Process design of hybrid temperature-antisolvent crystallization powered by machine learning

Luca Bosetti, Benedikt Winter, Andreas Locher, Lisanne Wittenberg, André Bardow

Energy & Process Systems Engineering, Department of Mechanical and Process Engineering, ETH Zürich, Tannenstrasse 3, 8092 Zurich, Switzerland.

abardow@ethz.ch

Abstract

The chemical industry's pursuit of a net-zero future necessitates energy-efficient separation technologies, with crystallization emerging as a potential solution due to its low energy requirements and ability to achieve high purities. However, the design of crystallization processes is hindered by the need for thermodynamic properties, which are usually obtained through expensive and time-intensive experiments. This study utilizes machine learning, specifically the SMILES to Properties Transformer (SPT) model, originally trained for estimating activity coefficients for vapor-liquid equilibria (Winter et al. 2023), to predict essential quantities for crystallization process development. Focused on the hybrid temperature-antisolvent crystallization of ibuprofen, a pharmaceutical-relevant system, this research showcases a novel design framework that identifies suitable solvents and operating conditions, supporting the efficient and sustainable development of crystallization processes.

**Keywords**: machine learning, thermodynamic property prediction, sustainable processes, computer-aided process design

* 1. Introduction

The global transition toward a net-zero economy necessitates the development of highly efficient and sustainable processes. Therein, separating components from complex mixtures has a critical and often economically burdensome role, particularly in the pharmaceutical and chemical sectors (Alder et al. 2016). To address the challenge of developing environmentally and economically favorable separations, computer-aided tools for process design are increasingly used. These tools empower the exploration of a broad space of process alternatives, facilitating the identification of feasible, environmentally friendly, and cost-effective pathways (Papadopoulos et al. 2018).

As a candidate separation technology, crystallization is particularly promising since it is usually operated at or near ambient temperatures while delivering solid products at high purity (Myerson 2002). Within the chemical and pharmaceutical industry, crystallization frequently serves as a final purification step, as exemplified by the styrene recycling process (Khandelwal 2022). Developing an efficient crystallization process requires a comprehensive understanding of the thermophysical properties associated with pure components and their interactions within mixtures. In particular, crystallization by an anti-solvent relies on accurate activity coefficients as well as enthalpies to achieve optimal process performance. If experimental data are unavailable or too expensive to attain, relying on predictive methods is the only alternative (Watson et al. 2022, Wang et al. 2020). The recent advent of modern machine learning models holds potential to overcome the lack of data requirements. However, the applicability of machine learning models for crystallization process design is currently unclear, although machine learning holds promise in improving design capabilities and promoting the adoption of crystallization as an efficient and sustainable process.

The present study analyzes a hybrid temperature-antisolvent crystallization process, where thermophysical properties are predicted by a machine-learning model (SMILES to Properties Transformer, SPT) (Winter et al. 2022, Winter et al. 2023). The use of a machine-learning algorithm with remarkable performance provides the possibility to systematically screen an extensive number of antisolvents and temperature levels. This screening approach enables the exploration of many process alternatives, which was hardly possible with conventional property prediction methods, and experimentally impossible. The final goal of the application of machine learning to property predictions is to screen many process alternatives, including the molecular degrees of freedom. Such an approach leads us to pinpoint energy-efficient, cost-effective and sustainable solutions, thereby facilitating the transition of the chemical and pharmaceutical industry toward a net-zero future.

* 1. Methods

Crystallization processes are based on the knowledge of the solid-liquid equilibrium. In crystallization, we define the “solute” as the component that we would like to crystallize, while “solvent” and “antisolvent” as the components that remain in the liquid phase. The expression for solubility can be derived from the chemical potential of the liquid and the solid phases and is most commonly employed in its simplified form as (Watson et al. 2021):

$$ln(x\_{sol}^{L}γ\_{sol}^{L})= \frac{-ΔH\_{sol}^{fus}}{R} \left(\frac{1}{T}-\frac{1}{T\_{sol}^{m}}\right) (1)$$

where $x\_{sol}^{L}$ is the liquid-phase molar fraction of the solute, $γ\_{sol}^{L}$ is the liquid-phase activity coefficient, $T$is the process’s operating temperature, $T\_{sol}^{m}$ is the melting temperature of the solute, and$ΔH\_{sol}^{fus}$is the enthalpy of fusion of the solute*.* It is worth noting that the liquid-phase activity coefficient depends on the composition of the liquid phase, thus the equilibrium mole fraction $x\_{sol}^{L}$ is obtained by implicitly solving equation (1).

The melting temperature and the enthalpy of fusion are pure component properties of the solute. Such properties should be known for components that need to be crystallized but are not always available in the early stages of development. Thus, we trained SPT to predict those properties. About 28’000 datapoints have been used as training data for the melting temperature, which were taken from Bradley et al. (2014). For the training on enthalpies of fusion, about 8800 datapoints were taken from Acree and Chickos 2016 and 2017. The activity coefficients are also predicted with SPT, which has been trained on vapor-liquid equilibrium data from the DDB (Dortmund Databank 2022), and from Brouwer et al. 2021, for a total of 200 000 datapoints for 800 components (Winter et al. 2023).

In the study, we assume a valuable solute dissolved in a solvent: the driving force of the process results from the variation of the temperature of the process, and from the addition of antisolvent to the system. The temperature at which the crystallizer is operated, the molecule used as antisolvent and the quantity of antisolvent used are treated as degree of freedom in the design process. The antisolvents are screened subject to the following constraints:

* the antisolvents should be liquid in a defined temperature range, i.e., $T\_{antisolv}^{boil}>T\_{max}$ and $T\_{antisolv}^{melt}<T\_{min}$, with $T\_{min}$ and $T\_{max}$ the minimum and maximum operating temperatures, respectively.
* the selected antisolvent shall not form any liquid phase split with the solvent considered.

For the estimation of the boiling point and of the liquid-liquid equilibrium, the SPT model has been trained to predict Antoine’s parameters (Winter et al, *in preparation*), and NRTL parameters, respectively (Winter et al. 2023). Key parameters to target sustainability are the yield of the product, the operation temperature of the reactor, the amount of antisolvent used, and the type of solvents used, according to Alder et al. 2016.

As a case study, we regard ibuprofen as the solute, ethanol as a solvent, and we screen different antisolvents to check for favorable crystallization conditions.

* 1. Results

The machine learning model SPT has been trained to predict melting temperatures and enthalpies of fusion. The predicted melting temperatures present a mean average error (MAE) of 29.3K for a total of 28’000 pure components, while the MAE for predicted enthalpies of fusion is 5.6 kJ/mol for a total of 8800 components. Experimental measurements should always be used if data are available, but SPT gives the possibility to broaden the design space by predicting properties not available from experiments.

The next step in predicting solid-liquid equilibrium is solving Equation 1 for the mole fraction of the components at equilibrium. The activity coefficients used are those obtained from training on vapor-liquid equilibrium (Winter et al. 2023). The predictions of solid-liquid equilibria for the case study of ibuprofen-ethanol-water are presented in Figure 1 for two temperature levels (solid lines). The dashed lines in Figure 1 refer to the values reported in Watson et al. 2021, who predicted activity coefficients with SAFT-γ Mie.



*Figure 1.* Ternary diagram of ibuprofen-ethanol-water system. The solid lines represent the solid-liquid equilibria at two temperature levels predicted with SPT. The dashed lines represent the corresponding literature data taken from Watson et al. 2022.

The solid-liquid equilibrium lines predicted by SPT correctly capture the temperature trend and the characteristic shape of the concentration profile. It is worth noting that SPT was originally trained on VLE data, so the relatively accurate results of SLE are very promising. Further training of the model on SLE data is expected to increase the accuracy of the predictions, but experimental data for solute-solvent-antisolvent mixtures at different temperatures are rare or not readily accessible.

* 1. Conclusions

Crystallization processes stand as a promising separation technique, offering innovative solutions to address the evolving demands of the pharmaceutical and chemical industries while striving for sustainability. The successful modeling of a hybrid temperature-antisolvent crystallization process necessitates a comprehensive characterization of the system, particularly the knowledge of the key pure component properties, e.g., melting temperatures and enthalpies of fusion, as well as mixture properties, e.g., activity coefficients.

The application of the machine-learning model SPT to crystallization presents a significant advancement in this context. SPT can be trained for melting properties, specifically melting temperatures and enthalpies of fusion, while also predicting activity coefficients, Antoine’s parameters and NRTL parameters. Notably, the solid-liquid equilibria predictions derived from SPT consistently align with the temperature and concentration trends established in the existing literature. This development provides the basis for improved design of crystallization processes.

**Acknowledgment**: The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101036854.

References

Acree, W., Chickos, J. S. (2016). Phase transition enthalpy measurements of organic and organometallic compounds. Sublimation, vaporization and fusion enthalpies from 1880 to 2015. Part 1. C1− C10. *Journal of Physical and Chemical Reference Data*, 45(3).

Acree, W., Chickos, J. S. (2017). Phase transition enthalpy measurements of organic and organometallic compounds and ionic liquids. Sublimation, vaporization, and fusion enthalpies from 1880 to 2015. Part 2. C11–C192. *Journal of Physical and Chemical Reference Data*, 46(1).

Alder, C. M., Hayler, J. D., Henderson, R. K., Redman, A. M., Shukla, L., Shuster, L. E., Sneddon, H. F. (2016). Updating and further expanding GSK's solvent sustainability guide. *Green Chemistry*, 18(13), 3879-3890.

Bradley, J. C., Lang, A., Williams, A. Jean-Claude Bradley double plus good (highly curated and validated) melting point dataset (2014). <https://doi.org/10.6084/m9.figshare.1031638.v1>, checked on 10/11/2023.

Brouwer, T., Kersten, S. R., Bargeman, G., Schuur, B. (2021). Trends in solvent impact on infinite dilution activity coefficients of solutes reviewed and visualized using an algorithm to support selection of solvents for greener fluid separations. *Separation and purification technology*, 272, 118727.

Dortmund Datenbank (2022). Available online at <http://www.ddbst.com/>, updated on 2022, checked on 2/6/2022.

Khandelwal, R. (2022). SuRe Styrene: a sustainable process to recover Ultra High Purity Styrene from waste polystyrene via chemical recycling. Sulzer Ltd 2022.

Meys, R., Kätelhön, A., Bachmann, M., Winter, B., Zibunas, C., Suh, S., Bardow, A. (2021). Achieving net-zero greenhouse gas emission plastics by a circular carbon economy. *Science*, 374(6563), 71-76.

Myerson, A. (2002). Handbook of industrial crystallization. Butterworth-Heinemann.

Papadopoulos, A. I., Tsivintzelis, I., Linke, P., Seferlis P. (2018) Computer-Aided Molecular Design: Fundamentals, Methods, and Applications. In: Reedijk, J. (Ed.) Elsevier Reference Module in Chemistry, Molecular Sciences and Chemical Engineering. Waltham, MA: Elsevier.

Wang, J., Zhu, L., Lakerveld, R. (2020). A hybrid framework for simultaneous process and solvent optimization of continuous anti-solvent crystallization with distillation for solvent recycling. *Processes*, 8(1), 63.

Watson, O. L., Jonuzaj, S., McGinty, J., Sefcik, J., Galindo, A., Jackson, G., Adjiman, C. S. (2021). Computer aided design of solvent blends for hybrid cooling and antisolvent crystallization of active pharmaceutical ingredients. *Organic Process Research & Development*, 25(5), 1123-1142.

Winter, B., Winter, C., Schilling, J., Bardow, A. (2022). A smile is all you need: predicting limiting activity coefficients from SMILES with natural language processing. *Digital Discovery*, 1(6), 859-869.

Winter, B., Winter, C., Esper, T., Schilling, J., Bardow, A. (2023). SPT-NRTL: A physics-guided machine learning model to predict thermodynamically consistent activity coefficients. *Fluid Phase Equilibria*, 568, 113731.