FRAMEWORKS FOR RESOURCE-EFFICIENT AND UNCERTAINTY-AWARE DIGITAL DESIGN OF CRYSTALLIZATION PROCESSES: FROM MECHANISTIC TO DATA-DRIVEN MODELING

Yash Barhate

Crystallization plays a central role in pharmaceutical manufacturing, serving both as an isolation step and as a means of controlling particle attributes critical to downstream processing and final product performance. Industrial goals increasingly emphasize resource-efficient design of crystallization processes that can consistently deliver critical quality attributes (CQAs). However, the complex and often system-specific interplay of dominant crystallization mechanisms makes process development highly challenging.

This thesis develops resource-efficient and uncertainty-aware digital design workflows that integrate targeted experimentation with mechanistic and/or data-driven models to support robust design of both batch and continuous crystallization systems. Motivated by the ongoing batch-to-continuous transition in the pharmaceutical industry, a mechanistic population balance model (PBM)-driven framework for continuous crystallization process development is first developed, wherein the notion of Monte Carlo-based uncertainty propagation is introduced to quantify model prediction uncertainty and evaluate it against experimental variability, thereby assessing the validity of the developed models.

A major challenge in PBM development lies in identifying the dominant crystallization mechanisms and the appropriate kinetic models to describe observed data. To address this, a systematic kinetic model development workflow is proposed to solve the model-discrimination problem, followed by model-based design of experiments (mb-DoE) to design informative experiments that improve model accuracy. These workflows were demonstrated on a semi-automated single-stage crystallizer operating in recycle, enabling significant savings in material and time during model calibration. In parallel, Bayesian optimization (BO)-assisted algorithms were integrated to accelerate nonlinear optimization problems in parameter estimation and mb-DoE, while also supporting likelihood-based characterization of parameter uncertainty.

Building on these foundations, model uncertainty is then incorporated directly into decision-making via stochastic optimization formulations for operating policy design. A surrogate-

assisted sample-average approximation strategy is employed to solve these problems efficiently, identifying operating strategies that are robust to both parametric and operational variability.

Recognizing that mechanistic PBMs may be impractical in some cases—due to resource constraints or the difficulty of capturing complex CQAs such as crystallinity—this thesis also proposes complementary data-driven workflows. Here, active learning strategies are combined with statistical machine learning models to guide efficient experimentation for design space exploration. Synthetic and experimental data augmentation approaches are further proposed to improve model accuracy when experimental budgets are limited. The trained machine learning surrogates are then employed for design space analysis and process optimization. In parallel, closed-loop BO workflows are explored to drive crystallization processes to optimize certain performance metrics iteratively.

Together, these contributions establish end-to-end digital design strategies—mechanistic where models are tractable, and data-driven where they are not—that reduce experimental burden, systematically account for uncertainty, and accelerate the adoption of digital tools in pharmaceutical crystallization.