Computers and Chemical Engineering 187 (2024) 108724

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

An optimization-free Fisher information driven approach for online design of experiments

Andrea Friso, Federico Galvanin *

Department of Chemical Engineering, University College London, Torrington Place, London, WC1E7JE, United Kingdom

ARTICLE INFO

Keywords:
Optimal design of experiments
Fisher information

ABSTRACT

Developing mathematical models used to elucidate reaction kinetics plays a crucial role in the design, control, and optimization of chemical processes. One of the most challenging tasks in kinetic model identification is the precise estimation of unknown kinetic model parameters. This challenge can be effectively addressed through the application of Model-Based Design of Experiments (MBDoE) techniques, which enable the design of experiments facilitating precise model parameter estimation with minimal runs and analytical resources. Nevertheless, MBDoE techniques rely on an optimization procedure that is susceptible to parametric uncertainty, making the design procedure computationally intensive and prone to issues of local optimality. MBDoE techniques are also employed in online procedures to expedite the identification of kinetic models in autonomous platforms. As a result, ensuring rapid convergence becomes imperative to mitigate numerical issues during operational processes. In this paper a Fisher Information Matrix Driven (FIMD) approach is introduced to tackle these challenges. The methodology integrates a sampling-based experimental design approach with experiment ranking based on FIM to select the most informative experiment at each iteration. The effectiveness of the proposed design methodology is examined and discussed via two different case studies of increasing complexity: a fed-batch reactor in which the fermentation of baker's yeast is carried out and a nucleophilic aromatic substitution in a flow reactor system.

1. Introduction

In chemical engineering, kinetic models are fundamental for quantitatively describing the progress of chemical reactions that occur in a reaction system (Bonvin et al., 2016). The availability of trustworthy models is an essential requirement for the design, simulation control, and optimization of dynamic systems. Systems can be described by dynamic, deterministic models that are usually expressed in the form of differential and algebraic equations (DAEs). When conventional model-building strategies are adopted (Asprey and Macchietto, 2002; Blau et al., 2008), reaction kinetics are modeled starting from a set of candidate models based on preliminary experimental observations and hypotheses on the reaction mechanism. The subsequent key step is to determine the best model for describing the system under analysis among the candidate sets. The standard sequential approach proposed by Asprey and Macchietto (2000) consists of three steps.

1. Preliminary analysis based on the practical identifiability and distinguishability of model structures (Zarrop, 1977). Practical identifiability of a kinetic model refers to the ability to uniquely determine the parameter values of the model using experimental data. Distinguishability refers to the ability to differentiate between two mathematical models based on their predicted outcomes (Asprey et al., 2000). It involves assessing whether the models produce significantly different results under the same conditions. This concept is crucial in scientific research for evaluating model validity and their ability to represent real-world phenomena.
2. Model-based design of experiments to discriminate among the candidate rival models that passed the first stage (MBDoE-MD, where MD stands for model discrimination) (Atkinson, 1970);
3. Model-based design of experiments to improve the precision of the identified parameters for the best model selected in Step 2 (MBDoE-PE, where PE stands for parameter estimation) (Atkinson et al., 2007).

If a model fails the identifiability test, this means that it is not possible to identify the complete set of model parameters from the available data. If no model passes the identifiability test in Step 1, it becomes necessary to propose a new set of candidate models or consider an alternative model reparametrization. In Step 2, a model structure is selected and the accuracy of the estimates is improved. To achieve this, it is necessary to conduct experiments to increase

* Corresponding author.
E-mail address: f.galvanin@ucl.ac.uk (F. Galvanin).

https://doi.org/10.1016/j.compchemeng.2024.108724
Received 15 January 2024; Received in revised form 3 May 2024; Accepted 7 May 2024
Available online 14 May 2024

the information available on model parameters. Experimental conditions are optimally computed according to the expected amount of information achievable for model discrimination (Step 2) or parameter estimation (Step 3) predicted by candidate models. In the current study, the focus is on the design of experiments to enhance parameter precision, assuming that the model structure, i.e. the set of relevant model equations, is known a priori. Before being able to apply MBDoE techniques, information on the system under analysis is needed which is provided by preliminary experiments. Several methods are available to obtain this initial information, for example performing a Design of Experiments (DoE) based on sampling techniques in the experimental design space to identify a set of candidate experiments that homogeneously explore the experimental design space, i.e. enabling space exploration (Montgomery, 2017). Obtaining preliminary data is crucial as it *(i)* allows to achieve a reliable estimate of model parameters to be used as initial guesses and *(ii)* allows to improve the precision of the parameter estimate from the early iterations of MBDoE. Given the importance of the initial information and consequently of preliminary experiments, various methodologies have been developed introducing indices to quantify the relative amount of information and to rank the information obtained from a set of preliminary experiments carried out on a system (Galvanin et al., 2016). In addition to these methodologies, different types of sampling methods have also been implemented which are able to provide a quantitative and qualitative representation of the experimental design space (Kusumo et al., 2020). In recent years, automated model identification platforms have been developed to identify the parameters of kinetic models online and speed up the identification procedure to reduce computational and experimental costs (Waldron et al., 2020; Pankajakshan et al., 2019). However, the parameter estimation procedure is complex and presents several challenges. The identification problem is often ill-conditioned and consequently, the objective function of MBDoE optimization may become difficult to compute (Quaglio et al., 2019). Another problem in automated platforms for model identification is the optimization procedure included in the MBDoE step. Often, optimization can be difficult for the following reasons:

- The system under analysis is structurally complex and requires candidate models characterized by a high number of parameters. Owing to the complexity given by the number of parameters of candidate models, the non-convexity of the optimization and numerical problems affecting the optimization procedure can lead to infeasible solutions.
- For systems characterized by a high number of state variables and design variables, the MBDoE procedure can become computationally very expensive, thus hindering the benefits of the online identification of kinetic models.

In this project, a new framework is proposed. The new approach is based on the evaluation of information from a set of candidate experiments using the expected Fisher Information Matrix (FIM). Differently from the traditional methodologies, this framework involves the exploration of the experimental design space and the evaluation of the expected information of each candidate experiment to determine the most informative experiment to run at each iteration. This eliminates the optimization step required in conventional MBDoE methods to identify the optimally informative experiment to be performed, making this framework computationally more efficient and suitable to be applied to systems characterized by multiple local optima. The paper is structured as follows: in Section 2 the standard framework and the proposed framework are presented, in Section 3 the two case studies and their results are described and in Section 4 the conclusion and future works are reported.

2. Proposed framework

Each candidate model in the MBDoE procedure can be formulated as a set of differential and algebraic equations (DAEs), where the measured variable $\mathbf{y}$, can be sampled at finite time instants.

$$\begin{aligned} &\mathbf{f}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{u}, \boldsymbol{\theta}, t) = 0 \\ &\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}, \boldsymbol{\nu}) \end{aligned} \tag{1}$$

In Eq. (1) $\mathbf{x} \in \mathbb{R}^{n_x}$, $\mathbf{u} \in \mathbb{R}^{n_u}$, $\mathbf{y} \in \mathbb{R}^{n_y}$, $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ and t are the vectors of the state variables representing the system, vector of manipulated variables, that in this work are defined as time-independent inputs, vector of the measured (output) variables, vector of the model parameters of the process, and time variable, respectively. The physical system is assumed to be subject to noise ($\boldsymbol{\nu} \in \mathbb{R}^{n_\nu}$). Each in-silico experimental measurement is obtained using the following expression:

$$\begin{aligned} &y = \hat{y} + \nu \\ &\nu \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{y}}) \end{aligned} \tag{2}$$

where ν is a normally distributed measurement error with zero mean and variance–covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{y}}$. The objective of the proposed methodology is to reduce the number of experiments required to estimate the model parameters in a statistically adequate way by performing experiments that lead to a precise estimation of the model parameters, i.e. minimum uncertainty. In this study three potential frameworks are analyzed for a precise identification of model parameters:

- DoE-based: this framework utilizes a systematic design of experiments to efficiently and informatively gather data. Employing a carefully planned sampling strategy, this approach aims to thoroughly explore the experimental design space, contributing to a comprehensive understanding and identification of the underlying model in the analyzed system.
- Standard MBDoE: this framework integrates mathematical models with experimental design to efficiently explore the input variable space, facilitating the identification and understanding of complex system models.
- FIM-driven: the proposed methodology combines sampling methods with a ranking of experiments based on the structure of the model and the experimental conditions analyzed.

For each framework, a schematic representation of the procedure is shown in Figs. 1, 2, and 3.

2.1. DoE approach

Standard DoE approaches (Box, 1980) are widely used. These DoE methods are based on sampling techniques to homogeneously explore the experimental design space to obtain the conditions for carrying out preliminary experiments and are particularly effective to acquire initial information. To enable exploration of the experimental design space, a screening experimental design as Latin Hypercube Sampling (LHS) has been used in this study. LHS is a powerful technique employed in DoE to efficiently explore the parameter space of the analyzed system (Ye, 1998; M. D. Mckay and Conover, 2000; Montgomery, 1984). This approach facilitates systematic sampling of the multidimensional parameter space, ensuring comprehensive exploration of the entire design space for a fixed number of experimental runs. Consequently, it emerges as a cost-effective strategy for experimentation. The Latin Hypercube Sampling (LHS) technique is utilized to generate a set of experimental points. Fig. 1 illustrates the framework implemented for the DoE.

The main steps of the DoE methodologies are:

1. Definition of the design space of the parameters: in this step, all the variables that can be manipulated and optimized during the

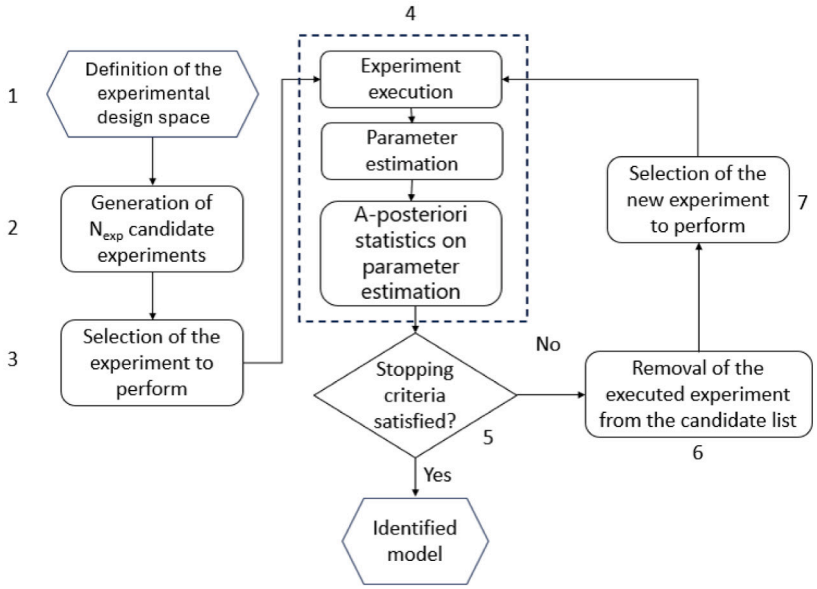

Fig. 1. Standard framework for DoE.

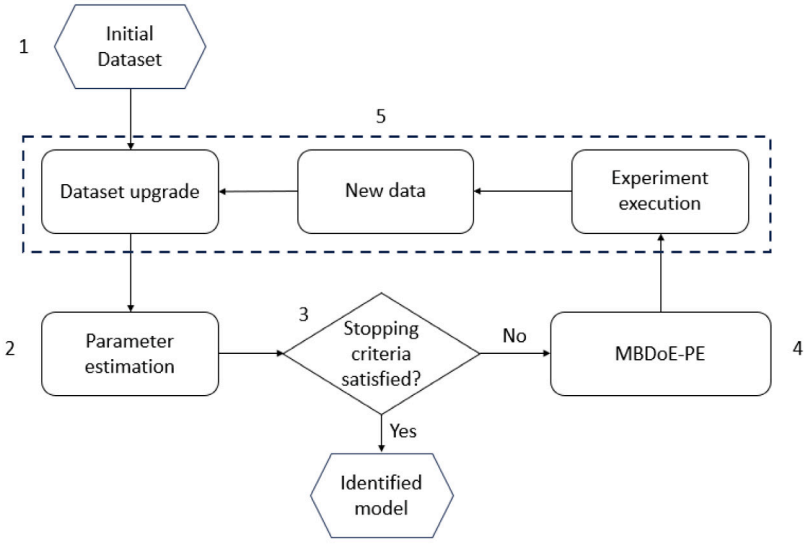

Fig. 2. Standard MBDoE framework.

process are collected in the experimental design vector (Eq. (3)).

$$\boldsymbol{\varphi} = [\mathbf{u}, \mathbf{x^0}, \tau_{exp}, \boldsymbol{t_{sp}}] \tag{3}$$

here $\mathbf{u}$ is the vector of the inputs which can be both time varying, $\mathbf{u(t)}$, and constant, $\mathbf{u}$, throughout the experiment, $\mathbf{x}^0$ is the vector of initial conditions, τ_{exp} is the experiment duration, and $\mathbf{t}_{sp}$ is the vector of sampling times (i.e. the set of time instants at which the measured responses are sampled), that is here assumed uniformly spaced. In this step, the upper/lower bounds on experimental conditions are defined. In this way the experimental design space within which to generate candidate experiments is created.

2. Generation of candidate experiments: LHS is used to generate candidate experiments to explore the space of operating conditions uniformly. In this work, the "maximin" criterion was used in which the minimum distance between candidate experiments is maximized (Johnson et al., 1990).
3. The first experiment is randomly selected.
4. Execution of the selected experiment and parameter estimation. This step is performed by maximizing the log-likelihood function reported in Eq. (4) (Bard, 1974).

$$\begin{aligned}\mathcal{L}(\boldsymbol{Y}|\boldsymbol{\theta}) = &-\frac{N}{2}\left[N_y \ln(2\pi) + \ln(\det(\boldsymbol{\Sigma}_y))\right] \\ &-\frac{1}{2}\sum_{i=1}^{N}\left[\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i(\boldsymbol{\theta})\right]^T \boldsymbol{\Sigma}_y^{-1}\left[\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i(\boldsymbol{\theta})\right]\end{aligned} \tag{4}$$

where $\hat{\boldsymbol{y}}_i$ represents the model prediction for the ith sample, $\boldsymbol{\Sigma}_y$ is the covariance of the measurement error for sample $\boldsymbol{y}$, N is the number of measurements and N_y is the number of measured output. The values of model parameters are those that optimize

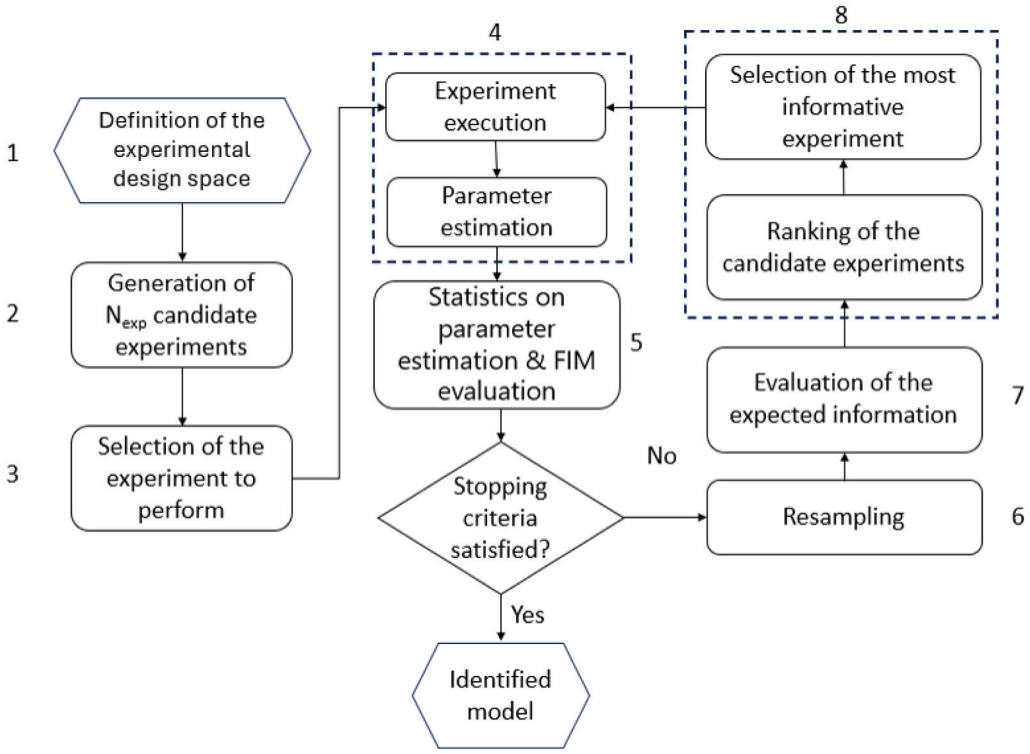

Fig. 3. Framework of the FIMD approach.

the expression in Eq. (4):

$$\hat{\boldsymbol{\theta}} = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}(\mathbf{Y}|\boldsymbol{\theta}) \tag{5}$$

Where $\mathbf{Y}$ is the $[N_y \times N]$ matrix of experimental measurements. The statistics on the parameters are calculated to assess whether the estimates are adequate to describe the system.

5. If the statistics on parameters are satisfied then the procedure can be stopped, otherwise a new experiment is conducted, but first, it is necessary to remove the conditions already tested from the set of candidate experiments (Steps 5-6-7 in Fig. 1).

The procedure described above and shown in Fig. 1 is an iterative procedure that requires a stopping criterion capable of interrupting the procedure when the values of the identified parameters are reliable. The stopping criteria used to stop the procedure can be the following:

1. Goodness of fit test on the model: this test is used to assess the fitting performance of the model under analysis and is based on a two-tailed χ^2 test (Draper and Smith, 2014). Under the hypothesis of the proposed model being exact, using Eq. (6) it is possible to evaluate the sum of normalized squared residuals associated with the fitting of the dataset:

$$\chi^2_Y = \sum_{i=1}^{N} \left[\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i(\hat{\boldsymbol{\theta}})\right]^T \boldsymbol{\Sigma}_y^{-1} \left[\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i(\hat{\boldsymbol{\theta}})\right] \tag{6}$$

The value obtained using Eq. (6) is compared with two reference values ($\chi^2_{5\%}$ and $\chi^2_{95\%}$) and if it lies between the two reference values the model is considered adequate. Whenever χ^2_Y is above the 95% percentile the model is falsified for underfitting. If χ^2_Y is below 5% percentile, the model is falsified for overfitting.

2. t-test on model parameters. The statistical quality of parameter estimates ($\hat{\boldsymbol{\theta}}$) can be assessed using a t-test with an adequate significance value (Walpole et al., 2016). The first step is to evaluate the t-values of all model parameters and compare them with the reference t-value:

$$\frac{\hat{\theta}_i}{t\left(\frac{1+\alpha}{2}\right)\sqrt{v_{\theta,ii}}} \geq t(\alpha) \quad \forall i = 1, \ldots, N_\theta \tag{7}$$

In Eq. (7) $t(\alpha)$ is the t-value obtained from a Student's t-distribution using the degrees of freedom of the system and the significance level α.

3. Experimental budget termination: once the maximum budget for conducting experiments has been reached, the experimental campaign must be stopped regardless of having identified the appropriate model to describe the system under analysis

In the case studies presented in Sections 3.1 and 3.2, the termination criterion (3) has been used to enable a comparative assessment of the efficacy of different experimental design approaches. Statistics for criteria 1 and 2 will still be calculated to assess the performance of the three proposed frameworks. This will enable the evaluation of the efficacy of both stopping criteria across various methodologies.

2.2. MBDoE approach

The procedure presented in Section 2.1 is commonly employed in preliminary studies of the system under analysis to initialize MBDoE frameworks. The information obtained is used as preliminary information to apply MBDoE. Fig. 2 illustrates a block diagram of the standard MBDoE framework.

The fundamental steps of the MBDoE methodology for parameter estimation are the following:

1. To perform a preliminary parameter estimation it is necessary to obtain preliminary knowledge on the system under analysis. This knowledge is usually provided by a preliminary set of data imported into the system.
2. Using the available model, a preliminary parameter estimation is conducted. This step is performed by maximizing the log-likelihood function (Eq. (4)).

3. Using the identified parameters a statistical analysis on the model is performed to assess the performance of the model. In this analysis, the uncertainty of the predicted parameters is evaluated, and the fitting of the model is checked. In MBDoE, the key statistic used in this step is the FIM, $\mathbf{H}_{\hat{\theta}}(\hat{\theta}, \boldsymbol{\varphi})$:

$$\mathbf{H}_{\hat{\theta}}(\hat{\theta}, \boldsymbol{\varphi}) = \left[\mathbf{V}_{\hat{\theta}}^{0}\right]^{-1} + \sum_{i=1}^{n_{sp}} \left(\frac{d\hat{\mathbf{y}}}{d\hat{\theta}}\right)_{i}^{T} \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \left(\frac{d\hat{\mathbf{y}}}{d\hat{\theta}}\right)_{i} = \left[\boldsymbol{V}(\hat{\theta}, \varphi)\right]^{-1} \quad (8)$$

where $\mathbf{V}_{\hat{\theta}}^{0}$ is the prior variance–covariance matrix of the model parameters, $\frac{d\hat{\mathbf{y}}}{d\hat{\theta}}$ is the sensitivity matrix, $\boldsymbol{\varphi}$ is the vector of the experimental conditions used to perform the experiment, and $\mathbf{V}(\hat{\theta}, \varphi)$ is the variance–covariance matrix, which is equal to the inverse of the FIM. $\mathbf{H}_{\hat{\theta}}$ is the metric used to quantify the amount of information that the experiment already performed provides on the model parameters.

4. Once $\mathbf{H}_{\hat{\theta}}$ is available the new experimental conditions to test are computed. This step is the core step of MBDoE, where the estimated parameters and their uncertainties are used to estimate the new experimental conditions. To identify the optimal experimental conditions it is necessary to maximize a metric of the FIM which can be the determinant, the trace, or the minimum eigenvalue:

$$\begin{aligned} &\boldsymbol{\varphi}_{new} = \max_{\hat{\theta}, \boldsymbol{\varphi}} [\Psi(\mathbf{H}(\hat{\theta}, \varphi))] \\ &\text{s.t.} \quad \varphi_{i,\text{lower bound}} \leq \varphi_i \geq \varphi_{i,\text{upper bound}} \end{aligned} \quad (9)$$

In this study, the D-optimality criterion (Pukelsheim, 2006) is used as metric Ψ.

5. The new experiment is performed and the dataset is updated.

In this study the stopping criterion used is the achievement of the maximum experimental budget, i.e. the maximum number of experiments. The other criteria, the goodness of fit test and the *t*-test are evaluated . The standard MBDoE framework involves two different optimization steps: (1) during maximum likelihood parameter estimation, to estimate the set of model parameters in model calibration (Eq. (5)), and (2) during optimal experimental design, used to identify the optimal experimental conditions to perform to reduce the variance of model parameters. This second optimization step can be computationally expensive, as it scales with the number of design variables of the system under analysis.

2.3. Proposed framework

In the subsequent section, the methodology proposed to address the challenges inherent in the MBDoE framework is outlined. This optimization-free approach, similar to traditional MBDoE techniques, evaluates expected information using a ranked-experiments approach, as detailed in Galvanin et al. (2015). For this reason, the methodology will be called Fisher Information matrix (FIM) Driven approach, or FIMD approach. The FIMD belongs to the class of sequential experimental design frameworks (Goujot et al., 2012; Duarte et al., 2019; Brendel et al., 2004) where model identification follows an iterative procedure. A block diagram of this method is depicted in Fig. 3.

The fundamental steps of the proposed FIMD framework are the following:

1. Definition of the experimental design space: in this step the same procedure used for the definition of the experimental design space used in the DoE framework is followed. A design vector is created using Eq. (3).
2. Definition of preliminary experiments to perform using DoE techniques. In this study, N_{exp} experiments are selected using LHS framework.
3. Selection of the first experiment to run: in this step, the first experiment to run is selected randomly, since there is no prior information on the system under analysis.
4. Execution of the experiment and parameter estimation: using Eq. (4) and Eq. (5) a parameter estimation is performed. The initial guess of the unknown parameters is updated with the new estimation ($\hat{\boldsymbol{\theta}}$).
5. Statistics on model parameters: statistics on the parameters are calculated to assess whether the values of the identified parameters are statistically reliable and to quantify the model adequacy using goodness of fit tests.
6. Re-sampling: this optional step is performed to enable experimental design space exploration during the procedure. In the resampling procedure the candidate experiments are generated using the LHS with 'maximin criterion' (Johnson et al., 1990). This step is added to minimize the possibility of running the same experiment multiple times and to improve the exploration of the experimental design space.
7. Evaluation of expected information: $\mathbf{H}_{\hat{\theta}}$ is calculated for each of the N_{exp} experimental points identified in Step 2 to compute the expected amount of information on parameters that can be achieved from each candidate experiment.
8. Ranking of candidate experiments: candidate experiments are ranked based on the computed information content using the concept of Relative Fisher Information (RFI) (Galvanin et al., 2015). The expression of the RFI is given by:

$$\text{RFI}_i = \frac{\|\mathbf{H}_i\|}{\sum_{i=1}^{N_{exp}} \|\mathbf{H}_i\|} = \frac{\|\mathbf{H}_i\|}{\|\mathbf{H}_{tot}\|} \quad (10)$$

In Eq. (10) the numerator is the expected information of the *i*th candidate experiment and the denominator is the total amount of information achievable from the full set of experiments.

9. Selection of the next experiment to perform: the next experiment (i + 1) to be performed is selected as the experiment providing the highest RFI:

$$\boldsymbol{\varphi}_{i+1} = \text{argmax}(\text{RFI}_i) \quad (11)$$

Also in this case, as in DoE and MBDoE, the proposed framework is an iterative approach, the stopping criterion is the achievement of the maximum experimental budget and the performance of the framework is evaluated using the metrics presented in Section 2.1.

3. Case studies

The proposed methodology (FIMD) is tested on two different case studies compared with the results obtained using standard methodologies. The first case study is a benchmark case study for the applications of MBDoE techniques, while the second case study is a more complex system characterized by a higher number of state variables, design variables and parameters. In each case study three experimental design approaches are compared:

- DoE: black box experimental design using LHS (Section 2.1);
- MBDoE: sequence of D-optimal experiments (Section 2.2);
- FIMD: proposed FIM-driven framework (Section 2.3).

Results are compared in terms of (a) profile of experimental design variables; (b) statistics obtained after parameter estimation after each iteration; (c) computational time. These case studies have been solved using Python, the library used to solve the set of differential equations is SciPy. The SciPy's 'odeint' solver is a tool for solving ordinary differential equations (ODEs) that employs an adaptive approach. It utilizes the LSODA algorithm, which combines implicit and explicit methods, allowing it to adjust dynamically to the complexity of the problem under analysis. This adaptability enables the solver to effectively handle a wide range of differential equations while maintaining both accuracy and computational efficiency. Moreover, its ability to automatically adjust integration precision based on the system's characteristics enhances its utility in scientific research and computational modeling.

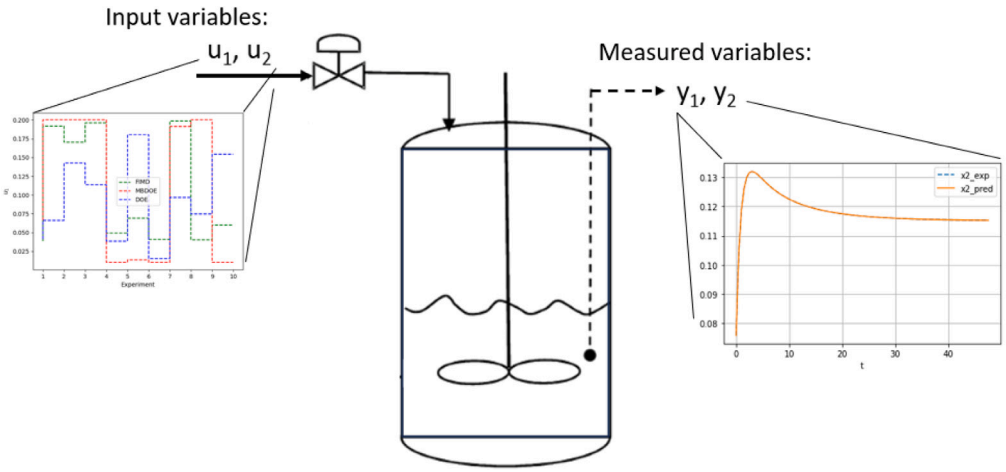

Fig. 4. Fed-batch reactor system under examination (Asprey and Macchietto, 2000). The figure on the left shows an example of the values of the variables manipulated during the different experiments, while on the right is the value of the measured variable y_1.

Table 1
Definition of experimental design space in terms of upper and lower bounds on design variables.

Parameter	Lower bound	Upper bound	Design variables
$x_1^0[g/l]$	1	10	Yes
u_1 $[h^{-1}]$	0.05	0.2	Yes
u_2 (g/l)	5	35	Constant

Table 2
Parameter values used for the in-silico experiments and initial guesses.

Parameter	True value	Initial guess
θ_1	0.31	5
θ_2	0.18	5
θ_3	0.55	5
θ_4	0.05	5

To perform the parameter estimation the scipy.minimize tool has been used. The L-BFGS-B algorithm, default in scipy.minimize, is a nonlinear optimization method tailored for minimizing scalar functions subject to linear or nonlinear constraints. It efficiently addresses optimization problems with variable bounds by combining the BFGS quasi-Newton method with limited-memory techniques (SciPy Developers, 2022).

3.1. Case study 1: Baker yeast growth model

The first proposed case study is the Baker yeast growth model appearing as a benchmark case study in a number of publications (Espie and Macchietto, 1989; Asprey et al., 2000; Asprey and Macchietto, 2000). Fig. 4 shows a schematic representation of the fed-batch reactor used as case study.

We assume that the model used to generate the in-silico data implements Monod-type kinetics and is described by the following set of differential equations

$$\begin{aligned}\frac{dx_1}{dt} &= \left(\frac{\theta_1 x_2}{\theta_2 + x_2} - u_1 - \theta_4\right) x_1 \\ \frac{dx_2}{dt} &= -\frac{\theta_1 x_2 x_1}{(\theta_2 + x_2)\theta_3} + u_1\left(u_2 - x_2\right)\end{aligned} \tag{12}$$

In Eq. (12) x_1 is the biomass concentration (g/l), x_2 is the substrate concentration (g/l), u_1 is the dilution factor (h^{-1}) and u_2 is the substrate concentration in the feed (g/l). In this model the manipulated variables are u_1 and u_2 and the model parameters are $\theta_1, \theta_2, \theta_3, \theta_4$. In Table 1 the experimental design space is defined in terms of upper and lower bounds on input variables and initial conditions.

The objective is to compare different experimental design strategies to precisely and accurately estimate the set of model parameters. Precision is assessed by evaluating the variance of model parameters at each iteration, while accuracy is assessed in terms of closeness to an assumed "true value" used to generate the in-silico experimental data. Table 2 shows the values of the parameters $\boldsymbol{\theta}$ used to simulate the experiments and the initial guess values. To generate the synthetic data, noise was simulated to render the generated data as close as possible to the experimentally obtained data. Specifically, a normal distribution with a mean of zero and a standard deviation of 0.5 g^2/l^2 was employed to emulate the measurement noise. The final value of the generated data is obtained using Eq. (2) using a variance–covariance matrix of measurement errors equal to:

$$\boldsymbol{\Sigma_y} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \tag{13}$$

In this case, the design vector, $\boldsymbol{\varphi}$, is composed of the initial state x_1^0 and the input u_1 while the initial value of the substrate concentration and the value of the substrate concentration in the feed (u_2) are kept constant and equal to 0.01 g/l and 5 g/l respectively.

$$\boldsymbol{\varphi} = [x_1^0, u_1] \tag{14}$$

The bounds used to define the design space for x_1 and u_1 are the same used in Asprey and Macchietto (2000) and reported in Table 1. To solve the differential equation of the kinetic model (Eq. (12)) has been used 10 time intervals so the discretization element is equal to 10.

3.1.1. Generation of candidate experiments

Both in the DoE framework and the FIM-driven one it is necessary to generate a first set of candidate experiments. To generate this set a Latin Hypercube Sampling (LHS) coupled with the bounds reported in Table 1 has been used. The criterion used to obtain a space filling sampling is the 'maximin' criterion (Johnson et al., 1990). The objective of this design is to maximize the minimum distance inter points:

$$\max\{\min[d(p_i, p_j)]\} \tag{15}$$

where $d(p_i, p_j)$ is the distance between two experimental points p_i and p_j.

3.1.2. Results case study 1

The sampling of the experimental design space obtained using this criterion is reported in Fig. 5a.

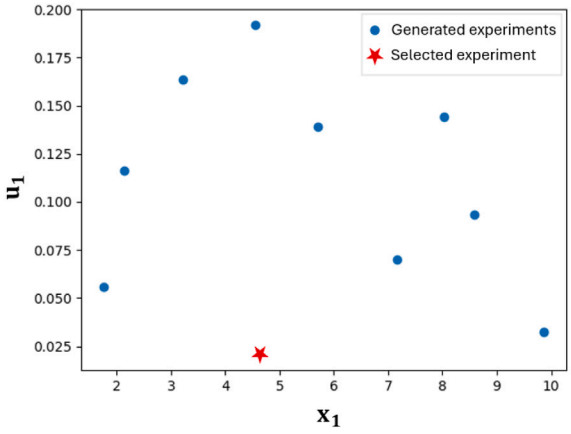

Fig. 5. Set of candidate experiments obtained using LHS, with the first experiment selected randomly.

Fig. 5 highlights the first randomly selected experiment. To ensure a fair comparison, the initial experiment is kept the same for all three methodologies. The profiles of the experimental design variables ($x_1(0)$ and u_1) in the experiment designed using DoE, FIMD, and MBDoE are reported in Fig. 6. The figures show how DoE and FIMD explore the experimental design space in a more effective way, while MBDoE either works in a narrower range of experimental conditions (Fig. 6(a)) or oscillates between two values (Fig. 6(b)).

In Figs. 7(a) and 7(b), the values of information metrics (determinant of FIM, Eq. (8)) of the single experiments performed and the cumulative information obtained are reported. Fig. 7(a) illustrates the comparison between the information obtained from the individual experiment evaluated using Eq. (8) (excluding the prior information given by the term $\left[\mathbf{V}_{\hat{\theta}}^{0}\right]^{-1}$). The experiments identified using the MBDoE methodology are those that provide the most information on parameters, and this is apparent by analyzing the total amount of information achieved from the experimental campaign (Fig. 7(b)). In the eighth and tenth experiments, it is evident that employing the MBDoE provides a higher amount of information compared to the other two methodologies, as shown in Fig. 7(a). This is due to the fact that while with MBDoE the experiment performed is always the most informative, with DoE the design of the experiment to be performed is fixed (it is generated at the start of the procedure, Fig. 5) and with FIMD the experiment selected is the most informative of the set of candidate experiments of that iteration, meaning that the average information contained in the eighth iteration was very low.

Figure shows the profiles of the estimates of the unknown parameters of the model (Eq. (12)). In Figs. 8 the comparison of the profiles of estimated (a) θ_1, (b) θ_2, (c) θ_3 and (d) θ_4 and their respective variance are reported. In these figures, the blue line represents the value of the identified parameters, the orange line is the value of the parameter used to generate the in-silico data and the blue shaded area represents the uncertainty of the prediction. From these figures it is possible to notice that the predictions obtained using the FIMD and the MBDoE are more precise than the predictions obtained using the DoE framework since the first experiment (the blue shaded area that represents the uncertainty of the prediction is lower for the FIMD and MBDoE).

In the following figure, the profiles of the variances of the different parameter predictions are compared. Fig. 9 shows the profiles of the variances of the four parameters. The profiles show that the variance evaluated using MBDoE decreases faster than the one evaluated using DoE and FIMD approach. The profiles of the variance evaluated using the FIMD approach exhibits a faster decrease compared to the profiles calculated using the DOE approach. This observation indicates that the FIMD achieves results comparable to those obtained with MBDoE but with improved performance compared to DoE as the variance on the predictions is smaller (Table 3). In Fig. 10 the correlation matrices of the identified parameters using the three frameworks are shown. In Fig. 10(a) the correlation matrix of the parameters obtained using the FIMD is shown, it is evident from this figure that the parameters θ_1 and θ_2 have a strong positive correlation as their correlation coefficient is greater than 0.8. Figs. 10(b) and 10(c) show the correlation matrices obtained with DoE and MBDoE, respectively. The correlations have been evaluated using the following equation:

$$R_{i,j} = \frac{v_{\hat{\theta}_{i,j}}}{\sqrt{v_{\hat{\theta}_{i,i}} v_{\hat{\theta}_{j,j}}}} \tag{16}$$

where $R_{i,j}$ are the values of the correlation matrix and $\boldsymbol{v}_{\hat{\theta}}$ is the variance covariance matrix obtained from the Fisher information matrix.

In Table 3 the values of the predictions obtained with DoE, MBDoE, and FIMD and their relative prediction variances obtained after 10 experiments are reported. This table shows that the variance in the predictions obtained with the FIMD is smaller than that obtained with the DoE and similar to that obtained with the MBDoE. This means that the estimates obtained with the proposed method are more precise than those obtained with the DoE and comparable with those obtained with the MBDoE. To assess the precision of the identified parameters, it is essential to examine the results of the t-test performed on the estimates (as discussed in Section 2.2). The t-values of all the model parameters after six performed experiments are compared with the reference t-value in Table 4.

The results presented in this table indicate that after six experiments estimates obtained using the FIMD and the MBDoE methodologies are statistically precise as they pass the t-test (their t-values are greater than the reference t-value evaluate with significance equal to 95%). However, the parameters obtained using the DoE methodology fail the t-test, making the parameters θ_2, θ_3, and θ_4 less accurate.

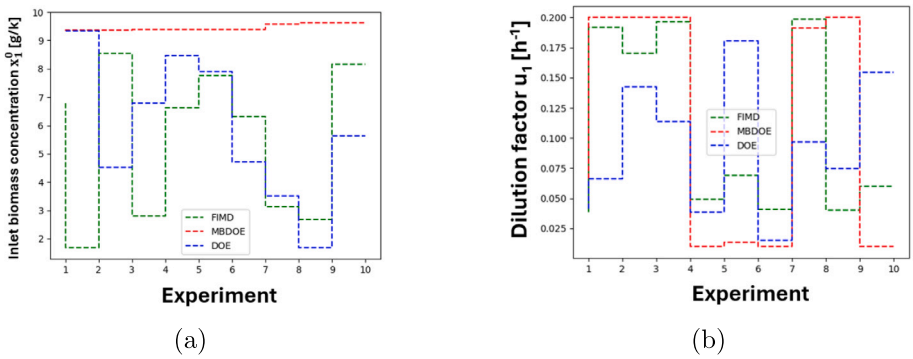

Fig. 6. Comparison of the experimental design variables using the FIMD, DoE, and MBDoE methodologies. (a) x_1^0; (b) u_1.

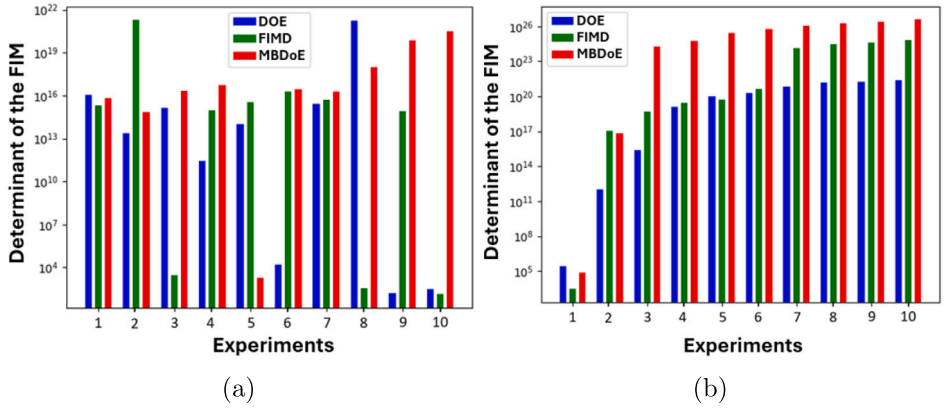

Fig. 7. Information iteratively obtained after each experiment designed using DoE (blue bars), MBDoE (red bars) and FIMD (green bars). (a) Information obtain from each single experiment; (b) Cumulative information obtained after every iteration of the three methods.

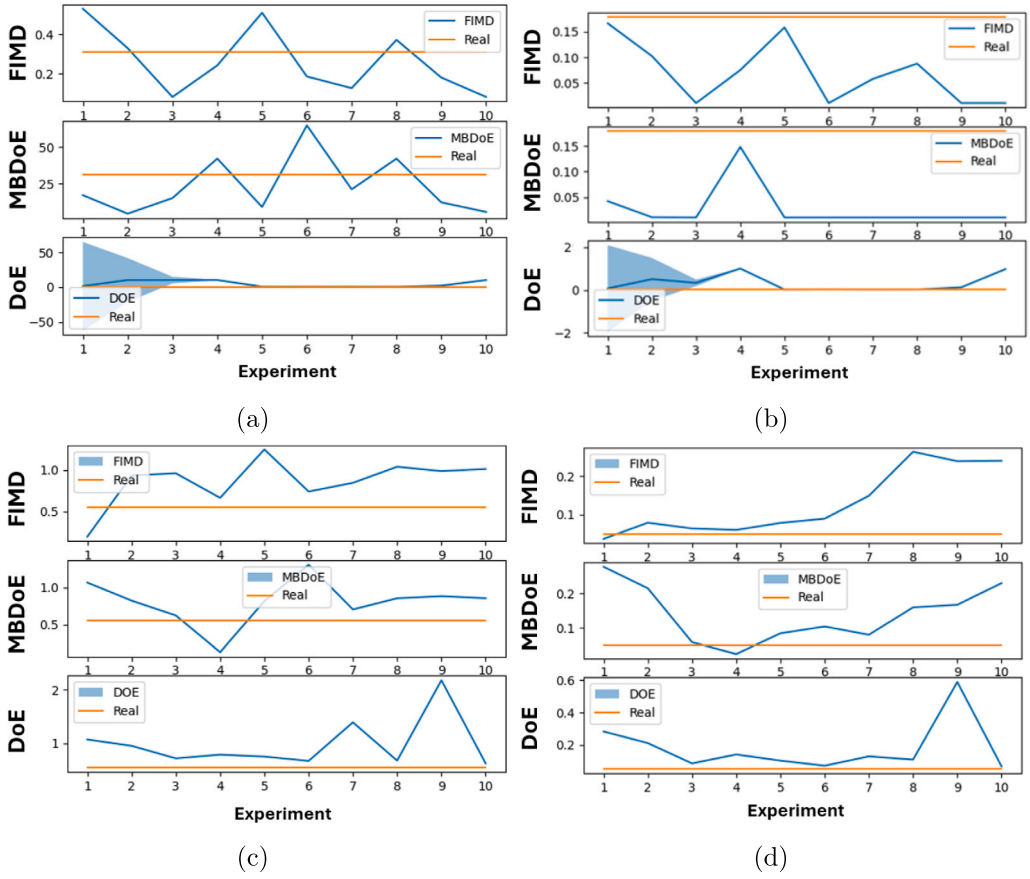

Fig. 8. Profiles of the identified model parameters in terms of estimated value (blue line), assumed true value (orange line) and uncertainty (blue shaded area). (a) θ_1, (b) θ_2, (c) θ_3, (d) θ_4.

Table 3
Parameters used to generate the in-silico data and values identified with DoE, MBDoE and FIMD.

Parameter	True value	DoE	MBDoE	FIMD
θ_1	0.31	$1.32 \pm 2 \cdot 10^{-2}$	$0.21 \pm 8 \cdot 10^{-3}$	$0.18 \pm 7 \cdot 10^{-3}$
θ_2	0.18	$1.76 \pm 9 \cdot 10^{-3}$	$0.02 \pm 7 \cdot 10^{-3}$	$0.01 \pm 9 \cdot 10^{-3}$
θ_3	0.55	$0.78 \pm 0.1 \cdot 10^{-5}$	$0.71 \pm 0.1 \cdot 10^{-5}$	$1.1 \pm 0.1 \cdot 10^{-5}$
θ_4	0.05	$0.07 \pm 0.23 \cdot 10^{-5}$	$0.21 \pm 0.13 \cdot 10^{-5}$	$0.23 \pm 0.19 \cdot 10^{-5}$

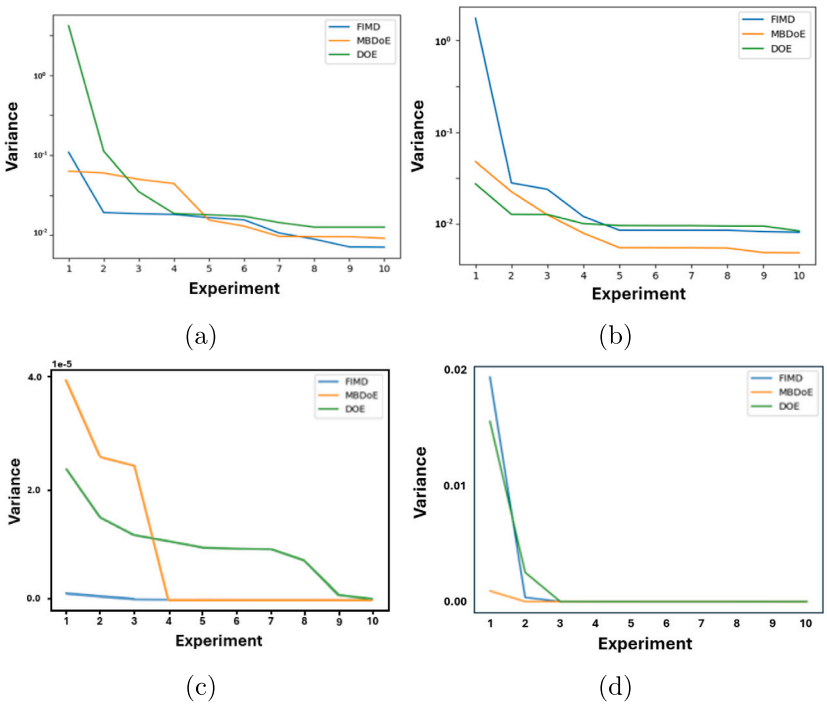

Fig. 9. Profiles of the variance of the predicted parameters (a) θ_1, (b) θ_2, (c) θ_3 and (d) θ_4.

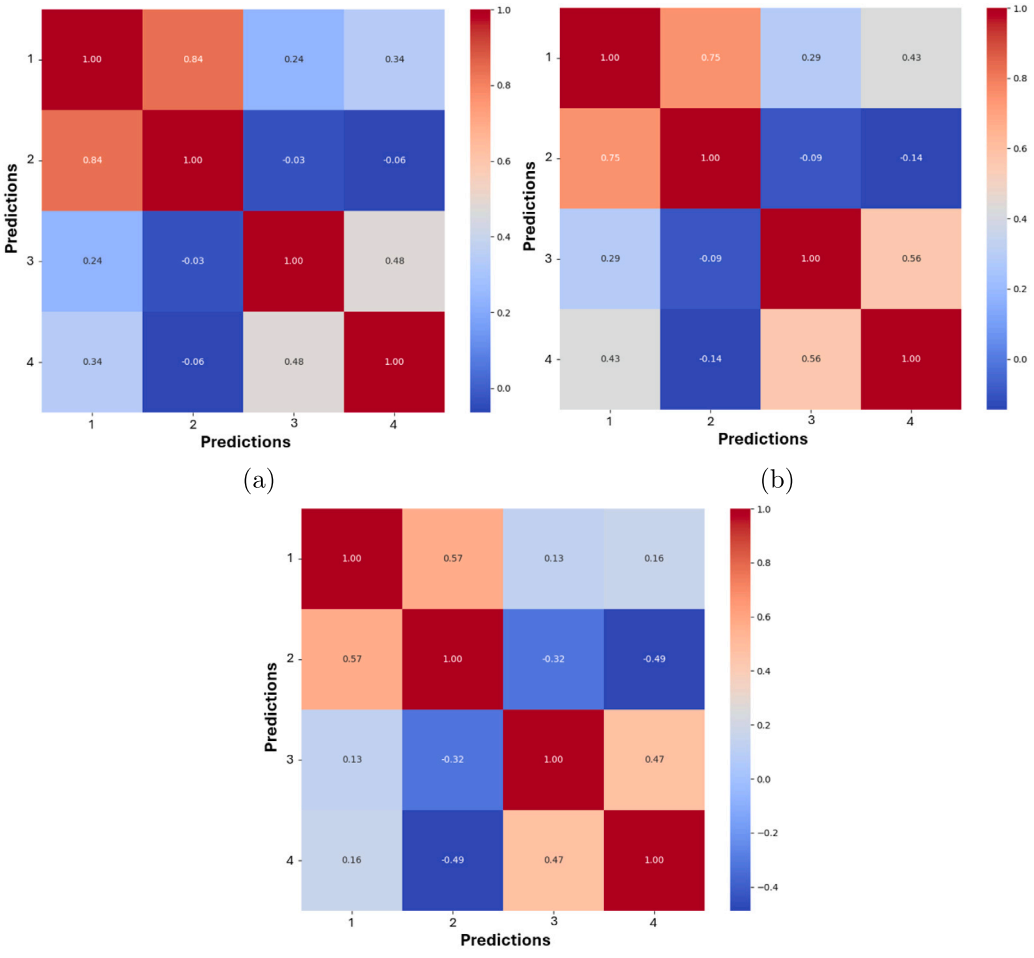

Fig. 10. Correlation matrices of the parameters predicted using (a) FIMD framework; (b) DoE and (c) MBDoE.

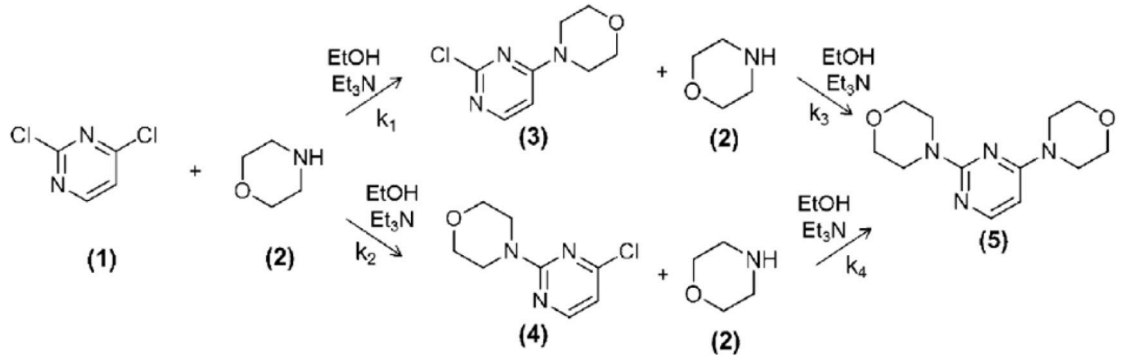

Fig. 11. Multistep S_NAr reaction (Reizman and Jensen, 2012).

Table 4
Results of the t-test after six experiments. Parameters failing the t-test are indicated in boldface.

Method	t-ref	t_{θ_1}	t_{θ_2}	t_{θ_3}	t_{θ_4}
DoE	1.83	2.68	**0.92**	**1.14**	**1.02**
FIMD	1.83	28	19	25	128
MBDoE	1.83	203	97	82	128

3.2. Case study 2: aromatic nucleophilic substitution

The second case study proposed consists of a nucleophilic aromatic substitution (S_NAr), described in the work of Reizman and Jensen (2012). The schematic of the reaction is depicted in Fig. 11.

This reaction involves five different measurable states: the 2,4-dychloropyrimidine (1) that reacts with the morpholine (2) in ethanol to form the desired 2-substituted aminopyrimidine (4) and the undesired 4-substituted aminopyrimidine (3) the final product of this reaction is the 2,4-substituted byproduct (5). The kinetic model used to perform the in-silico experiments follows the scheme reported in Fig. 11 under the assumption that the reaction system can be modeled as an ideal plug flow reactor (PFR). Considering these assumptions the kinetic model governing the generation and consumption of the different species is:

$$\frac{dc_1}{dt} = -k_1 c_1 c_2 - k_2 c_1 c_2 \tag{17}$$

$$\frac{dc_2}{dt} = -k_1 c_1 c_2 - k_2 c_1 c_2 - k_3 c_3 c_2 - k_4 c_4 c_2 \tag{18}$$

$$\frac{dc_3}{dt} = k_1 c_1 c_2 - k_3 c_3 c_2 \tag{19}$$

$$\frac{dc_4}{dt} = k_2 c_1 c_2 - k_4 c_4 c_2 \tag{20}$$

$$\frac{dc_5}{dt} = k_3 c_3 c_2 - k_4 c_4 c_2 \tag{21}$$

Eqs. (17)–(21) express the concentrations of the different components involved in the system in time. These equations are function of the temperature and of the initial concentrations, the rate constant k_i for the ith reaction is expressed in the following way:

$$k_i = A_i \exp\left(-\frac{E_{a,i}}{RT}\right) \quad i = 1, \dots, N_{react} \tag{22}$$

where A_i and E_i are the pre-exponential factor and the activation energy associated with the ith reaction that has to be determined using the experimental data and N_{react} is the number of reactions taking place. In Table 5 the values of the parameters used to generate the in-silico data and the initial guesses used in the identification procedure are reported. As in the previous case study, it was assumed that the measurement noise follows a normal distribution with mean equal to zero and standard deviation equal to 0.001.

Table 5
Parameter values used for the in-silico experiment and initial guesses.

Parameter		Value	Initial guess
A_1 [1/M s]	θ_1	0.712	1
E_1 [kJ/mol]	θ_2	$2.7{\cdot}10^4$	$1{\cdot}10^4$
A_2 [1/M s]	θ_3	1.389	1
E_2 [kJ/mol]	θ_4	$3.21{\cdot}10^4$	$1{\cdot}10^4$
A_3 [1/M s]	θ_5	4.237	1
E_3 [kJ/mol]	θ_6	$6.0{\cdot}10^4$	$1{\cdot}10^4$
A_4 [1/M s]	θ_7	3.853	1
E_4 [kJ/mol]	θ_8	$4.5{\cdot}10^4$	$1{\cdot}10^4$

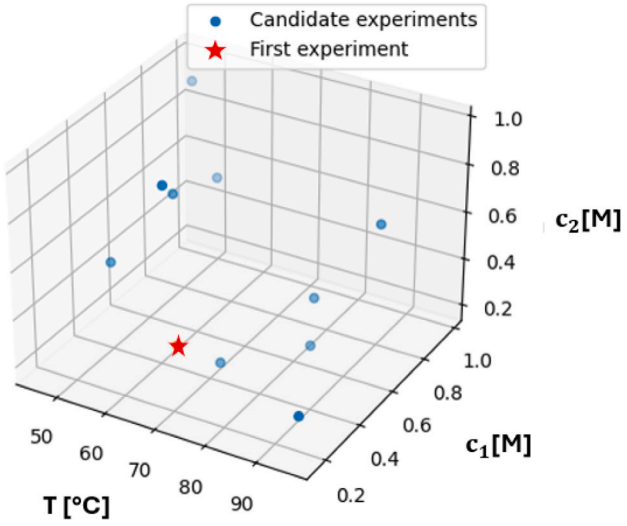

Fig. 12. Candidate experiment generated using the LHS and selected experiment.

Table 6
Definition of experimental design space.

Parameter	Lower bound	Upper bound
c_1	0.1	1
c_2	0.1	1
T [°C]	40	100

Table 6 shows the experimental bounds used to generate the candidate experiments with the LHS where c_1 and c_2 are the initial concentrations of the components 1 and 2, τ is the residence time (in this case it is kept constant and equal to 1200 s) and T is the reactor temperature in °C.

In this case, since the residence time is constant, the design vector $\boldsymbol{\varphi}$ is:

$$\boldsymbol{\varphi} = \left[c_1(0), c_2(0), T\right] \tag{23}$$

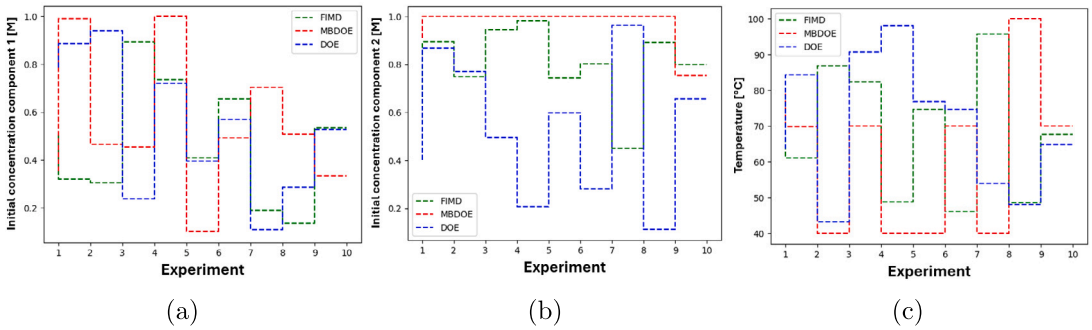

Fig. 13. A comparison of the experiments designed using the FIMD, DoE, and MBDoE methodologies. (a) Inlet 1; (b) Inlet 2; (c) Temperature.

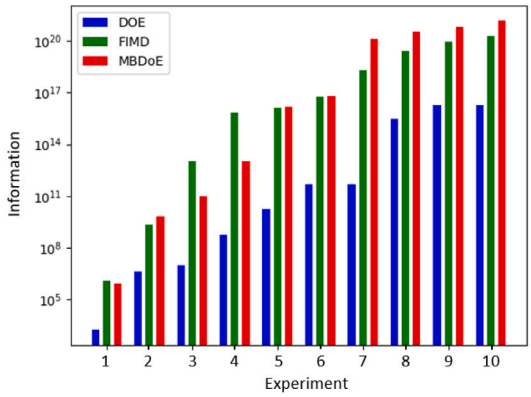

Fig. 14. Cumulative information calculated using Eq. (8) using the three different methodologies (FIMD, DoE, and MBDoE). The blue bar represents information from the DoE methodology, the green bar depicts total information from the FIMD approach, and the red one signifies information obtained using the standard MBDoE method.

The procedure followed to generate the candidate experiments is described in Section 3.1.1. To solve the differential equation of the kinetic model (Eq. (12)) has been used 10 time intervals so the discretization element is equal to 10. The following section presents the results obtained from the DoE, MBDoE, and FIMD methodologies to compare their relative performance.

3.2.1. Results case study 2

The distribution of the first set of experiments used for the three proposed framework generated using the bounds in Table 6 is illustrated in Fig. 12 with the first experiment to be performed.

Starting from the experiment shown in Fig. 12 the sequence of experiments for the three different frameworks described in the previous sections was carried out to identify the values of the eight unknown parameters. The profiles of the experimental design variables (inlet concentration of components 1 and 2, and temperature) in the experiment designed using DoE, FIMD, and MBDoE are reported in Fig. 13. The information achieved from each experimental campaign is reported in Fig. 14. The cumulative information acquired through the MBDoE methodology, evaluated from Eq. (8), is on the order of magnitude of 10^{21}, with the FIMD approach achieving around 10^{20} and the DoE framework yielding a total information of 10^{16}.

Fig. 14 illustrates the superior performance of MBDoE and FIMD over DoE, as evidenced by increased information obtained. While differences between MBDoE and FIMD are less pronounced, suggesting similar performance in complex systems, MBDoE yields slightly more information as experiments progress. This is reflected in the variance of model parameters. Subsequent figures show the profiles of the eight identified parameters and their variances, demonstrating variance reduction with fewer experiments compared to DoE.

In Fig. 17 the elements of the correlation matrix (Eq. (16)) of the identified parameters obtained using the three different methods are shown. Fig. 17(a) shows that parameters θ_6, θ_7 and θ_3, θ_1 are positively strongly correlated (the values of the correlation coefficients are greater that 0.7). Figs. 17(b) and 17(c) show that using DoE and MBDoE results in a similar parameter correlation pattern.

In Table 7 the values of the parameters identified using DoE, MBDoE, and FIMD and their respective variances are compared with the true values used to generate the in-silico data. Table 7 and Figs. 15–16 show that by adopting MBDoE and FIMD frameworks the variance of the estimates decreases faster than in DoE (this can be seen from the decrease of the blue shaded area in the figures and the variance values after 10 experiments in the table)

The rate of variance reduction translates into the precision of the estimates, as illustrated by the t-test results reported in Table 8.

The t-test reveals that after ten experiments all parameters obtained with FIMD and MBDoE are statistically reliable, while four out of eight parameters (A_1, A_2, E_2, E_4) obtained with the DoE approach failed the test, rendering them unreliable. This underlines the fact that using DoE requires more experiments than FIMD for precise parameter identification.

3.3. Computational burden

The significant advantage of the proposed methodology compared to MBDoE techniques is that it does not require any optimization to identify the optimal experiment to perform. This reflects on the computational costs and computational time necessary to precisely estimate the full set of parameters of the model under examination, allowing to explore the experimental design space without incurring into local optima.

In Table 9 the computational times required to reach a solution with the DoE, MBDoE, and the FIMD methodology are reported. Results were obtained using an Intel® Core™ i9-10885H @ CPU 2.40 GHz with 16.0 GB RAM. The reported computational times refer only to the times necessary for parameter estimation and experimental design to be performed, while in the case of DoE and FIMD, they do not take into account the time necessary to generate the first set of candidate experiments using LHS. Table 9 shows the time needed to reach a solution in Case study 1 and Case study 2.

The computational times in this table show that even though the complexity of the system under analysis, measured in terms of the number of parameters, inputs and measured states, is not significantly high, the time required to achieve a solution using MBDoE is more than four times greater than the time required by FIMD to reach a solution (Case study 1). In Case study 2, where the system is more complex (in terms of state variables, design variables and unknowns than the first case study) the difference in computational times increases significantly. In fact, if in Case study 1 the time required for MBDoE was approximately four times greater than the time required for FIMD, in Case study 2 the time required is approximately twenty times greater. The significant reduction in CPU time depends not only on the elimination of the optimization required to evaluate the new experimental conditions to

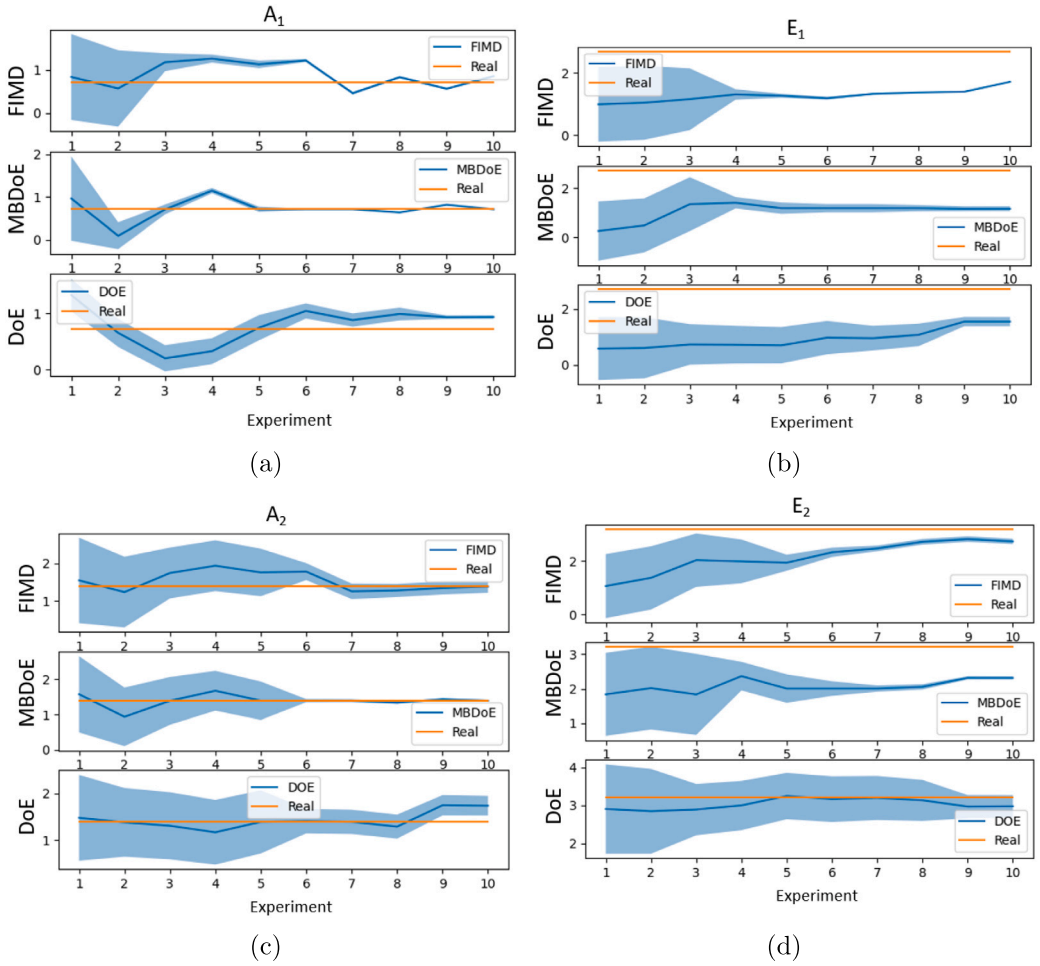

Fig. 15. Profiles of the identified model parameters in terms of estimated value (blue line), assumed true value (orange line), and uncertainty (blue shaded area). (a) A_1, (b) E_1, (c) A_2, (d) E_2.

Table 7
Parameter values used to generate the in-silico data and estimated values including 95% confidence intervals.

Parameter	True value	DoE	MBDoE	FIMD
A_1	0.712	0.931 ± 0.029	0.711 ± 0.006	0.855 ± 0.023
$E_1 \cdot 10^{-4}$	2.7	1.549 ± 0.164	1.159 ± 0.096	1.712 ± 0.019
A_2	1.389	1.737 ± 0.211	1.388 ± 0.024	1.380 ± 0.165
$E_2 \cdot 10^{-4}$	3.21	2.98 ± 0.29	2.313 ± 0.045	2.73 ± 0.10
A_3	4.237	4.711 ± 0.002	4.237 ± $1.9 \cdot 10^{-4}$	4.610 ± 0.001
$E_3 \cdot 10^{-4}$	6.0	3.928 ± 0.156	2.870 ± 0.017	4.736 ± 0.044
A_4	3.853	3.819 ± 0.009	3.853 ± 0.002	3.922 ± 0.009
$E_4 \cdot 10^{-4}$	4.50	4.45 ± 0.375	2.785 ± 0.010	4.262 ± 0.067

Table 8
t-values of the identified parameters after 10 experiments, parameters failing the t-test are indicated in boldface.

Method	t-ref	t_{A_1}	t_{E_1}	t_{A_2}	t_{E_2}	t_{A_3}	t_{E_3}	t_{A_4}	t_{E_4}
DoE	1.68	**1.46**	3.42	**1.65**	**1.43**	38.2	5.28	13.9	**1.22**
FIMD	1.68	2.91	5.25	2.46	4.34	79.6	20.8	26.8	12.9
MBDoE	1.68	6.45	6.23	5.67	6.43	101	20.9	38.3	12.7

be tested, but also on FIMD framework requiring less model evaluations than MBDoE. In the standard method, in fact, the model structure is evaluated in each optimization step, which greatly increases the computation time. The efficiency of FIMD becomes more pronounced as system complexity increases, with computation time for MBDoE approximately twenty times greater than the one for FIMD in the more complex Case 2. This emphasizes the scalability and effectiveness of the FIMD methodology, particularly in scenarios with higher-dimensional systems and increased unknown variables.

4. Conclusion and future work

This article introduced a novel Fisher Information Matrix Driven (FIMD) framework for rapid parameter identification, leveraging a ranking-based selection of candidate experiments. Comparative analyses with established methodologies, namely Design of Experiments (DoE) and Model-Based Design of Experiments (MBDoE), were conducted to test the performance of the proposed framework. The approach presented herein enables precise estimation of unknown model parameters, emphasizing a significant reduction in computational time

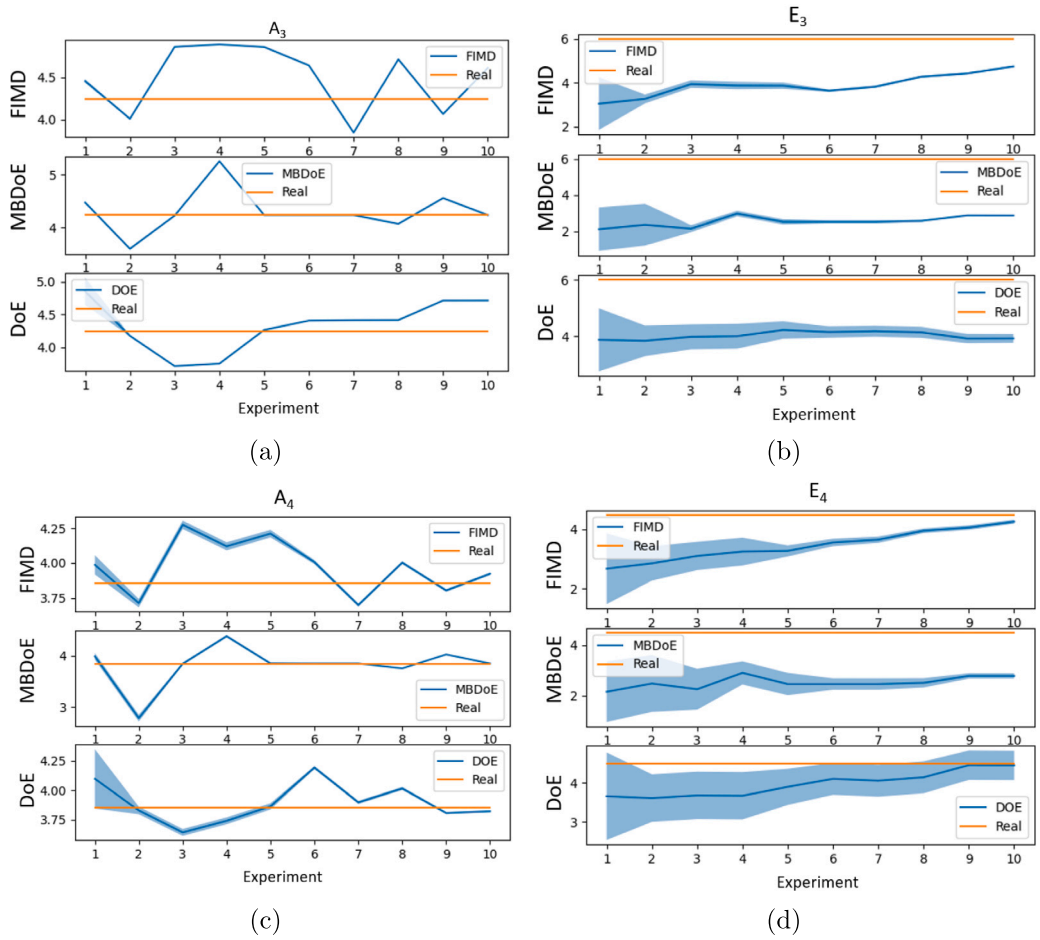

Fig. 16. Profiles of the identified model parameters in terms of estimated value (blue line), assumed true value (orange line) and uncertainty (blue shaded area). (a) A_3, (b) E_3, (d) A_4, (e) E_4.

Table 9
Comparison between computational times required by different experimental design approaches in the two case studies according to the model complexity (number of state variables n_x, number of inputs n_u and number of parameters n_θ).

	Case study 1	Case study 2
t_{FIMD} [s]	2.9	10
t_{MBDoE} [s]	13.5	233
n_x	2	5
n_u	2	3
n_θ	4	8

required for solution attainment. Two simulated (in-silico) case studies were employed to evaluate the effectiveness of the methodology in identifying parameters by performing the t-test on the identified parameters. In Case study 1, FIMD, DoE, and MBDoE were applied to a biochemical reactor. In this system it was possible to achieve results in terms of parameter precision comparable to a conventional MBDoE and significantly better than those obtained with a DoE. Notably, computational time requirements for solution attainment were comparable among the DoE and FIMD, with the FIMD framework achieving results in one-quarter of the time needed for MBDoE. The second case study, involving a reaction system, yielded even more promising outcomes in terms of statistical precision of parameter estimates. FIMD methodology was 20 times faster than a conventional MBDoE, but preserving a similar precision and accuracy in parameter estimation. These findings underscore the efficiency and accuracy gains achievable through the proposed FIMD methodology, establishing it as a valuable tool for expediting parameter identification in complex systems.

Future work will explore the practical applicability of the proposed methods to autonomous kinetic model identification platforms to effectively exploit the advantages deriving from the reduced computational time.

CRediT authorship contribution statement

Andrea Friso: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Federico Galvanin:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

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.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the Department of Chemical Engineering, University College London, United Kingdom and the EPSRC, United Kingdom Grant EP/R032807/1. The support is gratefully acknowledge.

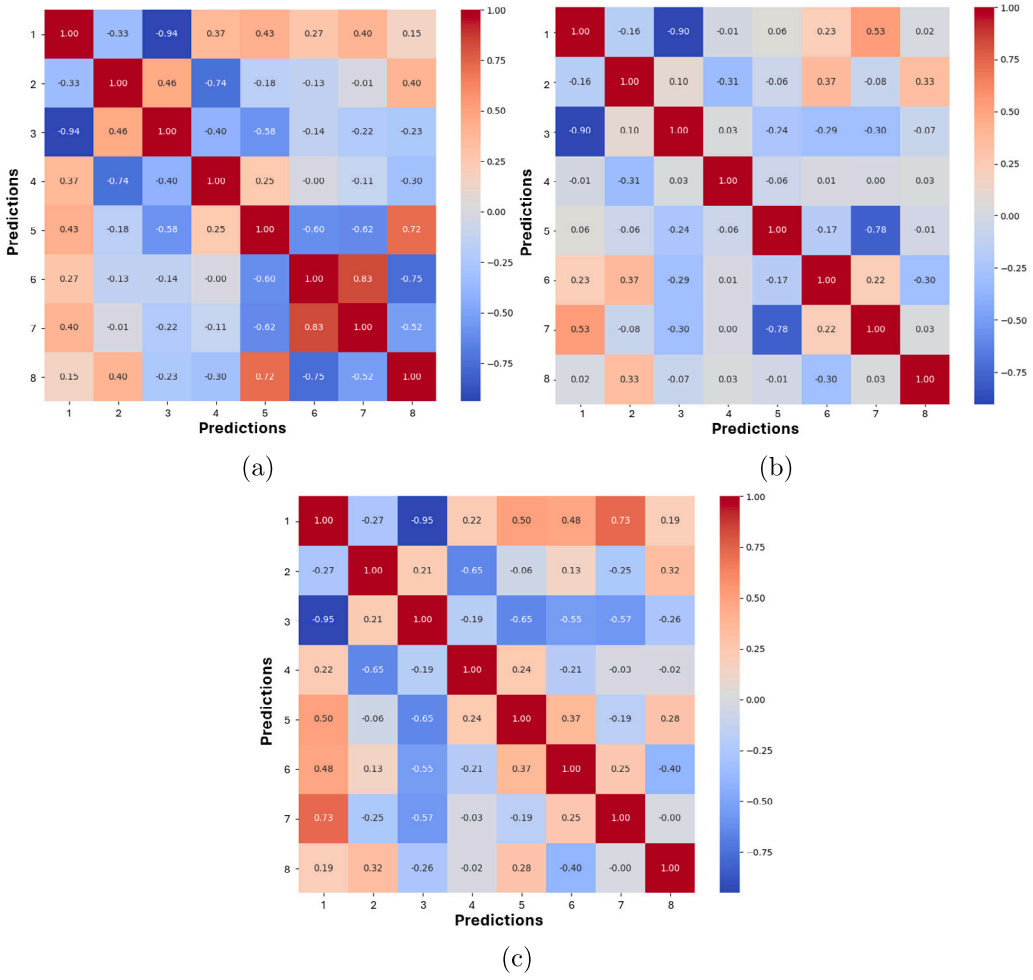

Fig. 17. Correlation matrices of the parameters predicted using (a) FIMD framework; (b) DoE and (c) MBDoE.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compchemeng.2024.108724.

References

Asprey, S., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. Comput. Chem. Eng. 24 (2), 1261–1267. http://dx.doi.org/10.1016/S0098-1354(00)00328-8, URL https://www.sciencedirect.com/science/article/pii/S0098135400003288.

Asprey, S., Macchietto, S., 2002. Designing robust optimal dynamic experiments. J. Process Control 12 (4), 545–556. http://dx.doi.org/10.1016/S0959-1524(01)00020-8, URL https://www.sciencedirect.com/science/article/pii/S0959152401000208.

Asprey, S., Macchietto, S., Pantelides, C., 2000. Robust optimal designs for dynamic experiments. IFAC Proc. Vol. 33 (10), 845–850. http://dx.doi.org/10.1016/S1474-6670(17)38645-7, URL https://www.sciencedirect.com/science/article/pii/S1474667017386457. IFAC Symposium on Advanced Control of Chemical Processes 2000, Pisa, Italy, 14-16 June 2000.

Atkinson, A.C., 1970. A method for discriminating between models. J. R. Stat. Soc. Ser. B Stat. Methodol. 32 (3), 323–345.

Atkinson, A., Donev, A., Tobias, R., 2007. Optimum Experimental Designs, with SAS. http://dx.doi.org/10.1093/oso/9780199296590.001.0001.

Bard, Y., 1974. Nonlinear parameter estimation. URL https://api.semanticscholar.org/CorpusID:5536860.

Blau, G., Lasinski, M., Orcun, S., Hsu, S.-H., Caruthers, J., Delgass, N., Venkatasubramanian, V., 2008. High fidelity mathematical model building with experimental data: A Bayesian approach. Comput. Chem. Eng. 32 (4), 971–989. http://dx.doi.org/10.1016/j.compchemeng.2007.04.008, URL https://www.sciencedirect.com/science/article/pii/S0098135407000968, Festschrift devoted to Rex Reklaitis on his 65th Birthday.

Bonvin, D., Georgakis, C., Pantelides, C.C., Barolo, M., Grover, M.A., Rodrigues, D., Schneider, R., Dochain, D., 2016. Linking models and experiments. Ind. Eng. Chem. Res. 55 (25), 6891–6903. http://dx.doi.org/10.1021/acs.iecr.5b04801.

Box, J.F., 1980. R. A. Fisher and the design of experiments, 1922–1926. Amer. Statist. 34 (1), 1–7, URL http://www.jstor.org/stable/2682986.

Brendel, M., Mhamdi, A., Bonvin, D., Marquardt, W., 2004. An incremental approach for the identification of reaction kinetics. IFAC Proc. Vol. 37 (1), 173–178. http://dx.doi.org/10.1016/S1474-6670(17)38727-X, URL https://www.sciencedirect.com/science/article/pii/S147466701738727X, 7th International Symposium on Advanced Control of Chemical Processes (ADCHEM 2003), Hong-Kong, 11-14 January 2004.

Draper, N., Smith, H., 2014. Applied regression analysis. In: Wiley Series in Probability and Statistics, Wiley, URL https://books.google.co.uk/books?id=byReBAAAQBAJ.

Duarte, B.P., Atkinson, A.C., Granjo, J.F., Oliveira, N.M., 2019. Calculating D-optimal designs for compartmental models with a Michaelis–Menten elimination rate. J. Process Control 83, 88–101. http://dx.doi.org/10.1016/j.jprocont.2019.09.001, URL https://www.sciencedirect.com/science/article/pii/S0959152419304020.

Espie, D., Macchietto, S., 1989. The optimal design of dynamic experiments. AIChE J. 35 (2), 223–229. http://dx.doi.org/10.1002/aic.690350206.

Galvanin, F., Cao, E., Al-Rifai, N., Gavriilidis, A., Dua, V., 2015. Model-based design of experiments for the identification of kinetic models in microreactor platforms. In: Gernaey, K.V., Huusom, J.K., Gani, R. (Eds.), 12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering. In: Computer Aided Chemical Engineering, vol. 37, Elsevier, pp. 323–328. http://dx.doi.org/10.1016/B978-0-444-63578-5.50049-9, URL https://www.sciencedirect.com/science/article/pii/B9780444635785500499.

Galvanin, F., Cao, E., Al-Rifai, N., Gavriilidis, A., Dua, V., 2016. A joint model-based experimental design approach for the identification of kinetic models in continuous flow laboratory reactors. Comput. Chem. Eng. 95, 202–215. http://dx.doi.org/10.1016/j.compchemeng.2016.05.009, URL https://www.sciencedirect.com/science/article/pii/S0098135416301673.

Goujot, D., Meyer, X., Courtois, F., 2012. Identification of a rice drying model with an improved sequential optimal design of experiments. J. Process Control 22

(1), 95–107. http://dx.doi.org/10.1016/j.jprocont.2011.10.003, URL https://www.sciencedirect.com/science/article/pii/S095915241100196X.

Johnson, M., Moore, L., Ylvisaker, D., 1990. Minimax and maximin distance designs. J. Statist. Plann. Inference 26 (2), 131–148. http://dx.doi.org/10.1016/0378-3758(90)90122-B, URL https://www.sciencedirect.com/science/article/pii/037837589090122B.

Kusumo, K.P., Gomoescu, L., Paulen, R., García Muñoz, S., Pantelides, C.C., Shah, N., Chachuat, B., 2020. Bayesian approach to probabilistic design space characterization: A nested sampling strategy. Ind. Eng. Chem. Res. 59 (6), 2396–2408. http://dx.doi.org/10.1021/acs.iecr.9b05006.

M. D. Mckay, R.J.B., Conover, W.J., 2000. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 42 (1), 55–61. http://dx.doi.org/10.1080/00401706.2000.10485979, URL https://www.tandfonline.com/doi/abs/10.1080/00401706.2000.10485979, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/00401706.2000.10485979.

Montgomery, D.C., 1984. Design and Analysis of Experiments.

Montgomery, D., 2017. Design and Analysis of Experiments. John Wiley & Sons, Incorporated, URL https://books.google.co.uk/books?id=Py7bDgAAQBAJ.

Pankajakshan, A., Waldron, C., Quaglio, M., Gavriilidis, A., Galvanin, F., 2019. A multi-objective optimal experimental design framework for enhancing the efficiency of online model identification platforms. Engineering 5 (6), 1049–1059. http://dx.doi.org/10.1016/j.eng.2019.10.003, URL https://www.sciencedirect.com/science/article/pii/S2095809919308537.

Pukelsheim, F., 2006. Optimal Design of Experiments. SIAM.

Quaglio, M., Waldron, C., Pankajakshan, A., Cao, E., Gavriilidis, A., Fraga, E.S., Galvanin, F., 2019. An online reparametrisation approach for robust parameter estimation in automated model identification platforms. Comput. Chem. Eng. 124, 270–284. http://dx.doi.org/10.1016/j.compchemeng.2019.01.010, URL https://www.sciencedirect.com/science/article/pii/S0098135418310603.

Reizman, B.J., Jensen, K.F., 2012. An automated continuous-flow platform for the estimation of multistep reaction kinetics. Organic Process Res. Dev. 16, 1770–1782, URL https://api.semanticscholar.org/CorpusID:101623469.

SciPy Developers, 2022. Scipy v1.8.0 reference guide. https://docs.scipy.org/doc/scipy/reference/. (Accessed 28 March 2024).

Waldron, C., Pankajakshan, A., Quaglio, M., Cao, E., Galvanin, F., Gavriilidis, A., 2020. Model-based design of transient flow experiments for the identification of kinetic parameters. React. Chem. Eng. 5, 112–123. http://dx.doi.org/10.1039/C9RE00342H.

Walpole, R., Myers, R., Myers, S., Ye, K., 2016. Probability & Statistics for Engineers & Scientists, Global Edition. Pearson Education, URL https://books.google.co.uk/books?id=Th3aDAAAQBAJ.

Ye, K.Q., 1998. Orthogonal column latin hypercubes and their application in computer experiments. J. Amer. Statist. Assoc. 93 (444), 1430–1439. http://dx.doi.org/10.1080/01621459.1998.10473803, URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1998.10473803, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1998.10473803.

Zarrop, M.B., 1977. Optimal experiment design for dynamic system identification. URL https://api.semanticscholar.org/CorpusID:29963436.