Model-Based Design of Experiments for Isotherm Model Identification in Preparative HPLC

Konstantinos Katsoulas, Maximillian Besenhard, Federico Galvanin, Luca Mazzei, Eva Sorensen\*

Department of Chemical Engineering, Sargent Centre for Process Systems Engineering, University College London, Torrington Place, London WC1E 7JE, UK

\*e.sorensen@ucl.ac.uk

Abstract

Chromatography is a pivotal purification process widely used in the pharmaceutical and fine chemical industries. Design of such processes are increasingly based on the use of mathematical models, within which the isotherm model equation is crucial. Traditionally, the values of the parameters present in the isotherm model are obtained through extensive experimentation such as Frontal Analysis (FA). This study focuses on identifying suitable isotherm models, and accurate values for their parameters, using Model-Based Design of Experiments (MBDoE) based on curve-fitting rather than FA. The methodology involves screening potential isotherm models, assessing model identifiability, selecting the model, and refining the precision of the parameter values. We implemented the methodology in a case study, benchmarking the MBDoE refinement against the traditional approach of factorial designs.

**Keywords**: Chromatography, Isotherm identification, Model-Based Design of Experiments, Parameter estimation.

* 1. Introduction

In the pharmaceutical industry, mathematical models for chromatography are an integral part of the concept of Quality-by-Design (QbD). Accurate models and model parameters, particularly related to adsorption isotherms, are essential for precise design. Nevertheless, obtaining the values of these parameters is challenging, often requiring costly and time-consuming experimental methods such as Frontal Analysis (FA). A common alternative is curve-fitting, also known as Inverse Method (IM), where one obtains the values of the parameters by fitting simulated chromatograms to experimental chromatograms based on a given isotherm model. IM requires less material, but users rarely consider the statistical accuracy of the parameter values obtained in terms of their precision and model structure (i.e. set of model equations), with most researchers focusing only on the goodness of the fit (Andrzejewska et al., 2009; Gétaz et al., 2013).

In selecting isotherm models, researchers often rely on intuition or experience, but methods exist for systematically identifying suitable model structures (Waldron et al., 2019). Also, experiments are rarely designed optimally, thus unnecessarily increasing the experimental effort. Model-Based Design of Experiments (MBDoE) is a powerful method for obtaining precise parameter estimates for a given model structure based on fewer experiments (Franceschini & Macchietto, 2008).

* 1. Theoretical background
     1. Parameter estimation and MBDoE theoretical background

Curve fitting consists in matching the results of the simulation with the corresponding experimental measurements. Fitting can be achieved by varying the parameter values to be estimated in order to maximise or minimise an objective function. This work focuses on maximising the log-likelihood objective function, (Bard, 1974):

(1)

where is a vector of the model parameters, is the normally distributed error of the measurement of each experiment, and are the residuals of the measurement of the experiment that represent the difference between the *ij*-th model prediction and the corresponding experimental measurement :

(2)

The goodness of fit of the simulated measurements can be evaluated through the test (Bard, 1974; Quaglio, 2020):

(3)

where is the vector of optimal parameter estimates. The Fisher Information Matrix (FIM) is a semi-definite matrix whose elements are estimated according to (Franceschini & Macchietto, 2008):

for and (4)

where is the sensitivity of the measurement of the experiment with respect to the parameters . From the FIM, we can acquire the variance-covariance matrix, , as the inverse of the former (Bard, 1974):

(5)

where is the initial FIM. The square root of the diagonal elements of the variance-covariance matrix, , yields the standard deviation of each parameter . The product between the standard deviation and the -value gives the confidence interval (according to the available Degrees of Freedom) of the parameters. For instance, the 95% confidence interval is given by (Bard, 1974):

for (6)

The 95% confidence interval of each parameter may, in turn, be utilised in the estimation of the -value of the parameters:

for (7)

The -test is utilised to evaluate the relative precision of the parameters (Bard, 1974; Quaglio, 2020). Note that if the elements of the FIM are linearly independent, that is the rank of the matrix is equal to the number of parameters to be estimated, then the model and its parameters are considered to be ***identifiable***. Naturally, the FIM is a local matrix; i.e., it depends on the collected measurements of the performed experiments:

(8)

where is the updated FIM based on the observed FIM, of experiments. Thus, although an experiment or a combination of experiments do not yield an identifiable model, there is the possibility that a combination of experiments might exist which renders a model and its parameters identifiable.

If the model is found to be identifiable, one can proceed with refining the model parameters using MBDoE. MBDoE aims to maximise properties of the FIM, or minimise properties of the variance-covariance matrix, to design the most informative experiments (Franceschini & Macchietto, 2008):

(9)

The most widely used criteria for MBDoE are the D-, E-, and A- optimal criteria that, in terms of the FIM, aim to maximise the determinant, the smallest eigenvalue, and the trace of the matrix, respectively.

* + 1. Mathematical model

The Equilibrium Dispersive Model (EDM) has been widely used in chromatography. It accounts for mass transfer resistances and dispersion by lumping these together in an apparent dispersion coefficient, . The mass balance of the model is then (Schmidt-Traub et al., 2020):

, (10)

where and are the concentrations in the mobile and stationary phases, respectively, is the hypothetical interstitial velocity given by , where , , and are the volumetric flow rate, cross-section of the column, and total porosity, respectively. The apparent dispersion coefficient depends on the velocity and efficiency of the column, that is , where is the column length and is the number of theoretical plates.

A Danckwerts boundary condition is considered at the inlet of the column (Katsoulas et al., 2023):

(11)

as well as at the outlet:

(12)

Initially, the concentration is zero everywhere within the column:

, (13)

Assuming that the concentration in the stationary phase is always in equilibrium with the concentration in the mobile phase, then the stationary concentration is a function of the mobile concentration, . In preparative conditions, however, non-linear isotherms, for instance of the Langmuir-type, are required. Assuming that the number of potential types of binding sites can vary, we proceeded with the following potential isotherms models (Carta, 2020):

, (14)

Regardless of the number of sites, we have found it necessary to implement a reparameterization for the first type of sites to prevent high correlation between model parameters:

(15)

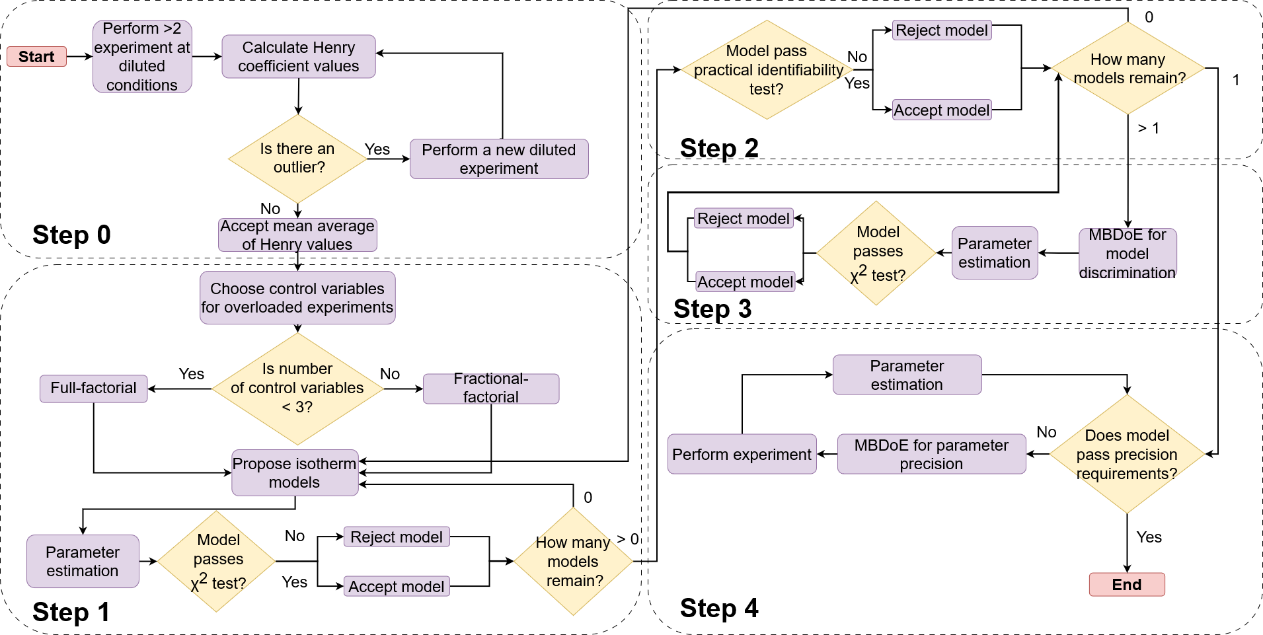
Note that in diluted conditions, the isotherm is represented by a linear function:

(16)

where for .

* 1. Methodology

The identification of potential isotherm models considered in this work follows the methodology outlined in Fig. 1 (adapted from Waldron et al., 2019). In Step 0, the Henry coefficient is estimated by performing experiments under diluted conditions, and these values are used as initial guesses for the parameter estimation later. Step 1 involves the design of factorial experiments that are used for screening purposes and are evaluated via the test. If the model structure passes the test, it is a candidate model for the system and it proceeds to the model identifiability test. The identifiability test is a local test that is conducted under a specified number of experiments, determined by the user. Using the preliminary parameter estimates acquired from the previous step, several factorial experiments are designed and performed *in-silico*. If the model passes the -test after parameter estimation, it is considered practically identifiable. In case there are more than one identifiable model, the methodology also involves a step of MBDoE for model discrimination. Further experimentation will eventually identify the most appropriate model, and we then proceed to improve the precision of its parameters, terminating the procedure when the effect of the parameter uncertainty on the model output, quantified by uncertainty analyses, is believed to be negligible, albeit this is judged differently in each application.

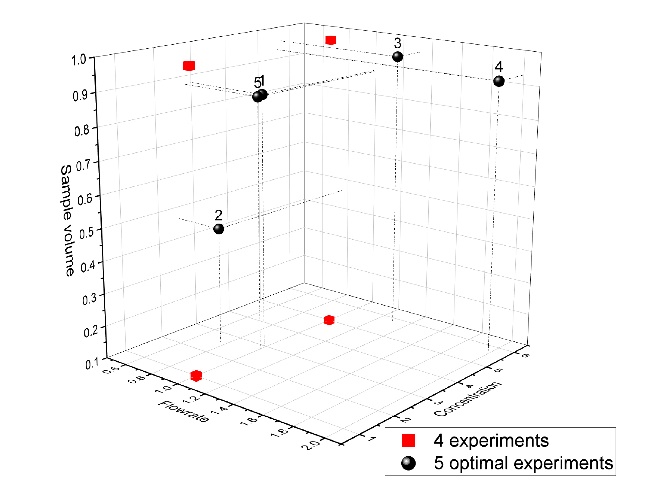


**Figure 1:** Flow diagram showing the methodology for isotherm model identification and parameter estimation.

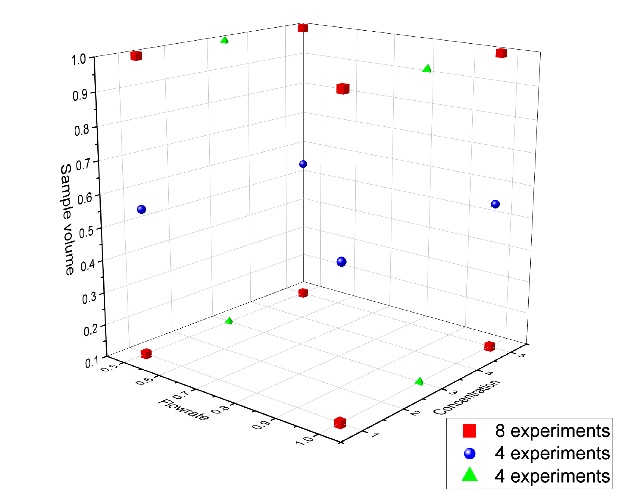
* 1. Case Studies

The methodology presented above is illustrated by a case study that investigates the isotherm model identification and estimation of its isotherm parameters for the separation of oligopeptide of tri-Leucine (LLL) in a Reversed Phase Liquid Chromatography (RPLC) column. For this case study, computer simulations are used to create what would normally be real experimental data. This is done so that the correct model and parameters are known for comparison. The bi-Langmuir isotherm was considered, and a Gaussian distributed error of 2% was introduced in the concentration measurements to simulate real-world applications. Plant parameters, as well as isotherm parameters, were based on the work of Andrzejewska et al. (2009) as follows: column length cm, column diameter mm, total porosity , and number of theoretical plates   
. The reader is referred to Table 1 for the real values of the isotherm parameters. The time-invariant controlled factors considered for the ‘experiments’ were the inlet concentration, the sample volume, and the volumetric flow rate. Simulations were performed in gPROMS® ModelBuilder, discretising the grid in 800 elements using the method of Orthogonal Collocations. The procedure began with Step 0, the estimation of Henry coefficient initial guesses by performing two experiments in diluted conditions. The next step, Step 1, involves designing factorial experiments to initiate the screening of the proposed models. Here, two different strategies were considered:

* Strategy A - 23 factorial using all three factors,
* Strategy B - 22 factorial using only the factors of inlet concentration and sample volume.



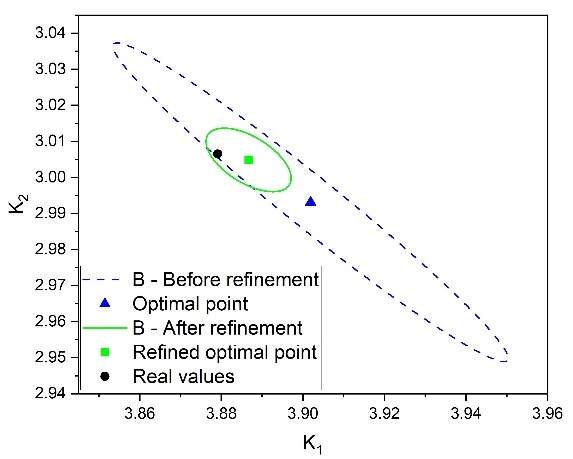
b)



a)

**Figure 2:** Illustration of the experimental conditions of the experiments performed by a) Strategy A and b) Strategy B.

Three models were proposed, the simple, bi-, and tri- Langmuir isotherms. (Note that the ‘real’ isotherm is the bi-Langmuir isotherm). From these, only the bi- and tri- Langmuir isotherms passed to the next step, and for both strategies. In Step 2, 27 total experiments were disbursed on a 27-factorial for each of the two remaining models. As expected, only the bi-Langmuir isotherm passes the identifiability test, and for both strategies, and thus the discrimination step is skipped. In the final step, the precision of the parameters was refined either by:



**Figure 3:** The 95% confidence ellipsoid between *K1* and *K2* in Strategy B before and after refinement.

* Strategy A – two rounds of 22 factorial, or,
* Strategy B – four D-Optimal and one E-Optimal experiments, sequentially.

Note that decisions on the usage of the different MBDoE criteria were taken based on statistics evaluation (i.e., correlation statistics). Fig. 2 (a) and (b) depict the total number of experiments in Strategy A and B, respectively. Note that the experiments designed by the 27-factorial and used for the identifiability test are not used in the parameter estimation procedure since they were experiments simulated using the preliminary estimates of each model and not the real values of the parameters. Table 1 summarises the estimated values of the parameters after improving their precision with Strategies A and B. Note that is 1.65 in both strategies. Although the optimal estimates in Strategy A are closer to the real values, we can only judge the quality of the estimates based on their relative precision. Strategy B employed fewer preliminary experiments than Strategy A while employing only 5 rounds of sequential MBDoE’s totalling 9 experiments. On the other hand, Strategy A required a total of 16 experiments designed by full factorials, yet the precision of the parameter estimates of the two strategies is equivalent. Fig. 3 shows the improvement of the precision of parameters and illustrated in the 95% confidence ellipsoids before and after refinement by MBDoE.

**Table 1:** The optimal parameter estimates and statistics after improving the precision.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | Total No. Experiments |
| Strategy A | 411.9/475.1 | 3.880.008 | 6.060.029 | 3.010.006 | 2.35 0.018 | 16  No MBDoE |
| -value |  | 504 | 211 | 485 | 134 |
| Strategy B | 119.3/199.2 | 3.890.008 | 6.040.02 | 3.010.007 | 2.330.024 | 9 |
| -value |  | 495 | 301 | 453 | 94 |
| Real values |  | 3.88 | 6.06 | 3.01 | 2.35 |  |

* 1. Concluding remarks

This study presents a step-by-step methodology for isotherm model identification in preparative chromatography. For the selected case study, we estimated the values of the isotherm parameters through curve-fitting following two distinct identification strategies. Refining parameter precision by MBDoE was benchmarked against factorial designs. MBDoE reduced the required number of experiments while maintaining precision. This work highlighted the benefits of MBDoE in making informed decisions in process development based on limited information.

Acknowledgement

The authors wish to acknowledge the financial support given to this research project by Eli Lilly and Company, and the Engineering and Physical Sciences Research Council (EPSRC), grant code EP/T005556/1.

References

Andrzejewska, A., Gritti, F., & Guiochon, G. (2009). *Journal of Chromatography A*, *1216*(18), 3992–4004.

Bard, Y. (1974). *Nonlinear Parameter Estimation*. Academic Press Inc.

Carta, G. (2020). *Protein chromatography: process development and scale-up* (A. Jungbauer, Ed.). Wiley-VCH.

Franceschini, G., & Macchietto, S. (2008). *Chemical Engineering Science*, *63*(19), 4846–4872.

Gétaz, D., Stroehlein, G., Butté, A., & Morbidelli, M. (2013). *Journal of Chromatography A*, *1284*, 69–79.

Katsoulas, K., Tirapelle, M., Sorensen, E., & Mazzei, L. (2023). *Journal of Chromatography A*, *1708, 464345*.

Quaglio, M. (2020). Novel techniques for kinetic model identification and improvement. In *Doctoral thesis, UCL (University College London).*

Schmidt-Traub, H., Schulte, M., & Seidel-Morgenstern, A. (2020). *Preparative Chromatography*. John Wiley & Sons, Incorporated.

Waldron, C., Pankajakshan, A., Quaglio, M., Cao, E., Galvanin, F., & Gavriilidis, A. (2019). *Industrial and Engineering Chemistry Research*, *58*(49), 22165–22177.