Novel Model-Based Design of Experiments Strategies for Batch and Continuous Crystallization of Pharmaceuticals

Xuming Yuan, Brahim Benyahia

Department of Chemical Engineering, Loughborough University LE11 3TU United Kingdom

Email: B.Benyahia@lboro.ac.uk

1. Motivation

The development of reliable mathematical models is critical for all R&D activities and current digitalisation endeavours in Pharma. Crystallization is a crucial separation technique as it is applied in more than 80% API manufacturing processes. However, the development of predictive and reliable models for crystallization processes is very challenging due the large sets of parameters, model structure and poor/insufficient experimental data. Additionally, the experiments can be costly but still sub-optimally designed. Therefore, this study aims to develop a rigorous method to identify and address issues pertaining to the design of information-rich experiments by incorporating the model structure, where structural identifiability and practical identifiability (estimability) analyses and model-based design of experiment (MBDoE) are systematically and effectively implemented.

2. Case Study

- Batch & continuous cooling crystallization process of paracetamol.
- Primary & secondary nucleation, growth & dissolution, agglomeration & breakage are considered.

Kinetics

Solubility	$C^* = p_0 + p_1 T + p_2 T^2$
Supersaturation	$S = C - C^*$
Primary Nucleation	$J_1 = k_{b1} S^{b_1}$
Secondary Nucleation	$J_2 = k_{b2} S^{b_2} \mu_2^{j_2}$
Growth	$G = k_g S^g$
Dissolution	$D_s = k_{ds}(-S)^{ds}$
Agglomeration Kernel	$Ker_{agg} = K_a(L_i^3 + L_j^3)$
Breakage Kernel	$Ker_{break} = K_b L_i^\gamma$

Batch

- **Mass balance**

Crystallization stage	$\frac{dC}{dt} = -3G\rho k_v \mu_2$
Dissolution stage	$\frac{dC}{dt} = 3D_s \rho k_v \mu_2$

- **Population balance equations**

Crystallization stage

$$\begin{cases} \frac{dN_1}{dt} = -\frac{G}{2\phi_1}N_1 + J_1 + J_2 + (B_1 - D_1)\phi_1 \\ \frac{dN_i}{dt} = -\frac{G}{2\phi_i}N_i + \frac{G}{2\phi_{i-1}}N_{i-1} + (B_i - D_i)\phi_i \\ \frac{dN_n}{dt} = -\frac{G}{2\phi_n}N_n + \frac{G}{2\phi_{n-1}}N_{n-1} + (B_n - D_n)\phi_n \end{cases}$$

Dissolution stage

$$\begin{cases} \frac{dN_1}{dt} = \frac{Ds}{2\phi_1}N_2 - N_1 \\ \frac{dN_i}{dt} = \frac{Ds}{2\phi_i}N_{i+1} - \frac{Ds}{2\phi_{i-1}}N_i \\ \frac{dN_n}{dt} = -\frac{Ds}{2\phi_{n-1}}N_n \end{cases}$$

Continuous

- **Mass balance**

Crystallization stage	$\frac{dC}{dt} = -3G\rho k_v \mu_2 + \frac{C_{in} - C}{\tau}$
Dissolution stage	$\frac{dC}{dt} = 3D_s \rho k_v \mu_2 + \frac{C_{in} - C}{\tau}$

- **Population balance equations**

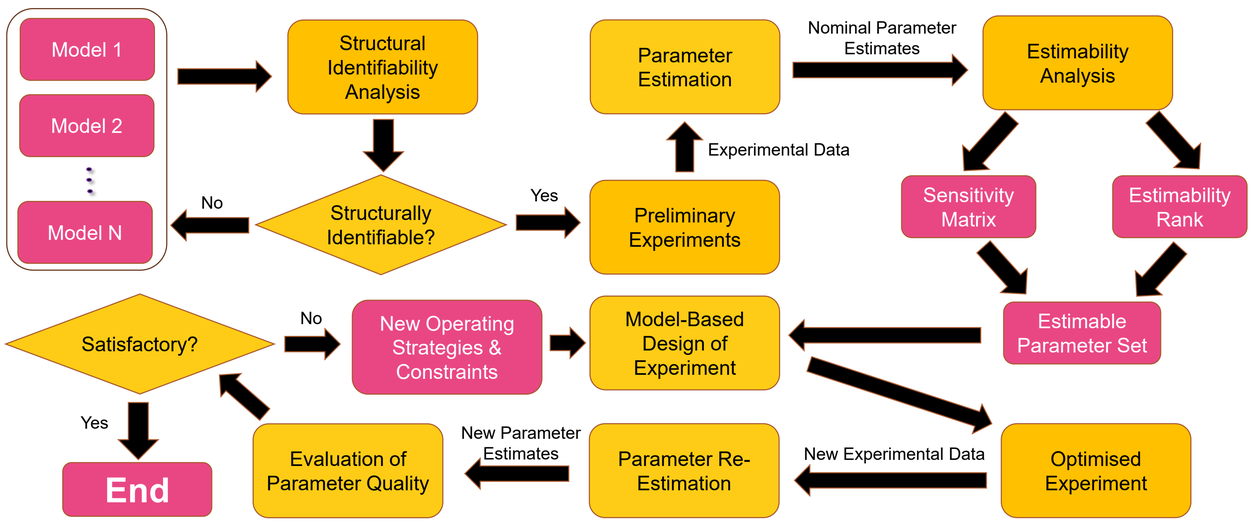
Crystallization stage

$$\begin{cases} \frac{dN_1}{dt} = -\frac{G}{2\phi_1}N_1 + J_1 + J_2 + (B_1 - D_1)\phi_1 + \frac{N_{1,in} - N_1}{\tau} \\ \frac{dN_i}{dt} = -\frac{G}{2\phi_i}N_i + \frac{G}{2\phi_{i-1}}N_{i-1} + (B_i - D_i)\phi_i + \frac{N_{i,in} - N_i}{\tau} \\ \frac{dN_n}{dt} = -\frac{G}{2\phi_n}N_n + \frac{G}{2\phi_{n-1}}N_{n-1} + (B_n - D_n)\phi_n + \frac{N_{n,in} - N_n}{\tau} \end{cases}$$

Dissolution stage

$$\begin{cases} \frac{dN_1}{dt} = \frac{Ds}{2\phi_1}N_2 - N_1 + \frac{N_{1,in} - N_1}{\tau} \\ \frac{dN_i}{dt} = \frac{Ds}{2\phi_i}N_{i+1} - \frac{Ds}{2\phi_{i-1}}N_i + \frac{N_{i,in} - N_i}{\tau} \\ \frac{dN_n}{dt} = -\frac{Ds}{2\phi_{n-1}}N_n + \frac{N_{n,in} - N_n}{\tau} \end{cases}$$

3. Methodology

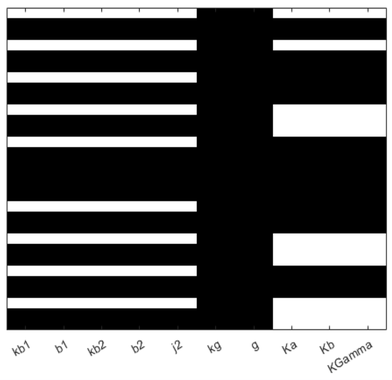


4. Structural Identifiability Analysis

Structural Identifiability: Whether the model parameters can be estimated uniquely from the given input (control) variables and measured outputs (observables) based on the model structure.

- Three observables investigated
  - Concentration (FTIR)
  - Mean crystal size (PVM)
  - Total crystal count (FBRM)
- Toolbox - GenSSI 2.0 on MATLAB R2021a based on a combination of generating series approach and identifiability tableaus
- One single observable does not guarantee structural identifiability
- A combination of any two - The model is structurally identifiable
- All the observables – Best identifiability performance

Identifiability Tableau – All Observables



5. Estimability Analysis

Estimability (Practical Identifiability): Whether few or all model parameters can be estimated accurately and precisely from the available experimental data.

Sensitivity matrix (Z): A matrix that reflects the sensitivity of measured outputs (y) respect to model parameters (θ) at different measurement times (t). Local sensitivity is applied in this study - normalized using the vector of nominal parameters and corresponding outputs.

$$z_{ij}(t_k) = \frac{\partial y_i(t_k)}{\partial \theta_j} \times \frac{\widehat{\theta_j}}{\widehat{y_i}\,(t_k)}$$

Sequential orthogonalization algorithm [1,2]: An algorithm that ranks the model parameters according to their estimability and identifies the subset of the most estimable parameters.

- 1. Select the parameter with the highest effect: find the index k such that:

$$k = argmax_i (\boldsymbol{Z_i})^T (\boldsymbol{Z_i}),\ i \in \boldsymbol{I_0} = \{1, \dots, n_p\}$$

$$\text{If } (\boldsymbol{Z_i})^T (\boldsymbol{Z_i}) \geq \lambda \text{ set } \boldsymbol{P_1} = \{p_k\} \text{ and } \boldsymbol{X_1} = \boldsymbol{Z_k}$$

Otherwise stop

- 2. Orthogonalization: compute the orthogonal projection of the matrix $\boldsymbol{Z}$:

$$\boldsymbol{R^j} = \left(\boldsymbol{I} - \boldsymbol{X_j}\left(\boldsymbol{X_j^T X_j}\right)^{-1} \boldsymbol{X_j^T}\right) \boldsymbol{Z}$$

- 3. Select the next parameter with the highest effect:

$$l = argmax_i (r_i^j)^T r_i^j,\ i \in \boldsymbol{I_j} = (\boldsymbol{I_{j-1}} - \{k, \dots\})$$

$$\text{If } (r_l^j)^T r_l^j \geq \lambda \text{ set } \boldsymbol{P_j} = \{\boldsymbol{P_{j-1}}, p_l\} \text{ and } \boldsymbol{X_{j+1}} = \{\boldsymbol{X_j}, \boldsymbol{Z_l}\}$$

Return to step 2

Otherwise stop

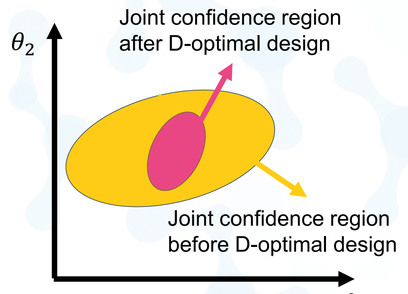
[1] Benyahia, B., Latifi, M.A., Fonteix, C., Pla, F., 2013. Emulsion copolymerization of styrene and butyl acrylate in the presence of a chain transfer agent. Part 2: Parameters estimability and confidence regions. Chem. Eng. Sci. 90, 110–118. doi: 10.1016/j.ces.2012.12.013

[2] Yao, K.Z., Shaw, B.M., Kou, B., McAuley, K.B. and Bacon, D.W. (2003). Modeling Ethylene/Butene Copolymerization with Multi-site Catalysts: Parameter Estimability and Experimental Design. *Polymer Reaction Engineering*, 11(3), pp.563–588. doi:https://doi.org/10.1081/pre-120024426.

6. A Single-Experiment MBDoE

D-optimal design is applied where the size of the joint confidence region of the model parameters is minimized. Four different temperature operating strategies were applied in the optimal experiments for the batch case, while the continuous case was subject to only temperature cycling without holds.

Temperature cycling was implemented in this study to obtain more information in one single experiment to make it more cost-effective. Seed properties, temperature control profile and sampling time were optimized. For the continuous case, the residence time was also optimized..



Mathematical Formulation of the Optimization Problem Applying Temperature Cycling with Holds (Batch Case)

$$\min_{R, \Delta t_R, t_{samp}, d_{seed}, \mu_{0,seed}} \det(FIM^{-1})$$

$$\text{s. t.} \quad -0.5 \leq R_i \leq 0\ (i = 1,3,5)$$

$$0 \leq R_i \leq 1\ (i = 2,4,6)$$

$$0 \leq \Delta t_{R,j}\ (1 \leq j \leq 12)$$

$$\sum_{j=1}^{12} \Delta t_{R,j} = 300$$

$$10 \leq 10 + \sum_{i=1}^{6} R_i \Delta t_{R,2i-1} \leq 45\ (1 \leq i \leq 6)$$

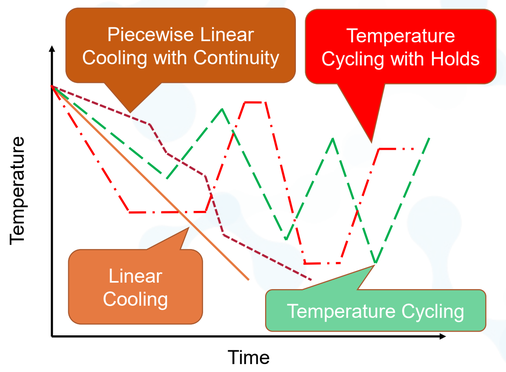
$$0 \leq t_{samp,1}$$

$$t_{samp,10} \leq 300$$

$$t_{samp,k} - t_{samp,k+1} \leq 0\ (1 \leq k \leq 9)$$

$$50 \leq d_{seed} \leq 150$$

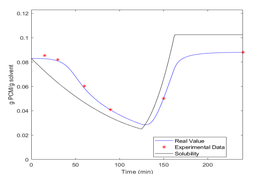
$$5 \times 10^2 \leq \mu_{0,seed} \leq 10^4$$



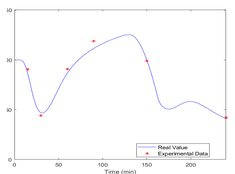
7. Results

Preliminary Experiment (Batch)

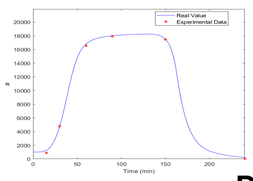
Concentration



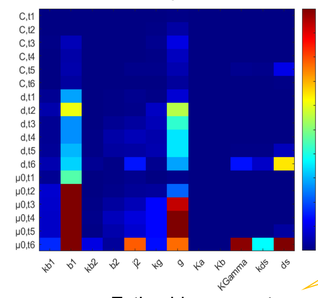
Mean Crystal Size



Total Crystal Count



Sensitivity Matrix



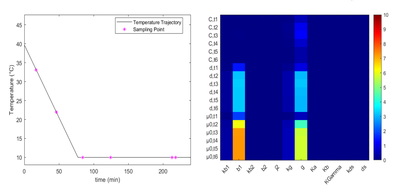
- Estimable parameters
- $\lambda = 0.01$

Estimability Rank

Rank	Parameter	Description
1	b_1	Primary nucleation power number
2	ds	Dissolution power number
3	g	Growth power number
4	γ	Breakage power number
5	j_2	Secondary nucleation power number of total surface area
6	k_g	Growth rate constant
7	k_{b1}	Primary nucleation rate constant
8	b_2	Secondary nucleation power number
9	k_{b2}	Secondary nucleation rate constant
10	k_{ds}	Dissolution rate constant
11	K_b	Breakage rate constant
12	K_a	Agglomeration rate constant

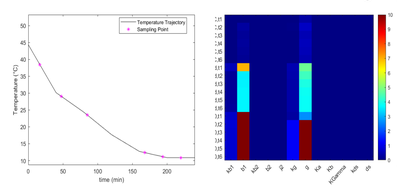
D-Optimal Experiments (Batch)

Linear Cooling



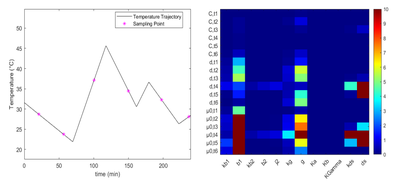
Rank	Estimable Parameters
1	b_1
2	ds
3	g
4	k_g
5	k_{ds}
6	k_{b1}

Piecewise Linear Cooling with Continuity



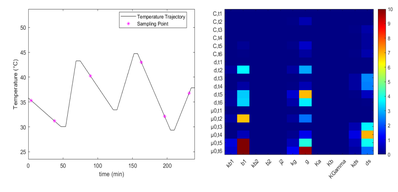
Rank	Estimable Parameters
1	b_1
2	ds
3	g
4	k_g
5	k_{b1}
6	k_{ds}

Temperature Cycling without Holds



Rank	Estimable Parameters
1	ds
2	b_1
3	g
4	k_{ds}
5	k_{b1}
6	k_{b2}
7	k_g
8	j_2

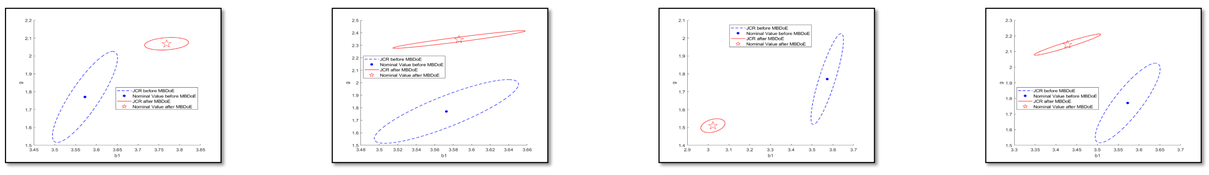
Temperature Cycling with Holds



Rank	Estimable Parameters
1	b_1
2	ds
3	g
4	k_{b2}
5	k_{b1}
6	k_g
7	k_{ds}

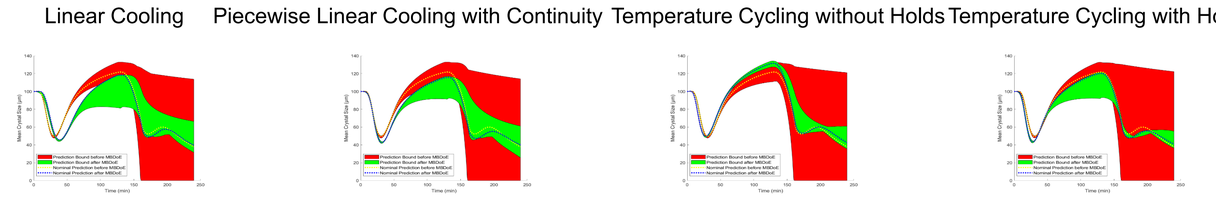
Comparison of the Joint Confidence Regions (Batch)

Linear Cooling | Piecewise Linear Cooling with Continuity | Temperature Cycling without Holds | Temperature Cycling with Holds



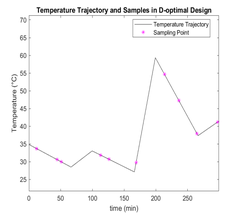
Comparison of the Confidence Bands (Batch)

Linear Cooling | Piecewise Linear Cooling with Continuity | Temperature Cycling without Holds | Temperature Cycling with Holds



Continuous Case Results

D-optimal experimental profile

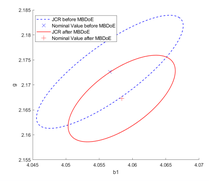


Estimability Rank (Left: Preliminary Experiment; Right: D-Optimal Experiment)

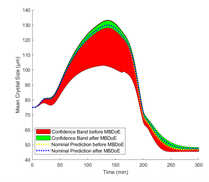
Estimable Parameters
b_1
g
ds
k_g
b_2
k_{ds}
k_{b1}
K_b

Estimable Parameters
b_1
g
ds
k_g
γ
b_2
k_{ds}
k_{b1}
k_{b2}

Comparison of Joint Confidence Regions



Comparison of Confidence Bands



8. Conclusions

- Structural identifiability analysis suggests the measurement of all the three outputs.
- A single experiment MBDoE was presented using different temperature operating strategies. Temperature cycling resulted in lower uncertainties compared to standard linear cooling.
- The MBDoE with continuous operation mode makes the optimal experiment more cost-effective as it incorporates more process dynamics in one single experiment.