Digital Design Strategies for Industrial Crystallisation Development

NMIS-Industry Doctorate Programme
Student: Mitchelle Mnemo
Institution(s): **University of Strathclyde**
Industry Sponsor(s): **Takeda**

1 Abstract

Crystallisation is a complex, multi-phase unit operation for separation and purification of products. For the pharmaceutical industry, active pharmaceutical ingredient (API) crystallisation is as a critical formulation step. Advanced formulations can be more challenging to crystallise, therefore require tighter control of API particle properties. Additionally, the scale-up process is complex. However, knowledge of the interconnections between the API, crystallisation behaviour and vessel hydrodynamics will allow development of a modelling framework and a scale-up strategy. The impact of this would be the realisation of a digital design for pharmaceutical manufacturing allowing the elimination of unexpected problems and increase process robustness for drug development.

2 Project Description

The work plan for this project is split into several work packages (WP).

WP1: (a) API and (b) vessel characterization: Screen API molecules to identify diverse chemical and mechanical properties. In addition, characterise a range of crystallisers (1.5mL-1,200L) using Computational Fluid Dynamics (CFD) to identify hydrodynamics conditions (i.e. turbulence, kinetic energy).

WP2: Test platform development: To design, build and characterise a lab scale development platform to explore hydrodynamic conditions while monitoring crystallisation processes. It will explore automation of the platform to maximise throughput.

WP3:Nucleation and crystal growth: Experimental to extract primary, secondary nucleation, crystal growth behaviour for selected APIs using PAT – offline and online analysis. Determine correlation between API, crystallizer design and behaviour through multivariate and machine learning approaches.

WP4: Design Strategy: Identify experiments and models critical to developing a scale- up process and deliver a proposed design strategy for process scale-up.

3 Data driven workflow

Data mining: CMAC in-house data and from literature. 24 molecules

Data preprocessing: Replacing missing values with average/most frequent

Data exploration and Visualisation: Selecting features of importance from data. Correlations between features.

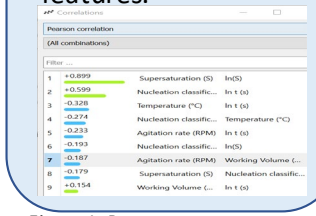

Figure 1. Pearson correlation combinations

Model training and testing:

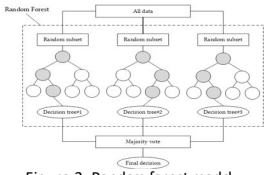

Figure 2. Random forest model

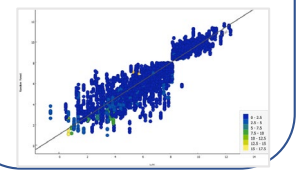

Figure 2. log induction time plotted against random forest model

4 Key Results

- Algorithm performance – Non-linear algorithms such Random Forest slightly outperform linear algorithms (i.e. Linear Regression, Ridge Regression)
- Feature importance – There is some correlation between features i.e. volume and agitation rate(RPM), non are highly correlated (i.e. >0.90)

5 Future Work

- Numerical descriptors representing the API – using Mordred descriptor calculator to extract important features (i.e. molecular weight, followed by ML testing and validation
- API selection and vessel characterisation using CFD models of planned crystallisers and hydrodynamic map
- Create simulated crystallization using kinetics obtained from database and analyse CFD flow patterns, shear rates and other kinetic features

[1] Moriwaki H, Tian YS, Kawashita N, Takagi T. Mordred: A molecular descriptor calculator. *J Cheminform,* **2018**;10(1):1–14