



A review on model-based design of experiments for parameter precision – Open challenges, trends and future perspectives

Margherita Geremia^a, Sandro Macchietto^b, Fabrizio Bezzo^{a,*}

^a CAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering, University of Padova, Via Marzolo 9, 35131 Padova PD, Italy

^b Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

ARTICLE INFO

Keywords:

Mathematical modelling
Model-based design of experiments
Parameter estimation
Model calibration
Process systems engineering

ABSTRACT

In the last decades, the systematic use of mathematical models has become pervasive within the chemical process engineering context. Experiment design plays a critical role for the rapid calibration and validation of mathematical models – whether mechanistic, data-driven or semiempirical models – with the ultimate goal of driving process development and optimisation. Since data is often limited, and running additional experiments to gather more information for model calibration might be expensive or impractical, model-based design of experiments (MBDoE) techniques have become increasingly relevant to support the model calibration task on a broad range of process engineering applications. In this review, we aim to provide a comprehensive overview of the advances in the field that have occurred since the previous work presented by Franceschini and Macchietto (2008), *Chemical Engineering Science*, 63, 4846–4872. We first provide the theoretical foundations behind the standard MBDoE problem formulation and highlight recent advances; we then thoroughly analyse limitations, open challenges, and current trends by discussing about 250 contributions in the field. Finally, we highlight future research directions that may enlarge and further enhance the robustness and reliability of MBDoE implementation in real industrial scenarios.

List of symbols

c	set of process constraints
d	disturbance
D	deviation matrix in extended Fisher information matrix
E	extended Fisher information matrix
$E_{\theta}(\hat{y}_i)$	unconditional expectation of the model output $\hat{y}_i$ over the joint distribution of all parameters
$E_{\theta_p}(\hat{y}_i \theta_p)$	conditional expectation of the output given a fixed value of θ_p , averaged over all the other parameters θ_p
$\bar{E}$	matrix of normed eigenvectors
f	vector field representing differential and algebraic equations
g	set of active constraints
H_{Bayes}	Bayesian counterpart of Fisher information matrix
H_{θ}	Fisher information matrix
H_{θ}^0	prior dynamic information matrix
$\bar{H}_{\theta}$	expected value of Fisher information matrix according to sigma point method
H_{θ}^{*-1}	Moore Penrose pseudoinverse of the Fisher information matrix
J_G	scalar index representing the prediction variance
$J_{G,\max}$	maximum value of G-optimality
$J_{G,\text{thr}}$	threshold chosen by the user

(continued on next column)

(continued)

List of symbols	
c	set of process constraints
k	scaling parameter in sigma points definition
$\hat{m}$	mean of a Gaussian Process
N_M	number of candidate models
N_{sp}	number of sampling intervals
N_y	number of measured variables
N_{θ}	number of model parameters object of estimation
P	weight matrix
$p(\hat{y})$	marginal likelihood
$p(\hat{y} \theta)$	likelihood function
$p(\theta)$	prior density over the parameters
$p(\theta \hat{y})$	posterior density over the parameters
r	maximum order of derivatives
R_{ij}	correlation factor between the pair of parameters θ_i, θ_j
$S_{i,j}$	i, j^{th} element of the inverse of the measurement error covariance matrix
S_i	first-order sensitivity index (of parameter i)
$S_{i,\text{TOT}}$	total effect sensitivity index (of parameter i)
$\mathbf{1}$	diagonal matrix
t	time

(continued on next page)

* Corresponding author.

E-mail address: fabrizio.bezzo@unipd.it (F. Bezzo).

(continued)

List of symbols	
$\mathbf{c}$	set of process constraints
t_0	initial time
$\mathbf{t}_{sp}$	vector of the output variables sampling times
$\mathbf{u}$	vector of control or input variables
σ_1	nonzero singular values in the diagonal matrix σ_1
σ_2	zero singular values
U_{KL}	Kullback-Leibler divergence from the likelihood to the evidence
$\mathbf{V}_y$	prediction variance
$\mathbf{V}_\theta$	variance-covariance matrix
σ_1	nonzero singular values in the diagonal matrix σ_1
σ_2	zero singular values
$\mathbf{W}^{MD}$	selection matrix to tailor the discrimination to a specific subset of candidate models
$\mathbf{w}^{PE}$	selection vector to minimize the parameter uncertainty of selected models
$w_{1,ii}$	weight coefficient associated to each model parameter θ_i
$w_{2,ij}$	weight coefficient associated with the pair θ_i, θ_j
$\mathbf{x}$	vector of state variables
$\mathbf{y}$	measurable output variables
$\hat{\mathbf{y}}$	vector of predicted output variables
$\mathbf{y}^0$	vector of initial conditions
$\hat{\mathbf{y}}^{flat}$	flat output vector

Greek letters

α	user selected constant variable in Eq. (41)
β	set of time-dependent backoff functions
ε^{MAX}	design optimality computed for parameter precision
ε^{MIN}	design optimality computed for model discrimination
ε_i^{λ}	specified lower bound for the increase of λ_i over the initial value λ_i^0
θ	vector of model parameter estimates
$\tilde{\theta}$	vector of model parameter estimates treated as stochastic variables
Θ	linear combination of the original parameter estimates
Λ	diagonal matrix containing the eigenvalues of $\mathbf{H}_\theta$
$\hat{\Sigma}$	variance of a Gaussian Process
π	sigma points
$\sigma_{y,i}^2$	estimated variance for the measured response y_i
τ	total duration of the experiment
φ	design vector
φ_{OPT}	optimal elements of the design vector
φ_{OPT}^{JD}	optimal experimental conditions for joint-MBDoE
φ_{OPT}^{MD}	optimal experimental conditions for model discrimination
ψ	selected metric of FIM
Ψ_H	Hunter and Rainer discrimination function
Ψ_B	Buzzi-Ferrari and Forzatti discrimination function

Acronyms

ANN	artificial neural network
DAE	differential and algebraic equations
DoE	design of experiments
FIM	Fisher information matrix
GP	Gaussian Process
GSA	global sensitivity analysis
HTE	high-throughput experimentation
MBDoE	model-based design of experiments
MC	Monte Carlo
OMBRE	Online Model-Based Redesign of Experiments
RFI	relative Fisher information
SDL	Self-Driving Laboratory

1. Introduction

Recent advances in computational methods within the field of process systems engineering have led to the growing adoption of mathematical models as suitable supporting tools for accelerating process development and optimisation (Pistikopoulos et al., 2021; Buede and Miller, 2024). Mechanistic models are typically formulated as systems of differential and algebraic equations (DAEs), with the model structure reflecting the causal mechanisms of the physical process (Bonvin et al.,

2016). Appropriate physics-based assumptions can provide a deep understanding of the underlying phenomena within the system under investigation and enable predicting dynamic behaviour beyond conditions previously observed.

The degree of knowledge and understanding of the process directly affects the model structure and its complexity in terms of both model equations and number of adjustable model parameters to be estimated from experimental measurements (Alizadeh et al., 2020). In this context, the modeller should investigate (i) whether it is possible to determine the values of all model parameters from measured data (i.e., model identifiability), (ii) whether the precision of the parameter estimates is statistically satisfactory, and, if different candidate models exist, (iii) which candidate most accurately predicts the real process behaviour. It is therefore crucial to design experimental campaigns so as to facilitate the parameter estimation and/or model discrimination activities. This review will mostly focus on the parameter estimation task.

The idea of optimal experiment design is not new and has been well established in the field of automation for many years (e.g., Lindqvist and Hjalmarsson, 2001; Hjalmarsson, 2005, 2008). Optimal experiment design is key for several reasons: (i) to aid parameter estimation under the assumption that the available model is structurally identifiable; (ii) to save material resources; (iii) to reduce measurements cost when lab assays are required; (iv) to lower the total time of the experimental campaign.

Traditional design of experiments (DoEs) approaches have their roots in the foundational work of Fisher (1925, 1934), who introduced key concepts like randomization, blocking, and factorial designs (Murphy, 2022). Box and Wilson (1951) viewed the output of an experiment as a response surface function of its inputs, with experiment design aiming at finding the optimum conditions to maximise this function. All these methods vary a subset of control (also input or manipulated) variables and analyse the corresponding outcomes (Box et al., 1978), and are often referred to as *black box* methods as they do not explicitly incorporate any model-based understanding of the system physics. In practice, they represent an attempt to improve on the simply random sampling of input space given by the Monte Carlo (MC) method (Metropolis and Ulam, 1949; Mascagni, 2015), or its more refined versions such as Sobol's sequences (Sobol, 1967). As a result, they can be time-consuming, resource-intensive, and may offer sub-optimal information. Also, any constraints on equipment, internal variables or outputs are not considered. They are not often applied in the context of mechanistic models; rather, they are commonly exploited to study the static relationship between process inputs and outputs. A significant advancement in this field is represented by the Design of Dynamic Experiments (Georgakis, 2013), which incorporates time-dependent input variations into the experimental conditions. However, even within this framework, the absence of a mechanistic model limits the ability to leverage physical knowledge of the system. More recently, the use of machine learning has become an increasingly popular trend, particularly in Bayesian parameter estimation (Nolan et al., 2021); however, significant amounts of data and extensive training are needed.

In contrast, the model-based design of experiments (MBDoE) approach emerges as a more powerful strategy to maximize the mechanistic informational content that can be gained from an experimental campaign (the following contributions provide some interesting reviews and general perspectives on the topic: Chaudhuri and Mykland, 1993; Franceschini and Macchietto, 2008a; Pronzato, 2008; Chakrabarty et al., 2013).

The MBDoE framework arises from the pioneering works on optimal experimental design by Bard (1974), and Goodwin and Payne (1977), who laid the theoretical basis for experiment planning in nonlinear and dynamic systems. The basic principle of MBDoE is to exploit the physical knowledge of the system under investigation – which is embedded into the available mathematical model – to guide the design of the experimental campaign such that collected measurements provide the maximum information value for model calibration and/or

discrimination. In other words, the model itself becomes a fundamental component of the experiment design, reducing the number of required experiments while optimising resource use. An optimisation framework is used for the solution of the resulting problem. [Espie and Macchietto \(1989\)](#) were the first to formulate the MBDoE for general models as an optimal control problem, while [Zullo \(1991\)](#), in an unpublished PhD thesis, developed much of the underlying computational machinery for computer-aided experimental design – although practical implementation proved challenging due to the numerical optimisation capabilities available at the time. Notably, several research groups have made long-standing contributions that have pioneered and advanced the formulation and application of MBDoE across multiple domains ([Lhomann et al., 1992](#); [Vanrolleghem et al., 1995](#); [Versyck et al., 1997](#); [Bauer et al., 2000](#); [Asprey and Macchietto, 2000](#); [Faller et al., 2003](#); [Bardow and Marquardt, 2004](#); [Banga and Balsa-Canto, 2008](#)), laying the groundwork for modern methodological developments. For more details concerning the early years of MBDoE development, please refer to the review papers mentioned previously. [Fig. 1](#) elucidates the differences in terms of information fluxes between traditional DoE and MBDoE for parameter estimation. Note that a key difference is that within the DoE framework, typically the “output” of the experiment design activity is an empirical model correlating input and output variables, whereas in MBDoE the mechanistic model is an “input” to the experiment design task.

It is worth highlighting that approaches originating from the control and automation fields, such as adaptive input design and dual control, fundamentally address similar problems as MBDoE. However, these approaches have distinct origins and typical use cases. Adaptive input design and dual control are primarily developed for relatively simple and often linear models, where the main objective is to balance real-time control and learning ([Hjalmarsson, 2005](#); [Stigter and Molenaar, 2012](#); [Goodwin and Sin, 2014](#)). In contrast, MBDoE focuses on the estimation of parameters in complex nonlinear dynamic models, commonly encountered in process systems engineering.

Over the last years, many contributions have advanced the theory and practice of MBDoE, addressing various strategies, challenges, and

case studies. With reference to the previous foundational review on MBDoE for parameter estimation by [Franceschini and Macchietto \(2008a\)](#), three critical directions of development within the MBDoE field were already identified, which still remain relevant today. Specifically:

- (i) Limitations arising from parameter uncertainties in the MBDoE framework, and the potential benefits of robust design techniques.
- (ii) Challenges arising in the estimation of models with large number of parameters, and the potential use of screening tools, to identify and exclude less influential model parameters, thereby reducing computational burden and avoiding numerical issues arising from increasing model complexity.
- (iii) The opportunity of using multi-objective optimisation problem formulations to integrate cost, performance and reliability, aiming at offering a more practical solution to practical industrial problems.

A new direction with respect to [Franceschini and Macchietto \(2008a\)](#) is the extension of MBDoE methodologies beyond their classical applications in process engineering, exploring new domains where model-based experiment design can provide significant benefits (as will be discussed in [Section 3.4](#)).

This review aims at providing a comprehensive overview of the key advances in the field, while also discussing future research opportunities that could further enhance the application of MBDoE in process systems engineering. The manuscript is organised as follows. In [Section 2](#), we recall the fundamentals regarding the MBDoE problem formulation for parameter estimation, and briefly touch on its application for model discrimination. Readers already familiar with these concepts may proceed directly to [Section 3](#), where we thoroughly analyse current trends and progress in MBDoE use and formulation. [Section 4](#) discusses future research directions that could be particularly beneficial to strengthen a more reliable adoption of MBDoE approaches in real case scenarios. Our conclusions are drawn in the final Section. [Fig. 2](#) presents a schematic

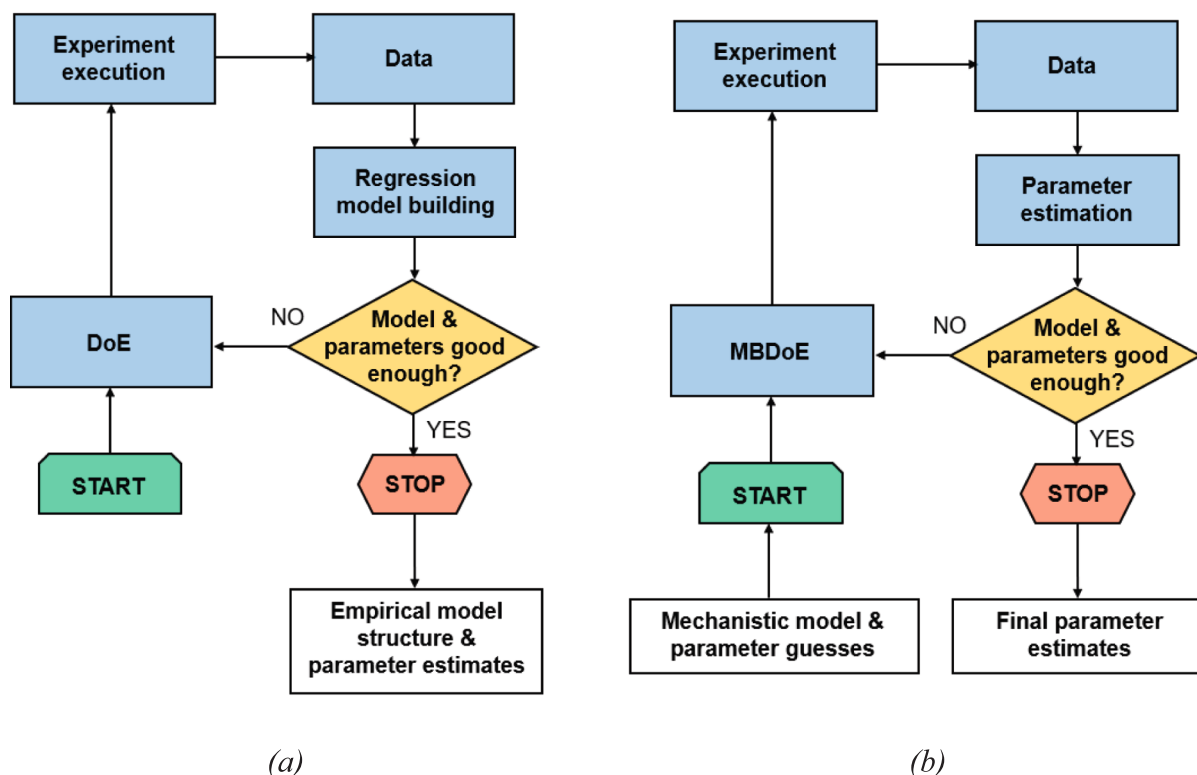


Fig. 1. Differences in terms of information fluxes between traditional DoE (a) and MBDoE for parameter estimation (b).

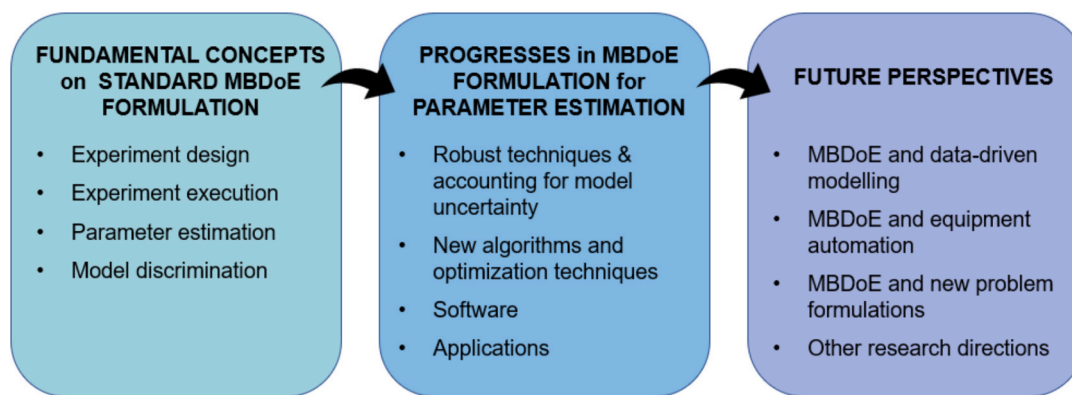


Fig. 2. Visual summary of the review structure.

overview of the article structure to aid in understanding the logical flow and navigating the review.

2. Fundamental concepts on standard MBDoE problem formulation

In the context of (optimal) experiment design, the activity of designing an experiment encompasses several critical decisions that must be made prior to the execution. The core of the design consists in determining the experimental variables (such as concentration of reactants and the flow of cooling water in a reactor) to be manipulated during the experimental campaign, while ensuring that they are consistent with the constraints of the system (Fisher, 1925; Box and Wilson, 1951). The design must be developed with a thorough understanding of the system being studied, ensuring that the chosen conditions guarantee meaningful data for the intended use, e.g., parameter estimation. An important aspect of the experiment design activity is the experimental budget, which reflects the financial and logistical issues that may limit the scope of the experiment – this is particularly relevant in the (bio)pharmaceutical field (Destro and Barolo, 2022). The costs associated with experiments can vary widely, depending on factors such as the need for specialized equipment and/or expensive raw materials, and time required to collect data. These limitations must be carefully addressed during the design phase in order to maximize the information obtained within the available resources. In practice, this involves optimising the allocation of resources and prioritizing experiments that offer the highest potential for obtaining informative and reliable results.

Focusing on the standard MBDoE procedure to design an experiment for parameter estimation, three key activities are required (Fig. 3): (i) the design of the experiment, based on the preliminary (or current) knowledge about the system; (ii) the execution of the experiment at the experimental conditions planned by design and collection of data; (iii) the (re)estimation and statistical assessment of the model parameters based on the new data.

A defining characteristic of the MBDoE methodology is that the core activity of the experiment design is driven by the model itself, which is used to predict the dynamic behaviour of the process under investigation and the expected information content obtained from the new experiment. In mathematical terms, the relationship between the process inputs and outputs can be described by the following set of DAEs:

$$\begin{cases} f(\mathbf{x}(t), \dot{\mathbf{x}}(t), \mathbf{u}(t), \theta, t) = 0 \\ \hat{\mathbf{y}}(t) = h(\mathbf{x}(t)) \end{cases}, \quad (1)$$

where $\mathbf{x}(t)$, $\mathbf{u}(t)$, and θ are (respectively) the vector of time-dependent state variables, the vector of controls or input variables that can be typically manipulated during the experiment, and the vector of the model parameter estimates; $\hat{\mathbf{y}}$ is the vector of predicted output variables, while vector field $\mathbf{f}$ represents all other DAEs. Vector θ contains the best current parameter estimates, which are characterized by a certain level of precision – that is generally defined by suitable statistical information. System (1) is typically subject to a number of process constraints, $\mathbf{C}$, that can be expressed as:

$$\mathbf{c} = \mathbf{x}(t) - \mathbf{g}(t) \leq 0, \quad (2)$$

where $\mathbf{g}(t)$ stands for set of active constraints, which typically include limits on both the outputs and internal states, as well as possible constraints on the control inputs.

A vector of initial conditions, $\mathbf{y}^0$, is required to solve the DAEs system:

$$\mathbf{y}^0 = \begin{cases} f(\mathbf{x}(t_0), \dot{\mathbf{x}}(t_0), \mathbf{u}(t_0), \theta, t_0) = 0 \\ \hat{\mathbf{y}}(t_0) = h(\mathbf{x}(t_0)) \end{cases}, \quad (3)$$

with t_0 being the initial time. The MBDoE framework for parameter estimation relies on the following assumptions about both the model and the physical system: (i) the model provides a structurally correct representation of the process under investigation; (ii) the values of all model

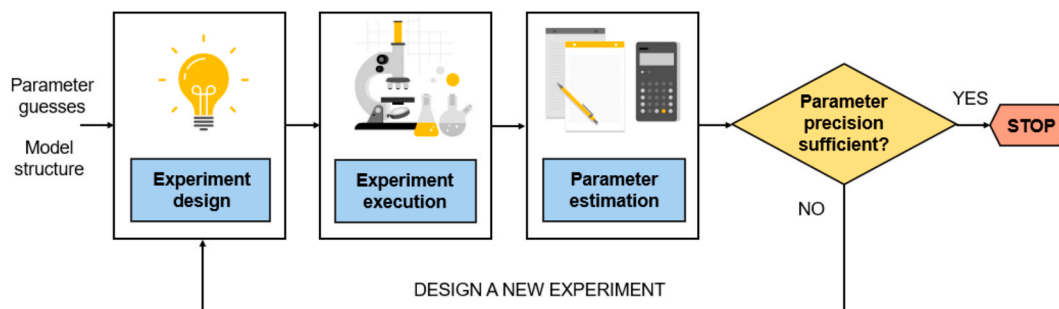


Fig. 3. MBDoE for parameter estimation: standard procedure. Adapted from Asprey and Macchietto (2000).

parameters can be determined from measured data (i.e., the model is identifiable (Miao et al., 2011)), and (iii) the physical system can be perfectly controlled during the experiment (i.e., the input variables can be manipulated exactly as planned during the experiment design phase). These assumptions, while foundational for traditional MBDoE theory, are often challenged in practice. The issue of process-model mismatch has recently gained increasing attention in the MBDoE literature and will be discussed in detail in Section 3.1.4. On the other hand, when structural or practical (because of noisy and/or scarce measurements) identifiability is compromised, model-based ranking and parameter selection techniques (e.g., sensitivity analysis (Saltelli et al., 2008) and estimability analysis (McLean and McAuley, 2012)) can still support design decisions. More generally, this issue reflects a fundamental trade-off between the mechanistic fidelity of the model and its identifiability based on available data (Walter and Pronzato, 1997), and highlights the broader need for MBDoE methodologies capable of coping with structurally uncertain or partially informed models.

2.1. Experiment design

A crucial point in the experiment design is the evaluation of the expected information predicted by the model. Its mathematical formulation is of paramount importance as the outcome of the optimisation of experiment design is a set of experimental conditions aimed at maximizing a particular measure of it.

2.1.1. Expected information

Given a set of preliminary experimental data, the expected information predicted by the model is quantified by the Fisher information matrix (FIM), $\mathbf{H}_\theta$ (Bard, 1974; Zullo, 1991):

$$\mathbf{H}_\theta(\theta) = \mathbf{H}_\theta^0 + \sum_{k=1}^{N_{sp}} \sum_{i=1}^{N_y} \sum_{j=1}^{N_y} s_{ij} \left[\frac{\partial \hat{y}_i(t_k)}{\partial \theta_l} \frac{\partial \hat{y}_j(t_k)}{\partial \theta_m} \right]_{l,m=1:N_\theta}, \quad (4)$$

where s_{ij} is the i,j th element of the inverse of the measurement error covariance matrix, N_{sp} is the number of sampling intervals, N_y is the number of measured variables, and N_θ is the length of θ . $\mathbf{H}_\theta^0$ is the prior dynamic information matrix, considering the statistical information about the parametric system before each trial is carried out. For instance, when the prior information can be expressed by simple bounds of variability for model parameters, a N_θ -dimensional hyper-rectangular region of variability can be built assuming a uniform distribution for the model parameters (Zullo, 1991). If no preliminary information is available at all, the term $\mathbf{H}_\theta^0$ in Eq. (4) can be neglected.

2.1.2. Design criteria

Here, we define what we consider to be the *best* experiment, which usually is the maximization of the expected information predicted by the model. This is typically carried out by maximizing a metric ψ of the Fisher information matrix, or more frequently, by minimizing a metric ψ of the variance–covariance matrix of model parameters (Franceschini and Macchietto, 2008). The variance–covariance matrix, $\mathbf{V}_\theta$, is mathematically defined as the inverse of FIM (Bard, 1974):

$$\mathbf{V}_\theta(\theta) = \left\{ \mathbf{H}_\theta^0 + \sum_{k=1}^{N_{sp}} \sum_{i=1}^{N_y} \sum_{j=1}^{N_y} s_{ij} \left[\frac{\partial \hat{y}_i(t_k)}{\partial \theta_l} \frac{\partial \hat{y}_j(t_k)}{\partial \theta_m} \right]_{l,m=1:N_\theta} \right\}^{-1}. \quad (5)$$

The optimal experimental settings are selected by acting on the elements of the design vector φ :

$$\varphi = [\mathbf{y}_0, \mathbf{u}(t), \mathbf{t}_{sp}, \tau]^T, \quad (6)$$

where $\mathbf{t}_{sp}$ is the vector of the output variables sampling times (assuming that, for some variables, only a limited number of measurements can be made available), and τ is the total duration of the experiment. The

optimisation problem to compute the optimal elements of design vector, φ_{OPT} , can be formulated based on the maximization of a certain metric, ψ , of the FIM as:

$$\varphi_{OPT} = \operatorname{argmax}_{\varphi} \{ \psi[\mathbf{H}_\theta(\theta, \varphi)] \}, \quad (7)$$

or, more frequently, based on the minimization of a certain metric, ψ , of the variance covariance matrix, as:

$$\varphi_{OPT} = \operatorname{argmin}_{\varphi} \{ \psi[\mathbf{V}_\theta(\theta, \varphi)] \}. \quad (8)$$

During the optimisation, the set of model parameters is fixed at the current estimates. As the quality of the parameter estimates is related to their confidence intervals, the most used design criteria to select the metric ψ are the so-called *alphabetic criteria* (Kiefer, 1959; Pukelsheim, 1993); details are reported in the Supplementary Material (Section S.1).

Standard MBDoE formulations can be implemented in a sequential or parallel manner, depending on the experimental setup and the computational resources available. Sequential designs allow for iterative refinement based on data collected at each step: after each experiment, parameter estimates are updated and the experimental design is re-optimised accordingly. This structure can improve robustness against model uncertainty, particularly during early stages of model development. However, the need to solve repeated optimisation problems makes sequential approaches computationally demanding over time. By contrast, parallel designs optimise multiple experiments simultaneously in a single optimisation problem. This can be advantageous when experimental facilities support simultaneous tests (e.g., high-throughput platforms), and no intermediate feedback is needed. When experiments are very long and elapsed time is limited, this may be the preferred option. However, the resulting optimisation problem includes a larger number of decision variables, proportional to the number of experiments or experimental units, which may increase computational complexity. In this regard, Sandrin et al. (2025) proposed a computational framework that enhances the tractability of parallel designs by combining effort-based optimisation over a discretised experimental domain with gradient-based refinement, allowing for globally optimal parallel designs under model uncertainty while maintaining computational tractability. Nonetheless, the trade-off between adaptability and optimisation complexity must be carefully considered based on the specific experimental and computational context (Galvanin et al., 2007).

2.2. Experiment execution

The new experiment is executed at the experimental conditions planned by design. The level of information that can be extracted from the experimental campaign is strongly related to the available equipment (control inputs and measurements). In the experimental practice, it should be possible to measure the control variables, $\mathbf{u}(t)$, and the set of system responses, $\mathbf{y}$, in each performed experiment. Note that the measured responses can be either “almost” continuous or sampled. The time resolution in capturing the behaviour of the system depends on the relationship between the sampling time and the dominant dynamics of the process. If the sampling time is sufficiently small compared to the system dynamics, the data can be treated as “almost” continuous; conversely, a low sampling rate may lead to quantization effects (Seborg et al., 2016).

For parameter estimation purposes, the set of measured variables should be as free as possible from the effect of disturbances, and from random and/or systematic measurements errors. The MBDoE procedure is built to improve the parameter estimation by exploiting the information embedded in the set of measured responses $\mathbf{y}$. To account for the distribution of the measurements errors, a variance model can be adopted to describe all the characteristics of the measurements system for the given experiment. Details are given in Section S.2 of the Supplementary Material.

2.3. Parameter estimation

The final step is the estimation of model parameters from the available experimental measurements. Several methods can be adopted, which are described in Section S.3 of the [Supplementary Material](#).

2.4. MBDoE for model discrimination

When different candidate models are available (i.e., models incorporating different assumption, and, thus, described by different mathematical structure), an MBDoE framework can be adopted to assist in the designing of optimal experiment for distinguishing between models responses. Although this review focusses on MBDoE for parameter estimation, a discussion on model discrimination is also necessary both for sake of completeness and because the two goals often intertwined and led to combined approaches.

The optimal set of experimental conditions for model discrimination, ϕ_{OPT}^{MD} , may be computed by maximizing a discrimination function, Ψ , that evaluates the difference between predictions of the candidate models (Vega-Ramon et al., 2024). The simplest discrimination function Ψ is that proposed by Hunter and Reiner (1967), which can be formulated as:

$$\Psi_H(\phi) = \sum_{M=1}^{N_M} \sum_{N=M+1}^{N_M} [(\hat{\mathbf{y}}_M(\phi) - \hat{\mathbf{y}}_N(\phi))^T (\hat{\mathbf{y}}_M(\phi) - \hat{\mathbf{y}}_N(\phi))], \quad (9)$$

where $\hat{\mathbf{y}}_M$ and $\hat{\mathbf{y}}_N$ are the set of responses predicted by models M and N among the N_M candidates, given the experimental conditions ϕ . A more commonly used formulation of Ψ is that presented by Buzzi-Ferraris and Forzatti (1983, 1984) and Schwaab et al. (2006):

$$\Psi_B(\phi) = \sum_{M=1}^{N_M} \sum_{N=M+1}^{N_M} [(\hat{\mathbf{y}}_M(\phi) - \hat{\mathbf{y}}_N(\phi))^T \mathbf{P} (\hat{\mathbf{y}}_M(\phi) - \hat{\mathbf{y}}_N(\phi))], \quad (10)$$

where $\mathbf{P}$ is a weight matrix which modulates the importance of the difference between the predictions of the candidates.

It is worth highlighting that the prediction capability of available candidate models strongly depends on the precision of the parameter estimates. To cope against this issue, Galvanin et al. (2016) and Deussen and Galvanin (2023) proposed a joint-MBDoE approach which aims at simultaneously improving parameter precision and discriminating among rival models through a multi-objective optimisation formulation. The optimal experimental set for the joint-MBDoE, ϕ_{OPT}^{JD} , is computed according to:

$$\phi_{OPT}^{JD} = \underset{\phi}{\operatorname{argmax}} \left\{ \sum_{M,N=1}^{N_M} w_{M,N}^{MD} P_M P_N \left[\sum_{i=1}^{N_y} \frac{(\hat{y}_{M,i} - \hat{y}_{N,i})^2}{\sigma_{y,i}^2} \right]_{N,M} \right\} \text{ s.t.} \quad (11)$$

$$\sum_{j=1}^{N_M} w_j^{PE} \frac{\|\mathbf{H}_{\theta j}\|}{N_M} \leq \varepsilon \text{ with } \varepsilon^{MIN} \leq \varepsilon \leq \varepsilon^{MAX},$$

where $\sigma_{y,i}^2$ is the estimated variance for the measured response y_i , $w_{M,N}^{MD}$ is the MN th element of a $N_M \times N_M$ dimensional selection matrix $\mathbf{W}^{MD}$ that is constituted by binary values [0,1] specified by the user to tailor the discrimination to a specific subset of candidates. w_j^{PE} is the j th element of the N_M dimensional selection vector $\mathbf{w}^{PE}$ that is constituted by binary values [0,1] specified by the user to minimize the parameter uncertainty of selected models. When $w_{M,N}^{MD} = 1$, the user is interested in discriminating between model M and N ; when $w_{M,N}^{MD} = 0$ the user is not interested in discriminating between model M and N . Similarly, when $w_N^{PE} = 1$, the user is interested in improving the parameter estimation for model N ; when $w_N^{PE} = 0$, model N is not considered in the optimal design. ε^{MIN} is the design optimality computed for model discrimination, while ε^{MAX} is that obtained for improving parameter estimation. The

optimisation in Eq. (11) is thus performed for a number of values of ε ranging in between ε^{MIN} and ε^{MAX} .

Schulze and Schenkendorf (2018, 2020) proposed an alternative approach for model discrimination based on the concept of differential flatness. A dynamic system of the form presented in Eq. (1) is said to be differentially flat if there exists a flat output vector, $\hat{\mathbf{y}}^{flat}(t)$, such that all system states and inputs can be expressed as explicit functions of it and a finite number of its time derivatives, up to a certain order r (Fliess et al., 1995). Mathematically, the flat output vector, $\hat{\mathbf{y}}^{flat}$, is expressed as:

$$\hat{\mathbf{y}}^{flat}(t) = \mathbf{h}^{flat}(\mathbf{x}(t)), \quad (12)$$

and the states and controls can be represented as explicit functions:

$$\mathbf{x}(t) = \boldsymbol{\zeta}_x(\hat{\mathbf{y}}^{flat}(t), \dot{\hat{\mathbf{y}}}^{flat}(t), \dots, \hat{\mathbf{y}}^{flat}(t)^{(r)}), \quad (13)$$

$$\mathbf{u}(t) = \boldsymbol{\zeta}_u(\hat{\mathbf{y}}^{flat}(t), \dot{\hat{\mathbf{y}}}^{flat}(t), \dots, \hat{\mathbf{y}}^{flat}(t)^{(r+1)}). \quad (14)$$

Flat systems are a special class of nonlinear systems for which such a representation exists. One of their key properties is that they allow trajectory planning and control synthesis to be carried out in a straightforward manner, without integrating the system dynamics. Moreover, for a system to be flat, the number of flat outputs must match the number of control inputs, i.e., $\dim \hat{\mathbf{y}}^{flat} = \dim \mathbf{u}$. In the context of MBDoE, the flatness property is exploited to compute analytical model-revealing inputs by maximizing the difference between flat outputs across different candidate models. This results in a computationally efficient solution to the model discrimination problem. However, the practical applicability of this method is limited by the fact that not all systems are flat – flatness is a structural property that must be verified a priori for each model under consideration.

Beyond traditional optimisation-based discrimination approaches, machine learning techniques have also been explored for model discrimination. They have gained increasing attention in recent years, as they have shown potential in identifying patterns from data without explicit model-based assumptions. For instance, Sangoi et al. (2024) developed a procedure that leverages artificial neural network (ANN) classifiers for rapid identification of kinetic model structures, integrated with optimal experiment design. The ANN-based approach involves training a classifier (Zou et al., 2009) on in silico-generated data for each candidate, enabling it to determine the most suitable kinetic model based on input features. A major advantage of the method is that it does not require explicit parameter estimation, as parameter uncertainty is inherently incorporated during the data generation step.

3. Progresses in MBDoE formulation for parameter estimation

In this Section we specifically focus on MBDoE application for parameter estimation. In the following, we discuss the current most relevant limitations and review recent developments aimed at addressing these issues. To facilitate comparison and understanding, Table 1 summarizes key references under each heading.

3.1. Development of robust techniques within standard MBDoE problem formulations and accounting for model uncertainty

3.1.1. Parametric uncertainty

A key issue in the standard MBDoE formulation is the uncertainty associated with the parameter estimates, which may significantly affect the ability to design informative experiments within the experimental space while satisfying all process constraints. The impact of parametric uncertainty influences both the reliability of model predictions, $\hat{\mathbf{y}}$, and the sensitivity of model responses with respect to parameter estimates (i.

e., the term $\frac{\partial \hat{y}_i(t_k)}{\partial \theta_l}$, appearing in Eq. (4) – this in turn affecting the

Table 1

Key references organised by topic.

Topic	References
Limitations in standard MBDoE problem formulations addressed	
Parametric uncertainty – Bayesian approach	Gibson et al. (2024); Cao et al. (2024); Rodrigues et al. (2022); Kalyanaraman et al. (2020); Overstall et al. (2020); Luo et al. (2015); Kalyanaraman et al. (2015); Terejanu et al. (2013).
Parametric uncertainty – Sigma point method	Mdluli et al. (2015); Telen et al. (2014)
Parametric uncertainty – Design objective based on worst-case metric	He et al. (2010); Barz et al. (2010)
Parametric uncertainty – Stochastic design	Nimmegeers et al. (2020); Mesbah and Streif (2015); Streif et al. (2014); Galvanin et al. (2010)
Parametric uncertainty – Global sensitivity analysis	Chu and Hahn (2010); He et al. (2009); Martinez et al. (2009); Brockmann et al. (2008); Chhatre et al. (2008); Yue et al. (2008).
Parametric uncertainty – Online experiment redesign	Pankajakshan et al. (2021); de-Luca et al. (2016); Galvanin et al. (2009a)
Parameter correlation	Kim and Lee (2018); Maheshwari et al. (2013a); Franceschini and Macchietto (2008b,c)
Ill-conditioning of the FIM	Quaglio et al. (2019); Shahmohammadi and McAuley (2019a,b); López et al. (2015); Barz et al. (2013); Kravaris et al. (2013); López et al. (2013); Li and Yeh (2012); Wu et al. (2008)
Effect of structural model-process mismatch	Ghodba et al. (2025); Petsagkourakis and Galvanin (2020); Quaglio et al. (2018); Galvanin et al. (2012)
New algorithms and optimisation techniques	
Local optimality	Sandrin et al. (2025); Friso and Galvanin (2024); Cenci et al. (2023); Kusumo et al. (2022a); Vanaret et al. (2021); Dürrbaum and Kroll, 2016
Multi-objective formulations	Wang et al. (2024); Bouchkira et al. (2023); Kusumo et al. (2022b); Pankajakshan et al. (2019); Nöh et al. (2018); Manesso et al. (2017); Tsay et al. (2017); Houska et al. (2015); Maheshwari et al. (2013a,b); Telen et al. (2012a); Franceschini and Macchietto (2008b,c).
Solution of the optimisation problems	Zhu et al. (2024); Bouchkira et al. (2023); Kaiser and Engell (2022, 2020); Tsay et al. (2017); Kuceroová et al. (2016); Steimel and Engell (2016); Kroll and Dürrbaum (2015); Telen et al. (2014b)

prediction of expected information. Assuming that the model structure is correct (i.e., no structural mismatch is present), the issue has been addressed through more robust design techniques. Fig. 4 illustrates how uncertainty in parameters affects uncertainty in predicted responses, leading to potential suboptimal design, while robust MBDoE techniques aim at addressing these issues.

3.1.1.1. Bayesian approach. A trending solution is the adoption of Bayesian design (Terejanu et al., 2013; Luo et al., 2015; Kalyanaraman et al., 2015; Kalyanaraman et al., 2020; Overstall, 2020; Rodrigues et al., 2022; Cao et al., 2024; Gibson et al., 2024). According to the Bayesian approach, the change in the information about the set of parameters estimates θ between the prior density over the parameters $p(\theta)$ and the posterior $p(\theta|\hat{\mathbf{y}})$ is given by Bayes' rule:

$$p(\theta|\hat{\mathbf{y}}) = \frac{p(\hat{\mathbf{y}}|\theta)p(\theta)}{p(\hat{\mathbf{y}})}, \quad (15)$$

where $p(\hat{\mathbf{y}}|\theta)$ is the likelihood function; $p(\hat{\mathbf{y}}) = \int p(\hat{\mathbf{y}}|\theta)p(\theta)d\theta$ is the marginal likelihood used for normalization. Within the MBDoE context, adopting a Bayesian approach aims at integrating the posterior distribution of model parameters $p(\theta|\hat{\mathbf{y}})$ rather than relying on point estimates. $p(\theta|\hat{\mathbf{y}})$ is used to represent the parametric uncertainty to account for the full range of possible parameter values. Then, the optimal experimental conditions are typically computed as to maximize the mutual information between θ and $\hat{\mathbf{y}}$, i.e., reducing the uncertainty in joint parametric distribution (Lindley, 1956). This measure is referred to Kullback-Leibler divergence from the likelihood to the evidence (Paulson et al., 2019):

$$U_{KL}(\theta) = \int \log\left(\frac{p(\hat{\mathbf{y}}|\theta)p(\theta)}{p(\hat{\mathbf{y}})}\right)p(\hat{\mathbf{y}}|\theta)d\hat{\mathbf{y}}. \quad (16)$$

The optimal experimental conditions, ϕ_{OPT} , are those maximizing the value of U_{KL} , i.e., the conditions that maximize the information gain about the parameters θ by minimizing the uncertainty in the posterior distribution.

Note that the optimisation problem can be also formulated by considering the Bayesian counterpart of the FIM (Wang and Yue, 2020) as:

$$\mathbf{H}_{\text{Bayes}}(\theta) = \int \mathbf{H}_{\theta}(\theta)p(\theta|\hat{\mathbf{y}})d\theta. \quad (17)$$

Although the Bayesian approach to MBDoE offers significant advantages in handling parameter uncertainty, some limitations must be considered. First, the computational complexity of integrating the selected optimisation function over the parameter space may be significant (Paulson et al., 2019). MC approaches have been proposed as a practical numerical solution within the Bayesian framework to approximate the integrals in Eq. (16) or (17) (He et al., 2009; Krausch et al., 2019), but their computational cost still can be prohibitive, especially for high-dimensional or non-linear models. Moreover, the choice of prior distribution $p(\theta)$ plays a critical role in shaping the results. Inaccurate or overly restrictive priors may lead to suboptimal designs, while non-informative priors may fail to adequately reflect prior knowledge, potentially resulting in inefficient experiments (He et al., 2009).

3.1.1.2. Sigma point method. An alternative approach to addressing parametric uncertainty has been proposed by Telen et al. (2014) and Mdluli et al. (2015). In this context, the FIM is efficiently computed using the Sigma Point method, which involves selecting a minimal set of strategically chosen values (sigma points) around each estimated model parameter, enabling an efficient evaluation of how such uncertainties influence optimal experiment design (Nimmegeers et al., 2017). The primary advantage of the Sigma Point method lies in its significantly lower computational cost compared to direct Bayesian approaches, which require evaluating high-dimensional integrals. In contrast to advanced computational techniques used within the Bayesian context (i.e., MC sampling), the Sigma Point framework provides more direct formulas that effectively leverage the problem structure. Specifically, for a current set of N_{θ} estimated model parameters with variance-covariance matrix $\mathbf{V}_{\theta}$, the evaluation of information requires $(2N_{\theta} + 1)$ computations, where the sigma points are defined as:

$$\pi_0 = \theta, \quad (18)$$

$$\pi_i = \theta + \sqrt{(N_{\theta} + k)\mathbf{V}_{\theta}}_{\theta i}, \quad (19)$$

$$\pi_j = \theta - \sqrt{(N_{\theta} + k)\mathbf{V}_{\theta}}_{\theta j - N_{\theta}} \text{ with } i = 1 : N_{\theta}, j = N_{\theta} + 1 : 2N_{\theta}, \quad (20)$$

where k is a scaling parameter which is typically fixed equal to $(3 - N_{\theta})$ (Julier and Uhlmann, 1996). The expected value of the FIM, $\bar{\mathbf{H}}_{\theta}(\theta)$, is finally computed as:

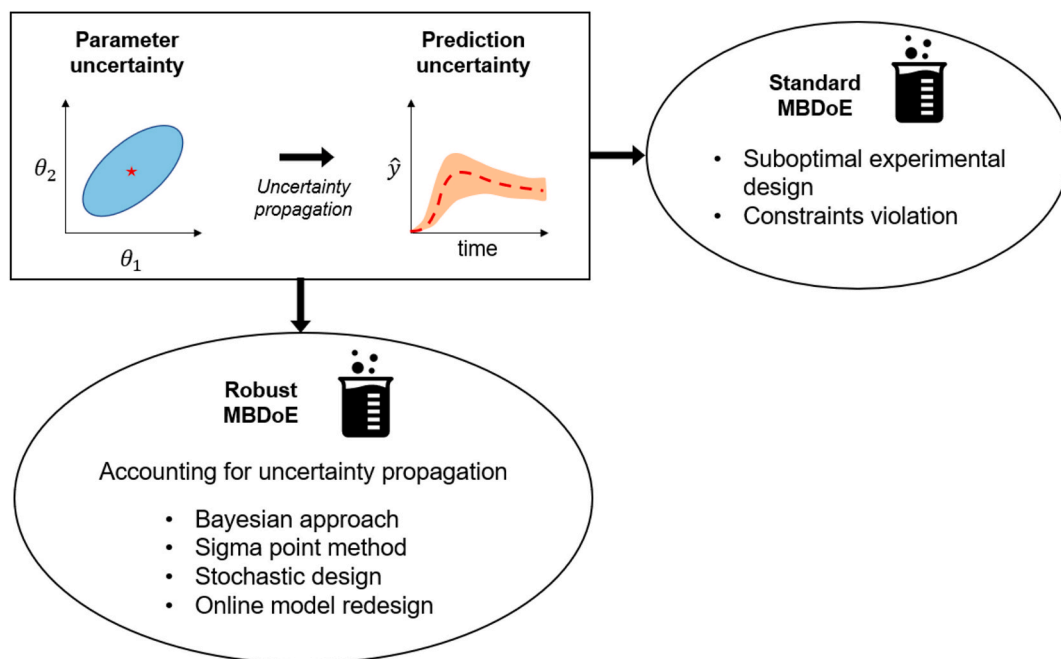


Fig. 4. Propagation of parametric uncertainty for a two-parameters model (θ_1 and θ_2) towards model prediction ($\hat{y}$) may lead to suboptimal experiment design and/or constraints violation, while robust MBDoE techniques aim at addressing these issues.

$$\bar{H}_\theta(\theta) = \frac{1}{N_\theta + k} k H_\theta(\pi_0) + \frac{1}{2} \sum_{i=1}^{2N_\theta} H_\theta(\pi_i). \quad (21)$$

3.1.1.3. Stochastic design. To account for the effect of parametric uncertainty in MBDoE, various stochastic approaches have been proposed. They involve considering all possible realizations of the set of model parameters within the expected uncertainty bounds and treating them as stochastic variables ($\tilde{\theta}$). Consequently, the design problem becomes inherently stochastic, with the solution represented as a stochastic design vector that satisfies model equations, optimality conditions, and feasibility constraints. Given that only one experiment is typically feasible at a time, the stochastic framework must be simplified to provide a unique optimal solution. The design objective can be evaluated using a conservative approach based on worst-case metric of parametric uncertainty (Walter and Pronzato, 1987; Asprey and Macchietto, 2002; Welsh and Kong, 2011). The general maximin design criterion based on FIM maximizes a selected metric of the FIM for the worst possible set of parameters, and is mathematically formulated as (He et al., 2010):

$$\varphi_{OPT} = \arg \max_{\varphi} \min_{\theta} \{ \psi[H_\theta(\theta, \varphi)] \}. \quad (22)$$

Alternatively, one can refer to the variance–covariance matrix, and use the minmax design criterion formulation (Barz et al., 2010):

$$\varphi_{OPT} = \arg \min_{\varphi} \max_{\theta} \{ \psi[V_\theta(\theta, \varphi)] \}. \quad (23)$$

While the worst-case approach ensures robustness by preparing for the most adverse parameter configurations, it often leads to overly conservative designs that may not be practical or efficient. In contrast, probabilistic experiment design approaches have been proposed, where the occurrence of different possible scenarios is considered with appropriate weighting (Galvanin et al. 2010; Streif et al., 2014; Mesbah and Streif, 2015; Nimmegeers et al., 2020).

Galvanin et al. (2010) proposed a backoff strategy that provides a balanced alternative by mitigating excessive conservatism and delivering solutions that remain robust within realistic uncertainty bounds. The computation of the FIM remains as defined in Eq. (4), while nominal constraints are adjusted using backoff margins to maintain the system

within a feasible region with a specified probability. This modifies Eq. (2) as follows:

$$c = x(t) - g(t) + \beta(\tilde{\theta}) \leq 0, \quad (24)$$

where β is a set of time-dependent backoff functions considering the effect of parametric uncertainty on the state variables at the designed experimental conditions. Backoff functions are quantified based on the predicted probability distribution of the state variables under parametric uncertainty, and reflect a specified confidence level. They are generally nonlinear, and their magnitude can be tuned by the user, allowing for variable degrees of conservatism depending on the criticality of each state variable, e.g., via MC simulations based on parametric or expected structural uncertainty. A key consideration is that large values may lead to overly conservative and poorly informative designs. One advantage of this approach is the ability to tailor backoff constraints for individual state variables and/or at specific times or conditions. Note that in this case, too, the reduction of the design space may lead to an unfeasible design.

Nimmegeers et al. (2020) presented an efficient approach for stochastic optimal experiment design based on polynomial chaos expansion, which is exploited to approximate the expected value of both the objective function and the process constraints as a sum of orthogonal polynomials (i.e., polynomials of which the inner product is equal to zero) through collocation points. These polynomials are a function of the uncertain variable for which a probability distribution is known (Mesbah and Streif, 2015; Nimmegeers et al., 2016).

3.1.1.4. Global sensitivity analysis. To further refine the robustness of the design under parametric uncertainty, the use of global sensitivity analysis (GSA) has been proposed as an alternative to traditional local sensitivity analysis indices in Eq. (4) (Brockmann et al., 2008; Chhatre et al., 2008; Yue et al., 2008; Martinez et al., 2009; He et al., 2009; Chu and Hahn, 2010). Instead of evaluating the effect of small perturbations around nominal parameter values, GSA considers the full range of parameter uncertainty and quantifies their contributions to output variability over the entire parameter space. Variance-based GSA, such as the Sobol's method (Sobol, 1993), are commonly preferred over

alternative approaches (e.g., Morris method) as they allow for a more thorough understanding of how uncertainty in parameters influences model predictions (Saltelli et al., 2008). The conditional variance characterizes the individual contribution of a parameter θ_l to the total variance of the predicted output $\hat{y}_i$, and is computed as:

$$\hat{y}_i(\theta) = \hat{y}_{i*} + \sum_{l=1}^{N_\theta} \hat{y}_{i,l}(\theta_l) + \sum_{1 < j < t < N_\theta} \hat{y}_{i,jt}(\theta_j, \theta_t) + \dots + \hat{y}_{i,12\dots N_\theta}(\theta_1, \dots, \theta_{N_\theta}), \quad (25)$$

where the partial functions are:

$$\hat{y}_{i*} = E_{\theta}(\hat{y}_i), \quad (26)$$

$$\hat{y}_{i,p}(\theta_p) = E_{\theta_p}(\hat{y}_i|\theta_p) - \sum_{s \neq p} \hat{y}_{i,s} - \hat{y}_{i*}, \quad (27)$$

with $E_{\theta}(\hat{y}_i)$ denoting the unconditional expectation of the model output $\hat{y}_i$ over the joint distribution of all parameters, while $\sim p$ indicating the complementary subset of p , i.e., θ_p represents the set of all parameters except θ_p . $E_{\theta_p}(\hat{y}_i|\theta_p)$ is, instead, the conditional expectation of the output given a fixed value of θ_p , averaged over all the other parameters θ_p ; this isolates the influence of parameter θ_p on the output while accounts for uncertainty in the remaining parameters.

The orthogonal property of the function decomposition (i.e., $\int \hat{y}_{i,p}(\theta_p) \hat{y}_{i,v}(\theta_v) p(\theta) d\theta = 0$ with $p \neq v$, p and v different subsets) allows to express the variance of the response of interest $V(\hat{y}_i)$ as:

$$V(\hat{y}_i) = \sum_{p=1}^{N_\theta} V_p + \sum_{1 < j < t < N_\theta} V_{jt} + \dots + V_{1\dots N_\theta}, \quad (28)$$

with

$$V_p = V_{\theta_p}(E_{\theta_{-p}}(\hat{y}_i|\theta_p)) - \sum_{s \neq p} V_s. \quad (29)$$

V_i , V_{it} , ..., $V_{123\dots N_\theta}$ are partial variances which describe the parameters effect on the variance of the response. According to Homma and Saltelli (1996), results from Sobol's GSA can be collected in two different metrics, i.e., the first-order effect index S_i , and the total effect index $S_{i,TOT}$. The first-order effect index, S_i , represents the direct effect contribution of each parameter l to the variance of the output: the higher the S_i value, the higher the influence of the l^{th} parameter on the output. With respect to y_i , it is defined as:

$$S_i = \frac{V_{\theta_i}(E_{\theta_{-i}}(\hat{y}_i|\theta_i))}{V(\hat{y}_i)}, \quad (30)$$

and can be substituted to the term $\frac{\partial \hat{y}_i(t_k)}{\partial \theta_l}$, appearing in Eq. (4), for each sampling time t_k . One advantage of this approach is the possibility of augmenting the problem formulation with additional constraints for each parameter on either individual S_i , or total effect index $S_{i,TOT}$ (i.e., designs where a specific parameter is above or below desired sensitivity thresholds).

3.1.1.5. Online Model-Based redesign of experiments. Alternative to traditional offline design approaches (one or more full experiments at a time), dynamic methodologies such as Online Model-Based Redesign of Experiments (OMBRE) offer innovative solutions for managing parameter uncertainty in real-time, i.e., during each experiment (Galvanin et al., 2009a). According to the traditional procedure for parameter estimation, parameters are estimated after the experiment execution. This means that the increasing information available as the experiment unfolds is not exploited until its completion, potentially leading to sub-optimal experimental conditions, particularly when initial parameter uncertainty is high and/or experiments are long. Instead, OMBRE

techniques enable real-time adaptation of experimental conditions as new data becomes available. The core principle of online redesign is to iteratively update the experiment design based on data collected during the experiment. After each experimental cycle (which is defined based on factors such as the elapsed time or number of samples collected), parameter estimates are refined, and experimental conditions are recalculated to maximize information gain – i.e., the FIM is updated iteratively as new measurements are integrated. Building on this foundation, De-Luca et al. (2016) introduced a more robust OMBRE methodology by proposing a new design criterion based on maximizing dynamic information gain. The method continuously adapts the experimental conditions to maximize the precision of model parameter estimates and includes an automatic mechanism to determine the optimal moments for redesigning the experiment, ensuring that a significant increase in information content is achieved before proceeding with parameter estimation and subsequent redesign. Specifically, the criterion is applied to the parameter with the highest uncertainty, using statistical indices like the t-value, which measures the precision of parameter estimates. More recent developments in the field (Pankajakshan et al., 2021) have led to the proposal of an adaptive OMBRE which dynamically determines when to re-optimize the experimental conditions based on the intermediate results of parameter estimation. The approach shows close similarity to parameter updates in online control, such as those implemented using the extended Kalman filter, where parameters are continuously updated based on real-time measurements.

The growing interest in real-time experimental design has also led to the development of adaptive strategies, such as adaptive input design and dual control, where system identification and control objectives are addressed simultaneously (Hjalmarsson, 2005; Stigter and Molenaar, 2012). These approaches integrate MBDoE with real-time feedback and control schemes and provide a rigorous foundation for managing uncertainty by continuously updating the experimental policy in response to new measurements. Recent contributions have advanced this viewpoint, proposing control-oriented frameworks for online MBDoE in various contexts. For instance, Lucia and Paulen (2014) introduced a robust model predictive control formulation that integrates experiment design; Qian et al. (2014) focused on closed-loop MBDoE for parameter estimation in thermal systems. Barz et al. (2016) proposed an adaptive input design approach for the identification of chromatography models; Jost et al. (2017) presented a feedback control algorithm that optimises both inputs and measurement times. Telen et al. (2018) analysed online vs. offline strategies for bioprocesses under uncertainty; Kim et al. (2022) applied an optimal experimental design strategy supported by model predictive control to automated cultivation platforms.

3.1.2. Parameter correlation

A crucial aspect in the formulation of the FIM is the potential correlation between parameters within the model. The FIM (Eq. (4)) is derived from the partial derivatives of model predictions with respect to parameter estimates, which are influenced not only by the individual values of the parameters but also by their interdependencies. Failure to account for these correlations may lead to misleading conclusions, as the covariance structure of the parameters affects both the distribution of possible parameter values and the reliability of model predictions. Indeed, it may be observed that high correlation is actually built-in in the choice of many traditional alphabetic design objectives. For example, the A and D-criteria are minimised when the parameter confidence ellipsoid collapses to a line (in 2D), and this is often obtained in practice (see Fig. S.1 in the Supplementary Material). In cases where parameters are highly correlated, information about one parameter may be shared with another, leading to an overestimation of the information available from the experiment (Franceschini and Macchiato, 2008b,c; Chu, Y., Hahn, J., 2012; Maheshwari et al., 2013; Yuan and Benyahia, 2023; Kim and Lee, 2018).

A typical way to calculate the correlation R_{ij} between the pair of

parameters θ_i , θ_j is to directly exploit the variance–covariance matrix itself (Maheshwari et al., 2013):

$$R_{ij} = \frac{V_{\theta_{ij}}}{\sqrt{V_{\theta_i} V_{\theta_j}}}. \quad (31)$$

Thus, to minimize the correlation among parameters one need to minimize the correlation among all parameters, e.g., by considering the sum of square values of all the correlation coefficients:

$$\min \sum_{i=1}^{N_\theta-1} \sum_{j=2, j \neq i}^{N_\theta} R_{ij}^2(\theta, \varphi). \quad (32)$$

The use of Eq. (32) as the objective for experiment design is only aimed at decreasing the correlation between parameters, whilst the confidence in parameter estimates may be very large (Issanchou et al. 2003). A multi-objective optimisation problem can be therefore formulated by coupling Eq. (32) with Eq. (7) or, alternatively, Eq. (8). Although Eqs. (7) and (8) already account for parameter interactions (variance–covariance matrix), Eq. (32) is introduced to explicitly minimize their correlation. Minimizing the sum of squared correlation coefficients helps in the reduction of parameter interdependence, improving numerical stability and reliability of parameter estimates. The selection of the optimal solution is based on identifying the Pareto front, which represents the set of non-dominated solutions offering the best trade-off between the objectives (Maheshwari et al., 2013).

As an alternative to the formulation of a multi-objective optimisation problem, Franceschini and Macchietto (2008b,c) proposed novel design criteria aimed at simultaneously including explicit measures of correlation, while targeting the experiments to the improvement of parameter precision. One way to achieve this is to minimise correlation in the objective function while requesting specified levels of parameter precision as problem constraints. In practice, specifying desired levels of parameter precision can be challenging, especially in the early stages of model development when limited information is available. These thresholds must therefore be chosen carefully, ideally based on engineering judgment, sensitivity analyses, or prior experimental evidence. If the required precision levels are too strict, the resulting optimisation problem may indeed become infeasible or lead to overly conservative experimental designs. A good practice at each design step, which delivers an understanding of the trade-offs between correlation and parameter precision, is to start with one objective and loose constraints, and then tighten as you go (and/or switch correlation and parameter precision as objective and constraints). To the purpose, the FIM (Eq. (4)) is first diagonalized as:

$$\mathbf{H}_\theta = \bar{\mathbf{E}} \mathbf{\Lambda} \bar{\mathbf{E}}^T, \quad (33)$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\mathbf{H}_\theta$ and $\bar{\mathbf{E}}$ is the matrix of corresponding normed eigenvectors. The new parameters $\Theta = \bar{\mathbf{E}}^T \theta$ are a linear combination of the original ones and represent principal components, from which information regarding their correlation structure can be derived. The information content is, then, expressed by means of the eigenvalues λ_i in $\mathbf{\Lambda}$ (associated with parameter θ_i) as constraints to the formulation of the optimisation problem (32):

$$\min \sum_{i=1}^{N_\theta-1} \sum_{j=2, j \neq i}^{N_\theta} R_{ij}^2(\theta, \varphi) \text{ s.t. } \lambda_i - \lambda_i^0 > \varepsilon_i^l \text{ with } i \in \{1, N_\theta\}, \quad (34)$$

where ε_i^l indicates a specified lower bound for the increase of λ_i over the initial value λ_i^0 calculated at a specific base point. The recognition that the best experiment has multiple requirements not easily reducible to a single scalar metric, led Franceschini and Macchietto (2008b,c) to suggest formulating the optimal DOE as a flexible combination of parameter precision (reflected by the ψ metric as in Eq. (7) or (8) or constraints as

in Eq. (34)) and correlation between specific parameter pairs (R_{ij} as in Eq. (34)), or overall ($\sum_{i=1}^{N_\theta-1} \sum_{j=2, j \neq i}^{N_\theta} R_{ij}^2(\theta, \varphi)$) as the argument in Eq. (34)). Depending on the problem at hand, either could be selected for minimisation or specified as a constraint, possibly changing at each design iteration. It was observed that many of the numerical or ill-conditioning problems usually associated with traditional formulations were solved. These issues typically arise from the poor conditioning of the FIM when parameters exhibit strong correlations, which can lead to instability during optimisation. The diagonalization of the FIM and the transformation of the parameters into principal components significantly reduce these correlations, resulting in a better-conditioned problem. Recent studies have demonstrated the effectiveness of this FIM approximation in enhancing numerical stability. For example, Karakida and Osawa (2020) and Soe and Sun (2024) highlighted how reducing the dimensionality of the FIM through diagonalization can lead to more stable problem formulations and improved computational efficiency.

Kim and Lee (2018) presented a different approach. After parameter estimation, weight coefficients are associated with each model parameter to account for both precision and correlation among them. Weight coefficients are used to define a new objective function for the next experiment design.

$$\varphi_{OPT} = \underset{\varphi}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N_\theta} w_{1,i} V_{\theta_{ii}} + \sum_{i \neq j}^{N_\theta} w_{2,ij} |R_{ij}| \right\}, \quad (35)$$

where $w_{1,ii}$ is the weight coefficient associated to each model parameter θ_i , and $w_{2,ij}$ is the weight coefficient associated with the pair θ_i , θ_j accounting for the effect of their correlation. The first term in the objective of Eq. (35) penalizes high variances to guide the selection of informative experiments aimed at improving parameter precision (i.e., minimizing parameters variance), while the second term minimizes the correlation among pairs. Choosing the weight coefficients is not a straightforward task, as their selection depends on model knowledge; moreover, in the presence of high correlation and large variance, the weights may not be independent. As noted earlier, trying with different values for the weights at each design step will typically give experimenters/modellers a better understanding of the trade-off between objectives, allows them to tailor their priorities and is a good practice.

Another approach consists in coupling the standard MBDoe formulation for parameter estimation (Eq. (7) or Eq. (8)) with estimability analysis (Wu et al., 2011), according to which only the most estimable parameters are considered – i.e., parameters that are the most effective towards model predictions and are little correlated each other (Chu and Hahn, 2012; Yuan and Benyahia, 2023).

3.1.3. Ill-conditioning of the FIM

A commonly encountered problem during the MBDoe procedure is the ill-conditioning of the FIM, which can make the matrix nearly non-invertible or lead to numerical instability during inversion. This issue arises when the matrix lacks full rank, typically due to the presence of many model parameters (White et al., 2016), potentially compounded by high correlations among them, or as a consequence of insufficiently informative experimental conditions. A common approach to address this issue is parameter subset selection, where only the most estimable parameters are included in the FIM (Wu et al., 2008; Barz et al., 2013; Kravaris et al., 2013; López et al., 2013; López et al., 2015). Orthogonalization-based parameter ranking for estimability analysis (Wu et al., 2011) has been successfully exploited in the so-called leave out approach to find and leave out problematic parameters – i.e., those parameters that are less influential with respect to model outputs and highly correlated – that make the FIM noninvertible (Shahmohammadi and McAuley, 2019a).

A common alternative to parameter subset selection for avoiding noninvertible FIM is to use a pseudoinverse of the original matrix. The

Moore Penrose pseudoinverse, $\mathbf{H}_\theta^{*-1}$, is the most adopted due to its straightforward computation via singular value decomposition (Li and Yeh, 2012; Shahmohammadi and McAuley, 2019a; Shahmohammadi and McAuley, 2019b):

$$\mathbf{H}_\theta = [\omega_1 \omega_2] \begin{bmatrix} \jmath_1 & 0 \\ 0 & 0 \end{bmatrix} [\nu_1 \nu_2], \quad (36)$$

where ω_1 and ω_2 are nonzero singular values in the diagonal matrix $\jmath_1$, ω_2 and ν_2 are zero singular values (or singular values below a prescribed threshold). The Moore-Penrose pseudoinverse for the FIM is defined as:

$$\mathbf{H}_\theta^{*-1} = \omega_1 \jmath_1^{-1} \nu_1. \quad (37)$$

While parameter subset selection ensures numerical stability and interpretability, the pseudoinverse retains all parameters in the analysis, albeit with reduced focus on practical identifiability, i.e., without analysing whether it is theoretically possible to uniquely determine the values of all model parameters from experimental measurements. The choice between these approaches depends on the specific objectives and constraints of the experiment design.

As an alternative, Quaglio et al. (2019) proposed a model identification approach of online reparameterization which is particularly useful when parameters are practically non-identifiable or highly correlated. The framework is based on two fundamental steps: (i) a primary parameter estimation which is required to assess whether it is possible to estimate all model parameters and their correlation; (ii) a parametrization update where the parameter space is transformed to a more robust parameter space aimed at reducing estimability issues. After the parametrization update, the parameter estimation is repeated within the new (robust) parameter space. Finally, parameter estimates are transformed back to the original space applying algebraic transformation. The algorithm has been demonstrated to effectively minimize the condition number of originally ill-conditioned problems. However, the effectiveness of the method may be limited, particularly when the model is highly nonlinear or has a high-dimensional parameter space, for which the transformation may become challenging. In these cases, the reparameterization process may lead to increased computational complexity and may not reliably avoid issues such as local optima or the failure to obtain a robust parameter space.

3.1.4. Effect of structural model-process mismatch

The selection of a reliable model should balance complexity and predictive capability, potentially leading to the development of model structures that embody a certain degree of misspecification. This in turn may impact on both the information that can be extracted from the FIM and the satisfaction of process constraints. Although the issue is critical in many experiment design applications, relatively few studies have directly addressed it.

A pioneering approach was developed by Galvanin et al. (2012), who proposed an online model-based redesign methodology where disturbances are explicitly estimated and incorporated into the experiment design process. In this approach, the system dynamics are augmented with a disturbance term $\mathbf{d}(t)$, leading to the following modified state-space representation:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \theta) + \mathbf{d}(t). \quad (38)$$

The term $\mathbf{d}(t)$ accounts for model-process mismatch and is estimated in real time during the experiment execution. The structure of $\mathbf{d}(t)$ is treated as an unknown input estimated online to capture discrepancies between the model and the actual process. It is assumed that $\mathbf{d}(t)$ smoothly varies over time, allowing for stable real-time estimation without imposing a specific functional form.

Note that the inclusion of $\mathbf{d}(t)$ may increase the degrees of freedom of the model, which could potentially impact the identifiability and statistical reliability of the parameter estimates. The experimental condi-

tions are then iteratively updated by solving an optimisation problem that minimizes the combined uncertainty in both θ and $\mathbf{d}(t)$. This strategy enables the simultaneous achievement of a statistically satisfactory estimation of the model parameters and the identification of potential discrepancies between the real system and the model.

To avoid miscalculating information during the experiment design phase, Quaglio et al. (2018) proposed an extended FIM formulation that accounts for model misspecification. Mathematically, the extended FIM, $\mathbf{E}$, is defined as the sum of the traditional FIM (Eq. (4)) and a deviation term $\mathbf{D}$:

$$\mathbf{E} = \mathbf{H}_\theta + \mathbf{D}, \quad (39)$$

$$\mathbf{D} = \sum_{k=1}^{N_p} \sum_{i=1}^{N_y} \sum_{j=1}^{N_y} s_{ij} (\hat{y}_i(t_k) - y_i(t_k)) \begin{bmatrix} \frac{\partial \hat{y}_i(t_k)}{\partial \theta_l} & \frac{\partial \hat{y}_j(t_k)}{\partial \theta_m} \end{bmatrix}_{l,m=1:1:N_\theta}, \quad (40)$$

with $y_i(t_k)$ representing the *true* i^{th} response at time t_k that will be measured during the experimental campaign. Since $y_i(t_k)$ cannot be known *a-priori*, it could be approximated using a support model, i.e., a data driven model. Note that because of the presence of matrix $\mathbf{D}$, matrix $\mathbf{E}$ (Eq. (39)) is not positive semidefinite and may therefore admit negative eigenvalues. The inversion of $\mathbf{E}$ loses significance and may also result in critical numerical problems, and, thus, design criteria must be based on the maximization of a certain direct scalar measure of $\mathbf{E}$. Another option could be to directly leverage the data-driven model for parameter updating.

As an alternative, Petsagkourakis and Galvanin (2020) proposed a novel algorithm based on Gaussian Processes (GP-MBDoE) where Gaussian Processes (GPs) are used to (i) approximate the model mismatch between the process and the model, and (ii) directly compute the expectation of the metric of the FIM and propagate the posterior distribution of parameters. The expected value of the objective is expressed as:

$$\varphi_{OPT} = \arg \min_{\theta} \left(\hat{m}(\varphi, \theta) + \alpha \sqrt{\hat{\Sigma}(\varphi, \theta)} \right), \quad (41)$$

where $\hat{m}(\varphi, \theta)$ and $\hat{\Sigma}(\varphi, \theta)$ are the mean and the variance of the GP which approximate the system under investigation after propagation of the uncertainty of parameter set θ ; α is a user-selected constant variable.

In spite of the potential effectiveness of both methods, it is worth noticing that large datasets should be available to train the empirical models of support. This requires the availability of experimental apparatus capable of generating extensive data and, thus, may be a strong limitation to the applicability in real-case scenarios. Recent advancements have also explored the integration of MBDoE with batch-to-batch optimisation and economic model predictive control to enhance robustness against model-mismatch and disturbances (Ghodba et al., 2025).

3.2. New algorithms and optimisation techniques

3.2.1. Local optimality

The standard MBDoE approach determines experimental conditions that optimise the objective function (Eq. (7) or, alternatively Eq. (8)). These optimal conditions are often restricted to small regions of high information content, which may result in limited exploration of the entire experiment design space. As a result, the model may exhibit poor predictive capabilities in unexplored regions.

To enhance space exploration and reduce model prediction uncertainty, Cenci et al. (2023) proposed a novel MBDoE formulation based on G-optimality mapping, which leverages a model prediction variance map to identify experimental conditions that minimize prediction uncertainty. Specifically, each experimental condition in the experiment design space is characterized in terms of prediction variance $\mathbf{V}_y$ (represented by a scalar index J_G) and information content (expressed

through a selected metric ψ of $\mathbf{H}_\theta$). For a given experimental point, J_G is computed as:

$$J_G = \sum_{j=1}^{N_y} \sum_{i=1}^{N_{sp}} \mathbf{V}_y(\theta, \varphi)_{j,i}. \quad (42)$$

The G-optimality-based criterion to be satisfied by the candidate design points is:

$$J_G \geq J_{G,thr} J_{G,max}, \quad (43)$$

where $J_{G,max}$ is the maximum value of G-optimality, and $J_{G,thr}$ is a threshold chosen by the user in between 0 and 1 – the closer $J_{G,thr}$ to 1, the fewer the candidate design points, since only those having the highest prediction variance are accepted; $J_{G,thr} = 0$ is equivalent to the standard MBDoe formulation. Among the selected candidate points satisfying Eq. (43), the experimental condition maximizing information is chosen as the optimal experiment. The procedure can be mathematically expressed as a constrained optimisation problem:

$$\varphi_{OPT} = \underset{\varphi}{\operatorname{argmin}} \{ \psi[\mathbf{V}_\theta(\theta, \varphi)] \} \text{ s.t. } J_G \geq J_{G,thr} J_{G,max}. \quad (44)$$

By incorporating a flexible threshold $J_{G,thr}$, the approach allows users to balance exploration and the maximization of information content in the experimental campaign. However, the selection of this value is critical, as it directly affects the number and quality of candidate points considered. Additionally, the grid-search algorithm may impose a high computational burden.

Friso and Galvanin (2024) proposed an alternative procedure that integrates a sampling-based experiment design approach with experiment ranking based on the FIM to iteratively select the most informative experiment. The procedure begins by generating N_{exp} exploratory experiments using a standard DoE method, such as Latin Hypercube Sampling. For each experimental point, the corresponding FIM is evaluated according to Eq. (4). The candidate experiments are then ranked based on their Relative Fisher Information (RFI), a normalized metric that quantifies the information content of each point relative to the entire experimental set. For the i^{th} experimental point, the RFI is computed as:

$$RFI_i = \frac{\|\mathbf{H}_{\theta,i}\|}{\sum_{j=1}^{N_{exp}} \|\mathbf{H}_{\theta,j}\|}. \quad (45)$$

The experiment with the highest RFI is selected as the most informative for the next iteration, ensuring that each step maximizes information gain. However, while this framework provides a powerful method for exploring the entire experiment design space, it heavily relies on the number and position of grid points. In the case of high-dimensional or complex models, it may require a large number of iterations to achieve a satisfactory level of precision in parameter estimates – potentially exceeding the computational effort of a standard MBDoe formulation.

Other methods for experiment selection involve attributing weights to a set of candidate experiments, where the optimal design is obtained by solving a convex optimisation problem over these weights (Yang et al., 2013). In such formulations, the weights themselves are the optimisation variables, and the optimality conditions can be established without requiring second derivatives. This reformulation, which applies also to nonlinear models, allows for the incorporation of previously conducted experiments and often improves computational tractability. In this context, Sagnol and Harman (2015) formulated the optimisation problem as a mixed-integer second-order cone program, enabling the computation of exact D-optimal designs. More recently, Seufert et al. (2021) propose a model-based experimental design approach for high-dimensional inputs that employs machine learning techniques to adaptively assign weights to candidate experiments, enabling efficient selection of the most informative trials while managing computational complexity. The weakness of this approach is that choice is limited to

candidate experiments proposed a-priori (not necessarily optimally informative, or even good).

An alternative approach to cope against the issue of space exploration is to discretise the entire experiment design space into a finite set of candidate experiments (Kusumo et al., 2022a; Sandrin et al., 2025; Vanaret et al., 2021; Dürrbaum and Kroll, 2016). For each candidate, the standard MBDoe framework (i.e., solving Eq. (7) or (8)) can be then applied thereby mitigating local optimality and/or suboptimality issues.

3.2.2. Multi-objective formulations

As already highlighted by Franceschini and Macchietto (2008a), during the experiment design different aspects of interest can be considered (e.g., correlation among parameters, directional information, costs...) and explicitly included in the optimal design formulation. Alternative to the single objective optimisation (Eq. (7) or Eq. (8)), the integration of various factors within the optimal design of experiments leads to the development of new multi-objective formulations (Telen et al., 2012a). The multi-objective framework requires the simultaneous solution of different objective problems, each expressed through a suitable objective function. Typically, no single solution exists; instead, the concept of Pareto optimality is employed to determine the set of optimal conditions. The Pareto front represents the set of solutions where no objective can be improved without worsening another, illustrating the trade-offs between the different objectives. Different methods can be adopted to evaluate the Pareto front, such as the weighted sum method (Das and Dennis, 1997), the normal boundary intersection (Das and Dennis, 1998), the normalized normal constraint (Messac et al., 2003) or evolutionary multi-objective methods (Rangaiah, 2009).

Regardless of the type of approach adopted to formulate the problem, many contributions focused on the integration of cost functions (Houska et al., 2015; Nöh et al., 2018; Pankajakshan et al. 2019; Wang et al., 2024), or risk functions (Kusumo et al., 2022b) within the standard MBDoe framework based on the maximization of the information content. As already discussed in Section 3.1.2, other works proposed multi-objective formulations to cope against the issue of parameter correlation (Franceschini and Macchietto, 2008b,c; Maheshwari et al., 2013a,b; Tsay et al., 2017). Bouchkira et al. (2023) combined standard D-optimal design with the maximization of Euclidian norms of the columns of the FIM to improve estimability. Considering a fed batch bioreactor, Telen et al. (2012b) included as extra objective the biomass production, allowing to study the possible trade-off between information content and production that typically occurs in real industrial environment. Manesso et al. (2017) maximized data information content using FIM-based metric while computing the model curvature to account for the effect of (strong) model nonlinearities within the optimal experiment design. More than two objectives were combined by Forte et al. (2017), i.e., the minimization of model predictions with respect to the nominal process operational point, experimental uncertainty of measured variables, and, if multiple candidate models are available, the maximization of the predictions of competing models.

Irrespective of the mathematical approach considered, the critical step in multi-objective formulations is the choice of an appropriate formulation for the specific problem at hand. This requires a deep understanding of the priorities, e.g., whether parameter precision is more important than managing (which) correlations, or whether exploring a broader design space is needed to understand the shape of the design surface. Additionally, it is always necessary to conduct a thorough analysis of the data produced at each experiment, and obtain a judicious understanding of the suggested designs and their evolution. While this process could eventually be automated, human intelligence continues to play a crucial role in ensuring that the chosen formulation is effective and the experimental results are properly interpreted.

3.2.3. Solution of the optimisation problem

When dealing with complex, nonlinear and high-dimensional models, the solution of the MBDoe optimisation problem may require

a high computational burden – and this is even more relevant in the case of multi-objective formulations. Various approaches have been proposed to streamline the optimisation solution, i.e., to improve the computational efficiency and to simplify the solution problem. Foundational work by Bauer et al. (2000) and Körkel et al. (2004) addressed these challenges by developing dedicated numerical methods for optimal experimental design in DAE systems, including the efficient generation of higher-order sensitivities and robust solution strategies. The computational burden of MBDoE methods is largely determined by the underlying optimisation algorithms employed. For instance, global optimisation techniques such as genetic algorithms or particle swarm optimisation are commonly used to handle multimodality, nonlinearity, and discontinuities in objective functions (Lei et al., 2017). While these methods increase the likelihood of identifying a global optimum, they often come at the cost of greater computational effort compared to local methods (e.g., gradient-based algorithms). Furthermore, the selection of the MBDoE approach (e.g., classical FIM-based criteria vs. Bayesian methods) also influences computational demand. On the other hand, MBDoE has already greatly benefitted from the significant improvements on general optimisation methods, and continues to do so.

The approximation of the original model using a surrogate is a powerful method to deal with computational challenges. For instance, Kucerová et al. (2016) demonstrated the effectiveness of polynomial chaos expansion in approximating nonlinear finite element simulations in a simple problem of two-dimensional, nonstationary linear heat transfer. Bouchkira et al. (2023) developed a polynomial response surface surrogate to mimic a complex first-principle model to predict the conversion rate of a phosphate ore digestion by phosphoric acid. Zhu et al. (2024) focused on piecewise affine surrogate-based optimisation for enhanced Menschutkin reaction kinetics. However, the use of surrogate models introduces a potential risk that the identified optimum may not correspond exactly to the true optimum of the original problem, due to the inherent approximations in the surrogate. To address this concern, several strategies can be employed to verify and enhance the reliability of surrogate-based optimisation. These include a posteriori validation using the full model at the predicted optimum (Forrester et al., 2008), cross-validation during surrogate training to assess generalization accuracy (Kuhn and Johnson, 2013), and trust-region approaches that restrict surrogate usage to regions where it has been demonstrated to be accurate enough (Conn et al., 2009). These techniques help quantify and control the approximation error, thus ensuring that the surrogate is reliable where it matters most, particularly in the neighbourhood of the optimal design.

Building on iterative design strategies, Steimel and Engell (2016) and Tsay et al. (2017) proposed a superstructure optimisation framework under parameter uncertainty. The method is based on a two-stage mixed integer optimisation problem: in the first stage, design and structural decisions are made, often represented by discrete variables (e.g., selecting specific experimental setups or configurations); the second stage addresses the operational adjustments and continuous optimisation, and uncertainty in parameter estimates is explicitly accounted. Kaiser and Engell (2020, 2022) refined the methodology to improve computational efficiency, introducing advanced uncertainty quantification techniques and demonstrated its robustness in the hydroaminomethylation of decene.

Another approach to reduce the computational burden involves selecting only a subset of parameters within the MBDoE framework. Techniques such as estimability analysis and parameter ranking, as discussed in Section 3.1.2, are particularly effective for this purpose. The selection of an effective numerical solver is also relevant. Telen et al. (2014b) introduced novel optimisation algorithms based on sequential semidefinite programming (Ramana, 1997), which offer specific advantages over sequential quadratic programming (Boggs and Tolle, 1995) within optimal experiment design. Specifically, the approach (i) guarantees a reduction in estimating uncertainty at the matrix level by incorporating a linear matrix inequality, (ii) allows easy formulation of

optimal designs in a direct control optimisation scheme, and (iii) simplifies matrix inversion problems, such as the computation of V_0 , through reformulation. Additionally, Kroll and Dürrbaum (2015) relaxed the optimisation solution based on D-optimal design criteria by applying determinant decomposition. The approach aims at reducing computational complexity in identifying optimal solutions, enhancing numerical stability, and improving the management of parameter variance.

While the literature presents a broad spectrum of advanced MBDoE formulations – including robust, stochastic, Bayesian, and multi-objective approaches – a critical evaluation reveals that their practical added value often remains difficult to quantify. For instance, Bayesian formulations offer a rigorous treatment of parameter uncertainty, but they typically require high-dimensional integration and remain computationally intractable for many real-case models. Similarly, multi-objective formulations attempt to balance competing goals (e.g., information content vs. parameter correlation), but frequently result in complex optimisation problems whose benefits over simpler scalar metrics are not consistently demonstrated. Only a limited number of contributions provide quantitative comparisons with classical MBDoE formulations in terms of reduction of experimental runs, improvement of parameter precision, or robustness against uncertainty. Moreover, while novel formulations tend to be methodologically elegant, they often rely on assumptions (e.g., full access to prior distributions, exact knowledge of experimental noise structure) that are challenging to meet in real applications. These limitations partially explain why many of these approaches remain confined to academic settings and simulation-based validations. More practically, the quest for arcane academic challenges has often run faster ahead than that for software useability and reliability.

Bridging the gap between methodological sophistication and real-world applicability remains a key challenge for future research.

3.3. Software

Despite advances in methods and algorithms, the implementation of MBDoE remains challenging, partly due to the paucity of general-purpose software packages (Agi et al., 2024). MBDoE necessitates the user to define specific model equations in the form of (1) and requires tools to be integrated with modelling platforms such as Python, MATLAB, AspenPlus, or GAMS. Table 2 lists available packages.

Currently, gPROMS® is one of the few commercial platforms that directly supports MBDoE activities for parameter estimation. While its capabilities are applicable across various industries, its usage is particularly prominent in pharmaceutical applications due to its integration with Quality by Design principles and specialized modelling tools for reaction and crystallization processes (Process Systems Enterprise, 2023).

For Python users, some open-source packages for MBDoE applications are available. Kusumo et al. (2022b) developed Pydex which provides an advanced framework for continuous experiment design for parameter estimation, using probabilistic methods to explore the experimental space and identify optimal configurations. Pydex stands out for the capability to handle uncertainties and operational constraints during the planning of experiments. The software operation is divided into two main phases: (i) probabilistic sampling of the experimental space, which helps the selection of promising configurations, and (ii) optimal experiment design, aimed at maximizing efficiency and minimizing costs. Another open-source Python package that should be mentioned is GPdoemd (Olofsson et al., 2019), which is specifically developed to design optimal experiments aimed at model discrimination. The core methodology of GPdoemd involves replacing complex models with GP surrogates. By leveraging GPs, the package facilitates approximate inference, enabling the marginalization of model parameters to account for uncertainty. However, Pydex and GPdoemd are not fully pre-configured packages, often require substantial user

Table 2

Comparison between the most popular open-source toolkits for MBDoE applications.

Software	Language	License	Application	Url Link
gPROMS®	gPROMS Modelling Language	Proprietary	Parameter precision	https://www.siemens.com/global/en/products/automation/industry-software/gproms-digital-process-design-and-operations.html
Pydex	Python	Massachusetts Institute of Technology (MIT)	Parameter precision	https://pydex.readthedocs.io/en/latest/
GPdoemd	Python	Berkeley Software Distribution (BSD) – 3	Model discrimination	https://www.sciencedirect.com/science/article/pii/S0098135419300468
Pyomo. DoE	Python	BSD – 3	Parameter precision	https://pyomo.readthedocs.io/en/6.8.0/contributed_packages/doe.html
MIDDoE	Python	MIT	Parameter precision & model discrimination	https://pypi.org/project/middoe/
VPLAN	C++	Academic (upon request)	Parameter precision	https://simopt.iwr.uni-heidelberg.de/vplan

customisation, and generally demand a strong understanding of both Python programming and the underlying MBDoE principles.

To make MBDoE methodologies more accessible, Wang and Dowling (2022) developed a comprehensive Python-based modelling package, Pyomo.DoE, which can be suitably exploited for parameter precision. The package integrates with a variety of solvers, making it versatile for different computational needs. It allows the user to easily define constraints, objectives and decision variables, and is suitable for both single and multi-objective optimisation formulations. Additionally, its ability to handle nonlinear, mixed-integer, and dynamic optimisation problems makes it useful for MBDoE applications that require complex model interactions and real-time optimisation.

More recently, Tabrizi et al. (2025) proposed a Python package, MIDDoE, which aims at tackling some flexibility limitations in the PYOMO array and model structure. The package offers MBDoE techniques tailored for both model discrimination and parameter estimation, and provides tools to assess model identifiability, including estimability analysis, while ensuring adaptability to process constraints.

Furthermore, the software VPLAN (<https://simopt.iwr.uni-heidelberg.de/vplan>) represents a important tool for MBDoE applications. It incorporates advanced numerical techniques for solving nonlinear optimisation problems for parameter precision and offers dedicated functionalities for uncertainty quantification and sensitivity analysis.

An essential computational technique that supports the implementation of MBDoE is automatic differentiation which enables efficient and accurate computation of gradients and sensitivities required for optimisation and uncertainty quantification. Several modern toolkits and frameworks, including those for dynamic optimisation, increasingly rely on automatic differentiation to ensure numerical robustness and scalability (Griewank and Walther, 2008). For example, Python libraries such as Autograd (Maclaurin et al., 2015), and PyTorch (Paszke et al.,

2019) offer automatic differentiation functionality that can be leveraged for MBDoE applications. Commercial modelling environments such as gPROMS® have also long incorporated built-in automatic differentiation features since their early commercial deployment in the early 2000 s.

In conclusion, the use of advanced toolkits and software represents a crucial step in accelerating the implementation of MBDoE techniques, streamlining a more efficient implementation for industrial purposes. In general, there is still the need for a user's friendly software, with a comprehensive availability of libraries of identification tools and MBDoE methods. We believe that one of the reasons that hindered a widespread utilisation of MBDoE techniques, in particular among industrial practitioners, is that their implementation and usage still requires specialistic skills and expertise.

3.4. MBDoE applications

MBDoE has been applied across a wide range of disciplines to improve parameter estimation and (less frequently) to assist in model discrimination. This Section provides an overview of its applications for parameter estimation, with some studies also integrating model discrimination, in various fields, relying on different data sources, i.e., either experimental or in silico. Experimental data refer to measurements obtained directly from laboratory or industrial experiments, while in silico studies rely on numerical simulations based on previously validated mathematical models (the purpose being a controlled demonstration of the methods rather than model development). The increasing number of studies based on experimental data underscores the growing effort to validate MBDoE methodologies in real-world settings. Table 3 summarizes the key publicly available contributions in various domains, categorized by different data types (experimental vs. in silico).

Several important trends emerge from Table 3:

- Kinetic reactions and biological systems are the most extensively studied fields, with a balanced mix of both theoretical and experimental studies (Barz et al., 2022). Here, MBDoE is primarily employed to refine parameter estimates only; however, some studies also incorporate model discrimination techniques among rival models (Hennecke et al., 2024; Han et al., 2016; Galvanin et al., 2015; Schaber et al., 2014), e.g., distinguishing between competing reaction mechanisms.
- Pharmaceutical manufacturing and pharmacodynamics leverage MBDoE to enhance process efficiency and product testing, while balancing regulatory constraints with experimental feasibility. Applications in this area are still limited – both considering in silico and experimental studies – and only focus on parameter estimation.
- Life science models (e.g., pharmacodynamics and physiological models) are key in biomedical research, for drug development, and personalized medicine. Models have been recently proposed and applied to aid in the diagnosis and characterization of diseases, but their complexity is such that they often require long and costly clinical tests. MBDoE has been demonstrated to be useful in designing minimal sets of clinical tests for parameter estimation. However, also in this context, in silico studies are predominant.
- Food microbiology and analytical chemistry show a significant number of cases employing real experimental data opposed to purely theoretical studies – all contributions concerning parameter estimation only. This suggests an easier practical implementation with respect to fields such as pharmaceutical manufacturing, where practitioners are generally more reluctant to trust complex available models.
- Electrochemical systems have been also studied, with specific focus on parameter precision. While research in this field remains predominantly in silico, some studies demonstrate the feasibility of applying MBDoE also in real-word applications.

Table 3

Case studies utilizing MBDoE for parameter estimation, with some studies also integrating model discrimination, across different fields categorized by data type (in silico vs. experimental).

Application field	Data type	References
Pharmaceutical science & manufacturing	In silico	Katsoulas et al. (2025); Geremia et al. (2024a,b); Yuan and Benyahia (2024); Geremia et al. (2023); Kilari et al. (2023); Yuan and Benyahia (2023)
	Experimental	Matsunami et al. (2023); Cenci et al. (2022); Pal et al. (2021); De-Luca et al. (2020)
Reaction kinetics	In silico	Yonge et al. (2024); Aspiron et al. (2020); Yue et al. (2013); Galvanin et al. (2009a)
	Experimental	Agunloye et al. (2024); Schaare et al. (2023); Waldron et al. (2020); Waldron et al. (2019); Jung et al. (2019); Jiang et al. (2019); Echtermeyer et al. (2017); Han et al. (2016); Galvanin et al. (2015); Schaber et al. (2014); Reizman and Jensen (2012); McMullen and Jensen (2011)
Biological systems and bioprocessing	In silico	Hennecke et al. (2024); Villaverde et al. (2022); Möller et al. (2019); Telen et al. (2017); Ruess and Lygeros (2015); Yu et al. (2015); Schenkendorf and Mangold (2011)
	Experimental	Saccardo et al. (2023); Moser et al. (2021); Barz et al. (2018); Luna and Martínez (2018); Cruz Bournazou et al. (2017); Bernardi et al. (2016); López et al. (2013); Flores-Sánchez et al. (2013); Valencia et al. (2013); Kontoravdi et al. (2010)
Analytical chemistry Life science models	Experimental	Barz et al. (2016); Barz et al. (2013)
	In silico	Taverna et al. (2019); Voß et al. (2019); Shotwell and Gray (2016); Maheshwari et al. (2015); Galvanin et al. (2011); Galvanin et al. (2009b)
Food microbiology	Experimental	Galvanin et al. (2024); Pargett et al. (2014)
	In silico	Garre et al. (2019)
Electrochemistry	Experimental	Longhi et al. (2018); Longhi et al. (2017); Neddermeyer et al. (2016); Longhi et al. (2016)
	In silico	Stroebel et al. (2024); Rhyu et al. (2024); Ciucci (2013); Tsikonis et al. (2011)
Other fields	Experimental	Park et al. (2018)
	In silico	Evripidou et al. (2024); Kaminski et al. (2016); Hametner et al. (2013)
	Experimental	Liu et al. (2025); Galvanin et al. (2016b); Stamati et al. (2015)

Other fields such as automotive engineering (Hametner et al., 2013), polymer science (Kaminski et al., 2016); separation processes (Evripidou et al., 2024; Galvanin et al., 2016b) and airpath modelling for engine control strategies (Stamati et al., 2015) are emerging. This highlights the versatility of MBDoE in tackling diverse modelling challenges in various contexts for improving parameter precision.

4. Future perspectives

In the evolving context of process systems engineering, MBDoE must adapt to meet the demands of increasingly complex and interdisciplinary modelling frameworks. In this section, we address emerging challenges and opportunities for future development related to the integration of MBDoE methodologies within advanced techniques. As shown in Fig. 5, future directions for MBDoE encompass several areas, including its integration with data-driven modelling frameworks, considering the impact of available equipment in real-world scenarios, and the development of new approaches for optimal design.

4.1. MBDoE and data-driven modelling

Recently, there has been a rapid advancement in data-driven modelling techniques and a progressive shift towards hybrid modelling approaches, where first-principle knowledge is complemented by data-driven components within a unified framework, with the goal to better capture the underlying system complexities (Sansana et al. 2021). However, this evolution inevitably raises several challenges concerning the integration of MBDoE methodologies with hybrid modelling paradigms.

4.1.1. Integration of MBDoE within hybrid modelling frameworks

Traditional MBDoE focuses on optimising the parameters of mechanistic models; in a data-driven DoE, the underlying system is described by some physics-free model, such as a neural network, with the optimisation consisting in some form of machine learning from just input–output data (data-driven). In hybrid frameworks, both mechanistic and data-driven components are used in combination to leverage their respective strengths: mechanistic models provide interpretability and physical consistency, but can be challenging to develop for complex systems, while data-driven models offer flexibility requiring no prior knowledge of the underlying principles of the system (only sufficient data for training). This makes them particularly useful when dealing with poorly understood processes or when first-principle modelling is impractical. Moreover, hybrid approaches offer a unique opportunity to integrate mechanistic knowledge with less quantitative data sources, including images, sound, and complex analytical results, expanding the applicability of MBDoE in innovative ways. When integrating MBDoE with hybrid models for parameter estimation this gives rise to several key challenges. The first challenge is to decide what form of data-driven models to use and how the two types of models are to be used together. One possible choice is to use the data-driven components to capture deviations from the mechanistic models, however this is not the only one. An additional complexity when optimising data-driven components, is the choice of which parameters (e.g., weights and biases) in the data-driven model to choose. These are crucial for making accurate predictions. The data-intensive nature of machine learning models raises two key concerns. The first is the number of experiments required for initial training and calibration, which can be substantial. The second is ensuring the experimental feasibility of the proposed designs, particularly in the early stages of training when extrapolation risks are high. Mechanistic models also face limitations in extrapolation, but they allow for the integration of physical constraints on parameters, input controls, and states, which helps mitigate feasibility issues – an advantage that is more challenging to implement in purely data-driven models (Azarpour et al., 2017; Aspiron et al., 2019). To address this issue, a recent framework proposed by Pankajakshan et al. (2025) combined Gaussian Process regression with active learning to optimise experimental conditions and manage the calibration of data-driven models effectively, which offers potential for optimising experiment design and ensures efficient use of resources.

Another critical challenge is the interaction between mechanistic and data-driven parameters. The interdependencies between the two components often complicate the identifiability of models and/or reduce interpretability. This necessitates the development of ad-hoc sensitivity analysis techniques and design criteria to guide experiments across both submodels (Selvarajan et al., 2022; Schweidtmann et al., 2024). As the hybrid model evolves with the accumulation of new data, adaptive learning techniques must be integrated into MBDoE frameworks to refine experiment design iteratively, ensuring that the data collection process remains optimal and evolves over time (Tian et al., 2020).

An additional issue is how to design experiments to maximize the mismatch between a given mechanistic model and the real system. This could serve two main purposes: (i) identifying which specific channel or equation fails to capture the real system behaviour, thereby providing

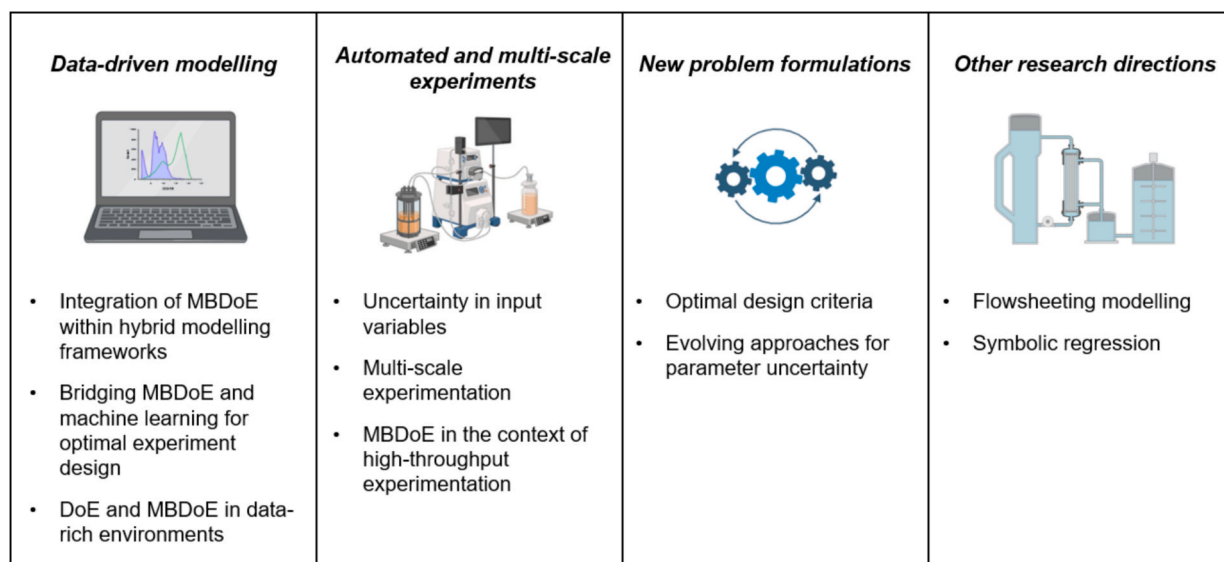


Fig. 5. Future directions in MBDoE, highlighting emerging challenges and opportunities for further exploration.

insights into its limitations, and (ii) enabling the optimal training of a suitable data-driven component to compensate for the observed mismatch. With concern to purpose (i), a promising research direction concerns the exploitation of MBDoE techniques not only to design specific experiments aimed at discriminating between competing model structures, but also to identify terms or pathways within the model equations that contribute most significantly to model inaccuracies or prediction failures (Meneghetti et al., 2014; Chen and Ierapetritou, 2020; Quaglio et al., 2020).

Literature on the topic of integrating MBDoE within hybrid modelling frameworks is scarce; however, a relevant attempt has been presented by Tappe et al. (2023). The work provides insights into limitations of traditional MBDoE concepts in calibrating physics-informed neural networks (PINNs) – where first-principle models are incorporated into the structure of artificial neural networks creating an hybrid framework in terms of model calibration – and proposes a physic loss function (Shukla et al., 2022) as suitable objective for optimal experiment design. While the work by Tappe et al. (2023) is limited to the specific case of PINNs and does not provide substantial details from a numerical/mathematical perspective, it provides some directions for future research. This includes exploring different types of hybrid model configurations, developing novel optimisation criteria, and advancing new techniques that could become pivotal for significant future innovations in MBDoE applications.

4.1.2. Bridging MBDoE and machine learning for optimal experiment design

Data-driven and machine learning techniques often face challenges related to data selection for model calibration and the risk of over-parameterization (Simon et al., 2023). Over-parameterization occurs when models have more parameters than necessary to describe the underlying data, which can lead to overfitting, reduced generalization, and inefficient use of experimental resources (too many experiments). This runs counter to the core principles of MBDoE, which focuses on optimising experiment design with minimal data (i.e., in data-poor situations). A critical question arises: can MBDoE itself be adapted to leverage machine learning techniques to optimise data selection and mitigate the risk of over-parameterization? Active learning and data selection strategies can play a vital role in addressing this challenge. By using algorithms that prioritize the most informative data points, these approaches could reduce the need for excessively large datasets while maintaining model performance. The potential of combining machine learning with

MBDoE could guide the design process in real-time, continuously adjusting to new data and minimizing overfitting risks.

Moreover, in data-scarce regimes, MBDoE provides a principled way to maximize the information gained from each experiment, enabling more efficient training of machine learning models. By strategically guiding data acquisition, MBDoE can enhance the ability of data-driven and hybrid models to even when only a limited number of experiments is feasible. This is especially valuable in biological and biopharmaceutical systems, where data collection is often costly and time-consuming (Destro and Barolo, 2022).

4.1.3. DoE and MBDoE in data-rich environments

Traditionally, MBDoE has been applied in data-scarce scenarios, where the careful design of experiments is crucial to maximize information gain from limited (experimental) data. Crucially, the model itself may be viewed as condensing much physics validated with prior data. In environments where large volumes of data are readily obtained (for example, when a large number of experiments are readily done), a deeper understanding of the relative advantages of black box DoE, MBDoE, and combined approaches (i.e., combination of DoE and MBDoE strategies) is critical. Traditional DoE focuses on the statistical design of experiments to explore the parameter space but does not take advantage of prior knowledge from the model. In contrast, MBDoE exploits the available model to design the experimental campaign, leading to more targeted and efficient experimentation. However, in data-rich environments, one may argue that the application of MBDoE may be unnecessary or offer limited benefits, as the abundance of data would reduce the need for optimisation of experiment design typically associated with MBDoE.

The true challenge lies in efficiently navigating the parameter space when data are abundant, avoiding redundancy, and ensuring that the experimental design continues to add value.

Combined approaches can integrate the strengths of both DoE and MBDoE approaches, allowing for a more dynamic exploration of the parameter space while refining the design based on data already collected (Agunloye et al., 2024a). In principle, this can offer the best of both worlds: the broad exploration capabilities of DoE and the efficiency of MBDoE. DoE can provide an initial framework for design, ensuring broad coverage of the experimental space, while MBDoE can refine the design by incorporating data-driven feedback and adjusting to model uncertainties as more data is collected. In other words, instead of being primarily used to explore the parameter space broadly, MBDoE can be

leveraged to refine designs dynamically, adjusting based on data already collected. This synergy allows for a more balanced approach. While starting with initial DoE to ensure broad exploration before refining with MBDoE is a standard recommendation, its implementation in data-rich environments may require adaptation. The abundance of data could shift the balance between exploration and exploitation, potentially allowing for a more flexible transition between DoE and MBDoE depending on real-time data availability and model refinement needs.

Importantly, MBDoE retains its distinctive advantage in scenarios where the cost of materials is significantly higher than the cost of measurements and experimental setups, availability of materials is scarce and/or experimental campaigns are time-consuming. This is particularly relevant in high-value industries such as biopharmaceuticals or specialty chemicals, where experiments are expensive and minimizing the number of runs is critical (Destro and Barolo, 2022). In such cases, the possibility to optimise scarce and costly measurements offers a tangible benefit over traditional DoE approaches, which typically assume a more balanced cost structure.

Future research will likely focus on how to balance the exploration of new areas of the parameter space with the exploitation of existing knowledge to optimise data collection, ensuring that experiments are both informative and computationally feasible.

4.2. MBDoE and automated and multi-scale experiments

The integration of MBDoE with experimental equipment presents several challenges when dealing with real-world industrial processes (as opposed to in-silico simulations). In traditional MBDoE approaches, it is often assumed that the control variables can be precisely manipulated and measured. However, this assumption does not always hold in practice, as real-world experimentation involves uncertainties and variations that must be accounted for. Moreover, as processes scale up or when high-throughput systems are employed, new considerations must be introduced into the experiment design framework to address equipment complexities and exploit the potential for automation. This section explores how MBDoE can be adapted to handle these challenges, ensuring its relevance in modern experimental practices.

4.2.1. Uncertainty in input variables

The standard MBDoE formulation assumes that the experimental control variables can be accurately set and measured. In practice, critical input factors are often difficult to determine or control precisely. Uncertainty may arise when exact values of certain variables entering the process cannot be determined with confidence (Elster, 2007). Instead, variability refers to unpredictable changes in critical inputs, such as alteration in the physical or chemical properties of raw materials, or the impact of external disturbances (Stauffer et al., 2018). Input uncertainty and variability can propagate through the model, affecting model predictions and parameter estimation (Abdi et al., 2024).

Up to now, robust design strategies have only been proposed in literature to account for the uncertainty associated with parameter estimates. However, the propagation of input uncertainties, particularly in nonlinear models, also impacts the reliability of model predictions, which in turn affects the prediction of expected information. One promising direction for future research is to extend robust design methodologies to account for uncertainty in critical input factors, enabling more reliable experiment design under real-world conditions where variability and uncertainty are inherently unavoidable. A critical yet underexplored aspect of incorporating input uncertainties into robust MBDoE frameworks concerns their distribution. In many cases, the exact distribution is not known a priori, which introduces an additional layer of complexity in their quantification and propagation. Frequently, uniform distributions are employed as a conservative approach, ensuring that all potential values within a specified range are equally probable (Petrik et al., 2013; Geremia et al., 2024b). Alternatively, Gaussian distributions are often assumed due to their simplicity

and widespread application in statistical modelling (Shewhart, 1929). However, these choices may not always reflect the underlying variability of real-world inputs. For example, inputs influenced by physical or chemical properties, such as raw material quality or environmental disturbances, may exhibit skewed or multimodal distributions (Sin et al., 2009). The choice of distribution has a direct impact on the quantification of prediction uncertainty and is likely to be carried out off-line, prior to any DoE, to establish the most appropriate distribution for each uncertain input. Furthermore, if the variabilities of the inputs are not independent, joint variability models would need to be considered. If certain process inputs, e.g., flowrate or temperature, can be measured during experiments, it may be possible to detect their actual experimental values (as opposed to their design setpoints). In such cases, it might be feasible to estimate their variance model in a similar way as is done for measurements of output values, enabling more accurate modelling of input uncertainty during the experiment. Addressing this gap requires further investigation into techniques for estimating distributions from sparse or noisy data.

Another challenge in incorporating input uncertainty within MBDoE formulation regards an increased computational demand, especially for nonlinear models and high-dimensional input spaces. To address the high dimensionality of input spaces, dimensionality reduction techniques such as Principal Component Analysis (Montgomery, 2013) or Partial Least Squares regression (Wold et al., 1983) could be integrated into robust MBDoE frameworks. These statistical methods identify the most influential combinations of input variables, thereby reducing the complexity of the design space while preserving the critical information required for accurate experiment design. However, the application of such techniques to MBDoE comes with significant limitations. In particular, while latent variables can help identify directions of maximum variance or covariance with the outputs, it is often challenging to map them unambiguously back to the original design space. Multiple combinations of inputs may correspond to the same latent representation, making it difficult to interpret and implement the resulting design in practice. This ambiguity can limit the practical utility of Principal Component Analysis or Partial Least Squares regression in experiment design, especially when precise input specification is required. Future work should address this issue, for example by exploring inverse mapping strategies or incorporating interpretability constraints into the dimensionality reduction process.

Additionally, stochastic optimisation (Powell, 2019), surrogate modelling (He et al., 2023), and parallel computing (Fox et al., 2014) may alleviate the computational burden, but further investigation is essential to ensure both scalability and accuracy. Global sensitivity techniques, which help assess the impact of different inputs on output measurements, could also be applied.

Equipment limitations should also be considered, as the capabilities and constraints of available instruments play a critical role in determining the quality and reliability of measurements. While optimal control techniques for sensor placement (e.g., those used for tubular reactors) have long been well-established, the recent work by Alen et al. (2025) have started to explore how management of inputs uncertainty can be extended to include equipment constraints, particularly in optimising sensor placement in weakly constrained settings. The proposed framework employs an algorithm for near-optimal experiment design through column subset selection, including a randomized technique that avoids adjoint computations, ensuring computational efficiency while incorporating system uncertainties. The framework directly addresses the impact of model errors and equipment constraints on sensor placement, reflecting the need of accounting for uncertainties and variability typically present in real-world systems.

4.2.2. Multi-scale experimentation

In process development and scale-up, experimentation is typically conducted across different equipment scales. Traditional MBDoE methodologies often assume a controlled and consistent experimental

setup, which works well in lab-scale experiments where the variables can be tightly regulated. However, as processes scale up to pilot or industrial levels, there are significant changes in equipment, environmental conditions, and operational constraints. These changes should be adequately accounted for, as they may potentially compromise the robustness and transferability of the developed models. Practically, two approaches can be followed for incorporating multi-scale experimentation within MBDoE frameworks: (i) designing and simultaneously executing multiple experiments for equipment at different scales, or (ii) sequentially scaling-up with model adaptation. The latter consists in first exploiting MBDoE at the laboratory scale for model discrimination, parameter estimation and statistical validation; then, the same model is used with equipment at larger scale trying to reconcile the differences, e. g., by designing further experiments specifically to characterize those scale-dependent variations.

One of the primary challenges in incorporating multi-scale experimentation into MBDoE frameworks is the heterogeneity of data generated at different scales. Variations in equipment design, operating conditions, and measurement precision can introduce systematic discrepancies that complicate model calibration and validation. Hierarchical modelling approaches can be employed, where scale-specific models are integrated within a unified framework that accounts for inter-scale variability (Gelman, 2006). Moreover, the design of experiments must consider scale-dependent factors, e.g., mixing efficiency, heat transfer rates, and reaction kinetics, which may behave differently across scales. Developing scale-bridging models to be integrated into MBDoE frameworks can facilitate the translation of experimental findings from one scale to another (Knap et al., 2016). The integration of multi-scale experimentation into MBDoE frameworks may also necessitate advancements in data fusion techniques. Combining heterogeneous data from various scales requires sophisticated statistical and machine learning methods to harmonize the data and extract meaningful insights. Techniques such as multi-fidelity modelling, where high-fidelity and low-fidelity data are combined to improve model predictions, can be also particularly useful in this context (Fernández-Godino, 2016). Future research should focus on developing standardized protocols and methodologies to account for the effect of multi-scale experimentation, e.g., establishing best practices for hierarchical modelling, scale-bridging techniques, and data fusion methods.

4.2.3. MBDoE in the context of high-throughput experimentation

The advent of high-throughput experimentation (HTE), driven by advancements in robotics and automation, has significantly transformed the landscape of experiment design (Mennen et al., 2019). High-throughput systems enable the rapid execution of many experiments in parallel, increasing the speed and scale at which data can be generated (Lu et al., 2024). Some examples of high-throughput systems include DIGIBAT (<https://www.imperial.ac.uk/digibat>), focused on material synthesis and electrochemistry for battery and energy research, AutoCHEM (<https://researchsupport.admin.ox.ac.uk/micromeritics-autochem-2920>), a fully automated chemisorption analyser, and the High-Throughput Experimentation Platform (<https://www.materials.ethz.ch/research/high-throughput-experimentation.html>), which specializes in the automation and optimisation of chemical reactions for materials and catalysis research.

A further evolution of HTE is represented by Self-Driving Laboratories (SDLs), which integrate machine learning with laboratory automation to conduct experiments in active learning loops. Autonomous labs leverage advanced robotics, automated modules, and artificial intelligence, with the main advantage of autonomously designing, executing, and analysing a wide range of experimental conditions. This enables an accelerated exploration of complex parameter spaces, uncovering experimental conditions that might otherwise be overlooked, while optimising resource utilization (Abolhasani and Kumaicheva, 2023; Baldea et al., 2025). One of the most significant benefits of SDLs is their potential to conduct experiments without human

intervention, thus eliminating bottlenecks associated with manual processes. Through continuous learning and integration with machine learning algorithms, SDLs can dynamically adjust experiment design based on real-time results, which optimises the exploration of new experimental configurations. This adaptive cycle enables a much faster progression of research and allows for the discovery of novel insights and data patterns, exceeding the capabilities of traditional experimental methods (Sadeghi et al., 2024).

While full automation is often envisioned, maintaining human oversight in critical decision-making steps remains essential to ensure reliability and industrial applicability (Scheurer and Reuter, 2025). As these technologies evolve, the potential for fully autonomous experimental cycles opens new possibilities for optimising experiment design, particularly in complex systems. Relevant contributions in biology and biotechnology have demonstrated the potential of MBDoE approaches in high-throughput systems. For example, Cruz Bournazou et al. (2017) and Barz et al. (2018) proposed an online MBDoE framework for adaptive redesign of parallel fed-batch cultivations of *Escherichia coli* in automated platforms. Further studies (Schmideder et al., 2015; Hernández Rodríguez et al., 2019; Möller et al., 2019; Habicher et al., 2020) demonstrated the scalability of model-based methods to multiple simultaneous bioreactor experiments with real-time data integration and parameter updating.

While HTE offers unparalleled efficiency in data collection, it also introduces new complexities. The large volume of data generated can be overwhelming, making it challenging to extract meaningful insights without a structured approach. Here, MBDoE may be incorporated to guide the experimental process, ensuring that even in high-throughput settings, experiments are designed to maximize information gain and resource efficiency (Abt et al., 2018; Cenci et al., 2023; Friso and Galvanin, 2024).

MBDoE can complement HTE by focusing on optimising experimental conditions to reduce redundancy and improve data quality. The challenge is to extend MBDoE to consider very large number of experiments. Galvanin et al. (2007) presented the first framework and formulation for extending traditional MBDoE to the case of design campaigns of parallel, sequential and mixed parallel/sequential sets experiments. For parameter precision, the alternative is often between extending the range of the design space explored to get improved model quality over a wider range of conditions, and carrying out multiple executions of the same experiment to get improved measurement variance data. More work is to be done on the selection of suitable objectives to balance the two and indeed developing overall more systematic procedures. A relevant point is maintaining computational viability. The sheer scale of data and the number of variables involved can lead to significant computational demands, especially in complex, nonlinear systems. A trade-off between the speed of HTE and the precision of MBDoE should be established to prioritize experiments based on their expected information yield relative to the computational effort required. Ensuring the scalability of MBDoE methods to handle the vast datasets produced by HTE also remains critical. Furthermore, the integration of MBDoE with automated systems requires robust algorithms capable of real-time decision-making and adaptability to unforeseen experimental outcomes. On the computational side, key bottlenecks lie in algorithm design, solution parallelization, and the need for faster data analysis. Leveraging the strengths of both MBDoE and HTE can help address these issues, providing a more efficient balance between computational complexity and accuracy, while ensuring that experiment design remains both efficient and insightful in an era of rapid data generation.

Looking forward, a promising direction lies in embedding MBDoE at the core of autonomous experimental cycles, where it can function not only as a design tool but also as a strategic decision engine. In SDLs, this role becomes even more critical, as MBDoE can continuously guide the exploration of experimental conditions based on real-time model updates and learning. When coupled with hybrid models that combine mechanistic insight with data-driven adaptability, such systems enable a

closed-loop environment where each experiment is informed by evolving understanding. In this context, SDLs may be seen as living realizations of process system engineering principles – merging modelling, optimisation, and control into an autonomous platform for intelligent experimental discovery.

4.3. MBDoE and new problem formulations

As experiment design becomes more sophisticated and data-rich, new formulations of objective functions, enhanced techniques for uncertainty quantification, and the integration of alternative optimisation goals are necessary to ensure that MBDoE remains effective. This section explores key areas where MBDoE can evolve, focusing on incorporating more complex information structures, leveraging Bayesian approaches, and considering broader objectives beyond parameter precision.

4.3.1. Optimal design criteria

Traditional MBDoE approaches typically rely on objective functions based on the FIM (Eq. (4)) to guide the design of experiments. However, an important question is whether traditional criteria are sufficient, or if additional information could further optimise experimental strategies. While the FIM effectively captures the sensitivity of model parameters, it overlooks other important aspects, such as correlations between variables and the intricate structure of data in non-linear or highly coupled systems. As discussed in Section 3, several methodologies have been developed to address limitations in the use of traditional FIM-based criteria, which include incorporating parameter uncertainty, accounting for structural mismatches in models, and introducing multi-objective optimisation frameworks. Despite these advances, significant challenges remain in fully capturing the complexity of real-world systems. One possible future trend is to focus on evaluating alternative criteria to guide the optimisation process; e.g., entropy-based methods (Huan et al., 2024). In the context of experimental design, entropy is used to quantify the spread or unpredictability in the parameter space, and experimental designs are chosen to explore regions of the parameter space where the uncertainty is high, thereby improving the overall efficiency of the design (Huan and Marzouk, 2013). These approaches typically aim at minimizing the expected posterior entropy of the parameter distribution after conducting an experiment. This is equivalent to maximizing the expected reduction in uncertainty, and is often implemented by maximizing the expected Kullback-Leibler divergence (Eq. (16)) from the likelihood to the evidence.

Additionally, integrating Bayesian approaches or leveraging techniques from information theory might offer new pathways to refine MBDoE methodologies, offering greater flexibility and adaptability in complex and uncertain environments. In a Bayesian framework, prior knowledge is formally represented through a prior distribution, which encodes the preliminary understanding of model parameters and associated uncertainties (Gelman and Shalizi, 2013). This can include, for instance, initial values and probabilistic distributions that describe uncertainty in specific aspects of the system before new data is collected. Once new experimental data is acquired, the prior distribution is updated to form a posterior distribution, which provides a refined estimate of the model parameters. While in traditional approaches prior knowledge is only embedded into the model structure and assumptions, the use of Bayesian methods facilitates a more explicit, mathematically rigorous, and standardized process for incorporating and updating this knowledge throughout the experimental design process. Given their potential, Bayesian methods can be used in a more standardized way to incorporate prior knowledge, update model predictions based on new data, and quantify uncertainty (see Section 3.1.1). For example, by using Bayesian optimisation or probabilistic modelling, MBDoE can be adapted to manage uncertainty in both model parameters and experimental outcomes. While these developments may offer more flexible and adaptive frameworks, it is important to recognise that having an intelligent human in the design loop – who is capable of understanding the

direction predicted by the design step, interpreting results, and, in particular, directing the evolution by reformulating objective functions and constraints – remains essential. The potential of advanced AI-based methods to support these high-cognitive level tasks presents one of the greatest challenges moving forward. Further research could investigate how these techniques can be efficiently integrated into MBDoE frameworks, with a focus on computational burden and real-world applicability.

4.3.2. Evolving approaches for parameter uncertainty

While MBDoE has made significant progresses in optimising parameter estimation, there is room for further sophistication. For instance, systems with intrinsic uncertainties, such as biological processes, where high uncertainty in model parameters is common, pose additional challenges.

In general, the question is how much parameter precision is necessary for a model to be useful? MBDoE can be re-formulated to account for the behaviour of such uncertain systems more effectively. One possibility is to focus on achieving an acceptable model performance, rather than aiming for exact parameter estimation. In other words, the idea is to quantify the maximum allowable parametric uncertainty such that model predictions fall within a predefined level of confidence (Geremia et al., 2023). An optimisation framework was recently proposed by Geremia et al. (2024a), where the precision of parameter estimates is explicitly optimised to target preset tolerance requirements on process key performance indicators and product critical quality attributes. Specifically, the confidence intervals of the estimated parameters are compared to the optimisation results, which defines the maximum allowable parameter uncertainty compatible with the desired model fidelity. If the confidence intervals are narrower than the allowable error, the model is accurate enough; otherwise, an additional MBDoE iteration is conducted, targeting the parameters that do not meet the criterion. Further studies may aim at expanding this approach to more complex and multi-scale systems, and deeply analyse how uncertainties propagate across different levels of the model structure. This could involve integrating robust optimisation techniques that consider worst-case scenarios, and ensuring model reliability even under extreme parameter variations.

Another promising avenue for future research lies in bridging MBDoE with advanced control techniques, which align with online experimental design strategies. These methodologies enable the simultaneous refinement of model parameters and real-time optimisation of input policies, offering a principled way to integrate learning and decision-making under uncertainty. The synergy between MBDoE and control theory is expected to play a key role in the development of next-generation experimental design frameworks (Hu et al., 2025), possibly exploiting automated/autonomous platforms.

4.4. Other research directions

The integration of MBDoE with emerging computational and data-driven techniques has the potential to automate the development of flowsheet models. Flowsheet models represent the interconnected unit operations in a process, along with material and energy streams, and are crucial for process optimisation and system understanding (Boukouvala et al., 2012). Traditional flowsheeting often relies on iterative cycles of experimentation and model refinement, but these can be resource-intensive and time-consuming, especially in complex industries like pharmaceutical manufacturing (Destro and Barolo, 2022).

By combining MBDoE with advanced methodologies, flowsheeting could possibly be automated more systematically. A possible direction is to combine the concept of superstructure used in flowsheeting synthesis, with a MBDoE perspective, whereby the components in the flowsheet (which are discrete decision variables) and operating conditions (which are continuous decision variables) constitute the design vector for the optimal experiments. This integration could reduce manual

intervention, improve model accuracy, and facilitate real-time optimisation of process flows. However, high-dimensional datasets and the complexity of physical mechanisms pose significant challenges in developing accurate and reliable flowsheets. In this context, MBDoE could guide the experiment design process while allowing for the incorporation of new data-driven techniques that enhance model precision and scalability. Another relevant point is that flowsheeting models are generally calibrated on a unit-by-unit base, i.e., fit-for-purpose data are generated for each unit sequentially. However, a global calibration approach, where all unit operations are considered simultaneously, can be effective, too (Geremia et al., 2023). Future works could further investigate pros and cons of different workflows that can be applied to calibrate flowsheeting models.

Symbolic regression has recently emerged as a powerful tool for deriving interpretable mathematical expressions directly from data, without requiring prior knowledge of the underlying physical mechanisms (Rogers et al., 2024). In the context of MBDoE, symbolic regression can aid in identifying governing equations for interconnected unit operations, making it a valuable tool for enhancing the adaptability and interpretability of process models.

By exploring a vast space of potential expressions, symbolic regression allows the identification of relationships between input and output features, even when the exact functional form of these relationships is unknown. When combined with MBDoE, this technique can systematically generate process models and optimise experimental conditions based on real-time data analysis. For example, Rogers et al. (2024) employed symbolic regression to derive mechanistic expressions for a personal care product, demonstrating the potential of this integrated approach. MBDoE was then used to discriminate among candidate models while optimising the process flow diagram, effectively balancing exploration and optimisation.

Similarly, other machine learning techniques, such as GPs or neural networks, could be employed to complement MBDoE by capturing system dynamics while maintaining interpretability. For instance, Rogers et al. (2025) proposed an MBDoE framework that integrates symbolic regression with neural network feature attribution to improve model structure identification and strategic sampling. These approaches could address challenges in model development, particularly when traditional first-principle modelling is limited by incomplete knowledge of the underlying mechanisms, shifting towards a more data-driven yet interpretable paradigm.

5. Conclusions

Over the years, MBDoE has been successfully applied to support experiment design, allowing to efficiently gather informative data for model calibration and parameter estimation. This review has explored the theoretical foundations of MBDoE, its standard formulations, and recent advances to address challenges such as parameter uncertainty, ill-conditioning, and multi-objective optimisation. Application across different fields have demonstrated the versatility and impact of MBDoE in experiment design. The final section of this review explored emerging perspectives that extend MBDoE beyond its traditional applications. The most significant points are summarized below.

- (i) Integration of MBDoE with hybrid modelling paradigms, where mechanistic models are combined with data-driven components to better capture the complexities of real-world systems. Challenges include the design of a combined architecture, problem formulation and algorithms, accounting for parameter interplay, computational scalability, and uncertainty quantification in hybrid models.
- (ii) Consideration of uncertainties in process inputs, which are completely overlooked in standard MBDoE formulations. Extending robust design strategies to account for variability/uncertainty in critical input factors will no doubt improve the

reliability of experiment designs in real-world conditions. New challenges include the accurate characterization of uncertainty distributions at appropriate levels, and associated computational burden. Advanced techniques such as dimensionality reduction, surrogate modelling, and stochastic optimisation have been suggested as potential solutions, but further investigation is needed to establish scalable and efficient frameworks.

- (iii) New formulations of the optimal design problem. Future trends should focus on incorporating more complex information structures, leveraging Bayesian approaches, and consider broader, more sophisticated design objectives beyond parameter precision (to include multiple objective and constraints, perhaps varying as experimentation and model building proceed).
- (iv) The combination of MBDoE with emerging computational techniques (e.g., symbolic regression) for the automatic development of process models. This integrated approach is expected to reduce manual intervention, enhance interpretability, and facilitate the systematic generation of process models, particularly in data-rich environments.
- (v) Design with high-throughput automated equipment. In this environment, data generation is no longer the bottleneck; instead, the focus shifts to selecting which experiments to run and how to effectively interpret the generated data. Developing strategies to optimise experimental selection and extract meaningful insights from large-scale automated experimentation will be a crucial aspect of future research.

Looking ahead, researchers have several exciting fields to explore. The development of robust MBDoE techniques that can integrate with data-driven modelling remains a priority. Second, addressing the computational challenges associated with high-dimensional systems and iterative design frameworks will require interdisciplinary collaboration, leveraging advancements in parallel computing, metaheuristic algorithms, and real-time data analytics. Moreover, extending MBDoE applicability to emerging fields, and focus on new formulations, could also broaden its impact.

MBDoE stands at a crossroads where traditional methodologies are progressively being refined to meet the demands of increasingly complex and interdisciplinary systems. By addressing the challenges outlined in this review and exploiting innovative computational and modelling techniques, MBDoE has the potential to remain at the forefront of experiment design, driving scientific and industrial advancements in the years to come.

A final comment is due. In spite of the considerable advances discussed, practical applicability of MBDoE approaches and large benefits demonstrated across a variety of applications, it is fair to note that their use in industry is still largely a niche event. In our view, this is less a consequence of appropriate formulations, good algorithms and useful solutions than appropriate practices and workflows. In reality, experiment execution, data analysis and model development are still largely done sequentially, often not even adjacent in time, and by personnel with little overlap in expertise. The opportunity for synergistic exploitation of models and data is thus missed. An ideal solution would be for MBDoE methods to be incorporated directly into lab control software or experimental equipment as integral support software. The advent of high throughput experimental platforms, autonomous experimental systems, and integration with artificial intelligence for improved usability and interpretation offer both a challenge and a great opportunity.

CRedit authorship contribution statement

Margherita Geremia: Methodology, Writing – original draft, Conceptualization. **Sandro Macchietto:** Writing – review & editing, Methodology. **Fabrizio Bezzo:** Supervision, Writing – review & editing, Conceptualization, Methodology.

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 would like to thank prof. Massimiliano Barolo (University of Padova) for critically reading the manuscript and providing useful insights and suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ces.2025.122347>.

Data availability

No data was used for the research described in the article.

References

- Abdi, K., Celse, B., McAuley, K., 2024. Propagating input uncertainties into parameter uncertainties and model prediction uncertainties – A review. *Can. J. Chem. Eng.* 102, 254–273.
- Abolhasani, M.A., Kumacheva, E., 2023. The rise of self-driving labs in chemical and materials sciences. *Nat. Synth.* 2, 483–492.
- Abt, V., Barz, T., Cruz-Bournazou, M.N., Herwig, C., Kroll, P., Möller, J., Pörtner, R., Schenkendorf, R., 2018. Model-based tools for optimal experiments in bioprocess engineering. *Curr. Opin. Chem. Eng.* 22, 244–252.
- Agi, D.T., Jones, K.D., Watson, M.J., Lynch, H.G., Dougher, M., Chen, X., Carlozo, M.N., Dowling, A.W., 2024. Computational toolkits for model-based design and optimization. In *Curr. Opin. Chem. Eng.* 43, 100994.
- Agunloye, E., Petsagkourakis, P., Yusuf, M., Labes, R., Chamberlain, T., Muller, F.L., Bourne, R.A., Galvanin, F., 2024a. Automated kinetic model identification via cloud services using model-based design of experiments. *Reac. Chem. Eng.* 9, 1859–1876.
- Agunloye, E., Yusuf, M., Chamberlain, T.W., Muller, F.L., Bourne, R.A., Galvanin, F., 2024b. Model-based Design of Experiments for the Identification of Kinetic Models of Amide Formation. *Comput. Aided Chem. Eng.* 53, 385–390.
- Alen, A., Díaz, H., Rao, V., Saibaba, A.K., 2025. Optimal sensor placement under model uncertainty in the weak-constraint 4D-Var framework. 10.48550/arXiv.2502.00150.
- Alizadeh, R., Allen, J.K., Mistree, F., 2020. Managing computational complexity using surrogate models: a critical review. *Res. Eng. Des.* 31, 275–298.
- Asprey, S.P., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. *Comput. Chem. Eng.* 24, 1261–1267.
- Asprey, S.P., Macchietto, S., 2002. Designing robust optimal dynamic experiments. *J. Process Control.* 12, 545–556.
- Asprion, N., Böttcher, R., Pack, R., Stavrou, M.E., Höller, J., Schwientek, C., Bortz, M., 2019. Gray-box modeling for the optimization of chemical processes. *Chem. Ingenieur Tech.* 91, 305–313.
- Asprion, N., Böttcher, R., Mairhofer, J., Yliruka, M., Höller, J., Schwientek, J., Vanaret, C., Bortz, M., 2020. Implementation and Application of Model-Based Design of Experiments in a Flowsheet Simulator. *J. Chem. Eng. Data* 65, 1135–1145.
- Azarpour, A., Borhani, T.N., Alwi, S.R.W., Manan, Z.A., Mutalib, M.I.A., 2017. A generic hybrid model development for process analysis of industrial fixed-bed catalytic reactors. *Chem. Eng. Res. Des.* 117, 149–167.
- Baldea, M., Georgiou, A.T., Gopaluni, B., Mercangöz, M., Pantelides, C.C., Sheth, K., Zavala, V.M., Georgakis, C., 2025. From automated to autonomous process operations. *Comput. Chem. Eng.* 196, 109064.
- Banga, J.R., Balsa-Canto, E., 2008. Parameter estimation and optimal experimental design. *Essays Biochem.* 45, 195–210.
- Bard, Y., 1974. *Nonlinear Parameter Estimation*. Academic Press, New York, NY (US).
- Bardow, A., Marquardt, W., 2004. Incremental and simultaneous identification of reaction kinetics: methods and comparison. *Chem. Eng. Sci.* 59, 2673–2684.
- Barz, T., Arellano-García, H., Wozny, G., 2010. Handling uncertainty in model-based optimal experimental design. *Ind. Eng. Chem. Res.* 49, 5702–5713.
- Barz, T., López, C., D.C., Arellano-García, H., Wozny, G., 2013. Experimental evaluation of an approach to online redesign of experiments for parameter determination. *AIChE J.* 59, 1981–1995.
- Barz, T., Bournazou, M.N.C., 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.
- Barz, T., Sommer, A., Wilms, T., Neubauer, P., Cruz Bournazou, M.N., 2018. Adaptive optimal operation of a parallel robotic liquid handling station. *IFAC-PapersOnline* 51, 765–770.
- Barz, T., Kager, J., Herwig, C., Neubauer, P., Bournazou, M.N.C., Galvanin, F., 2022. In: *Characterization of reactions and growth in automated continuous flow and bioreactor platforms—From linear DoE to model-based approaches*. Elsevier, pp. 273–319.
- Bauer, L., Bock, H.G., Körkel, S., Schlöder, J.P., 2000. Numerical methods for optimum experimental design in DAE systems. *J. Comput. Appl. Math.* 120, 1–25.
- Bernardi, A., Nikolau, A., Meneghesso, A., Morosinotto, T., Chachuat, B., Bezzo, F., 2016. High-fidelity modelling methodology of light-limited photosynthetic production in microalgae. *PLoS one* 11, e0152387.
- Boggs, P.T., Tolle, J.W., 1995. Sequential quadratic programming. *Acta Numer.* 4, 1–51.
- 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, 6891–6903.
- Bouchkira, I., Elmasaoui, S., Benjelloun, S., Latifi, A.M., 2023. Global estimability analysis and model-based design of experiments in surrogate modeling. *Comput. Aided Chem. Eng.* 52, 643–648.
- Boukouvala, F., Niotis, V., Ramachandran, R., Muzzio, F.J., Ierapetritou, M.G., 2012. An integrated approach for dynamic flowsheet modeling and sensitivity analysis of a continuous tablet manufacturing process. *Comput. Chem. Eng.* 42, 30–47.
- Box, G.E.P., Hunter, W.G., Hunter, J.S., 1978. *Statistics for experiments. An introduction to design, data analysis and model building*. John Wiley and Sons, New York, NY (US).
- Brockmann, D., Rosenwinkel, K.H., Morgenroth, E., 2008. Practical identifiability of biokinetic parameters of a model describing two-step nitrification in biofilms. *Biotechnol. Bioeng.* 101, 497–514.
- Buede, D.M., Miller, W.D., 2024. *The engineering design of systems: models and methods*. John Wiley and Sons, New York, NY (US).
- Buzzi-Ferraris, G., Forzatti, P., 1983a. Sequential experimental design procedure for discriminating among rival models. *Chem. Eng. Sci.* 38, 225–238.
- Buzzi-Ferraris, G., Forzatti, P., 1983b. Sequential experimental design for model discrimination in the case of multiple responses. *Chem. Eng. Sci.* 39, 81–96.
- Cao, Y., Su, F., Antwi-Afari, M.F., Lei, J., Wu, X., Liu, Y., 2024. Enhancing mix proportion design of low carbon concrete for shield segment using a combination of Bayesian optimization-NGBoost and NSGA-III algorithm. *J. Clean. Prod.* 465, 142746.
- Cenci, F., Bano, G., Christodoulou, C., Vueva, Y., Zomer, S., Barolo, M., Bezzo, F., Facco, P., 2022. Streamlining tablet lubrication design via model-based design of experiments. *Int. J. Pharm.* 614, 121435.
- Cenci, F., Pankajakshan, A., Facco, P., Galvanin, F., 2023. An exploratory model-based design of experiments approach to aid parameters identification and reduce model prediction uncertainty. *Comput. Chem. Eng.* 177, 108353.
- Chakrabarty, A., Buzzard, G.T., Rundell, A.E., 2013. Model-based design of experiments for cellular processes. *Wiley Interdiscip. Res. Syst. Biol. Med.* 5, 181–203.
- Chaudhuri, P., Mykland, P.A., 1993. Nonlinear experiments: Optimal design and inference based on likelihood. *J. Am. Stat. Assoc.* 88, 538–546.
- Chen, Y., Ierapetritou, M., 2020. A framework of hybrid model development with identification of plant-model mismatch. *AIChE J.* 66, e16996.
- Chhatre, S., Francis, R., Newcombe, A.R., Zhou, Y.H., Titchener-Hooker, N., King, J., Keshavarz-Moore, E., 2008. Global Sensitivity Analysis for the determination of parameter importance in bio-manufacturing processes. *Biotechnol. Appl. Biochem.* 51, 79–90.
- Chu, Y., Hahn, J., 2010. Quantitative optimal experimental design using global sensitivity analysis via quasi-linearization. *Ind. Eng. Chem. Res.* 49, 7782–7794.
- Chu, Y., Hahn, J., 2012. Generalization of a parameter set selection procedure based on orthogonal projections and the D-optimality criterion. *AIChE J.* 58, 2085–2096.
- Conn, A.R., Scheinberg, K., Vicente, L.N., 2009. *Introduction to Derivative-Free Optimization*. SIAM.
- Ciucci, F., 2013. Revisiting parameter identification in electrochemical impedance spectroscopy: Weighted least squares and optimal experimental design. *Electrochim. Acta* 87, 532–545.
- Cruz Bournazou, M.N., Barz, T., Nickel, D.B., Lopez, D.C., Glauche, F., Knepper, A., Neubauer, P., 2017. Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities. *Biotechnol. Bioeng.* 114, 610–619.
- Das, I., Dennis, J., 1997. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Struct. Optim.* 14, 63–69.
- Das, I., Dennis, J., 1998. Normal-boundary intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM J. Optim.* 8, 631–657.
- De-Luca, R., Galvanin, F., Bezzo, F., 2016. A methodology for direct exploitation of available information in the online model-based redesign of experiments. *Comput. Chem. Eng.* 91, 195–205.
- De-Luca, R., Bano, G., Tomba, E., Bezzo, F., Barolo, M., 2020. Accelerating the Development and Transfer of Freeze-Drying Operations for the Manufacturing of Biopharmaceuticals by Model-Based Design of Experiments. *Ind. Eng. Chem. Res.* 59, 20071–20085.
- Destro, F., Barolo, M., 2022. A review on the modernization of pharmaceutical development and manufacturing – Trends, perspectives, and the role of mathematical modeling. *Int. J. Pharm.* 620, 121715.
- Deussen, P., Galvanin, F., 2023. A joint model-based design of experiments approach for the identification of Kriging models in geological exploration. *Comput. Aided Chem. Eng.* 52, 789–794.
- Dürbaum, A., Kroll, A., 2016. On robust experiment design for identifying locally affine Takagi-Sugeno models. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 001844–001849. IEEE.
- Echtermeier, A., Amar, Y., Zakrzewski, J., Lapkin, A., 2017. Self-optimisation and model-based design of experiments for developing a C–H activation flow process. *Beilstein J. Org. Chem.* 13, 150–163.

- Elster, C., 2007. Calculation of uncertainty in the presence of prior knowledge. *Metrologia* 44, 111–116.
- Espie, D., Macchietto, S., 1989. The optimal design of dynamic experiments. *AIChE J.* 35, 223–229.
- Evripidou, N., Galvanin, F., Angeli, P., 2024. Mechanistic modelling of separating dispersions in pipes using model-based design of experiments techniques. *Chem. Eng. Sci.* 284, 119504.
- Faller, D., Klingmüller, U., Timmer, J., 2003. Simulation methods for optimal experimental design in systems biology. *Simulation* 79, 717–725.
- Fernández-Godino, M. G., 2016. Review of multi-fidelity models. *arXiv preprint arXiv:1609.07196*.
- Fisher, R.A., 1925. Statistical methods for research workers. Oliver and Boyd, Edinburgh.
- Fisher, R.A., 1934. The Design of Experiments. Oliver and Boyd, Edinburgh.
- Fliess, M., Lévine, J., Martin, P., Rouchon, P., 1995. Flatness and defect of non-linear systems: introductory theory and examples. *Int. J. Control* 61, 1327–1361.
- Flores-Sánchez, A., Flores-Tlacuahuac, A., Pedraza-Segura, L.L., 2013. Model-based experimental design to estimate kinetic parameters of the enzymatic hydrolysis of lignocellulose. *Ind. Eng. Chem. Res.* 52, 4834–4850.
- Forrester, A.L.J., Sobester, A., Keane, A.J., 2008. Engineering Design via Surrogate Modelling: A Practical Guide. Wiley, New York (US).
- Forté, E., von Harbou, E., Burger, J., Aspöron, N., Bortz, M., 2017. Optimal Design of Laboratory and Pilot-Plant Experiments Using Multiobjective Optimization. *Chemie-Ingenieur-Technik* 89, 645–654.
- Fox, G.C., Williams, R.D., Messina, G.C., 2014. Parallel computing works! Morgan Kaufmann Publishers, San Francisco (US).
- Franceschini, G., Macchietto, S., 2008a. Model-based design of experiments for parameter precision: State of the art. *Chem. Eng. Sci.* 63, 4846–4872.
- Franceschini, G., Macchietto, S., 2008b. Novel anticorrelation criteria for model-based experiment design: Theory and formulations. *AIChE J.* 54, 1009–1024.
- Franceschini, G., Macchietto, S., 2008c. Anti-correlation approach to model-based experiment design: Application to a biodiesel production process. *Ind. Eng. Chem. Res.* 47, 2331–2348.
- Friso, A., Galvanin, F., 2024. An optimization-free Fisher information driven approach for online design of experiments. *Comput. Chem. Eng.* 187, 108724.
- Galvanin, F., Macchietto, S., Bezzo, F., 2007. Model-based design of parallel experiments. *Ind. Eng. Chem. Res.* 46, 871–882.
- Galvanin, F., Barolo, M., Bezzo, F., 2009a. Online model-based redesign of experiments for parameter estimation in dynamic systems. *Ind. Eng. Chem. Res.* 48, 4415–4427.
- Galvanin, F., Barolo, M., Macchietto, S., Bezzo, F., 2009b. Optimal design of clinical tests for the identification of physiological models of type 1 diabetes mellitus. *Ind. Eng. Chem. Res.* 48, 1989–2002.
- Galvanin, F., Barolo, M., Bezzo, F., Macchietto, S., 2010. A backoff strategy for model-based experiment design under parametric uncertainty. *AIChE J.* 56, 2088–2102.
- Galvanin, F., Barolo, M., Macchietto, S., Bezzo, F., 2011. Optimal design of clinical tests for the identification of physiological models of type 1 diabetes in the presence of model mismatch. *Med. Biol. Eng. Comput.* 49, 263–277.
- Galvanin, F., Barolo, M., Pannocchia, G., Bezzo, F., 2012. Online model-based redesign of experiments with erratic models: a disturbance estimation approach. *Comput. Chem. Eng.* 42, 138–151.
- Galvanin, F., Cao, E., Al-Rifai, N., Dua, V., Gavrilidis, A., 2015. Optimal design of experiments for the identification of kinetic models of methanol oxidation over silver catalyst. *Chimica Oggi-Chemistry Today* 33, 51–56.
- Galvanin, F., Cao, E., Al-Rifai, N., Gavrilidis, A., Dua, V., 2016a. A joint model-based experimental design approach for the identification of kinetic models in continuous flow laboratory reactors. *Comput. Chem. Eng.* 95, 202–215.
- Galvanin, F., Marchesini, R., Barolo, M., Bezzo, F., Fidaleo, M., 2016b. Optimal design of experiments for parameter identification in electrodialysis models. *Chem. Eng. Res. Des.* 105, 107–119.
- Galvanin, F., Galletta, E., Bertomoro, A., Daidone, V., Casonato, A., 2024. Optimal design of infusion tests for the identification of physiological models of acquired von Willebrand syndrome. *Chem. Eng. Sci.* 286, 119660.
- Garre, A., Peñalver-Soto, J.L., Esnoz, A., Iguaz, A., Fernandez, P.S., Egea, J.A., 2019. On the use of in-silico simulations to support experimental design: A case study in microbial inactivation of foods. *PLoS ONE* 14, e0220683.
- Gelman, A., 2006. Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics* 48, 432–435.
- Gelman, A., Shalizi, C.R., 2013. Philosophy and the practice of Bayesian statistics. *Br. J. Math. Stat. Psychol.* 66, 8–38.
- Georgakis, C., 2013. Design of dynamic experiments: A data-driven methodology for the optimization of time-varying processes. *Ind. Eng. Chem. Res.* 52, 12369–12382.
- Geremia, M., Diab, S., Christodoulou, C., Bano, G., Barolo, M., Bezzo, F., 2023. A general procedure for the evaluation of the prediction fidelity of pharmaceutical systems models. *Chem. Eng. Sci.* 280, 118972.
- Geremia, M., Cisco, G., Diab, S., Bano, G., Bezzo, F., 2024a. Targeting fidelity of pharmaceutical systems models by optimization of precision on parameter estimates. *Comput. Chem. Eng.* 181, 108542.
- Geremia, M., Bezzo, F., Ierapetritou, M.G., 2024b. Design space determination of pharmaceutical processes: Effects of control strategies and uncertainty. *Europ. J. Pharm. Biopharm.* 194, 159–169.
- Ghodba, A., Richelle, A., McCreedy, C., Ricardez-Sandoval, L., Budman, H., 2025. A robust batch-to-batch optimization framework for pharmaceutical applications. *Comput. Chem. Eng.* 193, 108935.
- Gibson, L.A., Liu, F.F., Harris, T.J., McAuley, K.B., 2024. Simple Method for Incorporating Prior Parameter Information during Model-Based D-Optimal Design of Experiments for Multi-Response Chemical Process Models. *Ind. Eng. Chem. Res.* 63, 10651–10665.
- Goodwin, G.C., Payne, R.L., 1977. Dynamic System Identification: Experiment Design and Data Analysis. Academic Press.
- Goodwin, G.C., Sin, K.S., 2014. Adaptive filtering prediction and control. Courier Corporation.
- Griewank, A., Walther, A., 2008. Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation. SIAM.
- Habicher, T., Rauls, E.K.A., Egidio, F., Keil, T., Klein, T., Daub, A., Büchs, J., 2020. Establishing a fed-batch process for protease expression with *Bacillus licheniformis* in polymer-based controlled-release microtiter plates. *Biotechnol. J.* 15, 1900088.
- Hametner, C., Stadlbauer, M., Deregnaucourt, M., Jakubek, S., Winsel, T., 2013. Optimal experiment design based on local model networks and multilayer perceptron networks. *Eng. Appl. Artif. Intell.* 26, 251–261.
- He, F., Yue, H., Brown, M., 2009. Investigating Bayesian Robust Experimental Design with Principles of Global Sensitivity Analysis. *IFAC Proceedings Volumes* 43, 577–582.
- Han, L., Zhou, Z., Bollas, G.M., 2016. Model-based analysis of chemical-looping combustion experiments. Part I: Structural identifiability of kinetic models for NiO reduction. *AIChE J.* 62, 2419–2431.
- He, F., Brown, M., Yue, H., 2010. Maximin and Bayesian robust experimental design for measurement set selection in modelling biochemical regulatory systems. *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal* 20, 1059–1078.
- He, C., Zhang, Y., Gong, D., Ji, X., 2023. A review of surrogate-assisted evolutionary algorithms for expensive optimization problems. *Expert Syst. Appl.* 217, 119495.
- Hennecke, L., Schaare, L., Skiborowski, M., Liese, A., 2024. Closed-loop identification of enzyme kinetics applying model-based design of experiments. *Reac. Chem. Eng.* 9, 2984–2993.
- Hernández Rodríguez, T., Posch, C., Schmutzhard, J., Stettner, J., Weihs, C., Pörtner, R., Frahm, B., 2019. Predicting industrial-scale cell culture seed trains—A Bayesian framework for model fitting and parameter estimation, dealing with uncertainty in measurements and model parameters, applied to a nonlinear kinetic cell culture model, using an MCMC method. *Biotechnol. Bioeng.* 116, 2944–2959.
- Hjalmarsson, H., 2005. From experiment design to closed-loop control. *Automatica* 41, 393–438.
- Hjalmarsson, H., 2008. System identification of complex and structured systems. *Eur. J. Control* 15, 275–310.
- Homma, T., Saltelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.* 52, 1–17.
- Houska, B., Telen, D., Logist, F., Diehl, M., van Impe, J.F.M., 2015. An economic objective for the optimal experiment design of nonlinear dynamic processes. *Automatica* 51, 98–103.
- Hu, J., Zachariah, D., Wigren, T., Stoica, P., 2025. Adaptive Input Design for Nonlinear System Identification with Operational Constraints. *arXiv preprint arXiv:2502.20941*.
- Huan, X., Marzouk, Y.M., 2013. Simulation-based optimal Bayesian experimental design for nonlinear systems. *J. Comput. Phys.* 232, 288–317.
- Huan, X., Jagalur, J., Marzouk, Y., 2024. Optimal experimental design: Formulations and computations. *Acta Numerica* 33, 715–840.
- Hunter, W.G., Reiner, A.M., 1965. Designs for discriminating between two rival models. *Technometrics* 7, 307–323.
- Issanchou, S., Cognet, P., Cabassud, M., 2003. Precise parameter estimation for chemical batch reactions in heterogeneous medium. *Chem. Eng. Sci.* 58, 1805–1813.
- Jiang, Z., Portha, J.F., Commenge, J.M., Falk, L., 2019. Development and implementation of systematic model-development strategy using model-based experimental design. *Chem. Eng. Res. Des.* 146, 290–310.
- Jost, F., Sager, S., Thi-Thien Le, T., 2017. A feedback optimal control algorithm with optimal measurement time points. *Processes* 5, 10.
- Julier, S., Uhlmann, J., 1996. A general method for approximating nonlinear transformations probability distributions. Department of Engineering Science, University of Oxford, Tech. Rep.
- Jung, F., Janssen, F.A.L., Ksiazkiewicz, A., Caspari, A., Mhamdi, A., Pich, A., Mitsos, A., 2019. Identifiability analysis and parameter estimation of microgel synthesis: a set-membership approach. *Ind. Eng. Chem. Res.* 58, 13675–13685.
- Kaiser, S., Engell, S., 2020. Integrating Superstructure Optimization under Uncertainty and Optimal Experimental Design in early Stage Process Development. *Comput. Aided Chem. Eng.* 48, 799–804.
- Kaiser, S., Engell, S., 2022. Evaluating the Impact of Model Uncertainties in Superstructure Optimization to Reduce the Experimental Effort. *Comput. Aided Chem. Eng.* 49, 79–84.
- Kalyanaraman, J., Fan, Y., Labreche, Y., Lively, R.P., Kawajiri, Y., Realf, M.J., 2015. Bayesian estimation of parametric uncertainties, quantification and reduction using optimal design of experiments for CO₂ adsorption on amine sorbents. *Comput. Chem. Eng.* 81, 376–388.
- Kalyanaraman, J., Kawajiri, Y., Realf, M.J., 2020. Bayesian design of experiments for adsorption isotherm modeling. *Comput. Chem. Eng.* 135, 106774.
- Kaminski, S., Bardow, A., Leonhard, K., 2016. The trade-off between experimental effort and accuracy for determination of PCP-SAFT parameters. *Fluid Phase Equilib.* 428, 182–189.
- Karakida, R., Osawa, K., 2020. Understanding approximate fisher information for fast convergence of natural gradient descent in wide neural networks. *Adv. Neural Inf. Process. Syst.* 33, 10891–10901.
- Katsoulas, K., Galvanin, F., Mazzei, L., Besenhard, M., Sorensen, E., 2025. Model-based design of experiments for efficient and accurate isotherm model identification in High Performance Liquid Chromatography. *Comput. Chem. Eng.* 109021.
- Kiefer, J., 1959. Optimum experimental design. *J. Royal Stat. Soc.* 21, 272–319.

- Kilari, H., Barhate, Y., Kang, Y. S., Nagy, Z. K., 2023. A Systematic Framework for Iterative Model-Based Experimental Design of Batch and Continuous Crystallization Systems. In *Computer Aided Chemical Engineering*, 52, 1501–1506. Elsevier.
- Kim, J.H., Lee, J.M., 2018. Successive complementary model-based experimental designs for parameter estimation of fed-batch bioreactors. *Bioproc. Biosyst. Eng.* 41, 1767–1777.
- Kim, J.W., Krausch, N., Aizpuru, J., Barz, T., Lucia, S., Martínez, E.C., Neubauer, P., Bournazou, M.N.C., 2022. Model predictive control guided with optimal experimental design for pulse-based parallel cultivation. *IFAC-PapersOnLine* 55, 934–939.
- Knap, J., Spear, C., Leiter, K., Becker, R., Powell, D., 2016. A computational framework for scale-bridging in multi-scale simulations. *Int. J. Numer. Methods Eng.* 108, 1649–1666.
- Kontoravdi, C., Pistikopoulos, E.N., Mantalaris, A., 2010. Systematic development of predictive mathematical models for animal cell cultures. *Comput. Chem. Eng.* 34, 1192–1198.
- Körkel, S., Kostina, E., Bock, H.G., Schlöder, J.P., 2004. Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optim. Methods Softw.* 19, 327–338.
- Krausch, N., Barz, T., Sawatzki, A., Gruber, M., Kamel, S., Neubauer, P., Cruz Bournazou, M.N., 2019. Monte Carlo Simulations for the Analysis of Non-linear Parameter Confidence Intervals in Optimal Experimental Design. *Front. Bioeng. Biot.* 7, 122.
- Kravaris, C., Hahn, J., Chu, Y., 2013. Advances and selected recent developments in state and parameter estimation. *Comput. Chem. Eng.* 51, 111–123.
- Kroll, A., Dürrbaum, A., 2015. On joint optimal experiment design for identifying partition and local model parameters of Takagi-Sugeno models. *IFAC-PapersOnLine* 48, 1427–1432.
- Kuceroová, A., Sykora, J., Janouchová, E., Jarušková, D., Chleboún, J., 2016. Acceleration of robust experiment design using Sobol indices and polynomial chaos expansion. In: *Proceedings of the 7th International Workshop on Reliable Engineering Computing (REC)*, Bochum, Germany. 15–17.
- Kuhn, M., Johnson, K., 2013. *Applied predictive modeling*. Springer, New York (US).
- Kusumo, K.P., Kuriyan, K., Vaidyaraman, S., García-Muñoz, S., Shah, N., Chachuat, B., 2022a. Probabilistic framework for optimal experimental campaigns in the presence of operational constraints. *Reac. Chem. Eng.* 7, 2359–2374.
- Kusumo, K.P., Kuriyan, K., Vaidyaraman, S., García-Muñoz, S., Shah, N., Chachuat, B., 2022b. Risk mitigation in model-based experiment design: a continuous-effort approach to optimal campaigns. *Comput. Chem. Eng.* 159, 107680.
- Lei, G., Zhu, J., Guo, Y., Liu, C., Ma, B., 2017. A review of design optimization methods for electrical machines. *Energies* 10, 1962.
- Li, Y.H., Yeh, P.C., 2012. An interpretation of the Moore-Penrose generalized inverse of a singular Fisher information matrix. *IEEE Trans. Signal Process.* 60, 5532–5536.
- Lindley, D.V., 1956. On a measure of the information provided by an experiment. *Ann. Math. Stat.* 28, 986–1005.
- Lindqvist, K., Hjalmarsson, H., 2001. Identification for control: adaptive input design using convex optimisation. In: *Proc. 40th IEEE Conference on Decision and Control*, Orlando, Florida (USA).
- Liu, X., Estrada, L., Ouimet, J.A., McClure, M., Latulippe, D.R., Phillip, W.A., Dowling, A. W., 2025. Characterizing Transport Properties of Surface Charged Nanofiltration Membranes via Model-Based Data Analytics. *Ind. Eng. Chem. Res.* <https://doi.org/10.1021/acs.iecr.4c04763>.
- Lohmann, T., Bock, H.G., Schloeder, J.P., 1992. Numerical methods for parameter estimation and optimal experiment design in chemical reaction systems. *Ind. Eng. Chem. Res.* 31, 54–57.
- Longhi, D.A., Martins, W.F., da Silva, N.B., Carciofi, B.A.M., de Aragão, G.M.F., Laurindo, J.B., 2016. Estimation of the Temperature Dependent Growth Parameters of *Lactobacillus Viridescens* in Culture Medium with Two-step Modelling and Optimal Experimental Design Approaches. *Procedia Food Sci.* 7, 25–28.
- Longhi, D.A., Martins, W.F., da Silva, N.B., Carciofi, B.A.M., de Aragão, G.M.F., Laurindo, J.B., 2017. Optimal experimental design for improving the estimation of growth parameters of *Lactobacillus viridescens* from data under non-isothermal conditions. *Int. J. Food Microbiol.* 240, 57–62.
- Longhi, D.A., da Silva, N.B., Martins, W.F., Carciofi, B.A.M., de Aragão, G.M.F., Laurindo, J.B., 2018. Optimal experimental design to model spoilage bacteria growth in vacuum-packaged ham. *J. Food Eng.* 216, 20–26.
- López, C., D. C., Barz, T., Peñuela, M., Villegas, A., Ochoa, S., Wozny, G., 2013. Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production. *Biotechnol. prog.* 29, 1064–1082.
- López, C., D.C., Barz, T., Körkel, S., Wozny, G., 2015. Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. *Comput. Chem. Eng.* 77, 24–42.
- Luna, M.F., Martínez, E.C., 2018. Optimal design of dynamic experiments in the development of cybernetic models for bioreactors. *Chem. Eng. Res. Des.* 136, 334–346.
- Lu, J.M., Pan, J.Z., Mo, Y.M., Fang, Q., 2024. Automated Intelligent Platforms for High-Throughput Chemical Synthesis. *Artificial Intelligence*. Chemistry 100057.
- Lucia, S., Paulen, R., 2014. Robust nonlinear model predictive control with reduction of uncertainty via robust optimal experiment design. *IFAC Proceedings Volumes* 47, 1904–1909.
- Luo, L., Yao, Y., Gao, F., 2015. Iterative improvement of parameter estimation for model migration by means of sequential experiments. *Comput. Chem. Eng.* 73, 128–140.
- Maclaurin, D., Duvenaud, D., Adams, R. P., 2015. Autograd: Effortless gradients in numpy. In: *ICML 2015 AutoML workshop*, 238, 5.
- Maheshwari, V., Kandpal, M., Samavedham, L., 2013a. Multi-criteria optimization based experimental design for parameter estimation of a double feedback gene switching model. *IFAC-PapersOnLine* 12, 333–337.
- Maheshwari, V., Rangaiah, G.P., Samavedham, L., 2013b. Multiobjective framework for model-based design of experiments to improve parameter precision and minimize parameter correlation. *Ind. Eng. Chem. Res.* 52, 8289–8304.
- Maheshwari, V., Rangaiah, G.P., Lau, T., Samavedham, L., 2015. Application of design of experiments in hemodialysis: Optimal sampling protocol for β_2 -microglobulin kinetic model. *Chem. Eng. Sci.* 131, 84–90.
- Manesso, E., Sridharan, S., Gunawan, R., 2017. Multi-objective optimization of experiments using curvature and fisher information matrix. *Processes* 5, 63.
- Martinez, E.C., Cristaldi, M.D., Grau, R.J., 2009. Design of Dynamic Experiments in Modeling for Optimization of Batch Processes. *Ind. Eng. Chem. Res.* 48, 3453–3465.
- Mascagni, M., 2015. *Monte Carlo Methods: Early History and The Basics*, https://www.cs.fsu.edu/~mascagni/MC_Basics.pdf.
- Matsunami, K., Vandéputte, T., Jiménez, A.A.B., Peeters, M., Ghijs, M., Van Hauwermeiren, D., Stauffer, F., dos Santos Schultz, E., Nopens, I., De Beer, T., 2023. Validation of model-based design of experiments for continuous wet granulation and drying. *Int. J. Pharm.* 646, 123493.
- McLean, K.A.P., McAuley, K.B., 2012. Mathematical modelling in chemical processes—obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *Can. J. Chem. Eng.* 90, 351–366.
- McMullen, J.P., Jensen, K.F., 2011. Rapid determination of reaction kinetics with an automated microfluidic system. *Organic Process Research and Development* 15, 398–407.
- Mdluli, T., Buzzard, G.T., Rundell, A.E., 2015. Efficient optimization of stimuli for model-based design of experiments to resolve dynamical uncertainty. *PLOS Comput. Biol.* 11, 1004488.
- Meneghetti, N., Facco, P., Bezzo, F., Barolo, M., 2014. A methodology to diagnose process/model mismatch in first-principles models. *Ind. Eng. Chem. Res.* 53, 14002–14013.
- Mennen, S.M., Alhambra, C., Allen, C.L., Barberis, M., Berritt, S., Brandt, T.A., Zajac, M. A., 2019. The evolution of high-throughput experimentation in pharmaceutical development and perspectives on the future. *Org. Process Res. Dev.* 23, 1213–1242.
- Mesbah, A., Streif, S., 2015. Probabilistic approach to robust optimal experiment design with chance constraints. *IFAC-PapersOnLine* 28, 100–105.
- Messac, A., Ismail-Yahaya, A., Mattson, C., 2003. The normalized normal constraint method for generating the Pareto frontier. *Struct. Multidisciplinary Optim.* 25, 86–98.
- Metropolis, N., Ulam, S., 1949. The Monte Carlo Method. *J. Am. Stat. Assoc.* 44, 335–341.
- Miao, H., Xia, X., Perelson, A.S., Wu, H., 2011. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.* 53, 3–39.
- Möller, J., Kuchemüller, K.B., Steinmetz, T., Koopmann, K.S., Pörtner, R., 2019. Model-assisted design of experiments as a concept for knowledge-based bioprocess development. *Bioprocess Biosyst. Eng.* 42, 867–882.
- Montgomery, D.C., 2013. *Introduction to Statistical Quality Control*. Wiley, Hoboken, NJ (US).
- Moser, A., Kuchemüller, K.B., Deppe, S., Hernández Rodríguez, T., Frahm, B., Pörtner, R., Hass, V.C., Möller, J., 2021. Model-assisted DoE software: optimization of growth and biocatalysis in *Saccharomyces cerevisiae* bioprocesses. *Bioprocess Biosyst. Eng.* 44, 683–700.
- Murphy, T.J., 2022. Chapter 2 A Brief History of Experimental Design. *Statistical Design and Analysis of Experiments with R*, In JABSTB <https://tjmurphy.github.io/jabstb/>.
- Neddermeyer, F., Marhold, V., Menzel, C., Krämer, D., King, R., 2016. Modelling the production of soluble hydrogenase in *Ralstonia eutropha* by on-line optimal experimental design. *IFAC-PapersOnLine* 49, 627–632.
- Nimmegeers, P., Telen, D., Logist, F., Van Impe, J., 2016. Dynamic optimization of biological networks under parametric uncertainty. *BMC Syst. Biol.* 10, 86.
- Nimmegeers, P., Telen, D., van Impe, J., 2017. A sampling-based stochastic optimal experiment design formulation with application to the Williams-Otto reactor. *IFAC-PapersOnLine* 50, 8716–8721.
- Nimmegeers, P., Bhonsale, S., Telen, D., Impe, J.V., 2020. Optimal experiment design under parametric uncertainty: a comparison of a sensitivities based approach versus a polynomial chaos based stochastic approach. *Chem. Eng. Sci.* 221, 115651.
- Nöh, K., Niedenführ, S., Beyß, M., Wiechert, W., 2018. A Pareto approach to resolve the conflict between information gain and experimental costs: multiple-criteria design of carbon labeling experiments. *PLOS Comput. Biol.* 14, 1006533.
- Nolan, S., Smerzi, A., Pezzè, L., 2021. A machine learning approach to Bayesian parameter estimation. *npj Quantum Inf.* 7, 169.
- Olofsson, S., Hebing, L., Niedenführ, S., Deisenroth, M.P., Misener, R., 2019. GPdoemd: A Python package for design of experiments for model discrimination. *Comput. Chem. Eng.* 125, 54–70.
- Overstall, A.M., Woods, D.C., Parker, B.M., 2020. Bayesian Optimal Design for Ordinary Differential Equation Models with Application in Biological Science. *J. American Stat. Ass.* 115, 583–598.
- Pal, K., Szilagyi, B., Burcham, C.L., Jarmer, D.J., Nagy, Z.K., 2021. Iterative model-based experimental design for spherical agglomeration processes. *AIChE J.* 67, e17178.
- Pankajakshan, A., Petsagkourakis, P., Galvanin, F., 2021. A framework for adaptive online model-based redesign of experiments in dynamic systems. In: *Comput. Aided Chem. Eng.* 50. Elsevier, pp. 1293–1298.
- Pankajakshan, A., Waldron, C., Quaglio, M., Gavrilidis, A., Galvanin, F., 2019. A Multi-Objective Optimal Experimental Design Framework for Enhancing the Efficiency of Online Model Identification Platforms. *Eng.* 5, 1049–1059.
- Pankajakshan, A., Pal, S., Snead, N., Almeida, J., Besenhard, M.O., Abukhamees, S., Craig, D.Q.M., Gavrilidis, A., Mazzei, L., Galvanin, F., 2025. MLAPI: A framework

- for developing machine learning-guided drug particle syntheses in automated continuous flow platforms. *Chem. Eng. Sci.* 302, 120780.
- Pargett, M., Rundell, A.E., Buzzard, G.T., Umulis, D.M., 2014. Model-Based Analysis for Qualitative Data: An Application in *Drosophila* Germline Stem Cell Regulation. *PLoS Comput. Biol.* 10, e1003498.
- 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, A1309–A1323.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimselshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems* 32, 8024–8035.
- Paulson, J.A., Martin-Casas, M., Mesbah, A., 2019. Optimal Bayesian experiment design for nonlinear dynamic systems with chance constraints. *J. Process Control* 77, 155–171.
- Petrik, O., Silva, J.D.A., Moura, F.M.M.V.E., 2013. Impact of distribution choice for representing input variation: analysis of uncertainty in travel demand simulation in context of information shortage. *Transp. Res. J.* 2344, 40–48.
- Petsagkourakis, P., Galvanin, F., 2020. Backoff-Based Model-Based Design of Experiments Under Model Mismatch. In: *Comput. Aided Chem. Eng.*, 48 Elsevier, pp. 1777–1782.
- Pistikopoulos, E.N., Barbosa-Povoa, A., Lee, J.H., Misener, R., Mitsos, A., Reklaitis, G.V., Venkatasubramanian, V., You, F., Gani, R., 2021. Process systems engineering—the generation next? *Comput. Chem. Eng.* 147, 107252.
- Powell, W.B., 2019. A unified framework for stochastic optimization. *Eur. J. Oper. Res.* 275, 795–821.
- Process Systems Enterprise, gPROMS®, v. 4.2, 2023. <https://www.siemens.com/global/en/products/automation/industry-software/gproms-digital-process-design-and-operations/gproms-digital-applications.html>. Accessed on December 20, 2024.
- Pronzato, L., 2008. Optimal experimental design and some related control problems. *Automatica* 44, 303–325.
- Pukelsheim, F., 1993. *Optimal Design of Experiments*. SIAM, Philadelphia (US).
- Qian, J., Nadri, M., Moroşan, P.D., Dufour, P., 2014. In: Closed loop optimal experiment design for on-line parameter estimation. *IEEE*, pp. 1813–1818.
- Quaglio, M., Fraga, E.S., Galvanin, F., 2018. Model-based design of experiments in the presence of structural model uncertainty: an extended information matrix approach. *Chem. Eng. Res. Des.* 136, 129–143.
- Quaglio, M., Waldron, C., Pankajakshan, A., Cao, E., Gavrilidis, 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.
- Quaglio, M., Fraga, E.S., Galvanin, F., 2020. A diagnostic procedure for improving the structure of approximated kinetic models. *Comput. Chem. Eng.* 133, 106659.
- Ramana, M.V., 1997. An exact duality theory for semidefinite programming and its complexity implications. *Math. Program.* 77, 129–162.
- Rangaiah, G.P., 2009. *Multi-Objective Optimization – Techniques and Applications in Chemical Engineering*. World Scientific Publishing, Singapore.
- Reizman, B.J., Jensen, K.F., 2012. An automated continuous-flow platform for the estimation of multistep reaction kinetics. *Org. Process Res. Dev.* 16, 1770–1782.
- Rhyu, J., Zhuang, D., Bazant, M.Z., Braatz, R.D., 2024. Optimum Model-Based Design of Diagnostics Experiments (DOE) with Hybrid Pulse Power Characterization (HPPC) for Lithium-Ion Batteries. *J. Electrochem. Soc.* 171, 070544.
- Rodrigues, D., Makrygiorgos, G., Mesbah, A., 2022. Tractable global solutions to chance-constrained Bayesian optimal experiment design for arbitrary prior and noise distributions. *J. Process Control* 116, 1–18.
- Rogers, A.W., Lane, A., Mendoza, C., Watson, S., Kowalski, A., Martin, P., Zhang, D., 2024. Integrating knowledge-guided symbolic regression and model-based design of experiments to automate process flow diagram development. *Chem. Eng. Sci.* 300, 120580.
- Rogers, A.W., Lane, A., Mendoza, C., Watson, S., Kowalski, A., Martin, P., Zhang, D., 2025. Integrating Feature Attribution and Symbolic Regression for Automatic Model Structure Identification and Strategic Sampling. *Comput. Chem. Eng.* 109036.
- Ruess, J., Lygeros, J., 2015. Moment-based methods for parameter inference and experiment design for stochastic biochemical reaction networks. *ACM Transactions on Modeling and Computer Simulation* 25, 8.
- Saccardo, A., Felices-Rando, B., Sforza, E., Bezzo, F., 2023. Droop model identification via model-based design of experiments to describe microalgae nitrogen uptake in continuous photobioreactors. *Chem. Eng. J.* 468, 143577.
- Sadeghi, S., Canty, R.B., Mukhin, N., Xu, J., Delgado-Licona, F., Abolhasani, M., 2024. Engineering a Sustainable Future: Harnessing Automation, Robotics, and Artificial Intelligence with Self-Driving Laboratories. *ACS Sustainable Chemistry & Engineering* 12, 12695–12707.
- Sagnol, G., Harman, R., 2015. Computing exact D-optimal designs by mixed integer second-order cone programming. *Ann. Statist.* 43, 2198–2224.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. *Global Sensitivity Analysis. The Primer*, John Wiley and Sons, New York, NY (US).
- Sandrin, M., Pantelides, C.C., Chachuat, B., 2025. Methodological and Computational Framework for Model-based Design of Parallel Experiment Campaigns under Uncertainty. *J. Process Control*, Accepted.
- Sandrin, M., Pantelides, C.C., Chachuat, B., 2025. Methodological and computational framework for model-based design of parallel experiment campaigns under uncertainty. *J. Process Control* 152, 103465.
- Sangoi, E., Quaglio, M., Bezzo, F., Galvanin, F., 2024. An optimal experimental design framework for fast kinetic model identification based on artificial neural networks. *Comput. Chem. Eng.* 187, 108752.
- Sansana, J., Joswiak, M.N., Castillo, I., Wang, Z., Rendall, R., Chiang, L.H., Reis, M.S., 2021. Recent trends on hybrid modeling for Industry 4.0. *Comput. Chem. Eng.* 151, 107365.
- Schaare, L., Kuwertz, R., Heck, J., Skiborowski, M., 2023. Intensification of Kinetic Studies for a Multi-step Reaction in a Milli-structured Plate Reactor by using Model-based Design of Experiments. *Chem. Eng. Trans.* 99, 475–480.
- Schaber, S.D., Born, S.C., Jensen, K.F., Barton, P.I., 2014. Design, execution, and analysis of time-varying experiments for model discrimination and parameter estimation in microreactors. *Org. Process Res. Dev.* 18, 1461–1467.
- Schenkendorf, R., Mangold, M., 2011. Qualitative and quantitative optimal experimental design for parameter identification of a MAP kinase model. *IFAC-PapersOnline* 44, 11666–11671.
- Scheurer, C., Reuter, K., 2025. Role of the human-in-the-loop in emerging self-driving laboratories for heterogeneous catalysis. *Nature Catalysis* 8, 13–19.
- Schmideder, A., Severin, T.S., Cremer, J.H., Weuster-Botz, D., 2015. A novel milliliter-scale chemostat system for parallel cultivation of microorganisms in stirred-tank bioreactors. *J. Biotechnol.* 210, 19–24.
- Schulze, M., Schenkendorf, R., 2018. Flatness-Based Design of Experiments for Model Selection. *IFAC-PapersOnline* 51, 233–238.
- Schulze, M., Schenkendorf, R., 2020. Robust model selection: Flatness-based optimal experimental design for a biocatalytic reaction. *Processes* 8, 190.
- Schwaab, M., Silva, F.M., Queipo, C.A., Barreto Jr., A.G., Pinto, J.C., 2006. A new approach for sequential experimental design for model discrimination. *Chem. Eng. Sci.* 61, 5791–5806.
- Schweidtmann, A.M., Zhang, D., von Stosch, M., 2024. A review and perspective on hybrid modeling methodologies. *Digit. Chem. Eng.* 10, 100136.
- Seborg, D.E., Edgar, T.F., Mellichamp, D.A., Doyle, F.J., 2016. *Process dynamics and control*. John Wiley & Sons, Hoboken, NJ (US).
- Selvarajan, S., Tappe, A.A., Heiduk, C., Scholl, S., Schenkendorf, R., 2022. Process Model Inversion in the Data-Driven Engineering Context for Improved Parameter Sensitivities. *Processes* 10, 1764.
- Seufert, P., Schwihtek, J., Bortz, M., 2021. Model-based design of experiments for high-dimensional inputs supported by machine-learning methods. *Processes* 9, 508.
- Shahmohammadi, A., McAuley, K.B., 2019a. Sequential model-based A- and V-optimal design of experiments for building fundamental models of pharmaceutical production processes. *Comput. Chem. Eng.* 129, 106504.
- Shahmohammadi, A., McAuley, K.B., 2019b. Sequential Model-Based A-Optimal Design of Experiments When the Fisher Information Matrix Is Noninvertible. *Ind. Eng. Chem. Res.* 58, 1244–1261.
- Shewhart, W.A., 1929. *Control of quality of manufactured product*. Martino Fine Books, Milano, Italy.
- Shotwell, M.S., Gray, R.A., 2016. Estimability Analysis and Optimal Design in Dynamic Multi-scale Models of Cardiac Electrophysiology. *J. Agric. Biol. Environ. Stat.* 21, 261–276.
- Shukla, K., Xu, M., Trask, N., Karniadakis, G.E., 2022. Scalable algorithms for physics-informed neural and graph networks. *Data-Centric Eng.* 3, e24.
- Simon, J. B., Karkada, D., Ghosh, N., Belkin, M., 2023. More is better in modern machine learning: when infinite overparameterization is optimal and overfitting is obligatory. *arXiv preprint arXiv:2311.14646*.
- Sin, G., Gernaey, K.V., Lantz, A.E., 2009. Good modeling practice for PAT applications: Propagation of input uncertainty and sensitivity analysis. *Biotechnol. Prog.* 25, 1043–1053.
- Sobol, I.M., 1967. Distribution of points in a cube and approximate evaluation of integrals. *USSR Comput. Maths. Math. Phys.* 7, 86–112.
- Sobol, I.M., 1993. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.* 407–414.
- Soen, A., Sun, K., 2024. Trade-Offs of Diagonal Fisher Information Matrix Estimators. *arXiv preprint arXiv:2402.05379*.
- Stamati, I., Telen, I., Logis, F., van Derlinden, E., Hirsch, M., Passenbrunner, T., van Impe, J., 2015. Optimal Experiment Design for Calibrating an Airpath Model of a Diesel Engine. *SNE Simulation Notes Europe* 22, 121–128.
- Stauffer, F., Vanhoorne, V., Plicer, G., Chavez, P.F., Rome, S., Schubert, M.A., Aerts, L., De Beer, T., 2018. Raw material variability of an active pharmaceutical ingredient and its relevance for processability in secondary continuous pharmaceutical manufacturing. *Eur. J. Pharm. Biopharm.* 127, 92–103.
- Stein, J., Engell, S., 2016. Optimization-based support for process design under uncertainty: A case study. *AIChE J.* 62, 3404–3419.
- Stigter, J.D., Molenaar, J., 2012. Network inference via adaptive optimal design. *BMC Research Notes* 5, 1–11.
- Streif, S., Petzke, F., Mesbah, A., Findeisen, R., Braatz, R.D., 2014. Optimal experimental design for probabilistic model discrimination using polynomial chaos. *IFAC-PapersOnline* 47, 4103–4109.
- Stroebel, F., Schaeufl, F., Bohlen, O., Palm, H., 2024. Multi-Stage Optimal Experimental Design and Setup Strategies in Absence of System Pre-Knowledge. *IEEE Access*.
- Tabrizi, S.Z.B., Barbera, E., Leal da Silva, W.R., Bezzo, F., 2025. A Python/NumPy-based package to support model discrimination and identification. *Systems and Control Transactions* 4, 1282–1287.
- Tappe, A.A., Selvarajan, S., Heiduk, C., Scholl, S., Schenkendorf, R., 2023. PINN-based Design of Experiment Concept for Process Model Parameter Identification. *Comput. Aided Chem. Eng.* 52, 837–842.
- Taverna, B., Casonato, A., Bezzo, F., Galvanin, F., 2019. A framework for the optimal design of a minimum set of clinical trials to characterize von Willebrand disease. *Comp. Meth. Prog. Bio.* 179, 104989.

- Telen, D., Logist, F., van Derlinden, E., Tack, I., van Impe, J., 2012a. Optimal experiment design for dynamic bioprocesses: A multi-objective approach. *Chem. Eng. Sci.* 78, 82–97.
- Telen, D., Logist, F., van Derlinden, E., van Impe, J.F., 2012b. Evaluating the trade-offs in optimal experiment design using a multi-objective optimisation approach. *IFAC-PapersOnline* 16, 947–952.
- Telen, D., Vercommen, D., Logist, F., van Impe, J., 2014a. Robustifying optimal experiment design for nonlinear, dynamic (bio)chemical systems. *Comput. Chem. Eng.* 71, 415–425.
- Telen, D., Logist, F., Quirynen, R., Houska, B., Diehl, M., van Impe, J., 2014b. Optimal experiment design for nonlinear dynamic (bio)chemical systems using sequential semidefinite programming. *AIChE J.* 60, 1728–1739.
- Telen, D., Houska, B., Vallerio, M., Logist, F., van Impe, J., 2017. A study of integrated experiment design for NMPC applied to the Droop model. *Chem. Eng. Sci.* 160, 370–383.
- Telen, D., Nimmegheers, P., Van Impe, J., 2018. Uncertainty in optimal experiment design: comparing an online versus offline approaches. *IFAC-PapersOnline* 51, 771–776.
- Terejanu, D., Bryant, C.M., Miki, K., 2013. Bayesian optimal experimental design for the Shock-tube experiment. *J. Phys.: Conf. Ser.* 410, 012040.
- Tian, G., Zhou, Q., Birari, R., Qi, J., Qu, Z., 2020. A hybrid-learning algorithm for online dynamic state estimation in multimachine power systems. *IEEE transactions on neural networks and learning systems* 31, 5497–5508.
- Tsay, C., Pattison, R.C., Baldea, M., Weinstein, B., Hodson, S.J., Johnson, R.D., 2017. A superstructure-based design of experiments framework for simultaneous domain-restricted model identification and parameter estimation. *Comput. Chem. Eng.* 107, 408–426.
- Tsikonis, L., van Herle, J., Favrat, D., 2011. Investigating reliability on fuel cell model identification. Part I: A design of experiments approach. *Fuel Cells* 11, 850–865.
- Valencia, P., Cornejo, I., Almonacid, S., Teixeira, A.A., Simpson, R., 2013. Kinetic Parameter Determination for Enzyme Hydrolysis of Fish Protein Residue Using D-optimal Design. *Food Bioprocess Technol.* 6, 290–296.
- Vanaret, C., Seufert, P., Schwientek, J., Karpov, G., Ryzhakov, G., Oseledets, I., Aspiron, N., Bortz, M., 2021. Two-phase approaches to optimal model-based design of experiments: how many experiments and which ones? *Comput. Chem. Eng.* 146, 107218.
- Vanrolleghem, P. A., Van Daele, M., Dochain, D., 1995. Optimal experimental design for parameter estimation in activated sludge processes. In: *Proceedings of 1995 American Control Conference-ACC'95*, 1, 14–18. IEEE.
- Vega-Ramon, F., Schrecker, L., de Carvalho Servia, M.A., Hii, K.K.M., Zhang, D., 2024. Transient Flow-Assisted Kinetic Modelling and Reaction Network Identification for Pyrazole Synthesis. *Comput. Aided Chem. Eng.* 53, 55–60.
- Versyck, K.J., Claes, J.E., Van Impe, J.F., 1997. Practical identification of unstructured growth kinetics by application of optimal experimental design. *Biotech. Prog.* 13, 524–531.
- Villaverde, A.F., Pathirana, D., Fröhlich, F., Hasenauer, J., Banga, J.R., 2022. A protocol for dynamic model calibration. *Brief. Bioinform.* 23, 1–19.
- Voß, K., Pyta, L., Gesenhues, J., Mela, P., Schmitz-Rode, T., Abel, D., 2019. Towards a Model-Based Experimental Design of the Maturation Process of Biohybrid Heart Valves. *IFAC-PapersOnline* 52, 113–120.
- Waldron, C., Pankajakshan, A., Quaglio, M., Cao, E., Galvanin, F., Gavrilidis, A., 2019. An autonomous microreactor platform for the rapid identification of kinetic models. *Reac. Chem. Eng.* 4, 1623–1636.
- Waldron, C., Pankajakshan, A., Quaglio, M., Cao, E., Galvanin, F., Gavrilidis, A., 2020. Model-based design of transient flow experiments for the identification of kinetic parameters. *Reac. Chem. Eng.* 5, 112–123.
- Walter, E., Pronzato, L., 1987. Robust experiment design: between qualitative and quantitative identifiability. In: *Walter, E. (Ed.), Identifiability of Parametric Models*. Pergamon, Oxford.
- Walter, E., Pronzato, L., 1997. *Identification of Parametric Models from Experimental Data*. Springer.
- Wang, K., Yue, H., 2020. Robust sampling time designs for parametric uncertain systems. *IFAC-PapersOnline* 53, 622–627.
- Wang, J., Dowling, A.W., 2022. Pyomo.DOE: An open-source package for model-based design of experiments in Python. *AIChE J.* 68, e17813.
- Wang, J., Peng, Z., Hughes, R., Bhattacharyya, D., Neira, D. E. B., Dowling, A. W., 2024. Measure This, Not That: Optimizing the Cost and Model-Based Information Content of Measurements. *arXiv preprint arXiv:2406.09557*.
- Welsh, J.S., Kong, H., 2011. Robust experiment design through randomisation with chance constraints. 18th IFAC World Congress. *IFAC-PapersOnline* 44, 13197–13202.
- White, A., Tolman, M., Thames, H.D., Withers, H.R., Mason, K.A., Transtrum, M.K., 2016. The limitations of model-based experimental design and parameter estimation in sloppy systems. *PLoS Comput. Biol.* 12, e1005227.
- Wold, S., Martens, H., Wold, H., 1983. The Multivariate Calibration Problem in Chemistry Solved by the PLS Method. In: *Matrix Pencils*, 286–293.
- Wu, S., McLean, K.A.P., Harris, T.J., McAuley, K.B., 2011. Selection of optimal parameter set using estimability analysis and MSE-based model-selection criterion. *Int. J. Adv. Mech. Syst.* 3, 188–197.
- Wu, S., Harris, T.J., McAuley, K.B., 2008. The use of simplified or misspecified Models: linear case. *Can. J. Chem. Eng.* 85, 386–398.
- Yang, M., Biedermann, S., Tang, E., 2013. On optimal designs for nonlinear models: a general and efficient algorithm. *J. Am. Stat. Assoc.* 108, 1411–1420.
- Yonge, A., Gusmão, G.S., Fushimi, R., Medford, A.J., 2024. Model-Based Design of Experiments for Temporal Analysis of Products (TAP): A Simulated Case Study in Oxidative Propane Dehydrogenation. *Ind. Eng. Chem. Res.* 63, 4756–4770.
- Yu, H., Yue, H., Halling, P., 2015. Optimal experimental design for an enzymatic biodiesel production system. *IFAC-PapersOnline* 28, 1258–1263.
- Yuan, X., Benyahia, B., 2023. A Combined D-optimal and Estimability Model-Based Design of Experiments of a Batch Cooling Crystallization Process. *Comput. Aided Chem. Eng.* 52, 255–260.
- Yuan, X., Benyahia, B., 2024. A Holistic Approach for Model Discrimination, Multi-Objective Design of Experiment and Self-Optimization of Batch and Continuous Crystallization Processes. In: *Computer Aided Chemical Engineering*. 53, 391–396. Elsevier.
- Yue, H., Brown, M., He, F., Jia, J.F., Kell, D.B., 2008. Sensitivity Analysis and Robust Experimental Design of a Signal Transduction Pathway System. *Int. J. Chem. Kinet.* 40, 730–741.
- Yue, H., Halling, P., Yu, H., 2013. Model development and optimal experimental design of a kinetically controlled synthesis system. *IFAC-PapersOnline* 46, 327–332.
- Zhu, M., Mroz, A., Gui, L., Jelfs, K.E., Bemporad, A., Antonio, E.-C., Chanona, D.R., Lee, Y.S., 2024. Discrete and mixed-variable experimental design with surrogate-based approach. *Digit. Discov.* 3, 2589–2606.
- Zou, J., Han, Y., So, S.S., 2009. Overview of Artificial Neural Networks, Artificial Neural Networks: Methods and Applications. Ed. David J. Livingstone, Totowa NJ, Humana Press.
- Zullo, L., 1991. *Computer Aided Design of Experiments. An Engineering Approach*. The University of London, London (U.K.).