Systems & Control Transactions

Research Article
Foundations of Computer Aided Process Design (FOCAPD 2024)
Breckenridge, Colorado, USA. July 14-18, 2024
Peer Reviewed Conference Proceeding

PSE PRESS

Optimizing Batch Crystallization with Model-based Design of Experiments

Hailey G. Lynch[a], Aaron Bjarnason[b], Daniel J. Laky[a], Cameron J. Brown[b], and Alexander W. Dowling[a]*

[a] University of Notre Dame, Department of Chemical and Biomolecular Engineering, Notre Dame, Indiana 46556, USA
[b] University of Strathclyde, EPSRC Future Manufacturing Research Hub for Continuous Manufacturing and Advanced Crystallisation (CMAC), Glasgow G1 1RD, UK
* Corresponding Author: adowling@nd.edu.

ABSTRACT

Adaptive and self-optimizing intelligent systems such as digital twins are increasingly important in science and engineering. Digital twins utilize mathematical models to provide added precision to decision-making. However, physics-informed models are challenging to build, calibrate, and validate with existing data science methods. Model-based design of experiments (MBDoE) is a popular framework for optimizing data collection to maximize parameter precision in mathematical models and digital twins. In this work, we apply MBDoE, facilitated by the open-source package Pyomo.DoE, to train and validate mathematical models for batch crystallization. We quantitatively examined the estimability of the model parameters for experiments with different cooling rates. This analysis provides a quantitative explanation for the heuristic of using multiple experiments at different cooling rates.

Keywords: Model-based Design, Batch Crystallization, Intelligent Systems, Pyomo, Digital Twins

INTRODUCTION

Digital Twins and Model-based Design of Experiments

Spearheading the data-driven revolution, the industrial implementation of inexpensive sensors and other data-capturing technologies has increased the availability of information [1]. However, the collected data has significant analysis limitations due to existing software and modeling capabilities [2]. At the confluence of these challenges, digital twins, i.e., virtual representations of physical systems, are increasing across industries ranging from agriculture to medicine [3,4]. At the heart of many digital twins are science-based mathematical models with physically meaningful parameters estimated from sensor data, often in real-time. These science-based models are created, maintained, and updated using data science tools, including model selection, parameter estimation, sensitivity and uncertainty analysis, and model-based design of experiments (MBDoE). MBDoE supports improved parameter precision or model discrimination for science-based models with physically meaningful parameters [5] through the following form:

$$y_i = m(\boldsymbol{x}_i, \boldsymbol{\theta}) + \epsilon_i,\ \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2) \quad (1)$$

where y represents the i-th measurement from an experiment with an observation error ϵ_i following an assumed probability distribution, e.g., independent and identically normally distributed with mean 0 and variance σ_ϵ^2. Here, $m(\cdot,\cdot)$ denotes the mathematical model, $\boldsymbol{x}_i$ depicts the input and control variables for the i-th measurement, and $\boldsymbol{\theta}$ are the model parameters.

Training and validating a mathematical model using MBDoE is done sequentially. The model is updated and recalibrated before determining the next best set of experimental conditions to maximize information gain [6]. MBDoE aims to improve parameter precision in the estimated parameters by minimizing the associated uncertainty [5,7]. Encoding the mathematical model requires initial parameter values. Estimating the values for these parameters is done by using data collected from simulations of the experimental procedure. Model parameters are estimated from data using nonlinear regression:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \Psi := \frac{1}{2}\sum_i [y_i - m(\boldsymbol{x}_i, \boldsymbol{\theta})]^2 \quad (2)$$

For more sophisticated observation error assumptions,

maximum likelihood estimation, variational inference, or Bayesian methods are used instead of Eq. (2).

Next, we consider the Hessian matrix $\boldsymbol{H}$ to quantify the sensitivity of least-squares objective Ψ to the perturbations in $\boldsymbol{\theta}$:

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial^2 \Psi}{\partial \theta_1^2} & \cdots & \frac{\partial^2 \Psi}{\partial \theta_n \partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \Psi}{\partial \theta_1 \partial \theta_m} & \cdots & \frac{\partial^2 \Psi}{\partial \theta_m^2} \end{bmatrix} \tag{3}$$

When the residuals $y_i - m(\boldsymbol{x}_i, \boldsymbol{\theta})$ are small, $\boldsymbol{H} \approx \boldsymbol{Q}^T \boldsymbol{Q}$, where $\boldsymbol{Q}$ is the sensitivity of the model predictions with respect to the model parameters $\boldsymbol{\theta}$ [8]. The sensitivity matrix $\boldsymbol{Q}(\boldsymbol{\theta})$ is as follows:

$$\boldsymbol{Q}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial m(\boldsymbol{x}_1, \boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial m(\boldsymbol{x}_1, \boldsymbol{\theta})}{\partial \theta_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial m(\boldsymbol{x}_n, \boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial m(\boldsymbol{x}_n, \boldsymbol{\theta})}{\partial \theta_m} \end{bmatrix} \tag{4}$$

MBDoE uses the Fisher information matrix (FIM) $M_{\hat{\boldsymbol{\theta}}}$ to measure the information content about parameters $\boldsymbol{\theta}$ contained in proposed experiment(s) $\boldsymbol{x}_i$ for the model $m(\boldsymbol{x}_i, \boldsymbol{\theta})$:

$$\boldsymbol{M}_{\hat{\boldsymbol{\theta}}} \approx \Sigma_{\hat{\boldsymbol{\theta}}}^{-1} \approx \frac{1}{\sigma_\epsilon^2} (\boldsymbol{Q}^T \boldsymbol{Q}) \tag{5}$$

where σ_ϵ^2 is the variance of the measurement uncertainty ϵ and $\Sigma_{\hat{\boldsymbol{\theta}}}$ is the covariance of the estimated parameters. Finally, the MBDoE optimization formulation is given by:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmax}}\, \Phi[\boldsymbol{M}(\hat{\boldsymbol{\theta}}, \boldsymbol{x}) + \boldsymbol{M}_0] \tag{6}$$

where $\boldsymbol{x}^*$ are the optimal conditions for the next experiment considering prior information $\boldsymbol{M}_0$. $\Phi(\cdot)$ is a matrix operation such as determinant (D-optimality), trace (A-optimality) or smallest eigenvalue (E-optimality).

Let λ_i and $\mathbf{v}_i$ represent the eigenvalues and eigenvectors of $\boldsymbol{M}_{\hat{\boldsymbol{\theta}}}$. Because $\boldsymbol{M}_{\hat{\boldsymbol{\theta}}} \approx \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1}$, the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1}$ are approximately λ_i^{-1} and $\mathbf{v}_i$. The covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ is often represented as a confidence ellipsoid with axes $\lambda_i^{-1}\mathbf{v}_i$ [5].

Model $m(\cdot,\cdot)$ is identifiable if there exists a (possibly infinite) dataset for which it is possible to uniquely estimate the model parameters [9]. Structural issues, such as a parameter appearing only in a sum ($\theta_1 + \theta_2$) or product ($\theta_1 \times \theta_2$) will cause a model to be not identifiable; even with an infinite amount of data it is impossible to uniquely estimate both parameters (θ_1 and θ_2). In contrast, this short paper analyzes the practical identifiability, i.e., estimability, [10] of parameters in a batch crystallization model considering specific sets of experiments.

Crystallization

Crystallization is a fundamental unit operation for the purification of substances necessary for objectives relating to separations for critical mineral recovery [11], agriculture [12], food [13], fine chemicals [14], and pharmaceuticals [15]. Depending on the application (i.e., fertilizers, product quality and shelf life, fine chemicals, and active pharmaceutical ingredients), different critical quality attributes (i.e., crystal size, distribution) greatly impact the efficacy of the final product [16-19]. Tailored experimental procedures are required for each application to identify desired crystal properties. This remains difficult due to the highly nonlinear nature of crystallization dynamics and the need to control the operating conditions necessary for crystallization [20]. Due to the experimental design conditions of crystallization, open challenges include developing processes in confined volumes [21] as well as continuous manufacturing with time and material constraints [22].

The basic driving mechanisms for crystallization are nucleation and growth, which are thermodynamically favorable through a concentration gradient (i.e., supersaturation). Kinetic parameters for these processes are traditionally found through experimental procedures [23]. As such, determining the size and shape distribution of crystals is attainable through crystallization modeling [24]. Understanding crystallization processes with kinetic parameters is often achieved through population balance modeling [25]. However, estimating these parameters in crystallization models from experimental data is challenging due to the numerical complexities in a population balance equation (PBE) [26].

Crystallization DoE Heuristics

For *in-silico* processes, we focus on the heuristics in crystallization for designing improved experimental campaigns. It is common practice to estimate primary nucleation (k_{b_1} and b_1) through induction time measurements, i.e., unseeded with no cooling (β=0), then later estimating secondary nucleation (k_{b_2} and b_2) from seeded experiments. An aging process is necessary after the seeding process, and the decline of supersaturation can arise from growth only (k_g and g) or secondary nucleation and growth, which is based on the kinetics [27]. For unseeded experiments, a suitable seed for secondary nucleation is obtained from primary nucleation [28]. The growth of the crystal nuclei is necessary for detecting primary nucleation and activating secondary nucleation [29]. The rates for nucleation as well as crystal growth are usually dependent on the power law relations of supersaturation [30].

Paper Contribution

This paper uses practical identifiability analyses to provide mathematical insights into the above-mentioned crystallization modeling heuristics.

CRYSTALLIZATION MODEL

We now describe a simple batch crystallization model. We then analyze the practical identifiability of crystallization kinetics parameters $k_{b_1}, b_1, k_{b_2}, b_2, k_g$, and g (described below) using one or more unseeded experiment performed at different cooling rates β.

Population Balance Equation

A standard PBE represents the change in particle population for a specified volume. For the given problem, a one-dimensional PBE was used to describe the batch crystallization system with the form:

$$\frac{\partial n(L,t)}{\partial t} + \frac{\partial G n(L,t)}{\partial L} = B\delta(L - L_0) \quad (7)$$

where the number of crystals per crystal length and the volume of slurry are depicted by the number density function $n(L,t)$, G is the crystal growth rate, B is the nucleation rate, L is the length of the crystals, L_0 is the nuclei size, δ is the Dirac delta function, σ is relative supersaturation, and t represents the time. The kinetics for growth and nucleation of particles can be described with power laws:

$$\sigma = \frac{C - C_{sat}}{C_{sat}} \quad (8)$$

$$G = k_g \sigma^g \quad \text{if } \sigma > 0 \quad (9)$$

$$B_p = k_{b_1} \sigma^{b_1} \quad \text{if } \sigma > 0 \quad (10)$$

$$B_s = \varepsilon M_T k_{b_2} \sigma^{b_2} \quad \text{if } \sigma > 0 \quad (11)$$

$$B = B_p + B_s \quad (12)$$

Total nucleation rate (B) is often described with both primary (B_p) and secondary (B_s) mechanisms. Here k_g, k_{b_1}, k_{b_2} and g, b_1, b_2 are the rate constants and the power law exponents for growth, primary nucleation, and secondary nucleation, respectively. These six parameters are the focus of our analysis. M_T is the slurry density and ε is the power density. C_{sat} is the saturation concentration for the system temperature given in the equation below:

$$C_{sat} = A_1 \exp(A_2 T) \quad (13)$$

where A_1 and A_2 are parameters found through a solubility study. For an unseeded system (i.e., n=0), the initial and boundary conditions for this PBE are given by:

$$n(L,0) = 0 \quad (14)$$

$$n(0,t) = \frac{B\delta(L - L_0)}{G} \quad \text{if } \sigma > 0 \quad (15)$$

where we assume that the initial birth rate is negligible and growth rate is independent of crystal length. As time progresses, a solute mass balance equation is used to define the change in concentration inside the crystallizer such that:

$$\frac{dC}{dt} = -3\rho_c K_v \int_0^\infty nGL^2 dL \quad (16)$$

with C as the solute concentration, ρ_c as the solid density of a crystal, and K_v as the volumetric shape factor.

Method of Moments

To reduce the complexity of the PBE, the method of moments converted the partial differential equations of the PBE to a set of ordinary differential equations [32] by solving the equations in the model using the first four moments calculated through the population density function:

$$\frac{d\mu_0}{dt} = B \quad (17)$$

$$\frac{d\mu_i}{dt} = iG\mu_{i-1} \quad \forall i \in \{1,2,3,4\} \quad (18)$$

$$\frac{dC}{dt} = -\frac{dM_T}{dt} = -3\rho_c K_v G \mu_2 \quad (19)$$

where μ_i is the i-th moment of the crystal size distribution. The moments represent a physical significance per unit volume for total crystal number, total crystal length, total crystal surface area, and total crystal volume.

Model Implementation in Pyomo

A typical experiment was defined for a batch crystallization system in the Pyomo modeling environment. We fixed the parameters for a solid density of a crystal ρ_c, volume shape factor K_v, and the solubility parameters C_1 and C_2. For the linear cooling of the crystal suspension, an ordinary differential equation was used to describe temperature of the form:

$$\frac{dT}{dt} = -\beta \quad (20)$$

where β refers to the cooling rate. Empirical equations were used to define the growth and birth rate in the crystallizer [33,34]. The FIM $\boldsymbol{M} \in \mathbb{R}^{6\times6}$ was calculated using Pyomo.DoE [31] for the following six parameters: k_{b_1} and b_1 for primary nucleation, k_{b_2} and b_2 for secondary nucleation, and k_g and g for growth.

RESULTS

We simulated the batch crystallization experiment at four different cooling rates: β=0, β=0.1, β=0.2, and β=0.3 in °C min^{-1}. For each experiment, we calculated the FIM. Finally, we analyzed the practical identifiability of each experiment and several collections of multiple experiments. All experiments start as a 1% supersaturated solution.

Simulation Profiles

Figure 1 shows the temperature decreases linearly with time for β=0.1, 0.2, and 0.3 °C min^{-1} as expected. As noted previously, when applying cooling to a vessel, solubility of the solute generally decreases. Therefore, when the temperature of a solution changes, supersaturation is generated thermodynamically, causing the solute to crystallize out of solution.

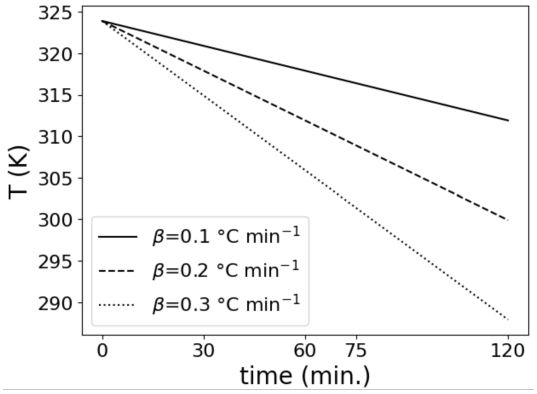

Figure 1. Change in temperature (K) over time for β=0.1, 0.2, and 0.3 (°C min^{-1}).

Subsequently, Figures 2 to 4 show the decrease in solute saturation concentration and operating concentration as a function of time for β=0.1, 0.2, and 0.3 °C min^{-1}. As expected, these concentrations decrease the fastest with the fastest cooling. This makes sense because the solubility will decrease as temperature decreases such that the solute separates from the solution as crystals form.

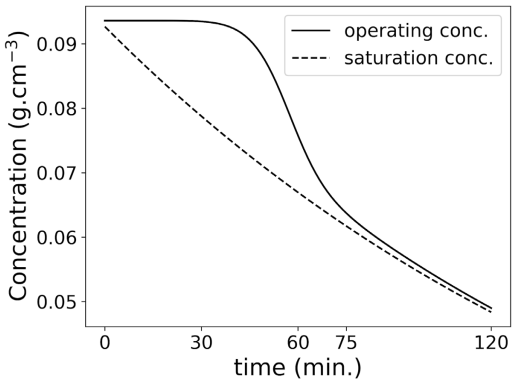

Figure 2. Operating and saturation concentration (g cm^{-3}) over time for β=0.1 (°C min^{-1}).

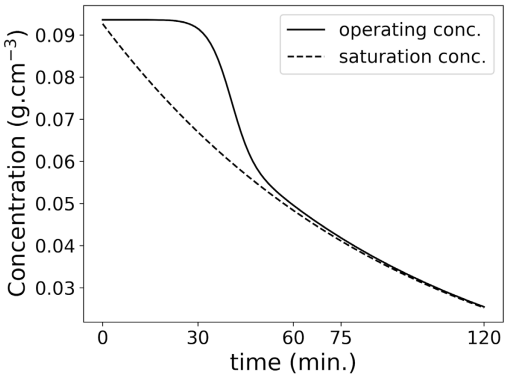

Figure 3. Operating and saturation concentration (g cm^{-3}) over time for β=0.2 (°C min^{-1}).

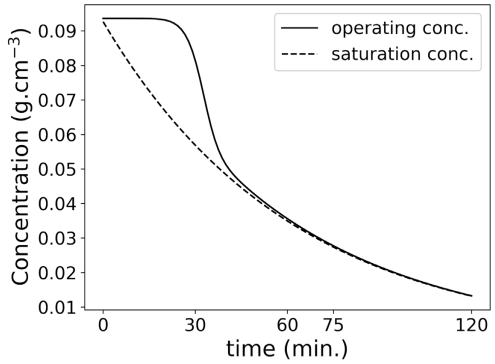

Figure 4. Operating and saturation concentration (g cm^{-3}) over time for β=0.3 (°C min^{-1}).

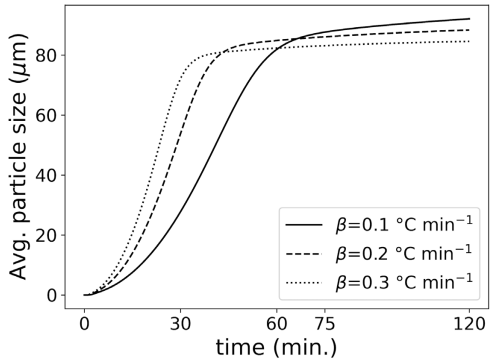

Figure 5. Change in average particle size (μm) over time for β=0.1, 0.2, and 0.3 (°C min^{-1}).

Figure 5 investigates the change in average particle size over time for three different cooling rates to understand the growth distribution. As the cooling rate increases, the average particle size grows at a faster rate. This is likely due to incubation time (time to first nuclei forming) decreasing as the cooling rate increases, resulting in a longer period of low supersaturation where growth is the dominant mechanism.

Eigendecomposition of the FIM

We now analyze the eigendecompositions of the FIMs from proposed experiments to explore the estimability of the model. Recall, the basis of eigenvectors is defined with respect to the model parameters. In Tables 1 to 6 (after references), each row (except the last) corresponds to a model parameter. Each column is an eigenvector. The last row reports the eigenvalues. Within each column, the largest absolute eigenvector element is highlighted in blue. For example, in Table 1, the fifth and sixth columns report $\lambda_5 = 9.270 \times 10^{-10}$ and $\lambda_6 = 2.622 \times 10^{-12}$.

Both eigenvalues are small, indicating significant practical identifiability issues. Specifically, the largest components of $\mathbf{v}_5$ and $\mathbf{v}_6$ correspond to the secondary nucleation parameters k_{b_2} and b_2. In other words, these eigenvectors are predominantly in the direction of k_{b_2} and

Table 1: Eigendecomposition of the FIM for the β=0 (no cooling) candidate experiment. The columns are eigenvectors, and the rows are model parameters. The last row are the eigenvalues. The largest absolute elements in each vector are highlighted in blue.

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$	$\mathbf{v}_5$	$\mathbf{v}_6$
k_{b_1}	7.261×10^{-2}	-6.303×10^{-1}	4.392×10^{-2}	-7.524×10^{-1}	1.716×10^{-1}	-2.966×10^{-3}
b_1	-8.827×10^{-2}	7.662×10^{-1}	-2.861×10^{-2}	-6.186×10^{-1}	1.469×10^{-1}	-2.069×10^{-3}
k_{b_2}	4.403×10^{-4}	-3.815×10^{-3}	-4.539×10^{-3}	1.758×10^{-1}	7.684×10^{-1}	6.154×10^{-1}
b_2	-2.953×10^{-4}	2.562×10^{-3}	2.497×10^{-3}	-1.417×10^{-1}	-5.988×10^{-1}	7.882×10^{-1}
k_g	7.922×10^{-1}	1.214×10^{-1}	5.979×10^{-1}	9.889×10^{-3}	9.894×10^{-4}	5.374×10^{-4}
g	5.995×10^{-1}	2.876×10^{-2}	-7.998×10^{-1}	-1.324×10^{-2}	-1.322×10^{-3}	-7.19×10^{-4}
λ_i	2.192×10^{3}	1.754×10^{1}	2.118×10^{-4}	1.052×10^{-7}	9.270×10^{-10}	2.622×10^{-12}

Table 2: Eigendecomposition of the FIM for the β=0.1 (°C min^{-1}) candidate experiment.

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$	$\mathbf{v}_5$	$\mathbf{v}_6$
k_{b_1}	-0.031	0.029	-0.226	0.279	-0.355	0.862
b_1	0.061	-0.084	0.493	-0.768	-0.298	0.260
k_{b_2}	-0.513	0.818	0.067	-0.038	-0.222	-0.107
b_2	0.117	-0.198	-0.092	0.163	-0.858	-0.419
k_g	-0.808	-0.495	-0.260	-0.186	0.006	-0.019
g	-0.256	-0.194	0.791	0.520	0.005	0.039
λ_i	5.753×10^{4}	4.905×10^{3}	2.357×10^{1}	9.083×10^{0}	1.441×10^{-2}	2.127×10^{-3}

Table 3: Eigendecomposition of the FIM for the β=0.2 (°C min^{-1}) candidate experiment.

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$	$\mathbf{v}_5$	$\mathbf{v}_6$
k_{b_1}	-0.025	0.009	0.318	-0.250	-0.095	-0.909
b_1	0.045	-0.032	-0.651	0.548	0.321	-0.413
k_{b_2}	-0.624	0.751	-0.020	-0.045	0.211	0.009
b_2	0.125	-0.165	0.206	-0.275	0.915	0.047
k_g	-0.743	-0.605	0.174	0.228	0.022	0.011
g	-0.203	-0.205	-0.634	-0.713	-0.080	-0.014
λ_i	6.317×10^{4}	3.206×10^{3}	2.081×10^{1}	1.897×10^{0}	3.127×10^{-2}	6.362×10^{-6}

b_2. Because λ_5 and λ_6 are small, the proposed experiment contains little information about these two parameters. On the other hand, we can see for the largest eigenvalues, λ_1 and λ_2, the predominant direction of $\mathbf{v}_1$ is growth (k_g), and $\mathbf{v}_2$ is primary nucleation (b_1), which contain the most information for this experiment. This makes sense since this experiment is not seeded, so we obtain information about primary nucleation b_1.

Table 2 shows a significant increase in the magnitude of the smallest eigenvalues and the number of eigenvalues above one for β=0.1 (°C min^{-1}) compared to β=0 (°C min^{-1}) (Table 1). This corresponds with an improved practical identifiability in the growth parameter (g). This is likely because using a non-zero cooling rate disposes the system to the thermodynamic driving force: solute supersaturation. As the cooling rate increases, the smallest eigenvalue decreases from order 10^{-3} (Table 2) to 10^{-6} (Table 3). In both tables, $\mathbf{v}_6$ is predominantly in the direction of k_{b_1}.

As the cooling rate increases (Table 4), the magnitude and primary direction of the eigenvalues and eigenvectors change. The smallest eigenvalue increases from order 10^{-6} (Table 3) to 10^{-5} (Table 4). Primary nucleation k_{b_1} remains the parameter with the least information for Tables 2 to 4.

Often, in crystallization studies, multiple cooling rates are used to identify growth. Thus, we consider two possible experiment campaigns: (1) two experiments at β=0.2 and β=0.3 and (2) three experiments at β=0.1, β=0.2, and β=0.3. For both scenarios, the FIMs are

summed before computing the eigendecomposition. Consistent with intuition, considering two experiments (Table 5) results in a larger smallest eigenvalue (order 10^{-3}) compared to each cooling rate separately (order 10^{-6} to 10^{-5}). Similarly, three experiments (Table 6) further increase the smallest eigenvalue (order 10^{-2}).

Optimality of the FIM

We now examine the A-, D-, and E-optimality measures of the FIMs from the proposed set of experiments to explore the information content of the cooling rates. Recall, the optimality criterion is defined by the trace (Eq. 21), determinant (Eq. 22), and smallest eigenvalue (Eq. 23) of the FIM:

$$\text{trace}(\boldsymbol{M}) = \sum_i \lambda_i \tag{21}$$

$$\det(\boldsymbol{M}) = \prod_i \lambda_i \tag{22}$$

$$\min\big(\text{eig}(\boldsymbol{M})\big) = \min\,(\lambda_1, \ldots, \lambda_n) \tag{23}$$

where λ_i is the eigenvalue of the FIM.

As denoted in Figure 6, the optimality measure has an overall increase for the experiments with the multiple cooling rates as opposed to the experiments with individual cooling rates. We can see that for no cooling (i.e., β=0), there is minimal optimal measure. However, for the individual cooling rates, A-optimality is nearly constant, D-optimality decreases slightly, and the optimality measure is marginally better for E-optimality with β=0.1 °C min^{-1}. As expected, when we consider two measurements, we have more information for A- and D-optimality. Overall, the experiment set with the three cooling rates contains the most information.

Computational Environment

The computational results were conducted using the following software versions: Python 3.10.12, Ipopt 3.13.2, Pyomo 6.7.1, and CasADi 3.6.5. The computer specifications include Windows 11 Version 22H2, Intel(R) Core (TM) i7-10510U CPU @ 1.80 GHz 2.30 GHz, and 16.0 GB RAM. For each individual cooling rate (β=0, β=0.1, β=0.2, and β=0.3), we found the average computational times for running the model, computing the FIM, and the eigendecomposition of the FIM which were approximately 0.078 s, 8.881 s, and 0.124 s respectively.

CONCLUSIONS

Through our analysis of the practical identifiability of parameters for a batch crystallization system, we found that growth appears to be the most estimable parameter as it consistently corresponds to the largest eigenvalues. As expected, the secondary nucleation parameters k_{b_2} and b_2 were not estimable in constant temperature experiments (β=0). If only one experiment is possible, a modest cooling rate (β=0.1 °C min^{-1}) is recommended to maximize estimability. However, performing experiments at multiple cooling rates is the most informative, as expected. Further analysis for the optimality measure of the FIM demonstrated an increase of information content when changing the cooling rate or adding multiple experiments.

In conclusion, MBDoE and practicable identifiability analysis provides a mathematical context for a common replicate crystallization heuristic. This analysis is limited to this model, and although the framework is general, the specific conclusions may not generalize. Analysis like that performed in this work demonstrates a mathematical impetus for fitting selective parameter sets for specific experiments. There is an opportunity to integrate these insights with existing crystallization parameter estimation workflows and tools [35].

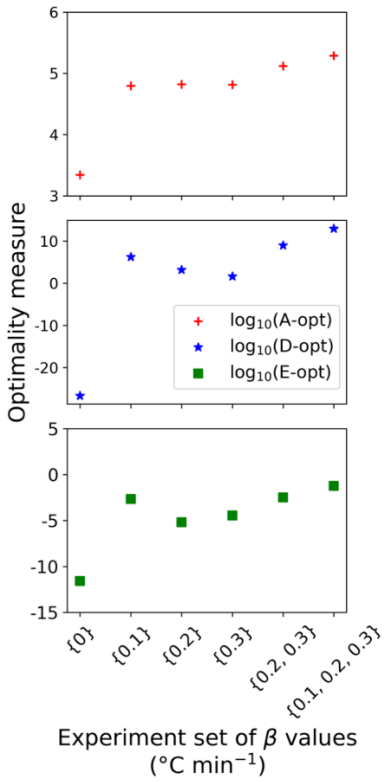

Figure 6. A-, D-, and E-optimality measures for six sets of candidate experiments define by their β values (°C min^{-1}).

ACKNOWLEDGEMENTS

H.G.L. acknowledges support from the Remick Graduate Fellowship in Engineering administered by the University of Notre Dame. A.B. acknowledges support from the University of Strathclyde Research Studentship Scheme, the University of Strathclyde and CMAC's Industry Membership for funding and supporting this work.

This effort was funded by the U.S. Department of

Energy's Process Optimization and Modeling for Minerals Sustainability (PrOMMiS) Initiative, supported by the Office of Fossil Energy and Carbon Management's Office of Resource Sustainability.

Disclaimer: This effort was funded by the U.S. Department of Energy, National Energy Technology Laboratory an agency of the United States Government, through a support contract. Neither the United States Government nor any agency thereof, nor any of its employees, nor the support contractor, nor any of their employees, makes any warranty, expressor implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

REFERENCES

1. Tao F, Q Qi. Make more digital twins. *Nature* 490-491 (2019)
2. Kusiak A. Smart manufacturing must embrace big data. *Nature* 23-25 (2017)
3. Pyliandis C, S Osinga, IN Athanasiadis. Introducing digital twins to agriculture. *Comput Electron Agric* 105942 (2021)
4. Björnsson B, et al. Digital twins to personalize medicine. Genome Med. 1-4 (2020)
5. Franceschini G, S Macchietto. Model-based design of experiments for parameter precision. *Chem.* 4846-4872 (2008)
6. Agarwal AK, ML Brisk. Sequential experimental design for precise parameter estimation. 1. Use of

Table 4: Eigendecomposition of the FIM for the β=0.3 (°C min^{-1}) candidate experiment.

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$	$\mathbf{v}_5$	$\mathbf{v}_6$
k_{b_1}	−0.027	−0.015	−0.418	0.140	−0.043	0.896
b_1	0.045	0.041	0.800	−0.389	0.101	0.441
k_{b_2}	−0.647	−0.735	−0.007	−0.137	0.147	−0.007
b_2	0.122	0.144	−0.353	−0.550	0.731	−0.038
k_g	−0.724	0.638	−0.053	−0.188	−0.173	−0.015
g	−0.198	0.171	0.239	0.687	0.634	0.031
λ_i	6.124×10^4	3.920×10^3	9.993×10^0	5.506×10^{-2}	9.359×10^{-3}	3.516×10^{-5}

Table 5: Eigendecomposition of the FIM for the β=0.2 + β=0.3 (°C min^{-1}) candidate experiments.

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$	$\mathbf{v}_5$	$\mathbf{v}_6$
k_{b_1}	−0.026	0.013	0.355	0.215	−0.053	0.908
b_1	0.045	−0.037	−0.706	−0.341	0.481	0.387
k_{b_2}	−0.635	0.744	−0.005	0.075	0.190	−0.034
b_2	0.124	−0.153	0.262	0.457	0.812	−0.158
k_g	−0.734	−0.622	0.139	−0.223	0.074	−0.009
g	−0.200	−0.186	−0.536	0.757	−0.255	0.012
λ_i	1.244×10^5	7.148×10^3	2.96×10^1	7.506×10^0	1.576×10^0	3.185×10^{-3}

Table 6: Eigendecomposition of the FIM for the β=0.1 + β=0.2 + β=0.3 (°C min^{-1}) candidate experiments.

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$	$\mathbf{v}_5$	$\mathbf{v}_6$
k_{b_1}	−0.028	0.015	0.143	0.332	0.191	0.912
b_1	0.051	−0.044	−0.462	−0.778	−0.157	0.391
k_{b_2}	−0.599	0.771	−0.087	−0.081	0.180	−0.026
b_2	0.122	−0.161	−0.363	−0.010	0.901	−0.122
k_g	−0.758	−0.579	−0.243	0.144	−0.098	−0.007
g	−0.218	−0.208	0.754	−0.508	0.291	0.003
λ_i	1.810×10^5	1.283×10^4	8.377×10^1	5.984×10^1	1.246×10^1	5.839×10^{-2}

reparameterization. *Ind. Eng. Chem. Process. Des. Dev.* 203-207 (1985)
7. Bazil JN, GT Buzzard, AE Rundell. A global parallel model based design of experiments method to minimize model output uncertainty. *Bull. Math. Biol.* 688-716 (2012)
8. Bard Y, Nonlinear Parameter Estimation. Academic Press (1974
9. Asprey SP, S Macchietto. Statistical tools for optimal dynamic model building. *Comput Chem Eng*. 1261-1267 (2000)
10. McLean KAP, KB McAuley. Mathematical modelling of chemical processes–obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *Can J Chem Eng*. 351-366 (2012)
11. Pramanik BK, et al. A critical review of membrane crystallization for the purification of water and recovery of minerals. *Rev. Environ. Sci. Biotechnol.* 411-439 (2016)
12. Moerman W, et al. Phosphate removal in agro-industry: pilot-and full-scale operation considerations of struvite crystallization. *Water Res.* 1887-1892 (2009)
13. Hartel RW. Advances in food crystallization. *Annu. Rev. food. Sci. technol.* 277-292 (2013)
14. Tung, HH, et al. Crystallization of Organic Compounds: An Industrial Perspective. John Wiley & Sons (2023).
15. Chen J, et al. Pharmaceutical crystallization. *Cryst. Growth Des.* 887-895 (2011)
16. Peng, L, et al. A comprehensive review of phosphorus recovery from wastewater by crystallization processes. *Chemosphere* 768-781 (2018)
17. Kiani H, DW Sun. Water crystallization and its importance to freezing of foods: A review. *Trends Food Sci. Tecnhol.* 407-426 (2011)
18. Carpenter K, WML Wood. Industrial crystallization for fine chemicals. *Adv. Powder Technol.* 657-672 (2004)
19. Variankaval N, AS Cote, MF Doherty. From form to function: Crystallization of active pharmaceutical ingredients. *AIChE J.* 1682-1688 (2008)
20. Ulrich J, P Frohberg. Problems, potentials and future of industrial crystallization. *Front Chem. Sci. Eng.* 1-8 (2013)
21. Meldrum FC, C O'Shaughnessy. Crystallization in confinement. *Adv. Mater.* 2001068 (2020)
22. Brown, CJ, et al. Enabling precision manufacturing of active pharmaceutical ingredients: workflow for seeded cooling continuous crystsallisations. *Mol. Syst. Des. Eng.* 518-549 (2018)
23. Chung, SH, DL Ma, RD Braatz. Optimal model-based experimental design in batch crystallization. *Chemometr Intell Lab Syst.* 83-90 (2000)
24. Jha SK, S Karthika, TK Radhakrishnan. Modelling and control of crystallization process. *Resour. Effic. Technol.* 94-100 (2017)
25. Randolph A, Theory of Particulate Processes: Analysis and Techniques of Continuous Crystallization. Elsevier (2012).
26. Nagy ZK, et al. Recent advances in the monitoring, modeling and control of crystallization systems. *Chem Eng Res Des.* 1903-1922 (2013)
27. He Y, et al. Seeding techniques and optimization of solution crystallization processes. *Org Process Res Dev.* 1839-1849 (2020)
28. Li H, Y Kawajiri, MA Grover, RW Rousseau. Modeling of nucleation and growth kinetics for unseeded batch cooling crystallization. *Ind. Eng. Chem. Res.* 4060-4073 (2017)
29. Cashmore A, et al. Rapid assessment of crystal nucleation and growth kinetics: comparison of seeded and unseeded experiments. *Cryst. Growth Des.* (2023)
30. Mullin JW, Crystallization. Elsevier (2001)
31. Wang J, AW Dowling. Pyomo.DOE: An open-source package for model-based design of experiments in Python. *AIChE J.* e17813 (2022)
32. Hulburt HM, S Katz. Some problems in particle technology: a statistical mechanical formulation. *Chem Eng Sci* 555-574 (1964)
33. Omar HM, S Rohani. Crystal population balance formulation and solution methods: a review. *Cryst. Growth Des.* 4028-4041 (2017)
34. Yazdanpanah N, ZK Nagy. The Handbook of Continuous Crystallization. The Royal Society of Chemistry (2020)
35. Szilagyi B, et al. Cross-pharma collaboration for the development of a simulation tool for the model-based digital design of pharmaceutical crystallization processes (CrySiV). *Cryst. Growth Des.* 6448-6464 (2021)

© 2024 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/

