Computers and Chemical Engineering 195 (2025) 109021

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/cace

Model-based design of experiments for efficient and accurate isotherm model identification in High Performance Liquid Chromatography

Konstantinos Katsoulas, Federico Galvanin, Luca Mazzei, Maximilian Besenhard, Eva Sorensen *

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

ARTICLE INFO

Keywords:
Modelling
High Performance Liquid Chromatography
Model-Based Design of Experiments
Parameter estimation
Isotherm identification

ABSTRACT

Chromatography is a key purification process in the pharmaceutical industry. The process design is based on knowledge of the adsorption isotherm that describes the separation within the chromatographic column. Although obtaining the values of isotherm model parameters has traditionally been the work of experimentalists, recently design methods based on mathematical models have emerged, and for these, accurate isotherm models and model parameter values are crucial. Different methods exist for parameter estimation, all depending on experiment execution. Model-Based Design of Experiments (MBDoE) can be used to optimally design experiments that maximise the information obtained from each experiment. In this work, we propose an MBDoE-based methodology that aims to identify the most suitable isotherm model, to estimate its parameters, and to evaluate its predictive capability. The methodology is tested on an in-silico case study where the performance is compared to that of traditional factorial design of experiments.

1. Introduction

High-Performance Liquid Chromatography (HPLC) is one of the most widely used methods for drug purification. In HPLC, analytes are introduced into a column where they separate spatially according to their different affinities towards the stationary phase (generally a solid adsorbent). The greater the affinity of an analyte towards the stationary phase, the higher the mean adsorption time of this analyte. When the adsorption–desorption occurs very rapidly compared to the other physico-chemical processes involved in the separation, the concentration of the analyte in the mobile phase can be assumed to be in equilibrium with the corresponding concentration in the solid phase, and this equilibrium is expressed mathematically through an adsorption isotherm. The use of mathematical models to analyse and design chromatographic processes has gained increased interest in recent decades as part of the quest towards Quality by Design (QbD) of pharmaceutical products (Yu, 2008; Grangeia et al., 2020; Besenhard et al., 2021). The accuracy of these models relies heavily upon the accuracy of the model parameters, in particular the adsorption isotherm parameters. However, identifying a suitable adsorption isotherm model and obtaining accurate values for the isotherm parameters is challenging and time consuming, often requiring extensive experimentation. One of the most commonly employed techniques for estimating isotherm parameters is the Frontal Analysis (FA) technique (Seidel-Morgenstern, 2004; Ahmad and Guiochon, 2006, 2007; Andrzejewska et al., 2009). In FA, the column is saturated with the analyte sample until a concentration plateau is reached; this then allows estimating the correlation between the mobile phase and the stationary phase concentrations. The procedure is repeated for several analyte concentrations, and then all the measurements are used to fit the isotherm parameters to the experimental data. The FA technique is rigorous and accurate but demanding in terms of material required, cost, and effort (Andrzejewska et al., 2009; Schmidt-Traub et al., 2020), and its use is therefore discouraged when dealing with expensive analytes such as peptides and monoclonal Antibodies (mAbs). An alternative approach to FA is the Inverse Method (IM), which is a curve-fitting exercise (James et al., 1999; Felinger et al., 2003; Seidel-Morgenstern, 2004; Andrzejewska et al., 2009; Gétaz et al., 2013; Åsberg et al., 2013). The IM, which is far less demanding compared to FA in terms of analyte requirements, minimises the difference between the numerically calculated predictions of the chromatogram and the corresponding experimental data by adjusting the isotherm model parameters. However, the parameter estimation procedure in many applications of curve-fitting in liquid chromatography lacks rigour, because it does not consider the model identifiability or the statistical significance of the parameter estimates. The statistical tools that are usually considered in the estimation are

* Corresponding author.
E-mail address: e.sorensen@ucl.ac.uk (E. Sorensen).

https://doi.org/10.1016/j.compchemeng.2025.109021
Received 29 October 2024; Received in revised form 13 January 2025; Accepted 21 January 2025
Available online 30 January 2025

the Fisher parameter and the residual sum of squares (Quiñones et al., 2000; Gritti and Guiochon, 2005; Kim et al., 2006; Andrzejewska et al., 2009), while more recent applications also consider the Akaike Information Criterion (AIC) (Marković et al., 2014; Roca et al., 2020; den Uijl et al., 2021). (Note that the Fisher parameter is different to the Fisher Information Matrix, which is discussed later in this paper.) Unfortunately, these measures only account for the goodness of fit and model complexity (i.e. AIC) of the overall model, and not for the precision of the parameter estimates themselves.

The uncertainty of isotherm parameters has been considered in the literature. For instance, Joshi et al. (2006) introduced a method for evaluating statistical tests in adsorption isotherm models. In their work, statistical properties such as the variance and the correlation of the isotherm parameters were derived from the Fisher Information Matrix (FIM). Similar work was conducted by Tolazzi et al. (2018), who also stressed the importance of evaluating the uncertainty of the isotherm parameters. Works by Kober et al. (2019, 2021) focused on retrieving precise isotherm parameters by optimally designing informative experiments. Note, however, that these approaches investigated the adsorption isotherm models without accounting for any other physico-chemical process besides adsorption and desorption, although chromatography involves many other additional physical phenomena such as convection along the column and mass transfer. Researchers have also shown increasing interest in the quantification of parametric uncertainty of chromatographic models using Bayesian inference and specifically Markov Chain Monte Carlo (MCMC) techniques (Briskot et al., 2019; He and Zhao, 2020; Yamamoto et al., 2021; Heymann et al., 2023; Sugiyama et al., 2024). Although MCMC can be used for parameter estimation, it can also be utilised in conjunction with another estimator (e.g. Maximum Likelihood) for uncertainty assessment (Heymann et al., 2023).

In addition to obtaining precise parameters for an isotherm model, the selection of which model to use is equally important. In chromatography modelling, selecting the most appropriate isotherm model often relies on intuition, prior expertise and the literature, rather than on systematic analysis. However, strategies are available to aid the identification of the most appropriate model. A model is considered identifiable when parameters can be uniquely determined from experimental observations (Franceschini and Macchietto, 2008b). If one or more candidate models are identifiable, model discrimination methodologies can be employed to select the most appropriate model (Schwaab et al., 2006; Dette et al., 2012; Galvanin et al., 2018). The model parameters can still be uncertain, even if a model adequately fits the experimental data and is identifiable. To reduce the parametric uncertainty, further experiments need to be designed and performed. Optimally designed experiments not only enable obtaining more precise estimates of model parameters with as few experiments as possible, but also assist in indicating which model is more accurate if more than one model alternative is available. Model-Based Design of Experiments (MBDoE) can be used to optimally design experiments that maximise the information obtained from each experiment (Franceschini and Macchietto, 2008b), and thus aid in reducing the parametric uncertainty. MBDoE can also be used to discriminate between different models by indicating experiments whose output can be simulated with high accuracy by potentially one of the candidate models. MBDoE is a good choice for reducing the uncertainty by improving the parameter precision, as well as for model discrimination, because of its efficiency and systematic nature that contribute to making informed decisions in experimental design, thus saving time and resources. MBDoE has been successfully used in various fields, such as bio-processing (Gadkar et al., 2005; Franceschini and Macchietto, 2007; Casey et al., 2007), kinetic model identification in micro-reactor systems (Waldron et al., 2019; Bawa et al., 2023; Pankajakshan et al., 2023), and recently in model-based development of batteries (Park et al., 2018; Andersson et al., 2022). Applications in chromatography are, however, limited. A few articles have reported the use of conventional Design of Experiment (DoE) techniques for optimisation purposes and design space analysis (Debrus et al., 2011; Shekhawat et al., 2019; Latrous, 2022). Specifically, in the review article of Latrous (2022), a number of applications were listed where conventional DoE techniques, such as full and fractional factorial design, have been used in chromatography for control variable (factor) screening and fine tuning for optimal separations. However, optimal experimental designs for liquid chromatography significantly advanced with the groundbreaking work of Barz et al. (2016), that introduced online re-design of breakthrough experiments, enabling the improvement of parametric precision of a pre-defined isotherm model.

In this work, we propose a systematic methodology for selecting a suitable isotherm model and to acquire precise model parameter values for HPLC models, while minimising the number of experiments required. We will evaluate the predictive ability of the model using statistical tests by employing factorial design techniques combined with MBDoE campaigns. To develop our methodology, we draw upon existing work in similar fields, such as the estimation of kinetic model parameters in microreactor systems (Waldron et al., 2019). The paper is organised as follows. Section 2 introduces the theoretical background of MBDoE, parameter estimation, chromatography modelling, as well as the methodology for model identification. Section 3 outlines a case study to which the methodology is applied, followed by the results and a discussion in Section 4. Section 5 concludes with the main findings and discusses potential future work. It will be shown that MBDoE can improve the precision of the estimated parameters using fewer experiments than the conventional DoE methods, thus making MBDoE an excellent vehicle to quickly achieve Quality by Design for pharmaceutical products.

2. Methodology

Before outlining our methodology for isotherm model identification, we will first consider the theoretical foundations of the estimator, statistical tests, and experimental design that are used (Section 2.1). Next follows a description of our approach to chromatography modelling (Section 2.2), and examples of isotherm models that are used in the case study (other models can of course also be used) (Section 2.3). The methodology for model identification is then outlined (Sections 2.4–2.5).

2.1. Theoretical background

This subsection outlines the notions of parameter estimation, the estimator of maximum likelihood, and the statistical tests that follow parameter estimation. Moreover, the concepts of parameter estimability and MBDoE are introduced.

2.1.1. Parameter estimation

A generic model for a chromatographic process consists of a system of Partial Differential and Algebraic Equations (PDAE) with variation in the axial direction only, an n_s dimensional vector of state variables (e.g. the mobile phase concentration within the column or the stationary phase concentration), $\mathbf{x}(t, z)$, and three n_s dimensional vectors of the derivatives of $\mathbf{x}(t, z)$, namely the first-order derivatives with respect to time, $\dot{\mathbf{x}}(t)$, and space, $\mathbf{x}_z(t, z)$, and the second-order derivative with respect to space, $\mathbf{x}_{zz}(t, z)$. It also involves an n_u dimensional vector of manipulated inputs (e.g. the fraction of the organic modifier), $\mathbf{u}(t)$, an n_w dimensional vector of time-invariant inputs (e.g. the inlet concentration), $\mathbf{w}$, a n_θ dimensional vector of model parameters (e.g. the isotherm parameters), $\boldsymbol{\theta}$, and an n_M dimensional vector of measured output variables, $\hat{\mathbf{y}}$. The measured variables vector groups the model outputs and can also be directly measured through experiments. In the paradigm of chromatography, such measured variable is the mobile phase concentration at the outlet as a function of time. In contrast, the mobile phase concentration within the column (i.e. everywhere but the outlet) is a state variable because it describes the state of the process

and it cannot be directly measured. Thus, we can write:

$$\mathbf{y} = \boldsymbol{f}[\mathbf{x}(t,z), \dot{\mathbf{x}}(t,z), \mathbf{x}_z(t,z), \mathbf{x}_{zz}(t,z), \mathbf{u}(t), \mathbf{w}, \boldsymbol{\theta}] \quad (1)$$

In order to simulate a process described by this model, one must first specify reliable values for the n_θ parameters $\boldsymbol{\theta}$. The parameter values can be estimated based on experimental data. Different estimators exist (e.g., Ordinary Least Squares Metzger et al., 2004; Marković et al., 2014, Maximum Likelihood Yan et al., 2015, Bayesian inference He and Zhao, 2020); in this work we will adopt the Maximum Likelihood, because this estimator also considers the variance of the parameter estimates. For a number of experiments, n_{exp}, with each experiment containing a number of measurements, n_m, the residuals ρ_{ij} of the ith experiment and the jth measurement are the difference between the model prediction f_{ij} and the corresponding experimental measurement y_{ij} (Bard, 1974):

$$\rho_{ij} = y_{ij} - f_{ij}(\boldsymbol{\theta}) \quad \text{for} \quad i = 1, \ldots, n_{exp} \quad \text{and} \quad j = 1, \ldots, n_m \quad (2)$$

In a Maximum Likelihood parameter estimation problem, we seek to estimate the parameter values of $\boldsymbol{\theta}$ that maximise the probability of the model predicting the experimentally measured variables accounting for a Gaussian distributed error, σ_{ij} (Bard, 1974). This probability is captured in the log-likelihood objective function, $\Phi(\boldsymbol{\theta})$, that is to be maximised:

$$\max_{\boldsymbol{\theta}}(\Phi(\boldsymbol{\theta})) = \max_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{n_{exp}} \sum_{j=1}^{n_m} \left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_{ij}^2) - \frac{1}{2}\left(\frac{\rho_{ij}}{\sigma_{ij}}\right)^2 \right] \right\} \quad (3)$$

Note that the $\mathbf{u}$ vector of manipulated inputs and the $\mathbf{w}$ vector of time-invariant controls in Eq. (1) are assumed to be perfectly controlled, and thus the associated potential systematic errors affecting the inputs are neglected. Therefore, the residuals only consider the measurement error of the system, which is assumed to be Gaussian distributed with a mean of 0 and a standard deviation equal to σ_{ij}. The parameter values obtained from the log-likelihood function (Φ) are called maximum likelihood estimate values, $\hat{\boldsymbol{\theta}}$ (Bard, 1974; Franceschini and Macchietto, 2008b; Waldron et al., 2019).

2.1.2. Goodness of fit

The model is evaluated based on its ability to produce model outputs that lie within the experimental measurements. However, the measurement noise must also be taken into account. The goodness of fit of the model is a measure of the comparison between the distribution of the residuals and a hypothetical distribution of the measurement error. This is conducted by employing the χ^2 test (Quaglio, 2020). This test aims at evaluating whether the null hypothesis is true; that is, whether the sum of the squared normalised residuals at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ follows a χ^2 distribution. Here, the hypothesis is that the sum of squared normalised residuals follows a χ^2 distribution. As mentioned, the distribution of the measurement error is assumed to be Gaussian with a mean of 0 and a standard deviation equal to σ_{ij}. Since the sum of squared normalised residuals is assumed to follow a χ^2 distribution, χ^2 is computed as (Bard, 1974):

$$\chi^2 = \sum_{i=1}^{n_{exp}} \sum_{j=1}^{n_m} \left(\frac{\rho_{ij}(\hat{\boldsymbol{\theta}})}{\sigma_{ij}} \right)^2 \quad (4)$$

The χ^2 value is then compared to a critical value, χ_c^2, that is a tabulated χ^2 distribution with a Degree of Freedom, DoF $= n_{exp} \times n_m - n_\theta$ and an α% significance level (e.g. 95%) (Bard, 1974). The significance level represents the maximum acceptable probability that one should observe an extreme result, although the null hypothesis is not rejected. For instance, in the 95% χ^2 test, if the χ^2 value is less than the 95% χ_c^2, the critical value from the statistical table, then accounting for the current evidence, the null hypothesis, that is the difference between the weighted residuals and expected weighted residuals is zero, cannot be rejected. If the opposite occurred, however, that is, if χ^2 is greater than χ_c^2, then the evidence would be sufficient to reject the null hypothesis. If the model passes the χ^2 test, then the model predictions based on the estimated values of the parameters are expected to fit the experimental measurements 95% of the times we were to repeat the same experiment.

2.1.3. Fisher information matrix and variance–covariance matrix of model parameters

To quantify the uncertainty of the parameter estimates, we consider their standard deviation, which indicates the magnitude of the error in our parameter estimates. The standard deviation estimation comes from the $n_\theta \times n_\theta$ variance–covariance matrix, $\mathbf{V}_\theta$, which is the first-order approximation of the inverse of the observed Fisher Information Matrix (FIM), $\mathbf{H}_\theta$ (Bard, 1974; Waldron et al., 2019), without accounting for any prior information coming from preliminary experiments or information on the range of the parameters:

$$\mathbf{V}_\theta \approx \mathbf{H}_\theta^{-1} \quad (5)$$

The FIM, $\mathbf{H}_\theta$, is a square matrix with k, l elements that are calculated as (Franceschini and Macchietto, 2008b; Quaglio, 2020):

$$\left[\mathbf{H}_\theta\right]_{kl} = \sum_{i=1}^{n_{exp}} \sum_{j=1}^{n_m} \frac{1}{\sigma_{ij}^2} \left(\frac{\partial f_{ij}}{\partial \theta_k} \frac{\partial f_{ij}}{\partial \theta_l} \right) \text{for} \quad k = 1, \ldots, n_\theta \quad \text{and} \quad l = 1, \ldots, n_\theta \quad (6)$$

where $\partial f_{ij}/\partial\theta_k$ and $\partial f_{ij}/\partial\theta_l$ are the sensitivities of the model prediction f_{ij} with respect to the kth and lth parameters, respectively. From the variance–covariance matrix, $\mathbf{V}_\theta$, we can obtain the parameter confidence intervals for the different significance levels (typically between 90% and 99%). For instance, the 95% confidence interval of the ith parameter is estimated by multiplying the square root of the variance element $\left[\mathbf{V}_\theta\right]_{ii}$ with the student t-value, $t_{ref}(95\%, \text{DoF})$:

$$95\%\ \text{confidence interval}_i = \sqrt{\left[\mathbf{V}_\theta\right]_{ii}} \times t_{ref}(95\%, \text{DoF}) \text{ for } i = 1, \ldots, n_\theta \quad (7)$$

The 95% confidence interval denotes that the probability of the true parameter values lying within this estimated region is 95%. The student t-value, which comes from statistical tables for the given number of DoF, is compared against the t_i-value of the ith parameter:

$$t_i \approx \frac{\hat{\theta}_i}{95\ \%\ \text{confidence interval}_i} \quad \text{for } i = 1, \ldots, n_\theta \quad (8)$$

If the t_i-value is greater than the corresponding student t_{ref}-value then the ith parameter is statistically different from zero at 95% confidence (Waldron et al., 2019; Bard, 1974; Quaglio, 2020). In other words, a t-value greater than t_{ref} indicates that the parameter is statistically significant when considering the experimental data because it provides an improvement over a simpler model without the parameter under evaluation.

2.1.4. Parameter estimability and practical identifiability

The dynamic sensitivities of the model output with respect to the model parameters quantify how sensitive the model output is to an infinitesimal change in the parameter values. The dynamic sensitivities of the model output are described by the sensitivity matrix, $\mathbf{Q}_s(\boldsymbol{\varphi})$, of dimensions $n_m \times n_\theta$, whose elements evaluated at a given design vector, $\boldsymbol{\varphi}$, a subset of the control variables $\boldsymbol{u}$ or $\boldsymbol{w}$, are computed as (Franceschini and Macchietto, 2008b):

$$\left[\mathbf{Q}_s(\boldsymbol{\varphi})\right]_{kl} = \frac{\partial f_{kl}}{\partial \theta_l} \quad \text{for} \quad k = 1, \ldots, n_m \quad \text{and} \quad l = 1, \ldots, n_\theta \quad (9)$$

A model is considered identifiable if parameters can be uniquely estimated from the data and the rank of the sensitivity matrix $\mathbf{Q}_s(\boldsymbol{\varphi})$ is equal to the number of model parameters (Franceschini and Macchietto, 2008b; Quaglio, 2020). If the rank of the sensitivity matrix is lower than the number of model parameters, then the elements of the matrix are not linearly independent, and the determinant of the FIM is zero, thus disabling any further parameter estimation activity (i.e. the confidence interval becomes zero). This practically means that the parameters are strongly correlated and they cannot be uniquely estimated.

The sensitivity matrix $\mathbf{Q}_s(\boldsymbol{\varphi})$ is a local matrix, i.e. it depends on the experiments and measurements collected. Thus, there might be another design vector, another $\boldsymbol{\varphi}$, that enables the estimation of independent parameters if the current design vector has failed. Since an infinite number of experiments is not feasible, the identifiability of a model is evaluated in a pre-defined experimental region (defined by a set of lower and upper bounds for $\boldsymbol{\varphi}$). If a model fulfils the identifiability criteria in this region, then it is denoted *practically* identifiable.

In our methodology, practical identifiability is evaluated based on the procedure developed by Waldron et al. (2019). The first step of the methodology is defining the experimental design space, as well as defining the measurement variance model. The next step involves determining the maximum number of experiments allowed. Waldron et al. (2019) claimed that the most rigorous approach is first designing a small number of experiments using a factorial or a Latin hypercube design followed by a sequential MBDoE; however, this approach can be computationally expensive. Therefore, in this work, we will implement the second proposed method of Waldron et al. (2019), where an initial set of experiments is designed to obtain information on the expected uncertainty of model parameters, then only if it is deemed necessary, will further optimal experiments be designed sequentially.

In terms of model identification, once the estimates and statistical information (i.e. χ^2- and t- values) have been acquired for each tested model, the models are evaluated one-by-one using the t-test. If all the t-values of a model are larger than the student t-value corresponding to the number of degrees of freedom of the experiments, then the model is considered practically identifiable and the model parameters are accepted. If the model fails the t-test, new experiments must be designed to further explore that model. These new experiments can be designed using either a standard factorial design or a more rigorous MBDoE procedure. If the model fails again after further experiments (up to a given maximum) have been considered, then the model is considered definitively unidentifiable under the given experimental budget in the investigated experimental design space, and is therefore rejected.

2.1.5. Model-based design of experiments for parameter precision

When using MBDoE for parameter estimation, the experiments are designed to minimise the confidence region of the parameter estimates, in other words, making the parameter estimates as accurate as possible. With MBDoE, the experiments are designed for conditions where the model output is the most sensitive to a change in the parameter values (Quaglio, 2020). The sensitivity of the model output with respect to the parameters is formulated through the FIM (Franceschini and Macchietto, 2008b). The expected FIM of the new experiments, $\hat{\mathbf{H}}'_\theta$, is calculated based on the following:

$$\hat{\mathbf{H}}'_\theta(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}) = \sum_{i=1}^{n_d} \hat{\mathbf{H}}_{\theta_i}(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}) + \mathbf{H}^0_\theta \tag{10}$$

where $\hat{\mathbf{H}}_\theta$ is the FIM to be estimated from the new experiments n_d, and $\mathbf{H}^0_\theta$ is the preliminary information matrix collected from experiments already performed or prior information on the range of variability of model parameters. The optimal design vector can then be obtained by locating values for $\boldsymbol{\varphi}$ that maximise scalar properties of the FIM, or minimise scalar properties of the FIM inverse, the variance–covariance matrix, that is:

$$\boldsymbol{\varphi} = \arg\max \psi_{criteria}\{\left[\hat{\mathbf{H}}_\theta(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi})\right]\} = \arg\min \psi_{criteria}\{\left[\hat{\mathbf{V}}_\theta(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi})\right]\} \tag{11}$$

The most commonly employed scalar properties, also called optimal design criteria, targeted for maximisation or minimisation, are the following Franceschini and Macchietto (2008b):

- **D-optimal criterion**: maximises the determinant of the FIM or minimises that of the variance–covariance matrix, thus minimising the volume of the confidence ellipsoid of the parameters.
- **E-optimal criterion**: maximises the smallest eigenvalue of the FIM or minimises the largest eigenvalue of the variance–covariance matrix, thus minimising the major axis of the confidence ellipsoid of the parameters.
- **A-optimal criterion**: minimises the trace of the variance–covariance matrix, thus minimising the rectangle that encloses the confidence ellipsoid of the parameters.

Note that in the case of the A-optimality the maximisation of the trace of the FIM does not coincide with the minimisation of the trace of its inverse, namely the variance–covariance matrix. The optimal criteria are not limited to those listed above, but there are also others used in the literature such as the G-, V-, and T- optimal designs (Shahmohammadi, 2019).

A visual interpretation of the above criteria is depicted in Fig. 1, which shows the joint confidence region between two parameters θ_1 and θ_2. The confidence region can be plotted for different confidence levels (typically 90, 95, or 99%). For the two-parameter case, the D-optimality considers the minimisation of the volume of the confidence region, the E-optimality considers the major axis of the confidence region, and A-optimality considers the hyper-rectangle enclosing the said region.

2.2. Mathematical model

A number of mathematical models have been used to describe HPLC, each of them corresponding to different levels of accuracy and complexity depending on how the physico-chemical phenomena are considered and on the nature of the analyte. In this work, the Equilibrium Dispersion Model (EDM) is used for simulating the elution profiles of a single component (solute/analyte) as it moves through a chromatographic column using a mobile phase consisting of a mixture of solvents (e.g. water and acetonitrile). The EDM assumes instantaneous adsorption–desorption kinetics and extremely fast intraparticle mass transfer kinetics (Schmidt-Traub et al., 2020; Katsoulas et al., 2023). The selection of the EDM is based on its relatively good accuracy in a wide range of applications combined with its relatively low computational cost. The mass balance for a generic analyte is given as (De Luca et al., 2020; Schmidt-Traub et al., 2020; Katsoulas et al., 2023):

$$\frac{\partial C_m}{\partial t} + F\frac{\partial q}{\partial t} + u_{int}\frac{\partial C_m}{\partial x} = \mathcal{D}_{app}\frac{\partial^2 C_m}{\partial x^2} \quad , \quad 0 < x < L \tag{12}$$

where C_m and q are the concentrations of the analyte in the mobile and stationary phases, respectively, F is the volumetric phase ratio, u_{int} is the hypothetical interstitial velocity, $\mathcal{D}_{app}$ is the apparent dispersion coefficient, and L is the column length. The volumetric phase ratio is the ratio between the volumes occupied by the mobile and stationary phases, $F = \epsilon_t/(1-\epsilon_t)$, where ϵ_t is the total porosity of the column. The hypothetical interstitial velocity is given by $Q/(S\epsilon_t)$, where Q is the mobile phase volume flow rate and S is the column cross section. The apparent dispersion coefficient lumps together the effects of hydrodynamic dispersion and mass transfer resistance in the stationary phase, with $\mathcal{D}_{app} = \frac{u_{int}L}{2N}$, where N is the number of theoretical plates. Note that in non-diluted conditions, the apparent dispersion coefficient is a function of the isotherm, and consequently of the concentration (Schmidt-Traub et al., 2020), and the dispersion coefficient is then not constant. In experiments where dispersion can be of importance, one therefore has to obtain values for the mass transfer coefficients and utilise a rigorous formulation for the apparent dispersion coefficient (Katsoulas et al., 2023). In order to reduce the computational burden, as well as not having to estimate values for the mass transfer coefficients, one may use a constant number of theoretical plates to calculate the coefficient, for instance by assuming that the number of plates is large enough to render the effects of dispersion insignificant (Andrzejewska et al., 2009).

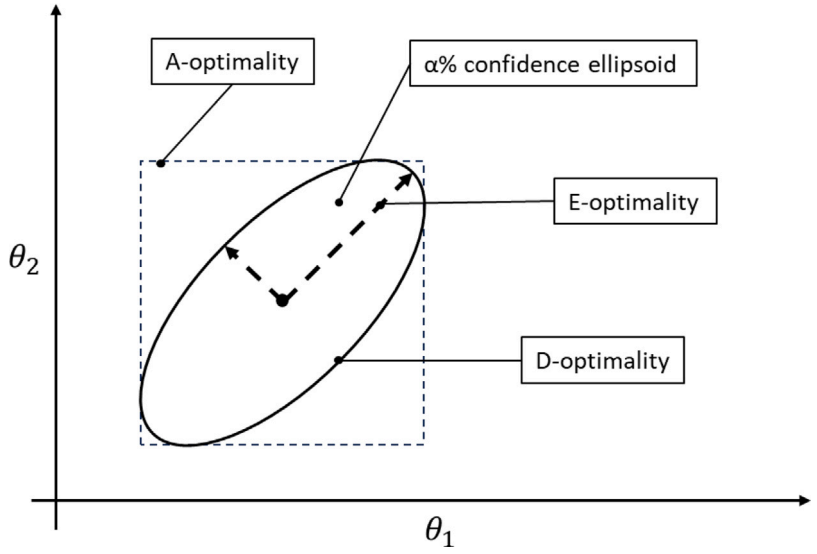


Fig. 1. Visual representation of the optimal design criteria on the α% confidence region formed by two parameters, θ_1 and θ_2.
Source: Adapted from Franceschini and Macchietto (2008b).

The stationary phase concentration, q, is expressed as a function of the mobile phase concentration C_m:

$$q = f(C_m) \tag{13}$$

This function is the adsorption isotherm that we need to identify and whose parameters we need to estimate. At the preparative scale of chromatography (non-diluted conditions), isotherms usually involve at least two parameters (Carta and Jungbauer, 2020). Different isotherm model alternatives will be discussed shortly.

To make the system of equations solvable, we need to couple the EDM with suitable boundary and initial conditions. At the inlet, the concentration equals the loading concentration. But as the analyte enters the column, it gets dispersed; therefore, the boundary condition at the inlet of the column ($x = 0$) reads (Schmidt-Traub et al., 2020):

$$C_{in} = C_m(x = 0,\ t) - \frac{D_{app}}{u_{int}} \frac{\partial C_m(x = 0,\ t)}{\partial x} \tag{14}$$

where C_{in} is the loading concentration at the column inlet. At the column outlet ($x = L$), the concentration is assumed not to change with respect to the immediate previous point in space, that is (Schmidt-Traub et al., 2020):

$$\frac{\partial C_m(x = L,\ t)}{\partial x} = 0 \tag{15}$$

This is an assumption that is often used in chromatography modelling and it has been found valid for chromatography applications owing to the very small boundary layer at the outlet (Katsoulas et al., 2023). The column is assumed to initially ($t = 0$) not contain any analyte, hence the initial condition is:

$$C_m(t = 0) = 0 \quad , \quad 0 < x < L \tag{16}$$

2.3. Isotherm models

The isotherm, Eq. (13), involves a function that expresses the correlation between the concentrations of the analyte in the mobile phase and in the stationary phase. In dilute conditions, where the analyte is introduced in very small amounts, the isotherm can be assumed to be linear (Carta and Jungbauer, 2020; Schmidt-Traub et al., 2020):

$$q = HC_m \tag{17}$$

where H is the Henry coefficient.

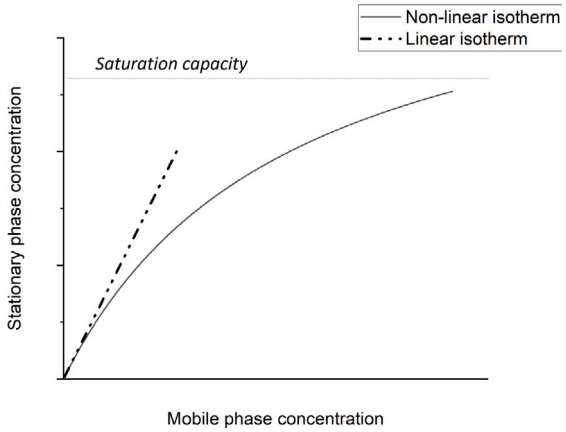


Fig. 2. Non-linear vs. linear isotherm representation as a plot of stationary phase analyte concentration as a function of mobile phase analyte concentration.

One of the most important characteristics of an analyte that elutes from a chromatographic column is its net retention time, which expresses the affinity of the analyte for the solid adsorbent (Schmidt-Traub et al., 2020). By evaluating the retention time, one can estimate the Henry coefficient:

$$H = \frac{t_R - t_{0,t}}{t_{0,t}} \frac{\epsilon_t}{1 - \epsilon_t} \tag{18}$$

where t_R is the overall retention time of the analyte in the column, $t_{0,t}$ is the retention time of an analyte that penetrates the pores of the stationary phase but does not adsorb. This isotherm works only in dilute conditions, where, as depicted in Fig. 2, the stationary and mobile phase concentrations are related linearly.

In preparative chromatography, the column is overloaded with analyte sample as large sample volumes at high concentrations are introduced. Modelling the adsorption behaviour at overloaded conditions requires the use of non-linear isotherms. Commonly employed non-linear isotherms are the Langmuir, BET, Freudlich, and Moreau isotherms (Gritti and Guiochon, 2004; Carta and Jungbauer, 2020), with the first being the most widely used. To illustrate the capabilities of the methodology presented in this work, we will focus on three Langmuir-type isotherms, the simple Langmuir, bi-Langmuir, and tri-Langmuir isotherms with two, four, and six parameters, respectively.

Note, however, that the methodology can be applied to any isotherm model. We have used the three different versions of the Langmuir isotherm as these are currently the most widely used and preferred by industry. The simple Langmuir isotherm reads:

$$q = \frac{K_1 C_m}{1 + \frac{K}{q_{sat,1}} C_m} \tag{19}$$

where in this case $\theta_{Langmuir} = [K_1, q_{sat,1}]$. The bi-Langmuir:

$$q = \frac{K_1 C_m}{1 + \frac{K_1}{q_{sat,1}} C_m} + \frac{K_2 C_m}{1 + \frac{K_2}{q_{sat,2}} C_m} \tag{20}$$

with $\theta_{bi-Langmuir} = [K_1, q_{sat,1}, K_2, q_{sat,2}]$. The tri-Langmuir:

$$q = \frac{K_1 C_m}{1 + \frac{K_1}{q_{sat,1}} C_m} + \frac{K_2 C_m}{1 + \frac{K_2}{q_{sat,2}} C_m} + \frac{K_3 C_m}{1 + \frac{K_3}{q_{sat,3}} C_m} \tag{21}$$

with $\theta_{tri-Langmuir} = [K_1, q_{sat,1}, K_2, q_{sat,2}, K_3, q_{sat,3}]$. Here $q_{sat,i}$ is the saturation capacity and K_i is the retention factor of the ith type of binding sites ($i = [1, 2, 3]$). As previously stated, in dilute conditions the isotherm is linear and we only consider the Henry coefficient (Eq. (18)), which relates to the retention factors of the binding sites as follows:

$$H_n = \sum_{i=1}^{n} K_i \quad \text{for} \quad n = 1, \ldots, 3 \tag{22}$$

In overloaded conditions, K_i is the retention factor and $q_{sat,i}$ is the saturation capacity that represent the adsorption behaviour of a binding site type of adsorbent. Different types of binding sites arise from variations in surface chemistry, pore size and functional group distributions throughout the column (Schmidt-Traub et al., 2020; Carta and Jungbauer, 2020). Consequently, the bi-Langmuir isotherm formulates the adsorption behaviour of an adsorbent that has two types of binding sites, with K_2 and $q_{sat,2}$ being the retention factor and saturation capacity of the second type of binding site, respectively; and equivalently for the tri-Langmuir isotherm. Essentially, each pair of retention factor and saturation capacity expresses the retention time on the binding site and the maximum amount of solute that can bind to the site, respectively.

2.4. Acquiring values for the Henry coefficient

The Henry coefficient, H, appears in Eqs. (19)–(21) and in most non-linear isotherms. Thus, obtaining values for this coefficient is crucial. In experiments under dilute conditions, one can estimate the Henry coefficient using Eq. (18), as for dilute conditions $K_1 = H$. This only requires recording the total retention time of an analyte introduced into the column, t_R, and the retention time of a non-retained tracer, $t_{0,t}$. When there is only one type of binding site, the Henry coefficient recorded at dilute conditions is equivalent to the retention factor of the only type of binding site, K_1. However, the value of H obtained in this manner is unreliable as a value for K_1, because using this value introduces additional uncertainty that may affect a potential subsequent estimation of $q_{sat,1}$. Therefore, in our procedure we use the Henry coefficient obtained at dilute conditions only as an initial estimate for K_1 and then we estimate K_1 as part of the overall estimation problem.

2.5. Experimental design and parameter estimation procedure

This work presents a step-by-step methodology to identify the most suitable isotherm model and estimate its associated isotherm parameters that best describe the adsorption–desorption equilibrium in an HPLC system. In the following, problems that may arise during the model identification or parameter estimation are identified, and practical solutions to those problems are proposed. Our methodology is based on a similar methodology for kinetic model identification (Waldron et al., 2019), but is adapted to the needs of isotherm models and the nature of chromatography. Fig. 3 depicts the proposed methodology, which will be described step-by-step in the following. In short, in Step 0, initial guesses for the parameters of the retention factor (see Eq. (22)) are obtained by performing experiments under dilute conditions. These values are needed in the first parameter estimation. The first parameter estimation takes place after non-linear screening experiments are designed and performed in Step 1 by factorial-design. With the use of these screening experiments we can propose candidate isotherm models and evaluate whether these models and initial parameters can fit the experimental data. If, at this preliminary stage, the candidate models and associated parameters fit the experiments, Step 2 evaluates the practical identifiability of these models under the current experimental budget. If at least one candidate model is identifiable we proceed with Step 3 to refine the precision of its parameters and minimise the parametric uncertainty.

2.5.1. Step 0: Estimate Henry coefficient values

As previously mentioned, it is useful to acquire Henry coefficient values that can later be used as part of the initial guesses to obtain the initial estimates of the retention factor parameters. Note that not all isotherms contain retention factors that are equivalent to the Henry coefficient of the linear isotherm (e.g. the BET isotherm); however, it is likely that one of the proposed isotherm expressions for separations of bio-molecules and fine chemicals would involve this retention factor (Carta and Jungbauer, 2020). Moreover, the Henry coefficient value also provides information on the net retention time of the molecule, i.e. the net time that the molecule is adsorbed in the solid (Schmidt-Traub et al., 2020).

The estimation of the Henry coefficient value starts with performing at least three experiments under dilute conditions, to ensure that the results are repeatable. The dilute conditions are guaranteed if a small sample volume at a low concentration is injected. Then, Eq. (18) is used to obtain the Henry coefficient value for each experiment. Note that the total porosity value must be known or already estimated. If the values of the coefficient from each of the initial experiments are similar, we then proceed using the mean average of the values. If there are outliers, another experiment in dilute conditions should be performed. This procedure should be continued until consistent results are obtained and an average value can be used.

2.5.2. Step 1: Goodness of fit test - Preliminary model screening

The first main step of the methodology involves conducting experiments that enable us to propose candidate isotherm models and to acquire initial estimates for the parameters of these models. It is assumed that the models are proposed by the user based on the user's previous experience, as model development is not included in our methodology.

The initial experiments can be designed by simple factorial design. After the experiments have been designed and performed, parameter estimation takes place for each of the proposed candidate models. These initial parameter estimates allow us to perform goodness of fit tests; in our case study, the 95% χ^2-test. Depending on whether or not a candidate model passes the test, it is either accepted for further investigation or rejected. If none of the candidate models passes the test, then new models must be considered.

2.5.3. Step 2: Practical model identifiability under an experimental budget

In this step, according to what was proposed in Section 2.1.4, factorial experiments can be simulated for the models that passed the initial screening using the initial parameter estimates obtained from Step 1. This means that the estimated parameter values are used in the given model to simulate the system at conditions determined by a standard factorial design. The practical model identifiability would then be evaluated for each model under a maximum experimental budget. The experimental budget refers to the total number of experiments we can allow to be performed in the lab in order to identify a model.

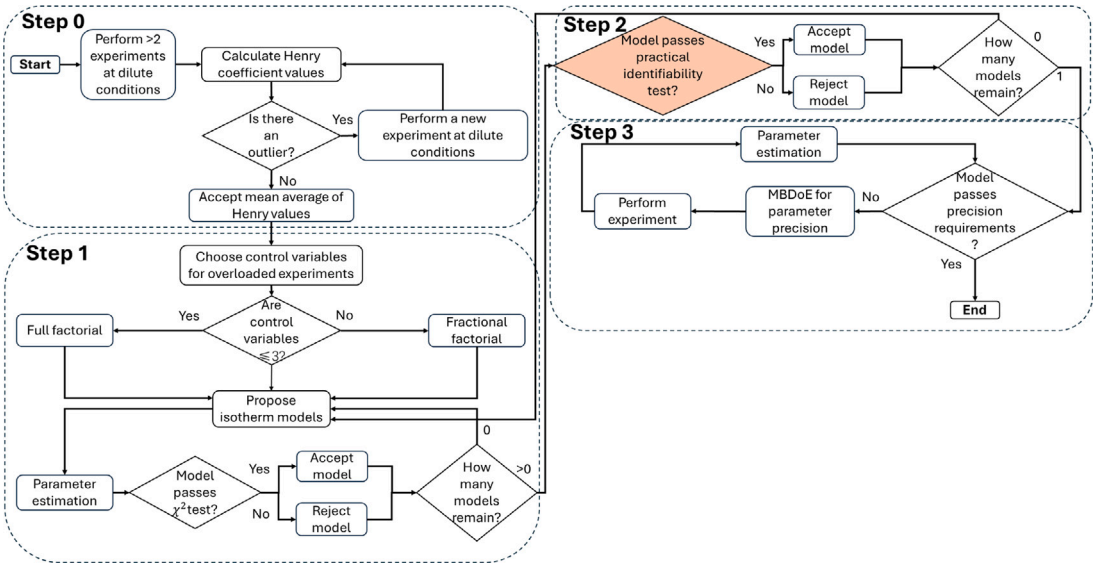


Fig. 3. Flow diagram showing the methodology for isotherm model identification and parameter estimation.

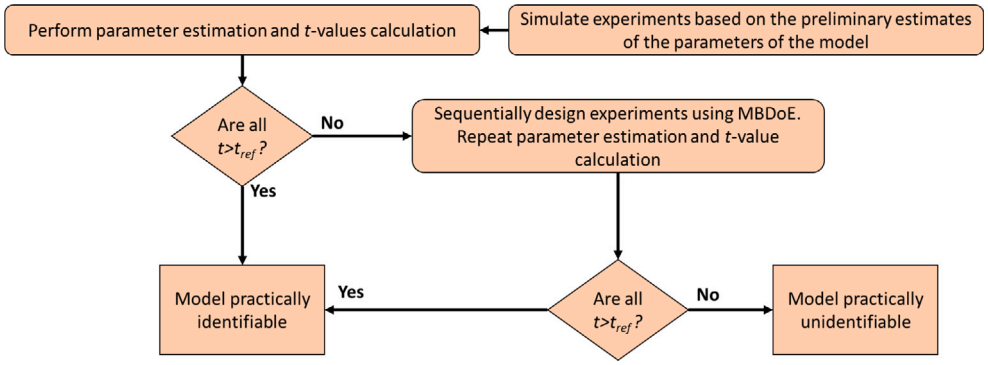


Fig. 4. Practical model identifiability flow diagram.
Source: Adapted from Waldron et al. (2019).

Although the practical identifiability test is evaluated under a maximum budget of in-silico experiments, it reflects whether a model can be identified under the same number of real experiments. In other words, the practical model identifiability test utilises simulations to answer the question "can one identify the proposed models under a certain number of experiments?". However, the test is still an approximation since there is a potential parametric mismatch between the preliminary and the real estimates."Real" experiments are conducted in all the other steps of this methodology, that is preliminary screening, and refinement of the parameter precision.

The practical identifiability is in our methodology assessed by the *t*-test. If a model passes the *t*-test following the first parameter estimation against the full factorial, it is considered practically identifiable. If a model fails the *t*-test, then MBDoE should be employed to design optimal experiments sequentially, unless the potential model is far from being identifiable within a reasonable experimental budget. For instance, if after a number of experiments the *t*-values are orders of magnitude off the t_{ref} values, this could indicate that a reasonable amendment (e.g., ∼ 20 percent) of the experiment budget would not render the model practically identifiable. After the MBDoE campaign, the parameter estimation would be repeated and the *t*-test would be re-evaluated. If all the *t*-values of the parameters are greater than the *t*-value of reference, then the model is considered practically identifiable, otherwise the model is practically unidentifiable. This procedure of testing the practical identifiability of a model was drawn from the work of Waldron et al. (2019) and is depicted in Fig. 4. Note that Fig. 4 delves into the practical model identifiability test, which is part of the general methodology depicted in Step 2 of Fig. 3. The merit of assessing practical model identifiability is two-fold. Firstly, it allows us to evaluate whether a satisfactory level of parametric precision (e.g., high *t*-values) can be achieved for a model structure within the defined design space, $\boldsymbol{\varphi}$. Secondly, by considering constraints on the experimental budget, we can obtain an approximate estimate of the total number of experiments required in the real-world laboratory setting.

2.5.4. Step 3: Parameter precision refinement with MBDoE

Step 2 has determined which candidate model is practically identifiable given the system under investigation and the experiments performed. This model is now carried forward to the last step, Step 3, to further improve the accuracy of the parameter values. Uncertainty is quantified through the *t*-test but even if the *t*-test is satisfied, the parametric uncertainty may have a large impact on the uncertainty of the output, i.e. the chromatogram. We quantify the uncertainty of the output by stochastically simulating chromatograms, using Monte Carlo simulations for a given number of scenarios, replacing the isotherm parameters with normal distributions that adhere to their estimated uncertainty. These rather inexpensive Monte Carlo simulations can be performed after each iteration of experiment design-parameter estimation along with the estimation of the *t*-values to evaluate whether the procedure can be terminated or not. If the parameteric uncertainty of the isotherm is sufficiently small, then the stochastically generated

chromatograms of the different scenarios will have no or very small deviation between them on the time axis. To be fully confident with the model verification and associated parameter values, experiments at the corners of the experimental design space should be performed. The final model validation also uses the χ^2-test to assess the goodness of fit, as previously used. Thus, the outcome of the uncertainty analyses as well as the design space corner validation determine if the precision requirements are met or not. If the model fails to pass the χ^2 test at this stage, the parameter refinement procedure must continue, i.e. further MBDoE experiments will be needed.

3. Case study development

Our proposed methodology aims to identify the isotherm model with the corresponding isotherm parameters that best describes a given chromatographic system. Because the accuracy of the model identified with our methodology is impossible to determine for an unknown system (since the values of the parameters are unknown), to illustrate our methodology we will use an in-silico case study for which the exact values of the parameters are known. The in-silico case study was generated using the EDM coupled with a bi-Langmuir isotherm. Using an in-silico system allows us to properly evaluate the performance of the methodology, because we know the true isotherm model. This would not have been possible if we had used a real experimental system.

Regardless of whether or not the case study was generated in-silico or from real experiments, in our methodology we use the EDM, coupled with candidate isotherm models, to estimate the model parameters, to perform the MBDoE, and for identification.

3.1. Re-parameterisation of Langmuir-type isotherm models

Parameter estimation is often hindered by various factors, one of which is the mathematical structure of the model, since this may be formulated in a way that prevents independent parameters from being estimated (Franceschini and Macchietto, 2008b). The main issue arises from high correlation between the parameters, which renders the Fisher Information Matrix (FIM) singular, thus impeding its estimation. To address this problem, several methods of re-parameterisation have been proposed in the literature, particularly for kinetic models involving Arrhenius-type expressions (Franceschini and Macchietto, 2007; Espie and Macchietto, 1988; Franceschini and Macchietto, 2008c,a; Schwaab and Pinto, 2008). Re-parameterisation has also been conducted on the inference of adsorption isotherm parameters using regression (Joshi et al., 2006; Marković et al., 2014); however, in chromatography, it has not yet been applied for curve fitting.

The Langmuir-type isotherms can be re-parameterised to increase the sensitivity of the model output with respect to the saturation capacity parameter, $q_{sat,1}$, i.e. for the low-energy sites of the Langmuir-type parameters, in order for the determinant of the FIM to increase in value, i.e. to reduce the parameter correlation. The steps taken to address this issue are outlined in the supplementary material. We implemented the following transformation:

$$\theta_1 = \ln(q_{sat,1}) \tag{23}$$

to Eqs. (19)–(21), which, after the transformation, read:

$$q = \frac{K_1 C_m}{1 + \frac{K}{exp(\theta_1)} C_m} \tag{24}$$

$$q = \frac{K_1 C_m}{1 + \frac{K_1}{exp(\theta_1)} C_m} + \frac{K_2 C_m}{1 + \frac{K_2}{q_{sat,2}} C_m} \tag{25}$$

$$q = \frac{K_1 C_m}{1 + \frac{K_1}{exp(\theta_1)} C_m} + \frac{K_2 C_m}{1 + \frac{K_2}{q_{sat,2}} C_m} + \frac{K_3 C_m}{1 + \frac{K_3}{q_{sat,3}} C_m} \tag{26}$$

Table 1
The bi-Langmuir parameter values used for the in-silico experiments.
Source: Taken from by Andrzejewska et al. (2009).

Model	Equation	K_1	$q_{sat,1}$ (θ_1)	K_2	$q_{sat,2}$
bi-Langmuir	$q = \frac{K_1 C_m}{1+\frac{K_1}{q_{sat,1}} C_m} + \frac{K_2 C_m}{1+\frac{K_2}{q_{sat,2}} C_m}$	3.88	426.7 (6.06)	3.01	2.35

The sensitivities of the model output, that is the outlet concentration, coupled with (a) the non re-parameterised and (b) the re-parameterised bi-Langmuir isotherm, with respect to the isotherm parameters are depicted in Figs. 5(a) and 5(b), respectively. As discussed in the previous section, the sensitivity of the outlet concentration with respect to the isotherm parameters is given by $\mathbf{Q}_s = \partial C_{out}/\partial\theta$ and is approximated by $\Delta C_{out}/\Delta\theta$, where θ is any of the isotherm parameters.

Fig. 5 suggests that replacing $q_{sat,1}$ with $\ln q_{sat,1}$ increased the sensitivity for the saturation capacity (Fig. 5 - orange line). In the original formulation of the model (Fig. 5(a)), the sensitivity (orange line) was an order of magnitude smaller than in the reformulated model (Fig. 5(b)), and therefore not observable on the sensitivity plot. The determinant of the FIM in the investigations conducted with the re-parameterised model increased drastically, by five orders of magnitude, thus rendering the parameters of the transformed model more estimable.

3.2. In-silico experiments considerations

Our in-silico system, which for the sake of illustrating our methodology takes the place of a real system, is inspired by the separation of tri-Leucine (LLL), a relatively small peptide. The process parameters, as well as the parameters for a bi-Langmuir isotherm, were inspired by the work of Andrzejewska et al. (2009), with the details given in the supplementary material. The parameter values are summarised in Table 1. This model and these parameter values are therefore the true values that our methodology should identify.

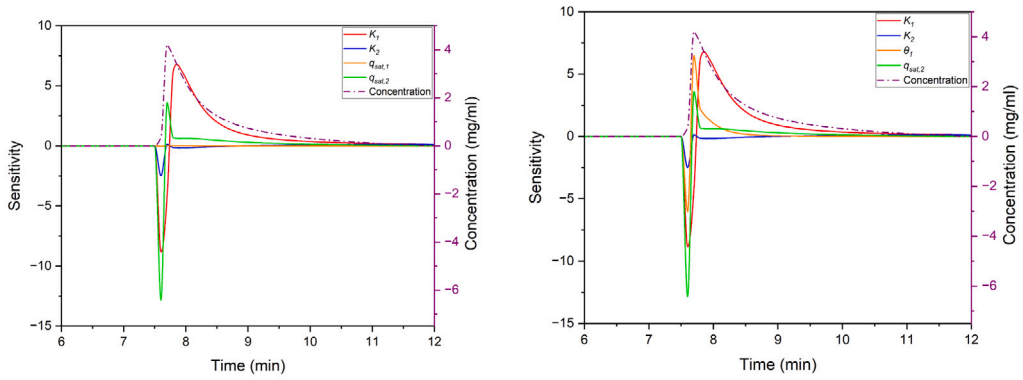
The case study was structured into cases A, B and C. In Cases A and B, the preliminary in-silico experiments in Step 1, as explained taking the place of real experiments for the sake of this investigation, were designed according to a full factorial design, while in Case C only according to a fractional factorial design, i.e. Case C involved fewer initial experiments. As will be shown shortly, these two different strategies led to different findings.

Moreover, in Case A we only employed standard factorial designs throughout the methodology while in Case B we considered MBDoE experiments after the preliminary experiments; this is to show the benefits of the MBDoE approach. The three case studies are summarised below:

- Case A: Factorial design for preliminary and precision experiments.
- Case B: Factorial design for preliminary experiments and MBDoE for precision experiments.
- Case C: Fractional factorial design for preliminary experiments and MBDoE for precision experiments.

Note that the fractional factorial design in Case C used for its screening experiments is actually against the proposed methodology, which argues that for a number of control variables (not exceeding three), the use of a full factorial should be favoured. Using the fractional factorial design enabled us to assess the impact of using fewer experiments, which may be favourable from a financial and time perspective, on the overall identification methodology.

The control variables considered (i.e. the time invariant inputs $\mathbf{w}$) were the flow rate, the inlet concentration, and the sample volume, that is the design vector will be $\boldsymbol{\varphi} \equiv \mathbf{w} = [Q, C_{in}, V]$. The measured variable was the concentration of the analyte at the outlet, $\mathbf{y} = C_{out}(t)$. To better represent a real system, we assumed *known* measurement error of 2% in the in-silico experiments, with this error being Gaussian distributed;



(a) Non re-parameterised bi-Langmuir isotherm (b) Re-parameterised bi-Langmuir isotherm

Fig. 5. Comparison between the sensitivities of the model output with respect to the parameters of the (a) non re-parameterised and (b) re-parameterised bi-Langmuir isotherm model. (The simulations were performed at inlet concentration C_{in} = 5 mg/ml, flow rate Q = 0.7 ml/min, sample volume V = 0.5 ml.)

Table 2
Levels of the control variables for use in factorial experiments.

Control variable	Symbol	Lower bound	Upper bound	Unit
Flow rate	Q	0.5	1	ml/min
Inlet concentration	C_{in}	1	5	mg/ml
Sample volume	V	0.1	1	ml

Table 3
Control variables in diluted conditions and the corresponding estimated values of the Henry coefficient (Q = 1 mg/ml).

	C_{in} (mg/ml)	V (ml)	H
Exp. 1	0.1	0.1	6.96
Exp. 2	0.5	0.01	6.94
Exp. 3	0.1	0.001	6.95

i.e. this error was added to the output of each in-silico experimental measurement, i.e. to $C_{out}(t)$. Note that in our in-silico experiments we did not account for any systematic error.

Cases A and B considered a bi-level full factorial design since we can control all three variables (flow rate, inlet concentration, and sample volume). In cases where the process involves more than three control variables, a fractional factorial design could be employed to maintain a low experimental effort. The bounds for each variable in the case study are summarised in Table 2.

The selection of the flow rate variation as a control variable is based on considerations of robustness and maximum pressure drop to ensure stability. Similarly, the choice of inlet concentration must be carefully determined to optimise the experimental conditions. Setting the concentration below a lower limit does not guarantee equilibrium in overloaded conditions (Schmidt-Traub et al., 2020). Overloaded conditions are important, because only in this region is the effect of saturation capacity appreciable. On the other hand, exceeding an upper limit gets too close to the peptide solubility in the mobile phase, which is 6 mg/ml for this case study (Andrzejewska et al., 2009). Additionally, the selection of an appropriate sample volume is crucial for obtaining meaningful data. The sample volume must be kept above a lower limit to ensure that the equilibrium condition remains non-linear, while exceeding the higher limit can result in wide plateaus in the chromatogram which is then not informative and will also cause problems in the parameter estimation as well as increase the experimental cost. With appropriate bounds on the control variables, reliable and informative experiments can be conducted that will simplify and facilitate the subsequent parameter estimation.

Note that the experimental design space was amended in the case of MBDoE-designed experiments. The upper and lower limits of the control variables took the values of Q = 0.1–2 ml/min, C_{in} = 1–6 mg/ml, and V = 0.1–1 ml. Note that the bounds for the flow rate and concentration were expanded while the limits of sample volume remained the same. MBDoE can recommend such amendment in the values of the control variables, allowing for experiments, in less experimentally robust (i.e. higher flow rate) and more resource-expensive (higher concentration) design regions, to retrieve more information about the system where the model is the most sensitive.

3.3. Numerical solution

The in-silico experiments, parameter estimation, and MBDoE were all performed in gPROMS®ModelBuilder (Barton and Pantelides, 1994; Process Systems Enterprise (PSE), 2022). The simulation grid was discretised according to the Orthogonal Collocation Finite Elements Method (OCFEM) of 2nd order in 800 elements. The selected optimisation solver was the Non-Linear Programming Sequential Quadratic Programming (NLPSQP) solver.

4. Results and discussion

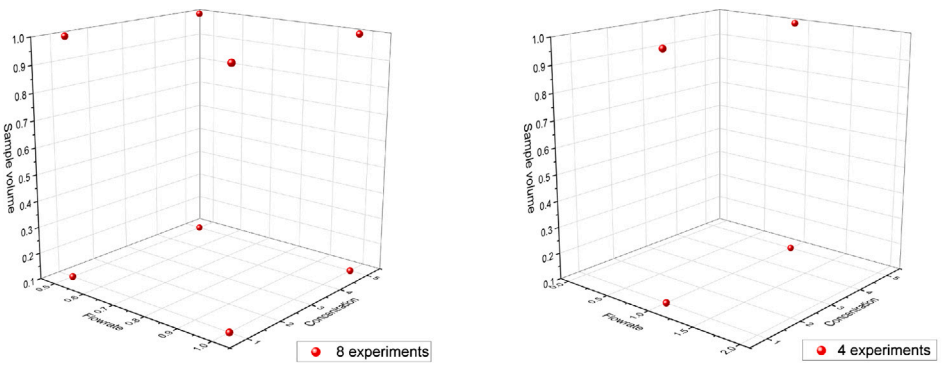
In this section, we are going to present the results on the case studies we investigated towards the identification of three isotherm models. Each case study leveraged alternative strategies and our aim was to clarify the merits of each of the strategies.

4.1. Step 0: Henry coefficient estimation

The first step of the methodology, which is in fact a pre-step, is to estimate the Henry coefficient, as this will be used as the initial estimate for K_1. For this, we conducted experiments under dilute conditions. The flow rate was kept constant throughout the experiments at Q = 1 ml/min. According to the proposed methodology, three experiments were initially performed. The experiments and the resulting Henry coefficient are summarised in Table 3. The value of the Henry coefficient is almost identical, therefore, as previously described, the mean average of the values was used in the following as the initial estimate for K_1.

4.2. Step 1: Goodness of fit - Preliminary model screening

In the first main step of the methodology, initial experiments were designed to enable the preliminary estimation of the parameters for the proposed isotherm models. According to the methodology flowchart in Fig. 3, since the number of control variables did not exceed three, experiments should be designed based on a full factorial design to identify the initial experiments for Cases A & B. To explore the impact of the number of initial experiments, a fractional factorial design was



(a) The full factorial used in Cases A & B (b) The fractional factorial used in Case C

Fig. 6. The initial experiments used in Step 1 of the proposed methodology.

Table 4
Preliminary estimates and 95% confidence intervals after an 8-factorial screening for Case studies A and B.

Model	Equation	χ^2	χ^2_c	Parameter estimates ± 95% confidence interval	Action
Langmuir	(24)	52,883	211.1	–	Reject
bi-Langmuir	(25)	179.67	209.04	$K_1 = 3.89 \pm 0.019$ $K_2 = 2.99 \pm 0.015$ $\theta_1 = 6.01 \pm 0.063$ $q_{sat,2} = 2.32 \pm 0.041$	Accept
tri-Langmuir	(26)	184.28	206.88	$K_1 = 3.81 \pm 0.135$ $K_2 = 1.78 \pm 7.32$ $K_3 = 1.31 \pm 7.20$ $\theta_1 = 6.28 \pm 0.392$ $q_{sat,2} = 1.13 \pm 5.83$ $q_{sat,3} = 1.46 \pm 5.24$	Accept

selected for Case C. Note that, theoretically, even a single experiment could be used to estimate the parameters for a given chromatographic model as long as the number of measurements in the experiment is greater than the number of model parameters. However, the use of a single experiment is meaningless, as it would yield very large uncertainties in the estimated model parameters. The model output, the analyte concentration in the mobile phase C_{out}, is expected to be the most sensitive to the variation of sample volume and concentration, and thus these two factors were selected for the fractional factorial design in Case C. The values of the control variables of the initially designed experiments are given in Fig. 6 for all three cases.

After performing the initial in-silico experiments and evaluating the resulting chromatograms, three isotherm models were considered as potential candidate models: the simple, bi- and tri-Langmuir isotherms. The selection of Langmuir-type isotherms was based on the shapes of the chromatographic peaks, which featured a tail, implying that the isotherm describing the equilibrium was of convex shape (Schmidt-Traub et al., 2020; Carta and Jungbauer, 2020).

Parameters were then estimated for the three candidate isotherm models using the initial experiments (Fig. 6) and the goodness of fit was evaluated based on the χ^2-test. For each isotherm model, the parameter vector $\boldsymbol{\theta}$ consists of the following parameters:

- $\boldsymbol{\theta}_{\text{Langmuir}} = [K_1, \theta_1]$
- $\boldsymbol{\theta}_{\text{bi-Langmuir}} = [K_1, \theta_1, K_2, q_{sat,2}]$
- $\boldsymbol{\theta}_{\text{tri-Langmuir}} = [K_1, \theta_1, K_2, q_{sat,2}, K_3, q_{sat,3}]$

Table 4 reports the χ^2 and the corresponding critical values χ^2_c after estimating parameters using the full or 8-factorial (Cases A & B), while Table 5 reports the values after the fractional or 4-factorial (Case C). From Tables 4 and 5, it is evident that the simple Langmuir isotherm fails to fit the experimental data by a big margin (χ^2 is much larger than χ^2_c). The contribution of each experiment to the χ^2 value of the Langmuir isotherm is provided in the supplementary material, which shows that the fit of the simulated peaks for experiment 1 and 3 of the 8-factorial explains the exceptionally high value of the χ^2 for the Langmuir isotherm. Replicate experiments could also be considered to ensure reproducibility and to rule out the case of outlier experiments that could induce the χ^2-test to fail. However, in our case, the Langmuir model failed to capture the trend of some of the experiments, as shown in Fig. 2b of the supplementary material. Hence, the lack of fit is not due to the measurement variance but rather because of a mismatch between the process underlying physics and the candidate model.

The bi- and tri-Langmuir isotherm models perform well in terms of goodness of fit, thus passing the test. However, we can observe that the 95% confidence intervals of the tri-Langmuir parameters indicate that the isotherm model may not be practically identifiable (Table 5). This is because the confidence intervals of K_2, K_3, $q_{sat,2}$ and $q_{sat,3}$ are very large compared to the optimal parameter value. These large values can be attributed to the fact that these parameters are strongly correlated. Note also that the χ^2_c values decrease with an increasing number of model parameters since that reduces the number of degrees of freedom for a given number of measurements.

4.3. Step 2: Practical model identifiability

The next step is the practical model identifiability test. Practical identifiability is only considered for the models that passed the χ^2 test in the preliminary screening, that is for the models that fit the preliminary experiments adequately.

As outlined in the methodology, to test the practical model identifiability, that is to evaluate whether it is possible to uniquely estimate parameters under an experimental budget, one must perform *simulated* experiments, designed based on the initial parameter estimates of the remaining candidate models. According to the practical identifiability methodology (Fig. 4), we set an experimental budget of 27 experiments

Table 5
Preliminary estimates and 95% confidence intervals after a 4-factorial screening for Case study C.

Model	Equation	χ^2	χ^2_c	Parameter estimates ± 95% confidence interval	Action
Langmuir	(24)	16,553	75.63	–	Reject
bi-Langmuir	(25)	37.39	73.32	$K_1 = 3.90 \pm 0.037$ $K_2 = 2.99 \pm 0.034$ $\theta_1 = 5.99 \pm 0.125$ $q_{sat,2} = 2.28 \pm 0.076$	Accept
tri-Langmuir	(26)	37.33	70.1	$K_1 = 3.90 \pm 0.212$ $K_2 = 1.79 \pm 1074$ $K_3 = 1.21 \pm 1074$ $\theta_1 = 6.00 \pm 0.455$ $q_{sat,2} = 1.29 \pm 828$ $q_{sat,3} = 1.00 \pm 827$	Accept

Table 6
Eigenvalues of the expected FIM under an experimental budget of 27-factorial experiments for different sets of preliminary estimates.

Estimate set	λ_1	λ_2	λ_3	λ_4
Case C	6.54×10^8	2.16×10^7	1.08×10^5	2.86×10^6
Cases A & B	6.54×10^8	2.17×10^7	1.01×10^5	2.75×10^6

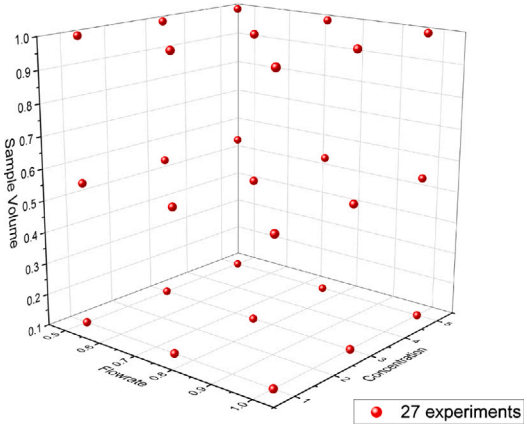


Fig. 7. 27-factorial design used in the practical identifiability step (Step 2).

that we chose to design with a full factorial altogether and perform the resulting (in-silico) experiments. While there were two candidate models after the initial screening, the bi-Langmuir and tri-Langmuir isotherms (after both 8-factorial (Cases A & B) and 4-factorial screening (Case C)), we only designed the 27-factorial once for each model, based only on the estimates acquired from the 8-factorial (Cases A & B), since the two sets of estimates (Cases A & B and Case C) of the two isotherm models were very similar so that we did not have to repeat another computationally expensive parameter estimation against 27 experiments. To assess whether we could use the same experiments produced by different preliminary estimates (i.e. that of Cases A & B and that of Case C) for practical identifiability purposes of all three case studies, we evaluated the expected FIM at the conditions dictated by the 27-factorial design using both sets of preliminary estimates. Note that that was a less computationally expensive procedure than having to perform the parameter estimation procedure for a new set of 27 experiments. From the expected FIM, we found that the eigenvalues and determinant of the expected FIM for the different preliminary estimates were highly similar; thus, we could justify using a single parameter estimation and its produced statistics to assess practical identifiability for all the case studies. For reference, in Table 6 we report the eigenvalues acquired for the different sets of estimates of the bi-Langmuir isotherm.

The conditions of the 27-factorial design are reported in Fig. 7. After performing the corresponding in-silico experiments and estimating the parameters, we proceed to the t-test. The test results are summarised in Table 7. The bi-Langmuir isotherm passed both the χ^2-test and the t-test, since all the parameter t-values were greater than the t_{ref}-value. However, the tri-Langmuir isotherm failed the χ^2-test, and the t-value criterion was satisfied only for two of the six parameters of the model. Since there is at least one identifiable model left from those initially proposed, in this case there is no need to amend the experimental budget, i.e. no need for further experiments. Moreover, as discussed in the Methodology section, if only the t-test failed, sequential MBDoE should be employed to design more informative experiments. Ultimately, in Cases A, B and C, the bi-Langmuir isotherm is the only model left (an expected outcome, since it was the "true" model used for the in-silico experiments).

4.4. Step 3: Parameter precision refinement with MBDoE

The final step of the methodology involves improving the precision of the parameters of the chosen model, here the bi-Langmuir isotherm model. So far, Cases A and B have been identical. In the following, we will consider two different approaches to refine the final parameter precision, Case A will use factorial design and Case B will use MBDoE. MBDoE will also be used to improve the precision after the 4-factorial screening for Case C. Table 8 summarises the final parameter values, 95% confidence intervals, t-values, and the total number of experiments required for each case, and Fig. 8 shows the values of the three control variables for each of the experiments (factorial and precision) in a 3-dimensional representation.

In Case A, the initial 8-factorial experiments (red squares in Fig. 8(a)) were first augmented by a 4-factorial set of experiments, adding an additional intermediate level in the sample volume control factor (blue dots in Fig. 8(a)), and a new parameter estimation took place. Although the parameters satisfied the t-test, the Monte Carlo simulation showed that chromatograms were still uncertain, and thus the precision requirements were still not met. Hence, another round of 4-factorial experiments was designed, this time adding an additional level of four experiments for the inlet concentration (green triangles in Fig. 8(a)). Note that each addition was added to the original 8-factorial design, that is the intermediate level of sample volume experiments remained bi-level after the addition of the new level for the factor of concentration. Switching between control variables to design fractional factorial experiments was done to explore more of the design space. After 16 total experiments, we could terminate the procedure.

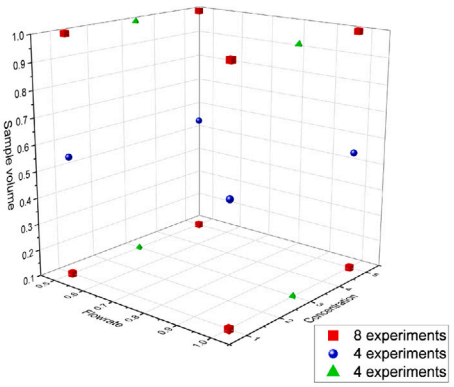
In Case B, instead of selecting experiments based on factorial design to improve the precision of the parameters, we employed MBDoE. Specifically, using the initial parameter estimates of the model obtained after Step 2, we proceeded to design two optimal experiments using the E-optimal criterion. This criterion was favoured over the D-optimal criterion because for the latter, regardless of the initial guesses, the expected FIM became singular after a few iterations. This behaviour could be attributed to parameter pairs being strongly correlated. Therefore, to assist in the decorrelation of the parameters,

Table 7
Statistics of the practical model identifiability test under an experimental budget of 27 experiments for all three case studies.

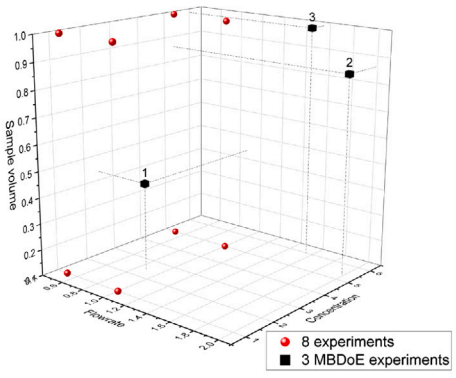
Model	χ^2/χ_c^2	K_1	θ_1	K_2	$q_{sat,2}$	K_3	$q_{sat,3}$	Action
bi-Langmuir	332.2/371.2	3.89 ± 0.009	6.01 ± 0.036	3.00 ± 0.007	2.32 ± 0.022	–	–	Accept
t-value ($t_{ref} = 1.65$)		418	166	451	105	–	–	
tri-Langmuir	1357.8/534.2	3.78 ± 0.082	6.36 ± 0.251	1.71 ± 51.8	1.33 ± 45.5	1.29 ± 51.8	1.32 ± 45.1	Reject
t-value ($t_{ref} = 1.65$)		45.7	25.3	0.03	0.02	0.03	0.03	

Table 8
Final optimal parameter estimates and corresponding statistics for all cases.

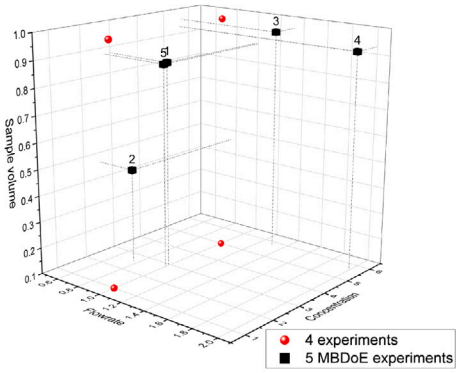
	χ^2/χ_c	$K_1 \pm 95\%$ conf. int.	$\theta_1 \pm 95\%$ conf. int.	$K_2 \pm 95\%$ conf. int.	$q_{sat,2\pm95\%}$ conf. int.	Total No. experiments
Case A	411.9/475.1	3.88 ± 0.008	6.06 ± 0.029	3.01 ± 0.006	2.35 ± 0.018	16
t-value ($t_{ref} = 1.65$)		504	211	485	134	MBDoE: No
Case B	227.2/275.0	3.89 ± 0.012	6.03 ± 0.039	3.00 ± 0.009	2.33 ± 0.027	11
t-value ($t_{ref} = 1.65$)		331	155	323	86	MBDoE: Yes
Case C	119.3/199.2	3.89 ± 0.008	6.04 ± 0.02	3.01 ± 0.007	2.33 ± 0.024	9
t-value ($t_{ref} = 1.65$)		495	301	453	94	MBDoE: Yes
Real values		3.88	6.06	3.01	2.35	



(a) 8-factorial and two rounds of 4-factorial experiments in Case Study A



(b) 8-factorial experiments and 3 optimally designed experiments in Case Study B



(c) 4-factorial experiments and 5 optimally designed experiments in Case Study C

Fig. 8. The experimental conditions of the factorial and the optimally designed experiments for all cases.

we selected the E-optimal criterion. After the two first iterations of MBDoE experiments, and since the correlation between the parameters was lower, a third, final experiment was designed using the A-Optimal criterion. This was selected as a more numerically stable alternative to the D-optimal criterion, since it minimises the trace of the variance–covariance matrix. We refer the reader to the article of Franceschini and Macchietto (2008b) where the merits of each optimal criterion are outlined thoroughly. The confidence ellipsoids between the K_1 and K_2 parameters after each MBDoE iteration are presented in Fig. 9. The first two iterations of E-optimal experiments shrinked the major axis of the confidence ellipsoid, while the minor axis remained almost identical, an outcome that reveals that the parameter correlation was reduced. The third A-optimal experiment succeeded in shrinking the confidence ellipsoid significantly. The procedure was terminated when the parameter estimates met the precision requirements.

In Case C, using the estimates from the initial fractional 4-factorial design, precision was improved with a total of five MBDoE-designed experiments that were executed sequentially as for Case B. The first four optimal experiments were designed employing the D-Optimal criterion. The fourth D-optimal experiment improved the precision of the parameters marginally or not at all; this can be seen in Fig. 10, where the confidence ellipsoids between the K_1 and K_2 parameters after the 3rd and 4th MBDoE iterations are almost identical. Therefore, the 5th and final experiment was designed using the E-optimal criterion. The

Table 9
Main decisions taken in each step of isotherm model identification for all three case studies.

	Factorial screening	Models accepted/proposed	Practical model identification	Number of models identified	Parameter precision
Case Study A	8-factorial	2/3	27-factorial	1	2 rounds 4-factorial
Case Study B	8-factorial	2/3	27-factorial	1	2 E-Optimal 1 A-Optimal
Case Study C	4-factorial	2/3	27-factorial	1	4 D-Optimal 1 E-Optimal

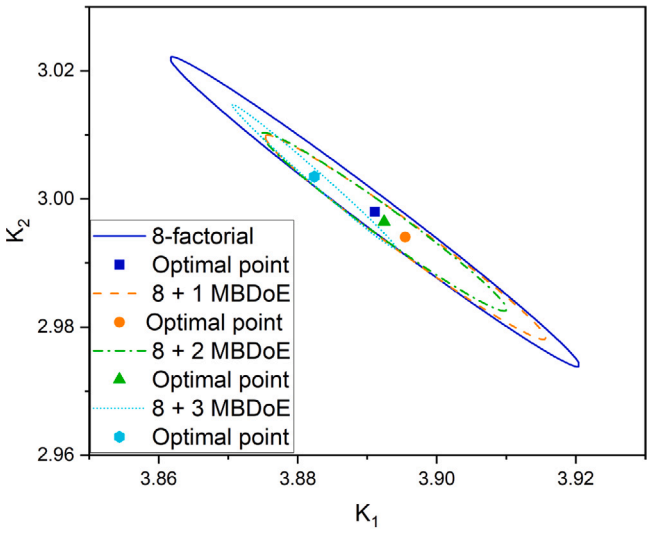


Fig. 9. Confidence ellipsoid between the K_1 and K_2 parameters after each MBDoE iteration in Case study B. The optimal points relate to the optimum following each additional MBDoE experiment.

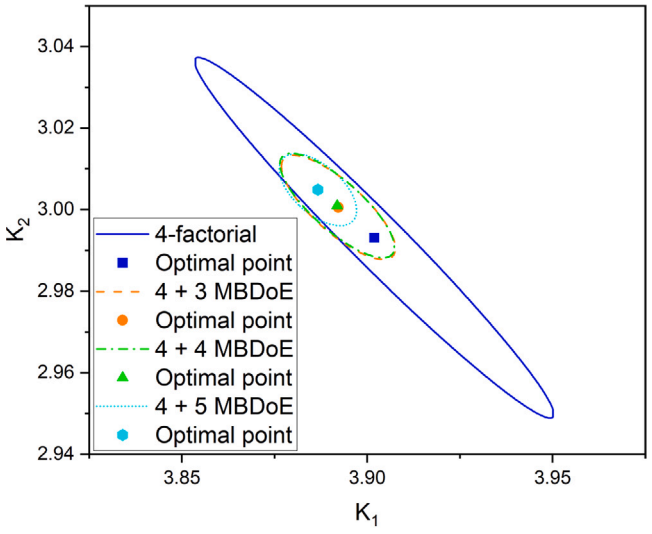


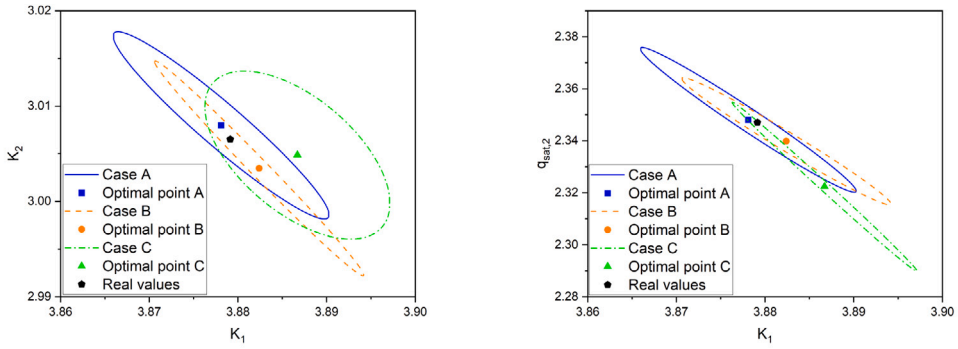
Fig. 10. The confidence ellipsoid between the K_1 and K_2 parameters after the 3rd, 4th, and 5th MBDoE iteration for Case C.

campaign was then terminated since the precision requirements were met, that is the t-test was satisfied and the Monte Carlo simulations produced chromatograms that are almost identical with no deviation on the time axis. Table 9 summarises the decisions taken from the preliminary screening to the precision refinement step for all three case studies.

From Table 8, it can be seen that the 16-factorial of Case A (8-factorial followed by 8-factorial) and the initial 8- or 4-factorials followed by MBDoE-designed experiments of Cases B and C yield parameters of equivalent precision. The 16-factorial in Case A led to the estimation of parameters that are in fact the closest to the real values from the in-silico case. Note that at no point in the procedure were these values actually used, and it is only because we used in-silico rather than real experiments that we are able to make this comparison. Although the 16-factorial yielded a good accuracy for this particular case study, this may not always be the case, because the factorial method is rather unpredictable as it does not rely on any insight into the process but just randomly assigns control variable values to each experiment. The merits of the factorial experiments lie in their simplicity, as no computational effort is required, but factorial experiments can potentially impede the parameter estimation process if the control variable values are unfavourable. For instance, values of low concentration and sample volume tend to produce non-informative experiments. Designing experiments with MBDoE, on the other hand, can assist in making more informed decisions on which experiments should be designed next and can focus on regions where the uncertainty is largest or where most information can be gained. Hence, parameters usually require fewer experiments to reach a certain level of confidence with MBDoE in relation to conventional DoE approaches.

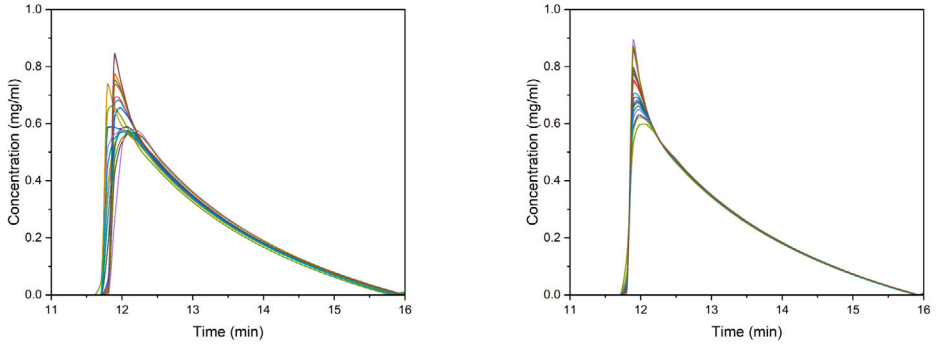
Although MBDoE involves some computational cost, this approach can potentially save significant material cost and experimental effort. In this work, through the application of MBDoE-based strategies, we refined the parametric precision (Cases B and C), thereby minimising the experimental effort necessary, relative to conventional DoE methodologies. Specifically, Cases B and C required 9 and 11 experiments, respectively, whilst Case A required 16 total experiments to reach a satisfactory precision. Additionally, Case C, which employed more MBDoE experiments (5 vs. 3 in Case B), required fewer overall experiments. Furthermore, the underlying mathematical rigour of experiments designed by MBDoE renders the experiments more relevant and contribute to making more informed decisions in the estimation procedure. Consequently, isotherm model identification with MBDoE should be favoured over methods that employ only factorial designs.

Fig. 11 shows the confidence ellipsoid formed between the final parameters (a) K_1 and K_2 and (b) K_1 and $q_{sat,2}$ for Cases A to C. These confidence ellipsoids are representative of the correlation type observed between the parameters of the bi-Langmuir isotherm model. The correlation appeared to be bi-modal between parameters, namely pairs of parameters with high or moderate correlation throughout the model identification procedure. Heat maps of the parameter correlations for all the case studies are given in the supplementary material. As observed from the confidence ellipsoid plots, the level of correlation can become very high, i.e. the major axes of the ellipsoids are significantly larger than the minor axes (Fig. 11(b), Case C), albeit there is no significant effect on the model identifiability, since the confidence ellipsoids are very small — i.e. the ellipsoid does not tend to become "spaghetti-like". Moreover, the magnitude of the confidence ellipsoids is *per se* not representative of the model output behaviour, because of parametric uncertainty. Therefore, uncertainty analyses will be conducted next to quantify the model output uncertainty owing to the uncertainty of its parameters.



(a) 95% confidence ellipsoid between parameters K_1 and K_2 (b) 95% confidence ellipsoid between parameters K_1 and $q_{sat,2}$

Fig. 11. 95% confidence ellipsoid between two pairs of parameters of the bi-Langmuir isotherm model for Cases A–C. The real values are the ones used in the in-silico experiments.



(a) After 4-factorial experiments (b) After 4-factorial and 5 MBDoE experiments

Fig. 12. Analysis of uncertainty in bi-Langmuir isotherm model parameter values for Case C (a) after the initial factorial screening (Step 1) and (b) after improving parameters precision (Step 3) ($Q = 0.5$ ml/min, inlet concentration $C_{in} = 1$ mg/ml, sample volume $V = 0.1$ ml).

4.5. Analysis of parametric uncertainty

The impact of the uncertainty in the bi-Langmuir isotherm model parameters on the model output, the analyte concentration at the outlet, was investigated for Case C by simulating 20 elution profiles where the parameter values were sampled from a normal distribution with a mean and a variance that emerged from the parameter estimation, thus enabling the generation of 20 different scenarios. Fig. 12(a) shows the uncertainty analysis with the estimates acquired from the initial 4-factorial screening (Step 1), while Fig. 12(b) shows the uncertainty analysis conducted after improving the precision of the initial estimates with 5 MBDoE experiments of the same case study (Step 3). The figure shows that, before refining parameters, simulations with uncertain parameter values can result in slightly different chromatograms with some level of deviation on the time axis. On the other hand, when the isotherm parameters were re-estimated after the optimally designed MBDoE experiments, the stochastically simulated chromatograms were almost identical. The aim of the Monte-Carlo simulations was to quantify the accuracy of the model estimation, i.e. how much was the offset on the time axis between the 20 stochastically generated chromatograms for the given parametric uncertainty. Minimising the uncertainty of the chromatogram can contribute to reducing the uncertainty in component purities for a separation of two or more components. Further analysis of uncertainty is provided in the supplementary material.

4.6. Analysis of model validation

To further analyse the accuracy of the model validation, we have also investigated the validity of the final parameters of each case study by evaluating the χ^2 test against eight experiments that were not used during parameter estimation. These experiments were at conditions at the bounds of the experimental design space and are depicted in Fig. 13. Note that by the term experimental design space, we refer to the expanded experimental design space that was used for MBDoE, which also includes potential linear regions of the adsorption isotherms as well as non-linear ones. Linear regions are areas of the design space where adsorption exhibits a linear behaviour and can best described by a linear isotherm. The results of the χ^2 test are presented in Table 10. Although eight experiments were initially selected, the experiment at $Q = 1$ ml/min, $C_{in} = 0.1$ mg/ml and $V = 0.1$ ml produced a very sharp peak yielding very few and noisy measurement points, and thus it was excluded from the validation procedure. According to the χ^2 values of each validation case (Table 10), the bi-Langmuir isotherm with parameters estimated from all three case studies fits the experimental data adequately. However, the bi-Langmuir of Case C has a χ^2 value close to the critical value. This can be explained due to the slight underfitting of the isotherm against experiments at low concentrations. This underfitting should not be of concern, because the isotherm passes the χ^2-test, and because the main purpose of the proposed methodology is to estimate parameters at preparative conditions, that is, for higher concentrations and sample volumes.

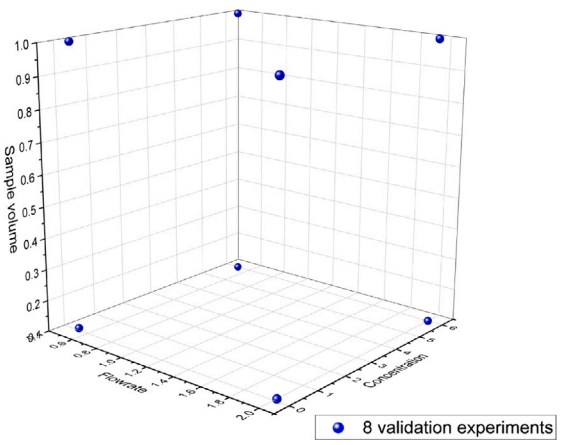


Fig. 13. Experiments selected at the bounds of the experimental design space used for model validation (Case C, bi-Langmuir isotherm model).

Table 10
The χ^2 test results of the model validation for the three case studies.

	χ^2/χ_c	Number of experiments	Action
Case study A bi-Langmuir	171.72/198.16	7	Accept
Case study B bi-Langmuir	171.49/198.16	7	Accept
Case study C bi-Langmuir	195.08/198.16	7	Accept

5. Conclusions

This work has proposed a methodology for isotherm model identification and model parameter estimation for preparative chromatography based on both factorial design and Model Based Design of Experiments (MBDoE). The step-by-step methodology includes the estimation of the Henry coefficient, initial screening of proposed models, evaluation of practical model identifiability, and final parameter value refinement. Future work will also assess the potential of incorporating MBDoE for model discrimination when there are two or more identifiable models in the system under investigation. An in-silico case study was used to illustrate the methodology. It was shown that our methodology is able to quickly and accurately identify the most appropriate model and the associated parameters and, more importantly, rigorously investigate their validity and accuracy. It was also shown that, compared to factorial design, MBDoE provides more insight, reduces material costs and experimental effort, and ensures a more robust parameter estimation.

CRediT authorship contribution statement

Konstantinos Katsoulas: Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Federico Galvanin:** Writing – review & editing, Methodology. **Luca Mazzei:** Writing – review & editing, Supervision. **Maximilian Besenhard:** Writing – review & editing, Supervision. **Eva Sorensen:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

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

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compchemeng.2025.109021.

Data availability

Data will be made available on request.

References

Åsberg, D., Leśko, M., Enmark, M., Samuelsson, J., Kaczmarski, K., Fornstedt, T., 2013. Fast estimation of adsorption isotherm parameters in gradient elution preparative liquid chromatography. I: The single component case. J. Chromatogr. A 1299, 64–70.

Ahmad, T., Guiochon, G., 2006. Effect of temperature on the adsorption behavior of tryptophan in reversed-phase liquid chromatography. J. Chromatogr. A 1129 (2), 174–188.

Ahmad, T., Guiochon, G., 2007. Numerical determination of the adsorption isotherms of tryptophan at different temperatures and mobile phase compositions. J. Chromatogr. A 1142 (2), 148–163.

Andersson, M., Streb, M., Ko, J.Y., Löfqvist Klass, V., Klett, M., Ekström, H., Johansson, M., Lindbergh, G., 2022. Parametrization of physics-based battery models from input–output data: A review of methodology and current research. J. Power Sources 521, 230859.

Andrzejewska, A., Gritti, F., Guiochon, G., 2009. Investigation of the adsorption mechanism of a peptide in reversed phase liquid chromatography, from pH controlled and uncontrolled solutions. J. Chromatogr. A 1216 (18), 3992–4004.

Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press Inc., New York.

Barton, P.I., Pantelides, C.C., 1994. Modeling of combined discrete/continuous processes. AIChE J. 40 (6), 966–979.

Barz, T., López C., D.C., Cruz Bournazou, M.N., Körkel, S., Walter, S.F., 2016. Real-time adaptive input design for the determination of competitive adsorption isotherms in liquid chromatography. Comput. Chem. Eng. 94, 104–116.

Bawa, S.G., Pankajakshan, A., Waldron, C., Cao, E., Galvanin, F., Gavriilidis, A., 2023. Rapid screening of kinetic models for methane total oxidation using an automated gas phase catalytic microreactor platform. Chemistry-Methods 3 (1).

Besenhard, M.O., Tsatse, A., Mazzei, L., Sorensen, E., 2021. Recent advances in modelling and control of liquid chromatography. Curr. Opin. Chem. Eng. 32, 100685.

Briskot, T., Stückler, F., Wittkopp, F., Williams, C., Yang, J., Konrad, S., Doninger, K., Griesbach, J., Bennecke, M., Hepbildikler, S., Hubbuch, J., 2019. Prediction uncertainty assessment of chromatography models using Bayesian inference. J. Chromatogr. A 1587, 101–110.

Carta, G., Jungbauer, A., 2020. Protein Chromatography. John Wiley & Sons, Incorporated, Newark.

Casey, F.P., Baird, D., Feng, Q., Gutenkunst, R.N., Waterfall, J.J., Myers, C.R., Brown, K.S., Cerione, R.A., Sethna, J.P., 2007. Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model. IET Syst. Biol. 1 (3), 190–202.

De Luca, C., Felletti, S., Macis, M., Cabri, W., Lievore, G., Chenet, T., Pasti, L., Morbidelli, M., Cavazzini, A., Catani, M., Ricci, A., 2020. Modeling the nonlinear behavior of a bioactive peptide in reversed-phase gradient elution chromatography. J. Chromatogr. A 1616, 460789.

Debrus, B., Lebrun, P., Ceccato, A., Caliaro, G., Rozet, E., Nistor, I., Oprean, R., Rupérez, F.J., Barbas, C., Boulanger, B., Hubert, P., 2011. Application of new methodologies based on design of experiments, independent component analysis and design space for robust optimization in liquid chromatography. Anal. Chim. Acta 691 (1–2), 33–42.

Dette, H., Melas, V.B., Shpilev, P., 2012. T-optimal designs for discrimination between two polynomial models. Ann. Statist. (1), 188–205.

Espie, D.M., Macchietto, S., 1988. Nonlinear transformations for parameter estimation. Ind. Eng. Chem. Res. 27 (11), 2175–2179.

Felinger, A., Cavazzini, A., Guiochon, G., 2003. Numerical determination of the competitive isotherm of enantiomers. J. Chromatogr. A 986 (2), 207–225.

Franceschini, G., Macchietto, S., 2007. Validation of a model for biodiesel production through model-based experiment design. Ind. Eng. Chem. Res. 46 (1), 220–232.

Franceschini, G., Macchietto, S., 2008a. Anti-correlation approach to model-based experiment design: Application to a biodiesel production process. Ind. Eng. Chem. Res. 47 (7), 2331–2348.
Franceschini, G., Macchietto, S., 2008b. Model-based design of experiments for parameter precision: State of the art. Chem. Eng. Sci. 63 (19), 4846–4872.
Franceschini, G., Macchietto, S., 2008c. Novel anticorrelation criteria for model-based experiment design: Theory and formulations. AIChE J. 54 (4), 1009–1024.
Gadkar, K.G., Gunawan, R., Doyle, F.J., 2005. Iterative approach to model identification of biological networks. BMC Bioinformatics 6 (1), 1–20.
Galvanin, F., Sankar, M., Cattaneo, S., Bethell, D., Dua, V., Hutchings, G.J., Gavriilidis, A., 2018. On the development of kinetic models for solvent-free benzyl alcohol oxidation over a gold-palladium catalyst. Chem. Eng. J. 342, 196–210.
Gétaz, D., Stroehlein, G., Butté, A., Morbidelli, M., 2013. Model-based design of peptide chromatographic purification processes. J. Chromatogr. A 1284, 69–79.
Grangeia, H.B., Silva, C., Simões, S.P., Reis, M.S., 2020. Quality by design in pharmaceutical manufacturing: A systematic review of current status, challenges and future perspectives. Eur. J. Pharmaceut. Biopharmaceut. 147, 19–37.
Gritti, F., Guiochon, G., 2004. Effect of the ionic strength of salts on retention and overloading behavior of ionizable compounds in reversed-phase liquid chromatography: I. XTerra-C18. J. Chromatogr. A 1033 (1), 43–55.
Gritti, F., Guiochon, G., 2005. Separation mechanism of nortriptyline and amytriptyline in RPLC. J. Chromatogr. A 1090, 39–57.
He, Q.L., Zhao, L., 2020. Bayesian inference based process design and uncertainty analysis of simulated moving bed chromatographic systems. Sep. Purif. Technol. 246, 116856, arXiv:1911.12133.
Heymann, W., Glaser, J., Schlegel, F., Johnson, W., Rolandi, P., von Lieres, E., 2023. Advanced error modeling and Bayesian uncertainty quantification in mechanistic liquid chromatography modeling. J. Chromatogr. A 1708, 464329.
James, F., Sepúlveda, M., Charton, F., Quinones, I., Guiochon, G., 1999. Determination of binary competitive equilibrium isotherms from the individual chromatographic band profiles. Chem. Eng. Sci. 54 (11), 1677–1696.
Joshi, M., Kremling, A., Seidel-Morgenstern, A., 2006. Model based statistical analysis of adsorption equilibrium data. Chem. Eng. Sci. 61 (23), 7805–7818.
Katsoulas, K., Tirapelle, M., Sorensen, E., Mazzei, L., 2023. On the apparent dispersion coefficient of the equilibrium dispersion model: An asymptotic analysis. J. Chromatogr. A 1708.
Kim, H., Kaczmarski, K., Guiochon, G., 2006. Thermodynamic analysis of the heterogenous binding sites of molecularly imprinted polymers. J. Chromatogr. A 1101, 136–152.
Kober, R., Schwaab, M., Steffani, E., Barbosa-Coutinho, E., Pinto, J.C., 2021. D-optimal experimental designs for precise parameter estimation of adsorption equilibrium models: initial concentration and solvent volume to adsorbent mass ratio as independent variables. Adsorption 27 (7), 1013–1022.
Kober, R., Schwaab, M., Steffani, E., Barbosa-Coutinho, E., Pinto, J.C., Alberton, A.L., 2019. D-optimal experimental designs for precise parameter estimation of adsorption equilibrium models. Chemometr. Intell. Lab. Syst. 192, 103823.
Latrous, L., 2022. Optimization and validation in liquid chromatography using design of experiments. Chem. Afr. 5 (3), 437–458.
Marković, D.D., Lekić, B.M., Rajaković-Ognjanović, V.N., Onjia, A.E., Rajaković, L.V., 2014. A new approach in regression analysis for modeling adsorption isotherms. Sci. World J. 2014.
Metzger, T., Didierjean, S., Maillet, D., 2004. Optimal experimental estimation of thermal dispersion coefficients in porous media. Int. J. Heat Mass Transfer 47 (14–16), 3341–3353.
Pankajakshan, A., Bawa, S.G., Gavriilidis, A., Galvanin, F., 2023. Autonomous kinetic model identification using optimal experimental design and retrospective data analysis: methane complete oxidation as a case study. React. Chem. Eng. 8 (12), 3000–3017.
Park, S., Kato, D., Gima, Z., Klein, R., Moura, S., 2018. Optimal experimental design for parameterization of an electrochemical lithium-ion battery model. J. Electrochem. Soc. 165 (7), A1309–A1323.
Process Systems Enterprise (PSE), 2022. gPROMS ModelBuilder. URL: www.psenterprise.com.
Quaglio, M., 2020. Novel Techniques for Kinetic Model Identification and Improvement (Ph.D. thesis). UCL (University College London).
Quiñones, I., Ford, J.C., Guiochon, G., 2000. Multisolute adsorption equilibria in a reversed-phase liquid chromatography system. Chem. Eng. Sci. 55 (5), 909–929.
Roca, L.S., Schoemaker, S.E., Pirok, B.W., Gargano, A.F., Schoenmakers, P.J., 2020. Accurate modelling of the retention behaviour of peptides in gradient-elution hydrophilic interaction liquid chromatography. J. Chromatogr. A 1614, 460650.
Schmidt-Traub, H., Schulte, M., Seidel-Morgenstern, A., 2020. Preparative Chromatography. John Wiley & Sons, Incorporated, Newark.
Schwaab, M., Pinto, J.C., 2008. Optimum reparameterization of power function models. Chem. Eng. Sci. 63 (18), 4631–4635.
Schwaab, M., Silva, F.M., Queipo, C.A., Barreto, A.G., Nele, M., Pinto, C., 2006. A new approach for sequential experimental design for model discrimination. Chem. Eng. Sci. 61, 5791–5806.
Seidel-Morgenstern, A., 2004. Experimental determination of single solute and competitive adsorption isotherms. J. Chromatogr. A 1037, 255–272.
Shahmohammadi, A., 2019. Model-Based Optimal Design of Experiments with Noninvertible Fisher Information Matrix (Ph.D. thesis). Queen's University (Canada).
Shekhawat, L.K., Godara, A., Kumar, V., Rathore, A.S., 2019. Design of experiments applications in bioprocessing: Chromatography process development using split design of experiments. Biotechnol. Prog. 35 (1), e2730.
Sugiyama, H., Yamamoto, Y., Suzuki, K., Yajima, T., Kawajiri, Y., 2024. Parameter estimation for reactive chromatography model by Bayesian inference and parallel sequential Monte Carlo. Chem. Eng. Res. Des. 203, 378–390.
Tolazzi, N., Steffani, E., Barbosa-Coutinho, E., Severo Júnior, J.B., Pinto, J.C., Schwaab, M., 2018. Adsorption equilibrium models: Computation of confidence regions of parameter estimates. Chem. Eng. Res. Des. 138, 144–157.
den Uijl, M.J., Schoenmakers, P.J., Pirok, B.W., van Bommel, M.R., 2021. Recent applications of retention modelling in liquid chromatography. J. Sep. Sci. 44 (1), 88–114.
Waldron, C., Pankajakshan, A., Quaglio, M., Cao, E., Galvanin, F., Gavriilidis, A., 2019. Closed-loop model-based design of experiments for kinetic model discrimination and parameter estimation: Benzoic acid esterification on a heterogeneous catalyst. Ind. Eng. Chem. Res. 58 (49), 22165–22177.
Yamamoto, Y., Yajima, T., Kawajiri, Y., 2021. Uncertainty quantification for chromatography model parameters by Bayesian inference using sequential Monte Carlo method. Chem. Eng. Res. Des. 175, 223–237.
Yan, F., Chu, Y., Zhang, K., Zhang, F., Bhandari, N., Ruan, G., Dai, Z., Liu, Y., Zhang, Z., Kan, A.T., Tomson, M.B., 2015. Determination of adsorption isotherm parameters with correlated errors by measurement error models. Chem. Eng. J. 281, 921–930.
Yu, L.X., 2008. Pharmaceutical quality by design: Product and process development, understanding, and control. Pharm. Res. 25 (4), 781–791.