

Science Guided Machine Learning for Conceptual Process Development: A Novel Heterogeneous Azeotropic Separation Case Study

Troy Gustke

Virginia Tech, Chemical Engineering

ABSTRACT

Hybrid modeling has stood out as stable method to apply machine learning (ML) to chemical process systems, as it can apply first-principles knowledge and constraints to data-driven methods for increased speed and precision. While substantial work has focused on leveraging operational data, the application of ML to conceptual process design remains limited, despite its strong influence on overall plant economics. Various data-driven methods have improved the speed and accuracy of process development, particularly for optimization; however, they are typically applied to established processes or predefined superstructures, leaving the conceptual process synthesis stage largely unaddressed. In this work, we integrate existing and novel science-guided ML methods into a process development framework to address challenging design problems. This approach integrates with existing process development procedures while benefiting from novel data-driven tools. These methods include modern ML thermodynamic predictive models, Bayesian parameter estimation and uncertainty quantification, science-guided ML/data analysis, and generative AI to improve the speed, depth, and nuance of developed rigorous process models. To appropriately demonstrate this framework with a challenging

design problem, we guide a case study for the heterogeneous azeotropic separation of 2-methyltetrahydrofuran (MeTHF), acetonitrile (ACN), and water; a mixture typically found in pharmaceutical solvent recovery systems. This system is not studied in literature and presents several key challenges including an ill-defined thermodynamic parameter regression, non-trivial process topology, and multi-objective solvent selection. We apply existing ML tools to solve each problem with novel approaches. In addition, we demonstrate the capabilities of modern large language models for process selection when supplied with thermodynamic data and in-the-loop subject matter experts. This case study demonstrates how the integration of data-driven tools enhances the traditional process development framework for conceptual design.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>

