Model-Based Design of Experiments for Efficient Modelling of Crystallisation Systems

Aaron Bjarnasona, Hailey Lynchb, Daniel Lakyb, Cameron Browna, and Alexander Dowlingb

aUniverity of Strathclyde, Glasgow, G1 1XQ, UK

aUniversity of Notre Dame, Notre Dame, IN 46556, USA

aaron.bjarnason@strath.ac.uk

Abstract

Model-based design of experiments has been demonstrated as a robust technique for developing reliable models in resource-limited settings (Wang and Dowling, 2022). Such models are in increasing demand in the pharmaceutical industries, where the rising popularity of ideas such as quality by design and continuous manufacture necessitate the development of digital decision-making tools (Yu et al., 2014).

Traditional experimental campaigns for crystallisation model development are often constructed a priori, relying heavily on experimenters’ judgement. In this simulation study, two crystallisation models are developed: one using traditional methods, and the other via a campaign of sequentially designed experiments using model-based techniques. The models describe a 1-D population balance tracking the evolution of the critical quality attributes of the crystalline product. Accurate prediction of these qualities is central to process design, therefore improving techniques to better quantify the model parameter set could in turn improve engineering decision making.

For the model-based experimental campaign, an open-source package Pyomo.DOE (Wang and Dowling, 2022) is used to train and validate the mathematical models. The effectiveness of the model-based experimental design approach is then compared with the traditional experimental campaign.

Automated experimental design reduces the experimentalists’ workload and tailors information gathering to model-relevant regions of the design space. Thus, we can deliver better models and/or reduce material and labour costs during the expensive pharmaceutical development phase.

**Keywords**: Model-based design of experiments, Crystallisation, Population balance modelling, Simulation study.

* 1. Introduction

Every industry aims to do more with less. And process modelling is no exception.  
Fewer experiments can mean less material usage and labour, reduced risk, lower cost, perhaps lower emissions and energy use (Franceschini and Macchietto, 2008), (Barz et al., 2022). But with fewer experiments comes less data, and when building models the less-is-more principle doesn’t often apply.

For process modellers, gaining more value from each experiment will be crucial in resolving this disparity. One technique to achieve this is model based design of experiments (MB-DoE) (Franceschini and Macchietto, 2008). The core concept is to assess the information we have and determine - according to an optimality criterion – which experimental conditions might give the most informative result with respect to model refinement. This enables greater model improvement per experiment or per measurement, resulting in a more efficient method for experimental campaign planning (Wang and Dowling, 2022). Therefore, we can work towards developing as-good or better models with less data.

This is good news for pharmaceutical crystallisation modellers, whose material is often both expensive and scarce - constraining the number of experiments performed. Some popular approaches such as frequentist Bayesian optimization - while often very successful - can leave users with little process understanding. Demonstrating such understanding can be crucial from a regulation perspective. So mechanistic models can help build system intuition and aid in making more holistic design decisions. While traditional model-building approaches involve relying heavily on expert intuition for experimental planning, the MB-DoE approach may offer a more democratised pathway.

A comparison of the two will be demonstrated in this study. Simulated experiments are used to generate virtual experimental measurements which in turn parameterise a six-parameter model. The MB-DoE approach then involves an estimability assessment and sensitivity analysis to determine which parameters should be targeted for optimal model refinement, and experimental design space selection.

While MB-DoE requires experience in numerical analysis outside many scientists’ and engineers’ skillset, software packages for this analysis are increasingly available. Indeed, the preliminary results for this study are handled using Pyomo.DOE (Wang and Dowling, 2022).

* 1. Methods

The main purpose of this work is to compare the efficacy of the traditional experimental campaign and the model-based design of experiments (MB-DoE) approach: a case study will be performed for each. By simulating experiments and taking ‘virtual measurements’ of solute concentration, temperature and crystal size we gain information to parameterise the model. A nominal experimental error is added to each measurement to mimic as closely as possible the experimental reality.

The overall problem can broadly be broken down into three categories: firstly, the description of the crystallisation system. Second, the parameterisation of the crystallisation model, and lastly the MB-DoE analysis.

* + 1. Crystallisation System

Crystallisation processes are typically represented using 1-D population balance modelling. By tracking the size distribution of the crystals formed as a result of the process conditions, we can design and optimise the critical quality attributes of the product. Equation 1 below describes the evolution of the crystal size distribution:

|  |  |
| --- | --- |
|  | (1) |

where is number density function which describes the number of crystals per crystal length and volume of slurry, is experimental time, is the characteristic size of the crystals, is crystal growth rate and nucleation (birth) rate of new crystals.

The initial and boundary conditions for solving the population balance equation (PBE) are given as:

|  |  |
| --- | --- |
|  | (2) |
|  | (3) |

If growth rate is assumed to be independent of crystal length and initial birth rate is considered negligible, the simplified form of PBE can be written as:

|  |  |
| --- | --- |
|  | (4) |

This PBE system is solved using the computationally light method of moments (MoM). The solute mass balance equation provides the change of concentration inside the crystalliser. The mass balance is written in ordinary differential equation form:

|  |  |
| --- | --- |
|  | (5) |

where is solute concentration, is the true density of crystals and is the volumetric shape factor.

For simplicity, the system is limited to three crystallisation phenomena: primary nucleation, secondary nucleation and crystal growth. Growth rate and birth rate are given by empirical equations (6) and (7):

|  |  |
| --- | --- |
|  | (6) |
|  | (7) |

where is the saturation concentration, is the stirring power density and is the slurry density. Kinetic parameters and are the growth rate constant and exponent, while , and , are primary (subscript ) and secondary (subscript ) nucleation rate constants and exponents respectively. Note that equation (7) therefore comprises the *total* nucleation (birth) rate, with primary and secondary nucleation terms on the left and right of the summation symbol respectively. Lastly, the systems of ODEs presented above are scaled before solving to bring the solution scales to similar orders of magnitude and reduce numerical difficulty. This also helps to reduce weighting in the parameter fit objective function.

* + 1. Parameterisation Problem Outline

With the crystallisation description defined, we now explore how the data gathered from the crystallisation simulations translates to populating the model with parameter estimations.

The parameter fit is determined by minimising in the following objective function:

|  |  |
| --- | --- |
|  | (8) |

where subscript denotes time values point, while and denote predicted measurement values for an estimated parameter set, and the 'true' experimental measurements against which they are compared, respectively. Note that the physical significance of is the diameter for a particle of average volume. Therefore, the three properties compared between simulations and 'measurements' are concentration, total crystal mass, and the average volume of those crystals.

Table 1 below outlines the numerical problem for parameter estimation. The range of possible values particularly for the constants and highlight the difficulty in finding accurate model parameter sets. With a multi-dimensional parameter space spanning several orders of magnitude, in all likelihood, many local optima will exist. This highlights the need for new, data-rich approaches to modernise crystallisation modelling.

Table 1 - Parameterisation Problem

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  |  |  |  |  |  |
| **True Value** |  |  |  | 1 | 1 | 2 |
| **Lower Bound** |  |  |  | 1 | 1 | 1 |
| **Upper Bound** |  |  |  | 3 | 4 | 5 |

* + 1. Model-Based Design of Experiments

For the 'traditional approach', a series of preordained experiments will be simulated. The experimental conditions are designed to target particular parameters. For example, a 'growth' experiment seeks mostly to gain information about the parameters which govern the crystal growth behaviour - and set the experimental conditions to match. A series of these experiments is followed by those targeting other key mechanisms.

The model-based design of experiments (MB-DoE) approach, however, starts with preliminary data (which for the purposes of the comparative study are the same as the first experiments from the traditional case) then designs experiments to shrink the parameter confidence interval using the D-optimality criterion. After each experiment, the full dataset is reassessed, and a new set of parameter estimates are generated. By recalculating the confidence intervals, the algorithm can decide which parameter space to target to shrink the confidence ellipse. A local sensitivity analysis then determines which region of the design space might provide information to shrink the ellipse in the desired fashion. A further optimization generates decision variables in the form of new experimental conditions. More information on the MB-DoE calculations can be found in (Wang and Dowling, 2022). Experimental conditions dictated or designed in each study consist of:

Table 2 - Experimental Design Decision Variables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Symbol** |  |  |  |  |  |  |
| **Experimental**  **Condition** | Initial Concentration | Initial Temperature | Linear Cooling Rate | Mass of Seed Crystals | Size of seed crystals | Stirring Power |

* 1. Conclusions

This study expects to find a noteworthy increase in the accuracy of parameter estimates using model-based experimental design versus traditional methods. The relative benefits are expected to be maximised at low-moderate levels of data generation – emphasising the power of model-based techniques in resource limited environments.

References

J. Wang and A. W. Dowling, “Pyomo.DOE: An open-source package for model-based design of

experiments in Python,”AIChE Journal, vol. 68, no. 12, pp. 1–24, 2022.

L. X. Yu, G. Amidon, M. A. Khan, S. W. Hoag, J. Polli, G. K. Raju, and J. Woodcock,

“Understanding pharmaceutical quality by design,” AAPS Journal, vol. 16, no. 4, pp. 771–783,

2014.

G. Franceschini and S. Macchietto, “Model-based design of experiments for parameter precision:

State of the art,” Chemical Engineering Science, vol. 63, no. 19, pp. 4846–4872, 2008.

T. Barz, J. Kager, C. Herwig, and P. Neubauer, Characterization of reactions and growth in

automated continuous flow and bioreactor platforms — From linear DoE to model-based

approaches. Elsevier Inc., 2022