Model-based Design of Experiments for the Identification of Kinetic Models of Amide Formation

Emmanuel Agunloye,a Muhammad Yusuf,b Thomas W. Chamberlain,b Frans L. Muller,b Richard A. Bourne,b Federico Galvanina\*

aDepartment of Chemical Engineering, University College London, WC1E 7JE London, United Kingdom

bInstitute of Process Resarch and Development, Schools of Chemistry and Chemical and Process Engineering, University of Leeds, LS2 9JT Leeds, United Kingdom

f.galvanin@ucl.ac.uk

Abstract

Model-based design of experiments (MBDoE) techniques have been applied to various process systems in the scientific community to optimally determine a minimum number of informative experiments to enable identification of a kinetic model structure with precisely determined parameters. The effectiveness of MBDoE techniques is however deeply affected by parametric uncertainty. To evaluate the effect of parametric uncertainty on model effectiveness, this work compares and evaluates two different MBDoE approaches: 1) LHS-MBDoE, where the Latin-hypercube sampling (LHS) precedes MBDoE application; and 2) robust MBDoE, where MBDoE techniques apply ab-initio via either the expected value or worst-case approach. Using experimental and in-silico data, MBDoE methodologies were tested on a pharmaceutically relevant reaction system involving homogeneous amide formation, which can be described using reversible chemical kinetics. The performances of the two MBDoE approaches were assessed using i) the lack-of-fit test; ii) the Student’s t‑test; and iii) the determinant of Fisher information matrix (FIM) as posterior scalar measure of information.

**Keywords**: Amide formation, Latin-hypercube sampling, kinetic models, model-based design of experiment, robust model-based design of experiment.

* 1. Introduction

Extensive experimentations occur across the pharmaceutical life cycle to characterise drug candidates and their processes to produce effective, safe and profitable drugs. Predominantly, the pharmaceutical industry employs statistical design of experiments (DoE) techniques, which include factorial and random designs (Destro and Barolo, 2022). DoE campaigns often require a large number of time-consuming experiments generated with limited information about the system, causing a waste of time and resources. Informative experiments can be designed using Model-based DoE (MBDoE) techniques (Franceschini and Macchietto, 2008), which are optimal experimental designs for gathering maximum information. MBDoE techniques have increasing applications in several fields, and across the entire product life cycle from development to manufacturing (Abt et al., 2018).

MBDoE application requires four key elements: 1) physics-based modelling, to account for the underlying knowledge about the process system; 2) design space characterisation, to specify the boundary values of process variables (such as input concentrations and process temperature), usually dictated by the equipment operating range; 3) definition of experimental design criteria, to state the objective to optimise from the experimental study; 4) testing criteria, to assess the performance of MBDoE (Franceschini and Macchietto, 2008). To design optimal experiments for parameter precision, MBDoE procedures can be initialised following two methodologies: conventional MBDoE and robust MBDoE (Asprey and Macchietto, 2002). The conventional methodology requires prior parameter estimates with their covariance matrix as initial guess, usually estimated from preliminary data obtained from experiments designed using statistical DoE, to compute the posterior Fisher information matrix in the design space. On the other hand, the robust methodology, which does not require prior parameter estimates, solves the experimental design problem in the uncertain parameter space to compute the expected information, which can be done via two approaches: expected value and worst-case approach. In this work, we compare and evaluate conventional and robust MBDoE methodologies using an optimal experimental design software platform called “SimBot” controlling remotely an automated smart flow reactor (“LabBot”) for kinetic model identification in the synthesis of a pharmaceutically relevant homogeneous amide formation.

* 1. Model identification algorithm

Fig. 1 shows a segment of the SimBot algorithm highlighting five modules: 1) model structure; 2) preliminary DoE; 3) parameter estimation; 4) MBDoE for parameter precision, and 5) model validation, where the precisely identified model can be tested.

A black rectangle with blue lines

Description automatically generated

Fig. 1: The SimBot framework for model identification using MBDoE for parameter precision.

The model structure defined in Module 1 can be formulated generally as a set of differential and algebraic equations (DAEs):

|  |  |
| --- | --- |
|  | (1) |
|  | (2) |

where is the elapsed time in a batch system or space time in a steady-state system. is the vector of state variables representing the process system and the first derivative of the state variables with respect to time. is the vector of control (manipulated) variables whereas is the vector of unknown model parameters to be estimated within a continuous realisable set ; is the vector of measured (output) variables.

Module 2 in the SimBot can be used to design preliminary experiments following two different methodologies: Latin-hypercube sampling (LHS) and robust MBDoE. Design of experiments in the LHS technique follows 4 steps: i) the LHS partitions the design space equally into hypercubes , ii) selects randomly hypercubes from this universal space while ensuring no two selected hypercubes ( and ) are in the same partitioned space, iii) then generates statistically random numbers within , one for each design space dimension, and iv) lastly multiplies each random number by the dimension range to situate a designed experiment in each selected hypercube . With the partitioning, the LHS explores the entire design space (“space filling design”). Robust MBDoE, on the other hand, directly employs the model in the design space without prior parameter estimates . Rather, this methodology employs an uncertain parameter space to design preliminary experiments. It is an ab-initio optimization technique to maximise a scalar measure, in this work the determinant, of the Fisher information matrix in the design space. The Fisher information matrix (FIM) for a sampled point in the design space can be expressed as (Quaglio et al., 2018):

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

where is the matrix of the sensitivity coefficients defined as:

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

and the standard deviation of the measurement error associated with the measurement of the *j*-th response variable (here assumed to be uncorrelated and independent from the sampled experiments); **φ** is the experimental design vector containing the decision variables being manipulated in the experiments. Two alternative robust MBDoE approaches have been implemented in the SimBot: 1) expected value and 2) worst case, expressed respectively as:

|  |  |
| --- | --- |
| ; | (5,6) |

In Module 3, the SimBot calculates the prior parameter estimates from preliminary DoEs by minimizing the negative log-likelihood function (Bard, 1974):

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

is the measurement covariance matrix. Parameter estimation performance is assessed by the module by using a lack-of-fit test and the parameter *t*-test given respectively as:

|  |  |
| --- | --- |
|  | (8) |
|  | (9) |

is the diagonal entry of the inverse of the Fisher information matrix **.** must be greater than and less than , the reference values being obtained at confidence level ( in this work) and degrees of freedom.

Module 4 for MBDoE is the optimal design for parameter precision algorithm that determines the impact of the SimBot software in any application. The algorithm employs the physics-based model with prior parameter estimates to search the design space for optimal design of experiment that would minimise the uncertainty region of model parameters (in other words, maximise parameter precision and model robustness) using a scalar measure of the Fisher information matrix. With denoting information from previous experiments, the expected marginal posterior Fisher information matrix covariance at a new experimental design can be obtained as:

|  |  |
| --- | --- |
|  | (10) |

A metric of cumulative Fisher information (i.e., the determinant) evaluated as can be used to assess the performance of alternative MBDoE approaches.

* 1. Pharmaceutical case study

To evaluate the relative performance of alternative MBDoE methodologies, we consider a case study related to homogeneous amide formation () in a flow reactor system, where an amine () reacts with an ester () reversibly as:

|  |  |
| --- | --- |
|  | (R1) |

where is an alcohol by-product.

This synthesis was conducted in a smart flow reactor situated at the University of Leeds and controlled remotely using the MBDoE-driven software operated from University College London in a cloud-based experimental platform. The process control variables are amine and ester inlet concentrations , residence time and the reactor temperature while the reactor exit concentration of amide is the response measured at steady state. The synthesis system operates as a plug flow reactor and can be modelled as a set of differential and algebraic equation with the design vector: and the four Arrhenius parameters describing the forward (f) and backward (b) reactions in R1 to estimate the set of model parameters .

* 1. Results

The information contents measured using the Fisher information matrix of the conducted experiments via the three approaches, namely LHS followed by MBDoE (LHS-MBDoE), robust worst-case (rMBDoE-WC), robust expected value (rMBDOE-EV) are shown in Fig. 2. The LHS-MBDoE methodology designed the least number of experiments (5 experiments) that precisely determined kinetic parameters in the amide formation model, followed by the worst-case robust MBDoE, which required the design of 7 experiments before reaching a statistically satisfactory parameter precision. The expected value approach, however, failed to reach the target parameter precision even after 10 experiments. Experimentations were conducted for the designed experiments in the LHS-MBDoE and data used in the initial parameter estimation (first 4 experiments) and parameter recalibration with 5 experiments (one additional experiment added to the first 4), the additional experiment being designed using a D-optimal MBDoE for parameter precision. With the amide formation model correctly identified using the LHS-MBDoE methodology, the model was simulated using the designed conditions from the robust MBDoE approaches and the simulation results were corrupted with noise generated randomly following a Gaussian distribution ( and ) to obtain in-silico data for the robust MBDoE approaches and save experimental resources in this study. Yielding reproducible results, the three approaches passed the lack-of-fit test in describing the data as shown in Table 1.

A graph of a number of numbers

Description automatically generated

Fig. 2: Cumulative Fisher information fraction with experimental runs for the MBDoE techniques: LHS-MBDoE satisfying performance test at Exp 5 with |FIM|= ; rMBDoE (wc) for robust MBDoE with worst-case satisfying performance test at Exp 6 with |FIM|= ; and rMBDoE (ev) for robust MBDoE with expected value fluctuating in information around |FIM|= .

Table 1: and *t-*values statistics

|  |  |  |  |
| --- | --- | --- | --- |
|  | LHS-MBDoE | rMBDoE (wc) | rMBDoE (ev) |
|  |  |  |  |
| () |  |  |  |
|  |  |  |  |

The 5 experiments from the LHS-MBDoE yielded the highest cumulative experimental information, the determinant of the cumulative Fisher information matrix in Fig. 2 being , an average value of per experiment. The experimental information of the last experiment designed using the MBDoE for parameter precision was with all parameter *t-*values after 5 experiments being larger than the reference . On the other hand, the 6 experiments in the worst-case robust MBDoE could only attain a determinant of in the cumulative Fisher information matrix, an average value of per experiment. Nevertheless, this still produced a precise estimation of model parameters, with *t-*values being larger than .

The maximum cumulative information attained at the end of 10 designed experiments using rMBDoE-EV is much lower compared to the other cases, with a determinant value of . Most of the experiments designed using this approach yielded little information about the amide model and the worst experiment in LHS-MBDoE contains more information than the cumulative model information achieved in the 10 experiments sequentially designed using rMBDoE-EV. Data from the 4 preliminary experiments designed by rMBDoE-EV carried nearly zero information about the model, initially estimating poor values for the parameters. Due to the high parameter correlation in the amide formation reversible kinetics reported in Eq. (R1), subsequent MBDoE for parameter precision could not improve poor performing parameter set without destabilizing their well performing counterpart, the technique therefore remaining stuck in improving only the first 2 parameters at the expense of the last 2: after 10 experiments the parameter *t-*values {}. As reported in Asprey and Macchietto (2002), the performance of the expected value approach in designing preliminary experiments depends on a known parameter distribution in the parameter space .

* 1. Conclusions

This paper evaluated and compared the performance of three MBDoE approaches in designing informative experiments in amide formation synthesis: 1) a Latin hypercube sampling for designing preliminary DoEs followed by MBDoE for parameter precision (LS-MBDoE); 2) a robust MBDoE using a worst case approach (rMBDoE-WC) employed ab-initio to design preliminary experiments and then subsequently to design further experiments for parameter precision; 3) a robust MBDoE using an expected value approach (rMBDoE-EV). The experimental information acquired from these methodologies was quantified using the determinant of the Fisher information matrix and the model parameter precision using the *t*-test. The LHS-MBDoE methodology combining exploration from the LHS and exploitation from the MBDoE performed most satisfactorily, attaining the highest cumulative information in the fewest runs of experimentation. None of the robust approaches matched this performance. Particularly, the expected values preliminary experiments ill-conditioned the model to design sub-optimal experiments that generated very low information on model parameters. Careful definition of prior model parameter information is therefore required before applying the expected value approach.

Acknowledgements

This project has received funding from the EPSRC (EP/R032807/1 and EP/X024016/1). The support is gratefully acknowledged.

References

F. Destro, M. Barolo, 2022. International Journal of Pharmaceutics 620

G. Franceschini, S. Macchietto, 2008, Model-based design of experiments for parameter precision: State of the art, Chemical Engineering Science, Volume 63, Issue 19, Pages 4846-4872, ISSN 0009-2509, <https://doi.org/10.1016/j.ces.2007.11.034>

M. Quaglio, E.S. Fraga, F. Galvanin, 2018. ChERD, Vol. 136, Pages 129 – 143, doi.org/10.1016/j.cherd.2018.04.041

S.P. Asprey, S. Macchietto, 2002. Designing robust optimal dynamic experiments, Journal of Process Control 12 (2002) 545–556

V. Abt, T. Barz, M.N. Cruz-Bournazou, C. Herwig, P. Kroll, J. Möller, R. Pörtner, R. Schenkendorf, R., 2018, Model-based tools for optimal experiments in bioprocess engineering, Current Opinion in Chemical Engineering, Volume 22, 2018, Pages 244-252, ISSN 2211-3398, <https://doi.org/10.1016/j.coche.2018.11.007>

Y. Bard, 1974, Nonlinear parameter estimation, Academic Press.