A Holistic Approach for Model Discrimination, Multi-Objective Design of Experiment and Self-Optimization of Batch and Continuous Crystallization Processes

Xuming Yuana , Brahim Benyahiaa\*

*aLoughborough University, Department of Chemical Engineering, Epinal Way, Loughborough, Leicstershire, LE11 3TU, UK*

*B.Benyahia@lboro.ac.uk*

Abstract

In this study, a systematic approach is proposed to determine the best mathematical model candidate and reduce model uncertainty at a lower experimental cost, followed by model-based self-optimization to maximize process performance and product quality. The proposed method starts with the structural identifiability analysis of all model candidates, where only the structurally identifiable models can survive. Using prior knowledge, if available, or data from initial experiments, a set of preliminary/nominal parameters are estimated for the survived models followed by model discrimination based on a F-test. If more than one model passes the test, model-discrimination design of experiment (MDDoE) will be performed to identify the best model. Next, the most estimable parameter set is revealed by the estimability analysis. Afterwards, optimal experimental designs are identified based on different seeding and temperature operating strategies in the context of multi-objective model-based design of experiment (MBDoE) formulation, where the D-optimality criterion of the estimable parameters and the experimental resources are optimized. Using the optimal experimental campaign, the model parameters of the best model are updated on the fly followed by model-based self-optimization to deliver the optimal operation and design strategies along with the design space, where the attainment of the targeted Critical Quality Attributes is guaranteed. A cooling crystallization process of paracetamol (batch/continuous) is used to demonstrate and validate the proposed approach. The results show that the best model can be systemically identified while minimizing the model uncertainties and maximizing prediction capabilities. Using the most reliable model, more robust and precise self-optimization and tracking of the design space can be at lower experimental cost, demonstrating the potential effectiveness of the proposed method in crystallization processes.

**Keywords**: Crystallization, Model Discrimination, Model-Based Design of Experiment, Self-Optimization, Estimability Analysis.

* 1. Introduction

Throughout the development stages of pharmaceutical processes, the availability of reliable mathematical models is critical for the effective exploration of the design space and the delivery of robust and cost-effective designs, operating and control strategies (Benyahia et al., 2021; Liu and Benyahia, 2022). In pharmaceutical processes, crystallization plays an important role as a separation and purification technique, which usually involves complex kinetics and mechanisms, including primary and secondary nucleation, growth, and dissolution, as well as agglomeration and breakage. Therefore, it is a common situation that a crystallization process can be described by various models, which usually involves a large set of parameters. Consequently, it is difficult to determine the most appropriate model, and it leads to the high requirement on the amount and quality of experimental data for model identification. Hence, it is challenging to build a reliable model for a crystallization process, especially at the early stage of the process development where data availability is limited.

To address such issues in the establishment of reliable crystallization models, Yuan and Benyahia (2023) proposed a methodology to investigate and identify the crystallization models comprehensively, which incorporated structural identifiability analysis, practical identifiability (estimability) analysis and model-based design of experiments (MBDoE). The proposed approach ensures that only the structurally identifiable models can survive and reduces the parameter uncertainty, while the estimability of the parameters can be assessed a priori or a posteriori. The implementation of the approach in the case of a seeded batch cooling crystallization process of paracetamol revealed the structural identifiability of the investigated model and the reduced uncertainty of the parameters through D-optimal design. However, the approach does not provide a further screening criterion for the structurally identifiable model candidates, and the estimability analysis is independent from the MBDoE. One potential complementary method is model-discrimination design of experiment (MDDoE), suggested by Sen et al. (2021) where the objective is to maximize the difference between two model candidates in the experiments to be designed. Another approach in Kilari et al.’s work (2023) applied a combination of Akaike information criterion (AIC) and F-test to determine the most appropriate model for the crystallization process. Nonetheless, the proposed approaches neglected investigating the structural and practical identifiability (estimability) of the models.

Hence, this paper provides a holistic approach for the development of cooling crystallization models, in which structural identifiability analysis, AIC-test and F-test, MDDoE, estimability analysis and MBDoE are systematically incorporated in the proposed methodology. This approach is implemented in an in-silico case study of the cooling crystallization of paracetamol to demonstrate the benefits of its application in the development of crystallization models. In addition, a following multi-objective optimization of the batch crystallization process is performed in the case study to show how the improved model is utilized to enhance the desired CQAs, where the results indicate the potential effectiveness of the proposed methodology in the establishment of reliable and predictive crystallization models.

* 1. Methodology

The framework of the proposed methodology is shown in Figure 1. At the start, several model candidates describing the same crystallization process are available. The first step is the investigation of the structural identifiability of the models (Test 1) based on a set of observable outputs, where only the structurally identifiable models can survive. Once the structurally identifiable models are recognized, preliminary experiments are conducted to generate the experimental data for preliminary parameter estimation, giving the nominal parameter sets. If any available prior knowledge of the nominal parameters exists, it can be introduced in the form of nominal values in which case the preliminary experiments are not required. The nominal parameters are then used to build the sensitivity matrices or covariance matrices. Next, AIC-test and F-test (Test 2) are carried out, which aim at screening the model that best fits the experimental data. If more than one model candidate passes the tests, MDDoE will be performed with the objective of maximizing the difference between the models. The resulting optimal experimental campaign will be conducted, and the parameters are re-estimated, and the best model candidate is determined by running Test 2 again. It should be noted that this MDDoE step can be conducted iteratively if one single optimal experiment is insufficient to discriminate the models. The estimability of the best model is analyzed later, revealing the most estimable parameters subset, which is used to compute the reduced Fisher information matrix (FIM) in the subsequent MBDoE. With the new experimental data collected from the optimized experimental profiles, the new parameter estimates are obtained from re-estimation. Finally, the uncertainties of the new estimates are evaluated (Test 3). A new MBDoE problem can be formulated and solved if the parameter uncertainty is higher than the targeted level. Otherwise, the workflow is terminated.



Figure 1. Proposed framework for Model Based Design of Experiments.

* + 1. Mathematical Models

A cooling crystallization process of paracetamol is selected as the case study in this paper. The real process involves primary and secondary nucleation, growth, and nucleation, as well as agglomeration and breakage. The crystallization processes are commonly simulated by using population balance equations (PBEs) and mass and energy balance equations. The finite difference approach is applied in this work to discretize the crystal size $L$ and the distribution $n(L)$ for the solution of the PBEs. The solubility of paracetamol ($C^{\*}$) follows a quadratic function of the temperature.

 $C^{\*}=p\_{0}+p\_{1}T+p\_{2}T^{2}$ (1)

The absolute supersaturation in the crystallization process is given by:

 $S=C-C^{\*}$ (2)

Two model candidates are investigated in this study, where the first model has the primary and secondary nucleation, growth, and dissolution terms, while the second model has the additional agglomeration and breakage terms. The kinetic equations and parameters of Model 1 and Model 2 are summarized in Table 1 and 2, respectively.

Table 1. Kinetic equations and parameters of Model 1.

|  |  |  |
| --- | --- | --- |
| **Kinetics** | **Equations** | **Parameters** |
| Primary Nucleation | $$J\_{1}=k\_{b1}S^{b\_{1}}$$ | $k\_{b1}$, $b\_{1}$ |
| Secondary Nucleation | $$J\_{2}=k\_{b2}S^{b\_{2}}μ\_{2}^{j\_{2}}$$ | $k\_{b2}$, $b\_{2}$, $j\_{2}$ |
| Growth | $$G=k\_{g}S^{g}$$ | $k\_{g}$, $g$ |
| Dissolution | $$D\_{s}=k\_{ds}(-S)^{ds}$$ | $k\_{ds}$, $ds$ |

Table 2. Kinetic equations and parameters of Model 2.

|  |  |  |
| --- | --- | --- |
| **Kinetics** | **Equations** | **Parameters** |
| Primary Nucleation | $$J\_{1}=k\_{b1}S^{b\_{1}}$$ | $k\_{b1}$, $b\_{1}$ |
| Secondary Nucleation | $$J\_{2}=k\_{b2}S^{b\_{2}}μ\_{2}^{j\_{2}}$$ | $k\_{b2}$, $b\_{2}$, $j\_{2}$ |
| Growth | $$G=k\_{g}S^{g}$$ | $k\_{g}$, $g$ |
| Dissolution | $$D\_{s}=k\_{ds}(-S)^{ds}$$ | $k\_{ds}$, $ds$ |
| Agglomeration Caused Birth | $$B\_{agg,i}=\frac{1}{2}K\_{a}L\_{i}^{5}∆L\_{i}\sum\_{j=1}^{i}\frac{n(L\_{i}^{3}-L\_{j}^{3})}{(L\_{i}^{3}-L\_{j}^{3})^{\frac{2}{3}}}N\_{j}$$ | $$K\_{a}$$ |
| Agglomeration Caused Death | $$D\_{agg,i}=K\_{a}n(L\_{i})∆L\_{i}\sum\_{j=1}^{N\_{L}}(L\_{i}^{3}+L\_{j}^{3})N\_{j}$$ |
| Breakage Caused Birth | $$B\_{brk,i}=2K\_{b}∆L\_{i}\sum\_{j=i+1}^{N\_{L}}L\_{j}^{γ-1}N\_{j}$$ | $K\_{b}$, $γ$ |
| Breakage Caused Death | $$D\_{brk,i}=K\_{b}L\_{i}^{γ}n(L\_{i})∆L\_{i}$$ |

For the continuous crystallization process, let $τ$ be the residence time. Considering both the PBEs and the mass balances, the crystallization stage of Models 1 and 2 are shown in Equations 3 and 4, respectively. The dissolution stage of both models only involves the dissolution kinetics, and they are not demonstrated here for brevity. The batch model can be obtained by simply removing the terms with $τ$ in the continuous model.



* + 1. Model Discrimination

The structural identifiability of both models is confirmed by using the Matlab toolbox GenSSI 2.0, developed by Chis, Banga and Balsa-Canto (2011), where the selected observables in this case study are the concentration, mean crystal size and crystal counts. Afterwards, the in-silico preliminary experiments as well as the corresponding parameter estimation are carried out and the models are assessed based on the AIC-test and F-test summarized in Table 3, which reveals that Model 1 is superior to Model 2 in this case study, as it has a lower AIC value, and the F-test strongly supports this model. Although the real process seems to exhibit some agglomeration and breakage, the effects are deemed not influential enough based on the preliminary experiments, thus the penalty term of the number of model parameters for Model 2 in AIC-test dominates, while Model 1 capture the main dynamics in the preliminary experiments with a simpler form. It is also worth noting that the computational effort for the parameter estimation for Model 2 is much higher than Model 1 (10 min against 4 h on a PC equipped with an 11th Gen Intel(R) Core (TM) i5-11400 CPU @ 2.60GHz Processor with 16GB RAM). Hence, Model 1 is the best model in this case study.

Table 3. AIC-test and F-test results for Model 1 and Model 2

|  |  |  |
| --- | --- | --- |
| **Model** | **AIC** | **p-value from F-test** |
| Model 1 | -181.3478 | 1 |
| Model 2 | 11.9647 |

* + 1. Estimability Analysis and Multi-Objective Model-Based Design of Experiment

The estimability analysis of the parameters in Model 1 based on the preliminary

experimental results in an estimability rank of the model parameters (Table 4), where all the parameters estimable except the secondary nucleation rate constant $k\_{b2}$. Afterwards,

a multi-objective MBDoE is performed, which aims at minimizing the determinant of the inverse of the FIM of the estimable parameter subset (D-optimal design) and the amount of paracetamol used in the experiment. Temperature cycling with holds and intermittent seed addition are implemented, and the decision variables include the cooling/heating rates and the durations, mean crystal size of the seed, seed addition amount and time, initial temperature, residence time and sampling times. The mathematical formulation of the optimization problem is shown in Equation 5. Figure 2 (a) and (b) show the Pareto front of the problem (the blue star represents the selected optimal profile for the experiment in this case study) and the corresponding optimal seed addition profile. The parameter estimation using both the preliminary data and the new data from the optimized experiment is performed, giving the new parameter estimates, and the reduction of parameter uncertainty is illustrated by the joint confidence regions before and after the MBDoE, where Figure 2 (c) gives an example of the parameters $k\_{g}$ and $ds$.

Table 4. Estimability ranking of the parameters in Model 2

|  |  |  |
| --- | --- | --- |
| **Rank** | **Parameter** | **Description** |
| 1 | $$ds$$ | Dissolution power number |
| 2 | $$b\_{1}$$ | Primary nucleation power number |
| 3 | $$g$$ | Growth power number |
| 4 | $$k\_{ds}$$ | Dissolution rate constant |
| 5 | $$k\_{b1}$$ | Primary nucleation rate constant |
| 6 | $$j\_{2}$$ | Secondary nucleation power number of total surface area |
| 7 | $$k\_{g}$$ | Growth rate constant |
| 8 | $$b\_{2}$$ | Secondary nucleation power number |
| 9 | $$k\_{b2}$$ | Secondary nucleation rate constant |



 

 (a) (b) (c)

Figure 2. MBDoE results (a) Pareto Front of the optimal solutions of the multi-objective MBDoE (b) Optimal seed addition profile (c) 95% Joint confidence regions of $k\_{g}$ and $ds$ before and after MBDoE

* + 1. Self-Optimization of the Continuous Cooling Crystallization Process

With the updated parameters, the self-optimization of the continuous cooling crystallization process is conducted subsequently, aiming at maximizing the mean crystal size at the end of the process, where the initial condition of the self-optimization stage is the final condition of the MBDoE stage. The mathematical formulation of the bilevel optimization problem is not given for the sake of brevity. The optimal temperature profile of the whole process (MBDoE and self-optimisation) is shown in Figure 3 (a), while the resulting mean crystal size trajectory is shown in Figure 3 (b). The confidence band of the mean crystal size is also shown in Figure 3 (c), where the narrower confidence band reveals the reduction in uncertainties thanks to the proposed MBDoE.



 (a) (b) (c)

Figure 3. Optimization results of the continuous cooling crystallization process (a) Optimal temperature profile (b) Mean crystal size trajectory (c) 95% confidence band of the mean crystal size trajectory.

* 1. Conclusions

A novel holistic approach is proposed for the development of crystallization models in this study, which incorporates model discrimination through structural identifiability analysis, AIC and F-test, MDDoE, estimability analysis and MBDoE. The proposed approach was implemented to a cooling crystallization process of paracetamol, where two model candidates were investigated. The results showed that the second model has superior performance. The subsequent multi-objective MBDoE combined with the estimability analysis successfully reduced the uncertainty of the parameters through the optimized experiment at a lower cost. The last step consists in the self-optimization which is aimed at finding the optimal profile that maximizes the mean crystal size. The results demonstrate the effectiveness of the proposed method in the establishment of reliable and predictive crystallization models.

References

Benyahia, B., Anandan, P.D. and Rielly, C. (2021). Robust Model-Based Reinforcement Learning Control of a Batch Crystallization Process. In: *9th ICSC*, 89-94.

Chis, O., Banga, J.R. and Balsa-Canto, E. (2011). GenSSI: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics*.

Fysikopoulos, D., Benyahia, B., Borsos, A., Nagy, Z.K. and Rielly, C.D. (2019). A framework for model reliability and estimability analysis of crystallization processes with multi-impurity multi-dimensional population balance models. *Comp. & Chem. Eng.*, 122, 275–292.

Liu, J. and Benyahia, B. (2022). Optimal start-up strategies of a combined cooling and antisolvent multistage continuous crystallization process. *Comp. & Chem. Eng.*, 159, 107671.

Sen, M., Arguelles, A.J., Stamatis, S.D., García-Muñoz, S. and Kolis, S.P. (2021). An optimization-based model discrimination framework for selecting an appropriate reaction kinetic model structure during early phase pharmaceutical process development. *Reaction Chemistry and Engineering*, 6(11), pp.2092–2103.

## Yuan, X. and Benyahia, B. (2023). A Combined D-optimal and Estimability Model-Based Design of Experiments of a Batch Cooling Crystallization Process. Comp. Aid. Chem. Eng., 52, 255-260.