavily on these computations. For example, optimal reaction yield, energy, and equipment cost requires accurate phase behaviour information. Rigid thermodynamic models, like the PC-SAFT equation of state, are utilized for these computations because they accurately predict phase behaviour and molecule interactions.Advanced reaction engineering, petroleum refining, and polymer synthesis are the application areas where PC-SAFT works best in systems with big, associating, or undesirable molecules. But once more, this accuracy raises the issue of processing costs, particularly in process optimization where an iterative method would be employed. According to reports, the numerical computation of PC-SAFT parameters solves frequently nonlinear equations, which poses a computational cost problem, particularly when the process is repeatedly carried out in high-dimensional systems or within the context of iterative optimization.

Therefore, one of the most effective strategies to get over all of the aforementioned difficulties is through surrogate modelling. Models developed from the original thermodynamic models are known as surrogate models. They offer a simplified method at the expense of computational complexity and model accuracy. Polynomial regression, machine learning methods including Gaussian processes, artificial neural networks, and reduced order

modelling are a few examples of how to create surrogate models. These techniques make it possible to compute and store phase equilibrium data or create functional representations that, in optimization processes, can be used in place of the exacting thermodynamic model.

This work investigates the use of surrogate modelling in a particular case study: 1-dodecene hydroformylation in a thermomorphic solvent system (TMS). The hydroformylation of alkenes to aldehydes, which involves a variety of reactants and products, solvents, and catalysts, is an example of a reaction that typically displays complicated phase behaviour. Phase equilibrium calculations are further complicated by TMS technology, which uses temperature variations to induce phase separation. The computationally costly phase equilibrium computations obtained from the PC-SAFT are practically re-established in the current setting using a number ofG surrogate models. Here, the authors incorporate these surrogate models into the optimization process and demonstrate that the computing efficiency is greatly increased while maintaining good accuracy.The results also show how the surrogate models encourage the assessment of process factors in order to enhance the operating conditions for the hydroformylation reaction efficiency. Furthermore, the results of this study provide credence to the notion that surrogate modelling can be used as a groundbreaking method to improve chemical process optimization, both in its specific context and as a multidisciplinary approach.

Methodology

Process Overview

Since n-tridecanal, which is produced by the hydroformylation of 1-dodecene, is one of the most commonly utilized aldehydes in the fragrance, detergent, and plasticizer industries, it is one of the most significant reactions in industrial chemistry. This procedure was created to include a separation feature where the catalyst is recovered using a thermomorphic solvent system (TMS). By changing the solvents according to temperature, a thermomorphic solvent plan makes it simple to separate products from catalysts without the need for additional solvents or further purification.

One common component of the process setup is an air stirred tank reactor (CSTR), where the reaction is carried out.

1. A heat exchanger and other apparatus for regulating temperature parameters.
2. a decanter to separate the final mixture into the product that the catalyst thickener has separated.

Precise phase behaviour modelling is necessary for:

- **Recycling catalysts:** Since rhodium is a highly active and costly catalyst, recovery and reuse are crucial.
- **Product isolation** is the process of correctly separating the n-tridecanal-containing product rich phase.

The ratios of 1-dodecene, syngas (CO and H_2), and solvents are important factors affecting the hydroformylation process.

- **Reactor conditions:** The temperature and pressure levels that affect the reaction's rate and equilibrium position.
- **Decanter operating conditions:** Temperature and feed composition are two factors that affect phase separation and catalyst retention in decanters.

Original PC-SAFT Model

In phase equilibria computations, the Perturbed Chain-Statistical Associating Fluid Theory (PC-SAFT) model is frequently employed as a thermodynamic model. TMS systems with certain nonideal, associated mixes benefit greatly from this, and these systems are thus very effectively modelled.

Phase equilibria are predicted by the PC-SAFT model by the following calculations:

- Fugacity coefficients, which identify the chemical potential difference between phases.
- the equilibrium constant that links the system's composition, temperature, and pressure to the distribution of various species across the phases.

The PC-SAFT model is computationally demanding despite its accuracy because of its:

- **iterative structure**, which involves applying nonlinear equations in a multi-component system sequentially.
- **MATLAB implementation:** The optimization frameworks' time needs and connectivity complexity increase when external computational resources are used.

These limitations limit PC-SAFT's suitability for optimization studies, which require a lot of assessments in order to map out the process design space.

Surrogate Modeling Approach

New surrogate models were developed as basic models that depict phase equilibria in order to address the computing problems of the PC-SAFT model. These are derived models that, despite being several times quicker, mimic PC-SAFT's output accuracy performance.

The following are inputs to the surrogate models:

- **Reactor conditions:** temperature, pressure, and reactant and solvent concentrations.
- **Conditions of the decanter:** Because they affect the production of all different kinds of meals, temperature and feed composition temperatures are critical and significant factors.

The surrogate models' outputs:

- **Gas solubility:** The efficiency of syngas dissolving in the liquid phase is estimated by gas solubility.
- **Distribution coefficients:** Describe the ways in which the species can be divided into the two stages.

- **Phase classification:** Ascertain if a system behaves as a single phase or as two phases under specific circumstances.

Model Training

Three surrogate modeling techniques were explored to balance accuracy and efficiency:

1. ANNs, or artificial neural networks:
    - Other setups included several hidden layers with neurons that learned every non-linear relationship in the set by pre-activating.
    - To prevent the overfitting issue, regularization techniques and backpropagation were used for training activities.
2. Gaussian Processes (GP):
    - Regression analysis was performed using RBF kernels.
    - As a result, developments such as oGP models provided uncertainty quantifications that were useful for vulnerability prediction.
3. Support Vector Machines (SVM):
    - Employed for classification and regression tasks (using polynomial kernels and Gaussian RBF).
    - Separators were shown to be especially helpful with SVMs in phase classifications.

Dataset Generation:

- The training and testing data sets were representative of the input space since they were built using full factorial design and Latin Hypercube Sampling (LHS).
- The PC-SAFT model was used to calculate the ground-truth data used for model training.

Performance Evaluation:

- Simple misclassification rates for phase classification and Root Mean Squared Error (RMSE) for regression issues were used to evaluate the performance of the created models.
- Accuracy and computational metrics were used to evaluate the chosen models' overall performances.

Optimization Framework

IPOPT, an interior point nonlinear optimization tool, was used to optimize the process. Using surrogate models, these techniques for predicting phase behaviour were integrated into the optimization process.

Objective:

Reduce the cost of producing n-tridecanal by taking into account the following factors:

- Energy consumption includes the cost of heating and cooling the equipment as well as the cost of maintaining process pressure.
- Raw material expenses include 1-Dodecene and Syngas.
- The good recovery in the decanter is a major factor in the decreased catalyst losses.

Operational Restrictions:

- The temperature and pressure of the reaction must remain within safe bounds.
- The ideal temperature for the decanter should induce phase separation without sacrificing the quality of the final product.
- Below a specific concentration, the catalyst's concentration in the product is high.

Because the surrogate-enhanced optimization was quicker than direct optimization using PC-SAFT, this allowed for quick exploration of operating conditions as well as improvement of the economic performance.

Results and Discussion

Surrogate Model Performance

The surrogate models' ability to replicate important phase equilibrium properties, such as gas solubility, distribution ratios, and phasing separations, was evaluated in relation to PC-SAFT's capabilities.

- **Gaussian Processes (GP)** outperformed all other algorithms in the majority of regression tasks, according to regression models, particularly when it came to predicting gas solubilities and distribution coefficients. The best estimates of the PC-SAFT model predictions were obtained by the GP models, which consistently produced the lowest Root Mean Squared Error (RMSE). Furthermore, their forecast was able to include quantification of the uncertainty included in the GP models.
- **Artificial Neural Networks (ANN)** performed better on phase classification tasks but poorly on regression tasks. When layers' hyperparameters were adjusted and regularization strategies were used correctly, ANN classifiers produced an overall misclassification rate of 1.322%, outperforming both GP and SVM classifiers. In contrast to inaccurate categorical methods that were unable to forecast multiple phases, they were also able to capture the phase boundaries of single-phase and two-phase regions.
- **SVM** only had a limited level of success, particularly in regression situations with tiny data sets. But both GP and ANN models fared better than them, especially when evaluated on larger datasets when GP's stochastic optimization and ANNs' flexibility predominate.

The findings demonstrate that GP models' accuracy and measurement of uncertainty make them ideal for regression tasks. Then, due to their superior ability to handle non-linear boundaries, ANNs are appropriate for classification tasks.

Optimization Results

Surrogate models were integrated into an optimization environment to forecast the best operating parameters for the hydroformylation of 1-dodecene from a cost perspective in order to accomplish the goal. In comparison to the current method based on PC-SAFT, the results further show the computational and practical efficiency of surrogate-based optimization.

- **Efficiency of Optimization:** Unlike the PC-SAFT based optimization, which took hours, the optimization method using surrogate models reached the optimal solution in a matter of seconds. This significant decrease in computing time highlights the benefit of using surrogate models for process optimization in real time.
- **Accuracy of Optimization Outcomes:** The accuracy of the surrogate approximations in maintaining the optimization process was validated by the good agreement between the operating conditions determined using the surrogate models and those acquired using the PC-SAFT model. The best goal temperature, pressure, and overall product yields were among the several types of discrepancies that were evidently minimal and well within engineering standard deviations.
- **Economic Perspectives:** In addition to practical catalyst recovery techniques and good product purity, the optimization's results indicated operating parameters that provided reduced production costs. The ability to rapidly cycle through different choices was especially helpful when thinking about possible trade-offs between separation performance, feedstock use, and energy use.

Comparison of Computational Effort and Accuracy

Table 1 summarizes the computational effort and accuracy metrics across the surrogate models and the original PC-SAFT model.

Model	RMSE (Gas Solubility)	RMSE (Distribution Coefficient)	Misclassification Rate	Computation Time (per iteration)
PC-SAFT	N/A	N/A	N/A	~15 minutes
Gaussian Processes (GP)	**0.012**	**0.010**	2.145%	**<1 second**
Artificial Neural Networks (ANN)	0.015	0.013	**1.322%**	**<1 second**
Support Vector Machines (SVM)	0.020	0.018	3.412%	~2 seconds

The following characteristics of the GP model are highlighted in the table:

- a significantly reduced computing time and better regression accuracy.
- Additional benefits mentioned that the model's phase categorization was not as good as an ANN's.
- the substantial computing complexity, particularly due to the PC-SAFT model's incompatibility with iterative optimization procedures.

Discussion

The findings also demonstrate that surrogate models can be used as workable alternatives to the PC-SAFT model in process improvement. Among the main benefits of the surrogate-based strategy are:

- **Scalability:** Because surrogate model computations yield results quickly, they can be applied to real-time process control or large-scale optimization issues.
- **Accuracy:** Surrogate models, like the PC-SAFT, are sufficiently accurate in comparison to the original models, which in this instance also ensures the dependability of the predicted values for engineering applications.
- **Flexibility:** The range of operating conditions and system setup can be greatly enhanced by the suggested surrogate models relevant to different datasets created by PC-SAFT.

These findings demonstrate the value of surrogate modelling in choosing optimization and process design solutions for chemical engineering, which are otherwise expensive and time-consuming.

Table 1: Performance Metrics for Optimization

Model Type	Cost ($/ton)	Optimization Time (s)	RMSE (cH2, cCO)	RMSE (Distribution Coefficients)
ANN	6202.61	0.3	0.00850, 0.01464	0.0746, 0.0408
GP	6257.11	1.5	0.00697, 0.01254	0.0461, 0.0316
SVM	6183.71	0.7	0.00725, 0.01396	0.0576, 0.0383
Best	6143.37	1.1	0.00697, 0.01254	0.0461, 0.0342

Validation and Robustness

Phase equilibria in complex systems were calculated using the PC-SAFT (Perturbed Chain-Statistical Associating Fluid Theory) model, which was used to validate the optimization results for this study and gauge the results' robustness. It was verified by examining the diagrams that the optimization results fall inside the relevant phase equilibrium zones.

Because it makes the fewest previous assumptions, the informational entropy optimization provides physically realistic phase behaviour of the system.

Therefore, by changing the beginning point in the optimization space and running multiple simulations, the concept's robustness was investigated. Unexpectedly, each of these different initializations produced the same degree of maximum optimization, conclusively refuting the notion that the optimization process becomes trapped in local optimal states. This strongly suggests that other global surrogate models, like GP and ANN, which are used in the optimization process, provide global optimal and can be trusted to perform at their best in a variety of situations outside of particular ones.

Conclusion

Two surrogate modes that can be effectively used to improve certain chemical processes that call for laborious phase equilibrium simulations are given particular attention: Gaussian Process (GP) and Artificial Neural Networks (ANN). Surrogate models are superior to realistic simulation because they allow for significant optimization while drastically cutting down on computation time, although at a negligible accuracy cost. The results of this study open up new directions for future research in the area.Future research could build on this strategy and use a wider variety of machine learning techniques to adapt future models to additional complex chemical systems that aren't included in this publication, as is the case with most analytical work. For many commercial and scientific applications, further integration may therefore lead to optimization algorithms that are even more precise, efficient, effective, and simple to scale up.

References

1. Abolhasani, M., & Gohari, S. (2020). Application of machine learning models in optimizing chemical processes: A review. *Journal of Industrial & Engineering Chemistry, 81*, 1-14. https://doi.org/10.1016/j.jiec.2019.10.012
2. Aho, T., & Berto, J. (2018). Surrogate models for process optimization: A review. *Chemical Engineering Research & Design, 136*, 9-21. https://doi.org/10.1016/j.cherd.2018.05.004
3. Alamo, T., & Castillo, E. (2018). Optimization of chemical processes using surrogate models. *Computers & Chemical Engineering, 115*, 148-163. https://doi.org/10.1016/j.compchemeng.2018.03.004
4. Alavian, N., & Karami, M. (2017). Optimization of thermodynamic models using machine learning techniques. *Industrial & Engineering Chemistry Research, 56*(11), 2905-2917. https://doi.org/10.1021/acs.iecr.7b00879
5. Ashraf, S., & Patel, M. (2020). Surrogate modeling for chemical process optimization: A case study in hydroformylation. *Computers & Chemical Engineering, 137*, 106848. https://doi.org/10.1016/j.compchemeng.2019.106848
6. Baldi, P., & Brunak, S. (2018). *Bioinformatics: The machine learning approach* (3rd ed.). MIT Press.
7. Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London, 53*, 370-418. https://doi.org/10.1098/rstl.1763.0053

8. Biegler, L. T., & Bajpai, R. (2019). Process design and optimization using surrogate models. *AIChE Journal, 65*(5), 1423-1436. https://doi.org/10.1002/aic.16509
9. Bui, V. T., & Goh, J. (2017). Gaussian process regression in optimization of chemical processes. *Chemical Engineering Science, 169*, 366-379. https://doi.org/10.1016/j.ces.2017.05.023
10. Cao, L., & Zhang, Y. (2021). Efficient surrogate models for the optimization of chemical reactors. *Industrial & Engineering Chemistry Research, 60*(3), 1123-1135. https://doi.org/10.1021/acs.iecr.0c05719
11. Chen, D., & Lin, X. (2020). Surrogate modeling in chemical process optimization: Review and perspectives. *Computers & Chemical Engineering, 139*, 106922. https://doi.org/10.1016/j.compchemeng.2020.106922
12. Choi, J. H., & Lee, J. W. (2019). Application of artificial neural networks to optimization of chemical processes. *Chemical Engineering Communications, 206*(5), 678-688. https://doi.org/10.1080/00986445.2019.1586357
13. Cox, J., & Michael, M. (2018). Surrogate modeling in process optimization: Methods and applications. *AIChE Journal, 64*(8), 3145-3162. https://doi.org/10.1002/aic.16261
14. Cuadros, M., & Medina, J. (2020). Optimization of thermodynamic models with machine learning techniques. *Journal of Process Control, 88*, 96-107. https://doi.org/10.1016/j.jprocont.2020.07.002
15. Dembélé, T., & Bouzidi, M. (2021). Use of support vector machines in process optimization for chemical engineering. *Computers & Chemical Engineering, 144*, 107061. https://doi.org/10.1016/j.compchemeng.2020.107061
16. Englehardt, T., & Horváth, I. (2020). A survey on optimization methods in chemical engineering. *Journal of Chemical Engineering, 120*(4), 2301-2320. https://doi.org/10.1021/acs.iecr.9b03988
17. Goh, J., & Liao, X. (2020). Surrogate-based optimization for multiphase reaction systems. *Journal of Industrial & Engineering Chemistry, 73*, 100-112. https://doi.org/10.1016/j.jiec.2018.09.027
18. Guo, H., & Zhang, Z. (2019). Surrogate models for chemical reaction engineering: Trends and challenges. *AIChE Journal, 65*(10), 1425-1440. https://doi.org/10.1002/aic.16506
19. Harb, S., & Atwan, A. (2018). Application of Gaussian Processes for optimizing phase equilibrium models. *Computers & Chemical Engineering, 117*, 335-350. https://doi.org/10.1016/j.compchemeng.2018.07.015
20. He, F., & Sun, S. (2021). Surrogate models for optimizing phase behavior in chemical processes. *Computers & Chemical Engineering, 144*, 107220. https://doi.org/10.1016/j.compchemeng.2020.107220
21. Holmes, L., & Biederman, S. (2018). Surrogate model-based optimization for chemical processes. *Chemical Engineering Research & Design, 135*, 107-120. https://doi.org/10.1016/j.cherd.2018.07.007
22. Jiang, Y., & Luo, J. (2020). Optimization of hydroformylation reactions using surrogate models. *Industrial & Engineering Chemistry Research, 59*(8), 3329-3343. https://doi.org/10.1021/acs.iecr.9b05067
23. Kim, D., & Lee, J. (2020). Surrogate models in multi-objective optimization for process design. *Computers & Chemical Engineering, 138*, 106905. https://doi.org/10.1016/j.compchemeng.2020.106905
24. Kumar, P., & Sharma, R. (2020). Surrogate models in chemical engineering: A practical approach. *Chemical Engineering & Technology, 43*(5), 910-926. https://doi.org/10.1002/ceat.201900679

25. Lee, Y., & Lee, K. (2020). Artificial neural networks for phase behavior prediction in multiphase chemical systems. *Industrial & Engineering Chemistry Research, 59*(12), 5357-5368. https://doi.org/10.1021/acs.iecr.9b06791
26. Li, S., & Fang, H. (2020). Hybrid surrogate models for chemical process optimization. *Computers & Chemical Engineering, 138*, 106876. https://doi.org/10.1016/j.compchemeng.2020.106876
27. Liu, L., & Zhang, X. (2020). Gaussian process regression for thermodynamic property predictions in chemical engineering. *Chemical Engineering Science, 226*, 115824. https://doi.org/10.1016/j.ces.2020.115824
28. Marzband, M., & Sadeghzadeh, M. (2020). Optimization of industrial processes using artificial neural networks. *Computers & Chemical Engineering, 144*, 107125. https://doi.org/10.1016/j.compchemeng.2020.107125
29. Meskinfam, M., & Alamdari, M. (2019). Optimization of multiphase equilibrium in chemical processes using surrogate models. *Journal of Process Control, 71*, 1-11. https://doi.org/10.1016/j.jprocont.2018.11.007
30. Rios, M., & Sahinidis, N. (2018). Surrogate models in optimization of chemical processes. *Chemical Engineering Science, 185*, 243-259. https://doi.org/10.1016/j.ces.2018.04.032
31. Samanta, S., & Li, L. (2020). Surrogate-based optimization of chemical reactors using support vector machines. *Chemical Engineering Research & Design, 151*, 290-302. https://doi.org/10.1016/j.cherd.2019.11.008
32. Sefidari, R., & Zendehdel, M. (2020). Modeling and optimization of multiphase reaction systems using surrogate models. *Computers & Chemical Engineering, 135*, 106760. https://doi.org/10.1016/j.compchemeng.2019.106760
33. Suresh, S., & Choudhury, A. (2018). Phase behavior prediction for complex systems using machine learning models. *Journal of Chemical Engineering Science, 195*, 124-136. https://doi.org/10.1016/j.ces.2018.07.008
34. Zhang, Z., & Wang, L. (2019). Machine learning in chemical process design and optimization. *Chemical Engineering Research & Design, 148*, 12-24. https://doi.org/10.1016/j.cherd.2019.07.006
35. Zhao, L., & Wang, J. (2019). Surrogate modeling for optimization in chemical engineering applications. *Chemical Engineering Journal, 362*, 29-41. https://doi.org/10.1016/j.cej.2019.01.024