



An exploratory model-based design of experiments approach to aid parameters identification and reduce model prediction uncertainty

Francesca Cenci^a, Arun Pankajakshan^b, Pierantonio Facco^a, Federico Galvanin^{b,*}

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

^b Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, United Kingdom

ARTICLE INFO

Keywords:

Model-based design of experiments
Trade-off between space exploration and information maximization
Model prediction uncertainty
Parameters precision
G-optimality
Maps of model prediction variance

ABSTRACT

The management of trade-off between experimental design space exploration and information maximization is still an open question in the field of optimal experimental design. In classical optimal experimental design methods, the uncertainty of model prediction throughout the design space is not always assessed after parameter identification and parameters precision maximization do not guarantee that the model prediction variance is minimized in the whole domain of model utilization. To tackle these issues, we propose a novel model-based design of experiments (MBDoE) method that enhances space exploration and reduces model prediction uncertainty by using a mapping of model prediction variance (G-optimality mapping). This explorative MBDoE (eMBDoE) named *G-map eMBDoE* is tested on two models of increasing complexity and compared against conventional factorial design of experiments, Latin Hypercube (LH) sampling and MBDoE methods. The results show that *G-map eMBDoE* is more efficient in exploring the experimental design space when compared to a standard MBDoE and outperforms classical design of experiments methods in terms of model prediction uncertainty reduction and parameters precision maximization.

1. Introduction

Models are widespread in process industries for a variety of applications, from process understanding to product and process optimization. For instance, simulations of a process model allow to evaluate the influence of the process conditions and/or disturbances (de Prada et al., 2019) on the variables of interest, while the use of modeling at all stages of product development enables the compliance of the product to the clients' need and the selection of the most convenient manufacturing route (Mihaluta et al., 2008). The development of predictive models to be used in model-based activities requires the identification of model structure, i.e. the set of model equations, and model calibration, i.e. the precise estimation of model parameters from experimental data.

Improving the quality of data and information generation has a direct impact on model identification activities, thus several design of experiments (DoE) techniques have been developed in the last century at the purpose. One of the first is factorial DoE (Montgomery, 2013), which requires a limited preliminary knowledge on the system, including response variables that are indicative of process conditions and experimental factors that may influence the responses. In factorial DoE, each factor is varied over discrete levels and experiments consist of all

combinations (or a fraction of them) of levels across all factors. The data thereby generated are used to calibrate an empirical model, usually including first or second-order terms, which is eventually refined in order to exclude uninfluential factors and/or to add higher order terms. Factorial DoE can be beneficial to many industrial sectors, e.g. pharmaceutical industries (Singh et al., 2005 Part I and II), food science and technology (Granato and de Araújo Calado, 2013), manufacturing industries (Czitrom, 1999). The use of factorial DoE has brought many advantages in the experimentation with respect to the commonly used 'one-factor-at-a-time' experimental design, allowing to: (i) select the factors that are more influential on the response; (ii) evaluate the effects of factors interaction on the response; (iii) reduce the experimental burden for experimental design exploration; (iv) identify regression models to be employed for the design, analysis and improvement of products and processes. However, DoE methods have some limitations too: (1) experimental conditions are designed all at once, without progressively updating the process knowledge as soon as data from a new experiment are available; thus, the experimenter is not taking advantage of the deeper process knowledge to improve the quality of the designed experiments; (2) conventional DoE approaches are typically used to calibrate empirical models, which are reliable in interpolation, but not always in extrapolation as they lack mechanistic process knowledge

* Corresponding author.

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

<https://doi.org/10.1016/j.compchemeng.2023.108353>

Received 4 April 2023; Received in revised form 9 July 2023; Accepted 11 July 2023

Available online 14 July 2023

0098-1354/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Nomenclature		V_{θ}	variance-covariance matrix
Symbols		V_{θ}^0	prior variance-covariance matrix of model parameters
$\det(\cdot)$	matrix determinant	\mathbf{x}	vector of state variables
\mathbf{f}	set of model equations	$\dot{\mathbf{x}}$	vector of first derivatives of state variables
\mathbf{h}	set of equations of measurable responses	\mathbf{y}	vector of measurable model responses
$\mathbf{H}_{\hat{\theta}}$	Fisher Information Matrix of a general model with estimated parameters $\hat{\theta}$	$\hat{\mathbf{y}}$	vector of measurable model responses predicted by the model
J_G	scalar value summarizing the G-optimality values calculated for all responses and all time points at a specific point in the design space	Greek symbols	
$J_{G,\min(/mean/max)}$	minimum (or mean or maximum) values of J_G calculate in the whole design space	α	significance level
$J_{G,\text{thr}}$	thresholds of G-optimality	θ	vector of model parameters
$L(\hat{\theta})$	negative log-likelihood function	$\hat{\theta}$	vector of estimated model parameters
N	total number of experimental measurements	$\theta_0, \theta_{\text{LB}}, \theta_{\text{UB}}$	vector of initial parameters values, lower bounds (LB) and upper bounds (UB) for parameters estimation
N_e	number of performed experiments	$\lambda_{\min(/max)}$	maximum (or maximum) eigenvalue
n_{sp}	number of sampling points considering all the performed experiments	σ_y	standard deviation of the measurement error
N_{sp_i}	number of sampling points in the i -th experiment	Σ_y	response variance-covariance matrix
N_u	number of control variables	ϕ	design vector
N_x	number of state variables	ψ	scalar measure of the FIM
N_y	number of response variables	Acronyms	
N_{θ}	number of model parameters	CI	confidence intervals
N_{ϕ}	number of possible experimental conditions within the discretised design space	DoE	factorial design of experiments
r	kinetic rate	eMBDoE	explorative MBDoE
t	time	FIM	Fisher information matrix
t_i	t -value for the i -th parameter	G-map	G-optimality map
\mathbf{t}_{sp}	vector of sampling points	G-map eMBDoE	explorative MBDoE based on G-optimality map
t_{ref}	reference t -value	H-map	Fisher information map
\mathbf{u}	vector of control variables	LB	lower bounds
\mathbf{V}_y	matrix of model prediction variances calculated by means of the G-optimality definition	LH	Latin Hypercube
		MBDoE	model-based design of experiments
		OFAT	one factor at a time
		SMSPE	supremum of the mean squared prediction error
		UB	upper bounds

(Garud et al., 2017; Duarte et al., 2004).

To overcome the aforementioned limitations, model-based design of experiments (MBDoE) techniques have been proposed that are centered on process knowledge (Espie and Macchietto, 1989). MBDoE is an optimal DoE technique because the outcome is a set of optimal experimental conditions (for example temperature, flow rate, feed composition in a flow reaction system) and it is *model-based* because a physics-based model is employed in the calculation of the optimal experimental conditions. Typically, MBDoE methods are applied for two main objectives in model identification: (1) MBDoE for model discrimination, where different candidate model structures are compared in order to select the most adequate one (Hunter and Reiner, 1965; Buzzi-Ferraris and Forzatti, 1983; Galvanin et al., 2016; Waldron et al., 2019); (2) MBDoE for improving parameter precision, which aims at minimizing the parameters uncertainty region (Galvanin et al., 2007; Franceschini and Macchietto, 2008; Waldron et al., 2020). Parameter estimation should be preceded by structural identifiability analysis to check the possibility of obtaining unique parameter estimates when noise-free experiments and absence of model uncertainty are assumed (Galvanin et al., 2013; Asprey and Macchietto, 2000). MBDoE for parameters precision will be described in detail because it is the main interest in this work.

MBDoE for parameter precision quantifies the information content of candidate experiments through the Fisher Information Matrix (FIM; Fisher, 1950), which is based on the calculation of partial derivatives of the model response with respect to every model parameter. Compared to factorial DoE, such calculation requires a deeper prior knowledge: a

model structure that is representative of the system; a set of initial values of model parameters to be refined with the experimental data. From the mathematical point of view, MBDoE can be described as an optimization problem where a scalar measure of the FIM is maximized. Depending on the scalar measure selected, a specific metric of the uncertainty region of model parameters is minimized; in other terms, parameters precision is maximized. Thus, fewer experiments are sufficient to identify statistically sound parameters, with great benefits in terms of time, labor and resources. Consequently, MBDoE has been successfully applied to both industry and research: in chemical processes, for example the production of aziridine through the C—H activation with a Pd-catalysis (Echtermeyer et al., 2017) or the execution of transient flow experiments to study the esterification of benzoic acid with ethanol (Waldron et al., 2020); in the production of renewably-sourced polymers like Cerenol, which has several applications in automotive, cosmetics and polymer specialties (Vo et al., 2021); in the pharmaceutical industry, e.g. the study of Michaelis-Menten kinetics for the production of a pharmaceutical agent (Shahmohammadi and McAuley, 2019); in civil engineering, e.g. for the determination of the optimal sensor locations to obtain the Young's moduli of tall structures (Reichert et al., 2021).

MBDoE methods determine experimental conditions that optimize a specific objective function and hence these optimal conditions are usually restricted in small regions of high information content. For instance, the optimal experiments to identify parameters of a kinetic model with two unknown parameters is made of a set of two distinct points; if several experiments are designed, they should fall in one of the two optimal conditions to avoid information loss (Box, 1968). However, this

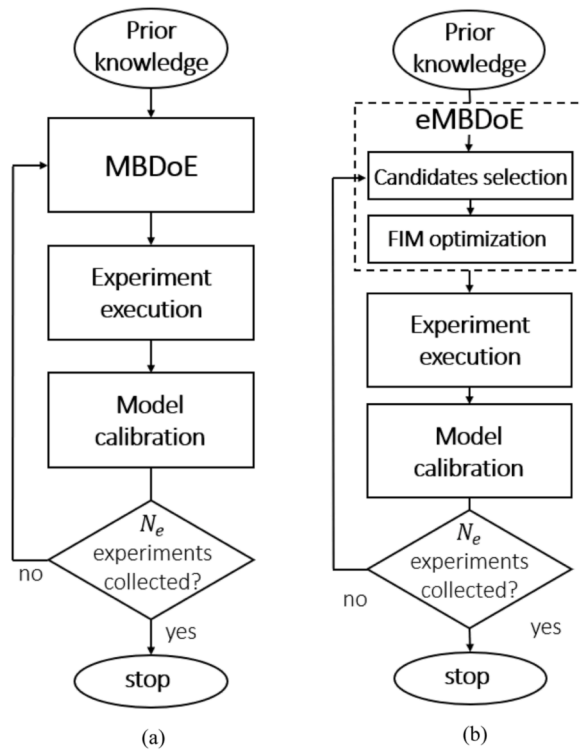


Fig. 1. Workflow of a conventional sequential MBDoE (a) and of a sequential G-map eMBDoE (b).

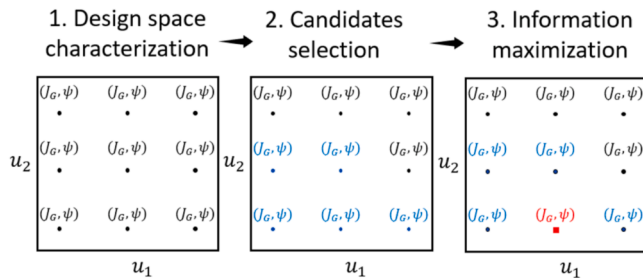


Fig. 2. Schematic representation of the novel eMBDoE method based on G-maps. Grids refer to two general control variables u_1 and u_2 ; J_G and ψ indicate the G-optimality index and the FIM scalar measure selected; blue points represent experimental conditions satisfying the condition on G-optimality; the red square represent the point of maximum information among the candidate design points.

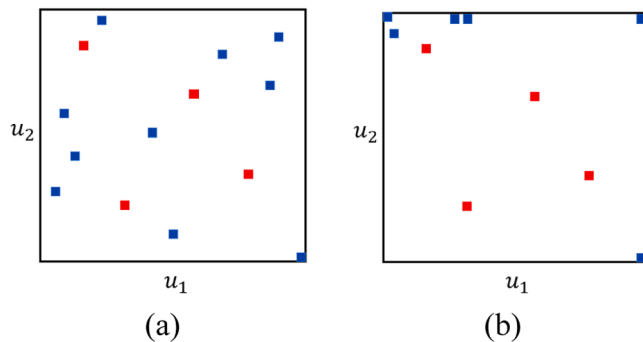


Fig. 3. Graphical representation of design space exploration for (a) an explorative, space-filling design; (b) an exploitative design (MBDoE).

Table 1

Case study 1 (algebraic model): Initial settings.

Model type	Algebraic
Inputs	u_1, u_2 $u_1 \in [-10, 10]$ $u_2 \in [-10, 10]$
Outputs	y $\sigma_y = 5$
Parameters	$\theta_{\text{true}} = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5] = [3.5, -2, 1.7, 1.1, 8]$ $\theta_0 = [1, 1, 1, 1, 1]$ $\theta_{\text{LB}} = [-10, -10, -10, -10, -10]$ $\theta_{\text{UB}} = [10, 10, 10, 10, 10]$

feature of MBDoE may lead to a scarce exploration of the design space, which in turn may result in poor predictive capability of the model, particularly in unexplored regions of the experimental design space.

The minimization of parameters uncertainty ensured by MBDoE data does not necessarily imply that the entire design space is characterized by a minimization of model prediction uncertainty. In literature, approaches have been proposed to evaluate the regions of model reliability within the design space based on different criteria to evaluate the prediction error. Dasgupta and Mukhopadhyay (2021) built a map of supremum of the mean squared prediction error (SMSPE) using a kriging interpolating technique, while Quaglio et al. (2018) mapped the design space through a reliability function that depends mainly on the difference between predicted and measured responses.

As an alternative, model prediction uncertainty can be quantified in terms of model prediction variance using the so-called “G-optimality” (Smith, 1918; Kiefer and Wolfowitz, 1959), a metric which can be evaluated in the whole design space to detect regions of model reliability without increasing the experimental burden. G-optimality has been explored in an MBDoE context; for instance, it has been used as an objective function in order to determine the optimal experimental conditions that minimize the response prediction variance (Smith, 1918; Kiefer and Wolfowitz, 1959). Moreover, the relationship between D-optimality and G-optimality (namely, between maximization of the FIM determinant and minimization of G-optimality, respectively) has been analysed for different types of models in order to define the specific conditions under which the equivalence of the two criteria holds. For example, the equivalence of D-optimality and G-optimality is demonstrated by Kiefer and Wolfowitz (1960) for linear models with homoscedastic errors. Instead, Wong (1995) demonstrates that the equivalence between D- and G-optimality rarely holds in case of heteroscedastic models. Prus (2019) discusses the features of G-optimal designs with random coefficient regression models and states that the equivalence with D-optimal designs does not hold in general with this type of models. In addition, the classical G-optimal criterion for MBDoE is modified by Stigler (1971) in order to allow for a few experiments suitable to check the adequacy of model structure (namely, to assess process-model mismatch). However, to the author’s knowledge G-optimality has never been used to enhance space exploration of MBDoE designs, to precisely estimate parameters and reduce model prediction variance in the whole design space with the minimum experimental effort. Furthermore, a formal description of the relation between G-optimality and other MBDoE criteria for non-linear systems with general variance models is still lacking in the scientific community. Therefore, a numerical approach that is not strictly related to a specific type of model is adopted in this work and validated with simulated data. In this paper, we tackle these issues by proposing a general new technique that integrates G-optimality maps into the conventional MBDoE optimization framework for parameter estimation.

Section 2 of this paper illustrates the mathematical modeling of state of the art MBDoE and of the proposed method, together with the indices

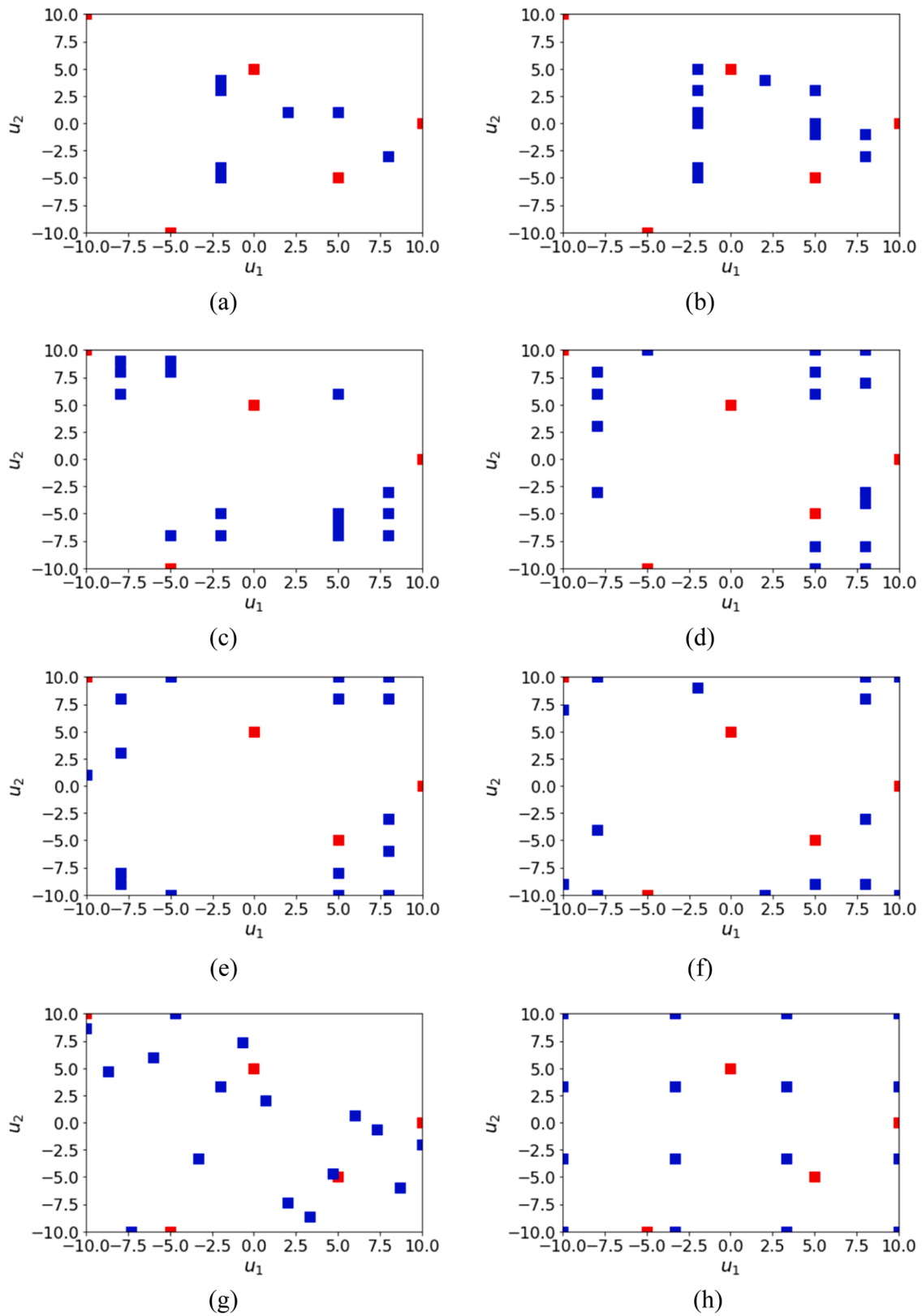


Fig. 4. Design space with the experiments selected by: (a) MBDoE; (b) G-map eMBDoE $J_{G,thr} = 0.25$; (c) G-map eMBDoE $J_{G,thr} = 0.50$; (d) G-map eMBDoE $J_{G,thr} = 0.65$; (e) G-map eMBDoE $J_{G,thr} = 0.75$; (f) G-map eMBDoE $J_{G,thr} = 0.85$; (g) Latin Hypercube; (h) 4^2 full factorial DoE. Red squares indicate the 5 preliminary experiments.

to assess model performance. Section 3 shows the results of the application of G-map eMBDoE to two simulated case studies, including a comparison with exploitation-based methods, i.e. MBDoE, and exploration-based ones, i.e. factorial DoE and Latin Hypercube (LH) sampling. Finally, in Section 4 conclusions are drawn and future works are proposed.

2. Model-based design of experiments (MBDoE)

Model-based design of experiments methods for parameters precision require the model structure in order to quantify the expected information content of an experiment. The structure of a general differential and algebraic model can be represented as:

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

where \mathbf{f} is a set of model equations, \mathbf{x} and $\dot{\mathbf{x}}$ are N_x -dimensional vectors of state variables and their first derivatives respectively, \mathbf{u} is a N_u -dimensional vector of control variables, t is time, $\boldsymbol{\theta}$ is a N_θ -dimensional vector of model parameters, \mathbf{y} is a N_y -dimensional vector of response variables that are measurable.

Parameter estimates (indicated as $\hat{\boldsymbol{\theta}}$) from experimental data are computed by minimizing the difference between measured responses (\mathbf{y}) and predicted responses ($\hat{\mathbf{y}}$) through the negative log-likelihood function $L(\hat{\boldsymbol{\theta}})$ (Bard, 1974)

$$L(\hat{\boldsymbol{\theta}}) = \frac{N}{2} \log(2\pi) + \frac{n_{sp}}{2} \log(\det|\boldsymbol{\Sigma}_y|) + \frac{1}{2} \sum_{i=1}^{n_{sp}} [\mathbf{y}_i - \hat{\mathbf{y}}_i(\hat{\boldsymbol{\theta}})]^T \boldsymbol{\Sigma}_y^{-1} [\mathbf{y}_i - \hat{\mathbf{y}}_i(\hat{\boldsymbol{\theta}})], \quad (2)$$

where $\boldsymbol{\Sigma}_y$ is the variance-covariance matrix of measurement error, n_{sp} is the number of sampling points considering all the N_e performed experiments, namely $n_{sp} = \sum_{i=1}^{N_e} N_{sp_i}$ (N_{sp_i} is the number of sampling points in the i -th experiment), N is the total number of experimental measurements calculated as $N = \sum_{i=1}^{N_e} N_{sp_i} N_y$. When experimental data are collected, the variance terms in $\boldsymbol{\Sigma}_y$ can be calculated as the square of the pooled standard deviations (Killeen, 2005). However, not all experiments are equally able to provide estimates $\hat{\boldsymbol{\theta}}$ with enough statistical precision, because this depends on their information content. The information content of experiments is evaluated through the Fisher Information Matrix (FIM) \mathbf{H}_θ which, for dynamic systems, can be expressed as Zullo (1991):

$$\mathbf{H}_\theta(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}) = [\mathbf{V}_\theta^0]^{-1} + \sum_{i=1}^{n_{sp}} \left(\frac{d\hat{\mathbf{y}}_i}{d\hat{\boldsymbol{\theta}}} \right)^T \boldsymbol{\Sigma}_y^{-1} \left(\frac{d\hat{\mathbf{y}}_i}{d\hat{\boldsymbol{\theta}}} \right), \quad (3)$$

where \mathbf{V}_θ^0 is the $N_\theta \times N_\theta$ prior variance-covariance matrix of model parameters, while $\left(\frac{d\hat{\mathbf{y}}_i}{d\hat{\boldsymbol{\theta}}} \right)$ is the $N_y \times N_\theta$ matrix with first-order derivatives of model responses with respect to the parameters at time point i .

Based on Cramer-Rao Theorem, the inverse of the FIM represents a lower limit for the variance-covariance matrix (\mathbf{V}_θ) of the parameters (Bard, 1974):

$$\mathbf{V}_\theta(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}) \geq [\mathbf{H}_\theta(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi})]^{-1}. \quad (4)$$

In other terms, Eq. (4) provides an upper limit to parameters precision and when the equality holds, parameters are defined efficient (Bard, 1974). Finally, the variance-covariance matrix (\mathbf{V}_θ) can be approximated as the inverse of the FIM by using the first term Taylor expansion (Bard, 1974).

MBDoE for parameter identification is an optimization problem that aims at minimizing the parametric uncertainty (represented by \mathbf{V}_θ) by maximizing a scalar measure ($\psi(\mathbf{H}_\theta)$) of the FIM. To this purpose, the so-called alphabetical criteria (Pukelsheim, 1993) are widely used:

maximization of the FIM determinant or minimization of \mathbf{V}_θ determinant (D-optimal criterion); maximization of FIM trace or minimization of \mathbf{V}_θ trace (A-optimal criterion); maximization of the FIM minimum eigenvalue or minimization of the \mathbf{V}_θ maximum eigenvalue (E-optimal criterion); minimization of the ratio between maximum and minimum FIM eigenvalue (modified E-optimal). The optimization problem is formulated as:

$$\boldsymbol{\varphi}_{\text{opt}} = \underset{\boldsymbol{\varphi}}{\operatorname{argmin}} \psi(\mathbf{V}_\theta), \quad (5)$$

with $\psi(\mathbf{V}_\theta) = \psi_E(\mathbf{V}_\theta)$ in the case of E-optimal design:

$$\psi_E(\mathbf{V}_\theta) = \lambda_{\max}(\mathbf{V}_\theta), \quad (6)$$

where $\boldsymbol{\varphi}$ is the “design vector”, which contains the set of control variables that define the experimental conditions, while λ_{\max} refers to the maximum eigenvalue of the variance-covariance matrix \mathbf{V}_θ .

Instead, conventional G-optimal MBDoE minimizes the maximum prediction variance, which can be expressed as (Kiefer and Wolfowitz, 1959):

$$\boldsymbol{\varphi}_{\text{opt}} = \underset{\boldsymbol{\varphi}}{\operatorname{argmin}} \psi(\mathbf{V}_y), \quad (7)$$

If all N_y responses can be characterized through N_{sp_i} sampling points, \mathbf{V}_y is a $N_y N_{sp_i} \times N_y N_{sp_i}$ matrix containing the estimated variance of each response at each time point. Its j -th element $\mathbf{V}_y(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi})|_{j,i}$ is calculated as:

$$\mathbf{V}_y(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi})|_{j,i} = \left(\frac{d\hat{\mathbf{y}}_j}{d\hat{\boldsymbol{\theta}}} \right)_i^T [\mathbf{H}_\theta]^{-1} \left(\frac{d\hat{\mathbf{y}}_j}{d\hat{\boldsymbol{\theta}}} \right)_i, \text{ for } j = 1, \dots, N_y; i = 1, \dots, N_{sp_i} \quad (8)$$

where $\left(\frac{d\hat{\mathbf{y}}_j}{d\hat{\boldsymbol{\theta}}} \right)_i$ is the $N_\theta \times 1$ vector of first derivatives of $\hat{\mathbf{y}}_j$ with respect to the full set of model parameters at sampling point i , while \mathbf{H}_θ is the Fisher information matrix of Eq. (3). The scalar index $\psi(\mathbf{V}_y)$ in Eq. (7) is usually the largest diagonal element of \mathbf{V}_y .

Conventional MBDoE is strictly related to the identification of the optimum, therefore the design space can be poorly explored in favor of many replicated points around regions of high information content. Although this minimizes the experimental burden to obtain precise parameters, the predictive capability of models can be improved if it is validated over a wider region of the design space. We aim at overcoming these limitations by proposing a novel explorative MBDoE methodology based on the mapping of model prediction uncertainty throughout the whole design space (Section 2.1).

2.1. Explorative mbdoe (eMBDoE) based on G-optimality maps

The novel MBDoE method proposed in this work aims at enhancing space exploration, precisely estimating model parameters and minimizing model prediction variance across the whole design space with a minimum experimental effort. To this aim, the calculation of model prediction variance is included within the MBDoE optimization framework. More specifically, mapping of G-optimality values is performed to obtain an explorative MBDoE (eMBDoE) method; therefore, the novel method is named *G-map eMBDoE*.

Fig. 1a shows the standard sequential procedure for MBDoE (Espie and Macchietto, 1989; Asprey and Macchietto, 2000), where optimal experimental design, experiment execution and model calibration are carried out sequentially in the design of N_e experiments.

Similarly, Fig. 1b shows the sequential procedure of the proposed G-map eMBDoE, in which optimal experimental design is carried out using the novel G-map eMBDoE method. The following Sections 2.1.1-2.1.3 provide more details on each step of the proposed G-map eMBDoE procedure.

Table 2

Number of distinct design points for each scenario compared in the study.

Scenario	No. distinct design points
MBDoE	7
eMBDoE, thr:0.25	12
eMBDoE, thr:0.50	15
eMBDoE, thr:0.65	16
eMBDoE, thr:0.75	16
eMBDoE, thr:0.85	14
LH	16
DoE	16

2.1.1. Prior knowledge

The prior knowledge is defined as the information needed to initialize the MBDoE or eMBDoE procedure. It includes:

- Structurally identifiable models in the form of Eq. (1);
- A set of preliminary experiments to obtain initial parameter estimates to initialize the MBDoE or G-map eMBDoE procedure. Preliminary experiments are designed using DoE methods, for instance factorial DoE or LH;
- Upper and lower bounds for each control variable included in the design vector φ ;
- Variance-covariance matrix of measurement error Σ_y .

This information allows to calculate both FIM (Eq. (3)) and G-optimality (Eq. 8).

2.1.2. eMBDoE design

In the sequential MBDoE procedure shown in Fig. 1a, the optimal design φ_{opt} is given by the solution to the optimization problem (Eq. (5)). Whereas, in the G-map eMBDoE shown in Fig. 1b, the most informative

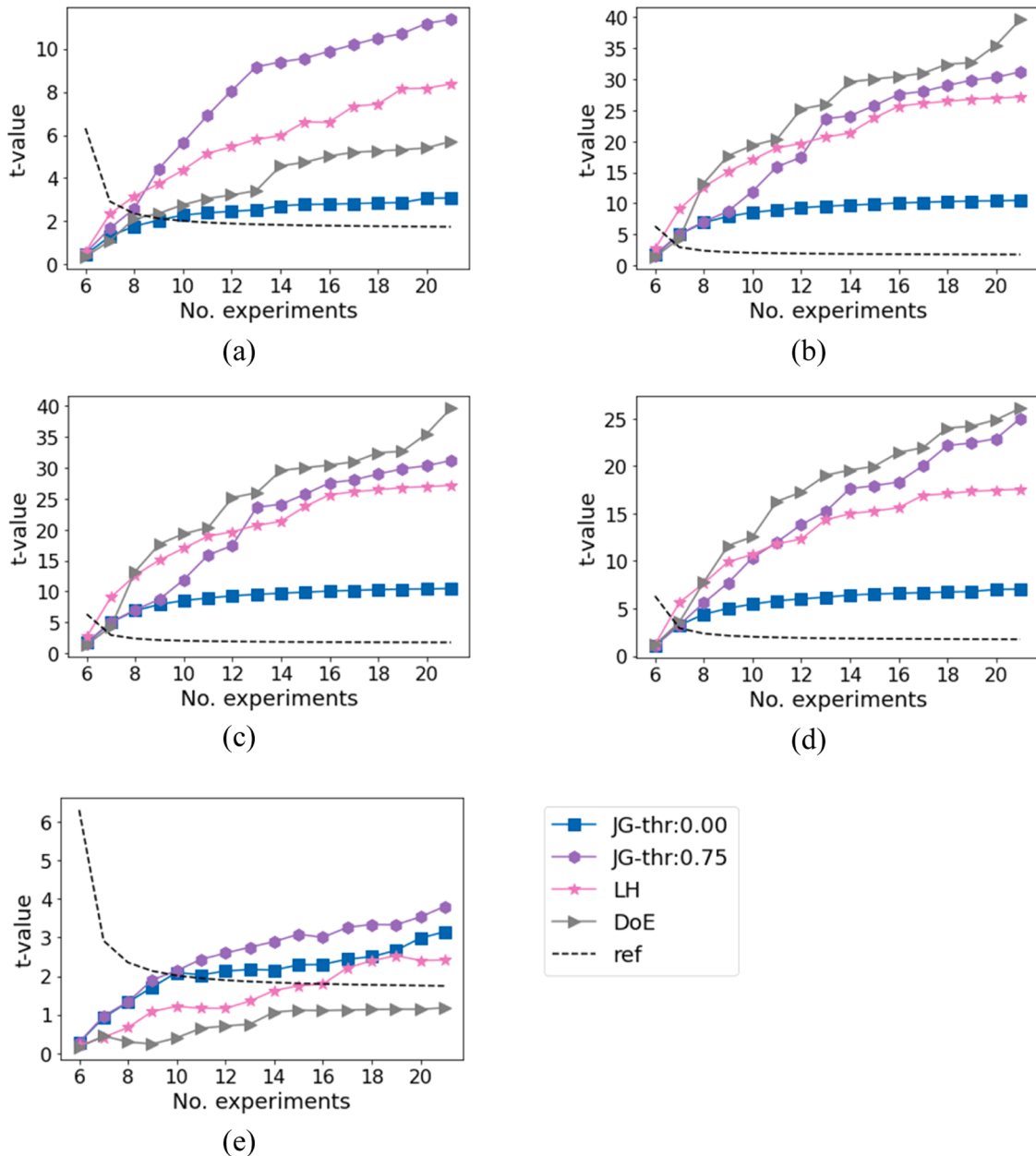


Fig. 5. Profiles of t-values calculated with: MBDoE ($J_{G, \text{thr}}=0.00$); G-map eMBDoE ($J_{G, \text{thr}}=0.75$); Latin Hypercube (LH); 4^2 full factorial DoE. Figures (a)-(e) show results of parameters 1–5, respectively. t-values are compared against the reference t-value ('ref' in the legend). Only t-values referred to the 16 optimal/explorative data are shown.

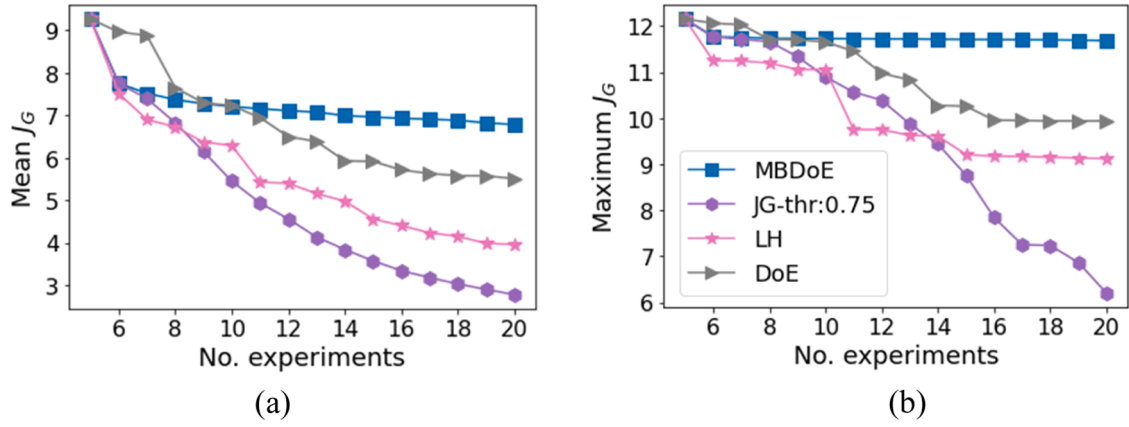


Fig. 6. Profiles of scalar indices of G-optimality including (a) mean G-optimality; (b) maximum G-optimality calculated for: MBDoE ($J_{G, \text{thr}}=0.00$); G-map eMBDoE ($J_{G, \text{thr}}=0.75$); Latin Hypercube (LH); 4^2 full factorial DoE.

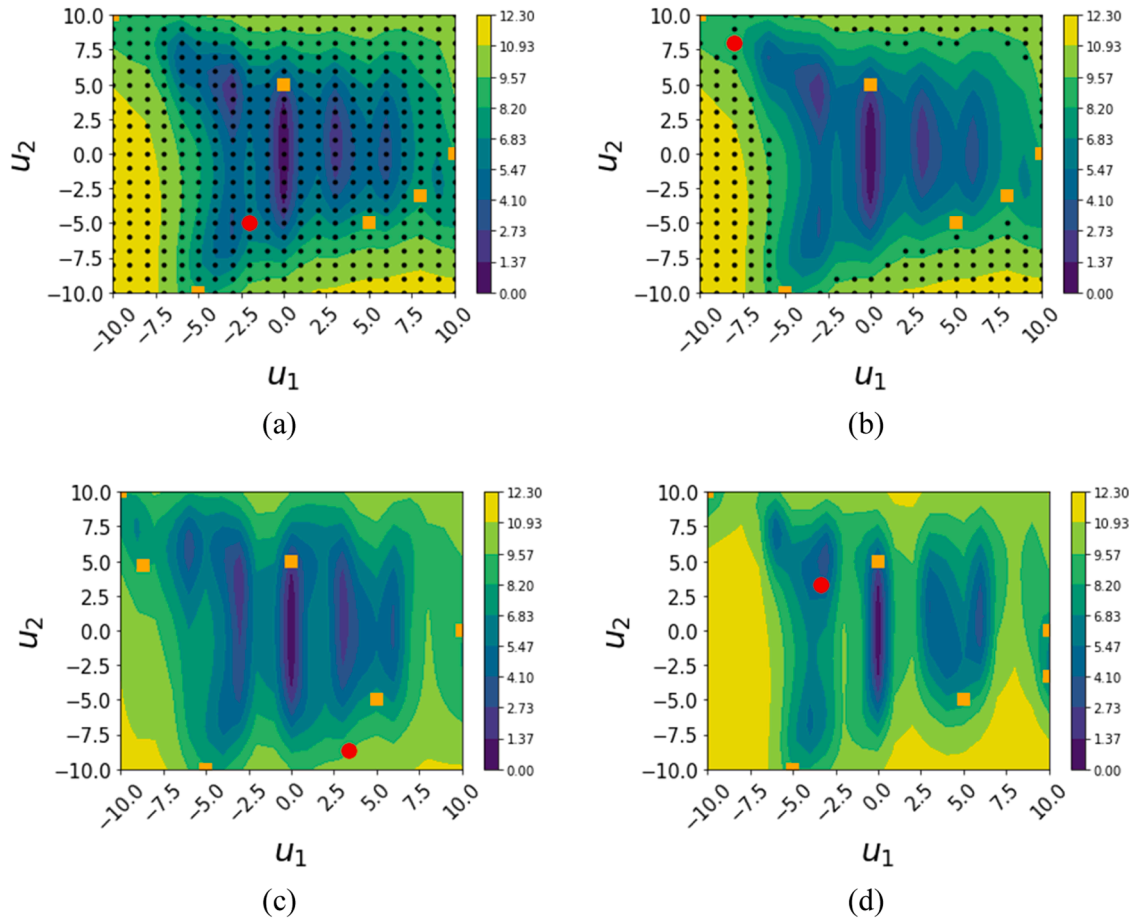


Fig. 7. G-maps generated after 6 calibration experiments designed with: (a) MBDoE, (b) eMBDoE and $J_{G, \text{thr}}=0.75$, (c) LH, (d) 4^2 full factorial DoE. The candidate design points of MBDoE and eMBDoE are represented with black dots, while already measured experiments are indicated with orange squares. Finally, the red point indicates the experiment selected at this iteration.

experiment is evaluated using an additional step which is described below and illustrated in Fig. 2.

- each experimental condition in the design space is characterized in terms of model prediction variance, represented by scalar indices J_G , which leads to a map of G-optimality named *G-maps*. Similarly, every point in the design space is characterized in terms of information

content, represented by the scalar measure ψ , generating a map of FIM-based information named *H-map* (step 1 of Fig. 2);

- only experiments that satisfy a threshold $J_{G, \text{thr}}$ on model prediction variance represented by J_G are retained for the subsequent optimization (blue points in step 2 of Fig. 2);
- among these candidates, the experimental condition maximizing information is chosen as the optimal experiment (red square in step 3 of Fig. 2).

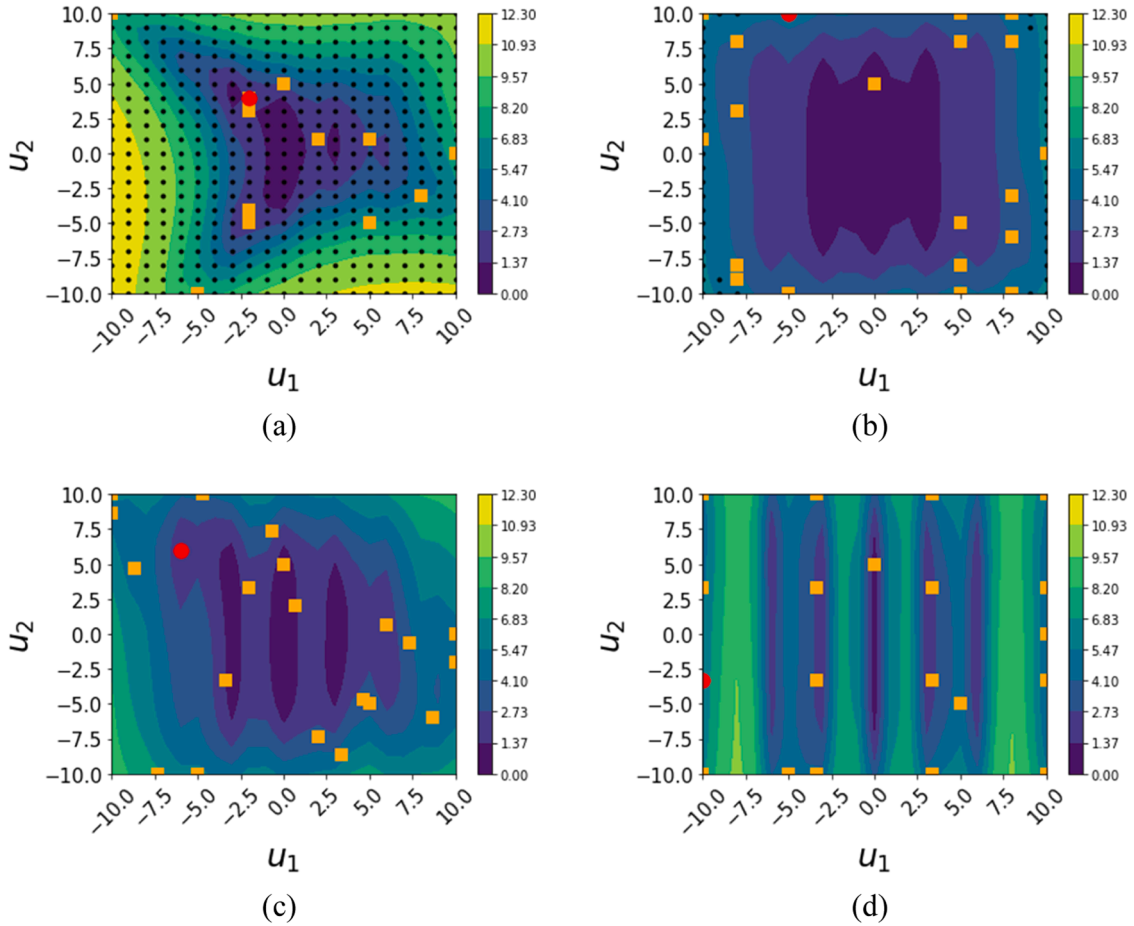


Fig. 8. G-maps generated after 20 calibration experiments designed with: (a) MBDoE, (b) eMBDoE and $J_{G, \text{thr}} = 0.75$, (c) LH, (d) 4^2 full factorial DoE. The candidate design points of MBDoE and eMBDoE are represented with black dots, while already measured experiments are indicated with orange squares. Finally, the red point indicates the experiment selected at this iteration.

This procedure has two main degrees of freedom:

- (1) the definition of the scalar index J_G ;
- (2) the definition of the G-optimality-based requirement to be satisfied.

In this work, we set the following:

- (a) for a given point in the grid (i.e., for a given $\boldsymbol{\varphi}$), the prediction variance $\mathbf{V}_y(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi})|_{j,i}$ of each response at every time point is calculated and then summed to obtain a single scalar J_G :

$$J_G = \sum_{j=1}^{N_y} \sum_{i=1}^{N_{sp}} \mathbf{V}_y(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi})|_{j,i} \quad (9)$$

- (b) the G-optimality-based criterion to be satisfied by the candidate design points is:

$$J_G \geq J_{G, \text{thr}} J_{G, \text{max}}, \quad (10)$$

where $J_{G, \text{max}}$ is the maximum value of G-optimality in the grid, while $J_{G, \text{thr}}$ is a threshold chosen by the user such that: $0 \leq J_{G, \text{thr}} \leq 1$. More specifically, $J_{G, \text{thr}} = 0$ means that all points in the grid are candidates to solve Eq. (5), therefore the design becomes equivalent to a standard MBDoE. The closer $J_{G, \text{thr}}$ gets to 1, the fewer are the remaining design candidate design points, since only points having the highest model prediction variance are accepted.

The above optimal experimental design procedure of G-map eMBDoE can be translated into a constrained optimization problem described by:

$$\begin{aligned} \boldsymbol{\varphi}_{\text{opt}} &= \underset{\boldsymbol{\varphi}}{\text{argmin}} \Psi(\mathbf{V}_{\boldsymbol{\theta}}) \\ \boldsymbol{\varphi}_{\text{s.t.}} \quad &J_G \geq J_{G, \text{thr}} J_{G, \text{max}}. \end{aligned} \quad (11)$$

Therefore, information is maximized considering only the candidate design points having $J_G \geq J_{G, \text{thr}} J_{G, \text{max}}$. In this paper, the grid-search approach employed to solve Eq. (11) does not impact on the final result, since the grid is so fine that an optimization over continuous variables would provide almost identical results (as shown by ad hoc simulations, omitted here for sake of conciseness). If the system dimensionality increases, the computational burden to generate the grids also increases. Therefore, it may be convenient to build grids to initialize the procedure and to set the result obtained as an initial guess for further optimization over continuous variables.

2.1.3. Experiment execution and iterative model calibration

Once the new experiment $\boldsymbol{\varphi}_{\text{opt}}$ is designed, it can be executed either in the physical process or in the simulated one. The new acquired measurement is added to the calibration dataset collected up to the previous iteration and model parameters are estimated using the maximum likelihood method (Eq. (2)). Then, a criterion is assessed in order to decide whether to continue the design and model calibration or not. This criterion is user-defined and can be based on model performance (such as parameter precision or model prediction accuracy) or on the maximum allowed experimental budget. The latter is used in this paper, which implies the experimental campaign is terminated when the experimental budget of N_e experiments is reached. The performance of

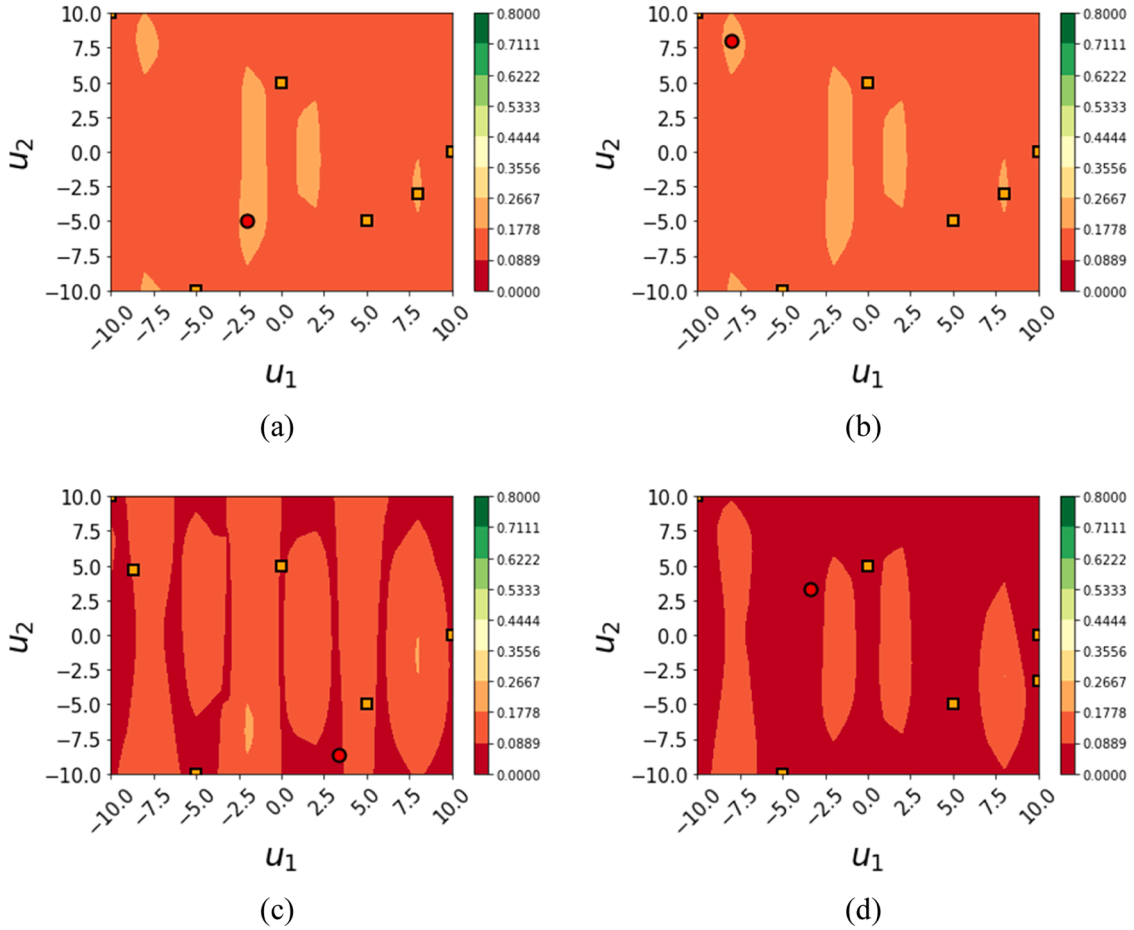


Fig. 9. H-maps generated with 6 calibration experiments (orange squares). Different methods are compared: (a) MBDoE; (b) eMBDoE with $J_{G, \text{thr}} = 0.75$; (c) LH and (d) 4^2 full factorial DoE. The red dot represents the experiment design at the current iteration.

the model at every iteration of the sequential procedure is assessed after the model calibration step. The types of analyses used for this performance evaluation are described in [Section 2.2](#)

2.2. Model calibration analysis

After calibrating the model with the new experiments designed and executed at the i -th iteration, the performance of the optimal experimental design procedure is assessed considering:

- Space exploration;
- Parameter precision;
- Model prediction variance throughout the design space;
- FIM-based optimality metrics used to solve [Eq. \(5\)](#) throughout the design space. [Asprey and Macchietto \(2000\)](#) suggested E-optimality as the most effective criterion to use for the model presented in [Section 3.2](#), being particularly effective on reducing parametric uncertainty when a sequential experimental design approach (as in this work) is adopted ([Galvanin et al., 2007](#)). Therefore, this criterion is used for both case study 1 and 2 for comparison purposes. However, ongoing work is showing that the advantages of G-map eMBDoE over conventional design techniques hold true also when different optimality criteria are used.

The corresponding analysis methods will be detailed in the following sub-sections.

2.2.1. Space exploration

The profile of each control variable is visualized to qualitatively compare the level of space exploration of the proposed experimental design techniques (eMBDoE, MBDoE or LH). As illustrated in [Fig. 3](#), two control variables u_1 and u_2 are represented in the x-axis and y-axis, respectively; red squares indicate the preliminary experiments used to initialize the procedure; the blue squares represent the subsequent (eMBDoE, MBDoE or LH) experiments added iteratively in the sequential procedure. [Fig. 3a](#) shows an example of an *exploratory design*, i.e. a design that covers the entire design space, typical of space-filling design methods like LH, while [Fig. 3b](#) shows an example of an *exploitative design* where new experiments (blue squares) are designed in a limited region of high information content, frequently encountered in MBDoE applications.

Since optimal experimental design methods tend to select replicated design points, i.e. optimal experiments with the same design vector ϕ_{opt} , the different methods are compared in terms of number of distinct design points ϕ_{opt} (see [Tables 2 and 4 of Section 3](#) for further details) to measure space exploration.

2.2.2. Parameter precision

Parameter precision is assessed through a t -test. At a given iteration of the sequential procedure, the Fisher information \mathbf{H}_θ associated to the collected calibration data is calculated. Then, the corresponding parameters variance-covariance matrix \mathbf{V}_θ is calculated from [Eq. \(4\)](#) and t -values are calculated for each iteration as:

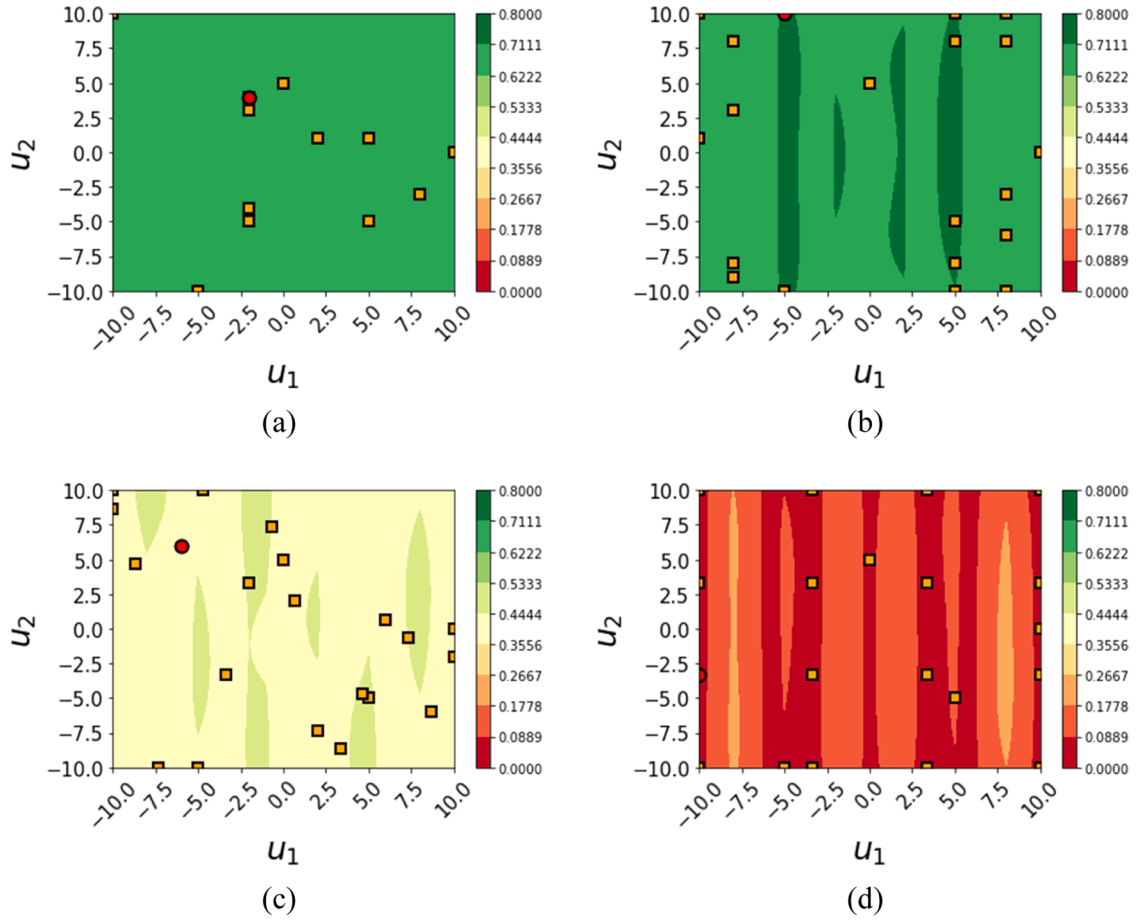


Fig. 10. H-maps generated with 20 calibration experiments (orange squares). Different methods are compared: (a) MBDoE; (b) eMBDoE with $J_{G, \text{thr}} = 0.75$; (c) LH and (d) 4^2 full factorial DoE. The red dot represents the experiment design at the current iteration.

Table 3

Case study 2 (dynamic differential model): initial settings.

Model type	Differential
Inputs	u_1, u_2 $u_1 \in [0.05, 0.20]$ $u_2 \in [5.0, 35.0]$
Outputs	$y = [x_1, x_2]$
- standard deviation of measurement errors	$\sigma_y = [1.0, 1.0] \text{gL}^{-1}$
- initial conditions	$y_0 = [x_1(0), x_2(0)] = [1.0, 0.01] \text{gL}^{-1}$
- sampling points	$t_{\text{sp}} = [7, 14, 21] \text{h}$
Parameters	
- true values	$\theta_{\text{true}} = [\theta_1, \theta_2, \theta_3, \theta_4] = [0.31, 0.18, 0.55, 0.05]$
- initial values	$\theta_0 = [5.0, 5.0, 5.0, 5.0]$
- lower bounds	$\theta_{\text{LB}} = [-20, -20, -20, -20]$
- upper bounds	$\theta_{\text{UB}} = [20, 20, 20, 20]$

$$t_i = \frac{\hat{\theta}_i}{t_{1-\frac{\alpha}{2}}(N - N_0) \sqrt{\mathbf{V}_{\theta_{ii}}}} \quad i = 1, \dots, N_\theta \quad (12)$$

where $t_{1-\frac{\alpha}{2}}(N - N_0)$ is the Student t -value with significance level α and $(N - N_0)$ degrees of freedom, while $\mathbf{V}_{\theta_{ii}}$ is the i -th diagonal element of the variance-covariance matrix. A parameter is statistically precise if:

$$t_i > t_{\text{ref}} \quad i = 1, \dots, N_\theta \quad (13)$$

with $t_{\text{ref}} = t_{1-\alpha}(N - N_0)$.

2.2.3. Maps of G -optimality and information content

At every iteration of the eMBDoE sequential procedure, two maps are built and compared:

- a G -optimality map (G -map), where J_G is calculated at every point of the grid and displayed as a contour plot.
- an information map (H -map), where ψ is calculated at every point of the grid and displayed as a contour plot.

The comparison between the two maps is useful to better understand which are the regions that would be selected based only on ψ (i.e. regions of maximum information) and how much the method will move the design points from those regions by changing the threshold on J_G .

In addition to building G -maps using J_G values, the distribution of G -optimality values (J_G values) in the entire experimental design space at each iteration of the G -map eMBDoE procedure are represented by the following scalar indices:

$$J_{G, \min} = \min(J_{G,i}), \quad i = 1, \dots, N_\phi \quad (14)$$

$$J_{G, \text{mean}} = \text{mean}(J_{G,i}) = \frac{\sum_{i=1}^{N_\phi} J_{G,i}}{N_\phi}, \quad i = 1, \dots, N_\phi \quad (15)$$

$$J_{G, \max} = \max(J_{G,i}), \quad i = 1, \dots, N_\phi \quad (16)$$

where $J_{G, \min}$, $J_{G, \text{mean}}$ and $J_{G, \max}$ are the minimum, mean and maximum values of J_G , considering all the N_ϕ points in the grid.

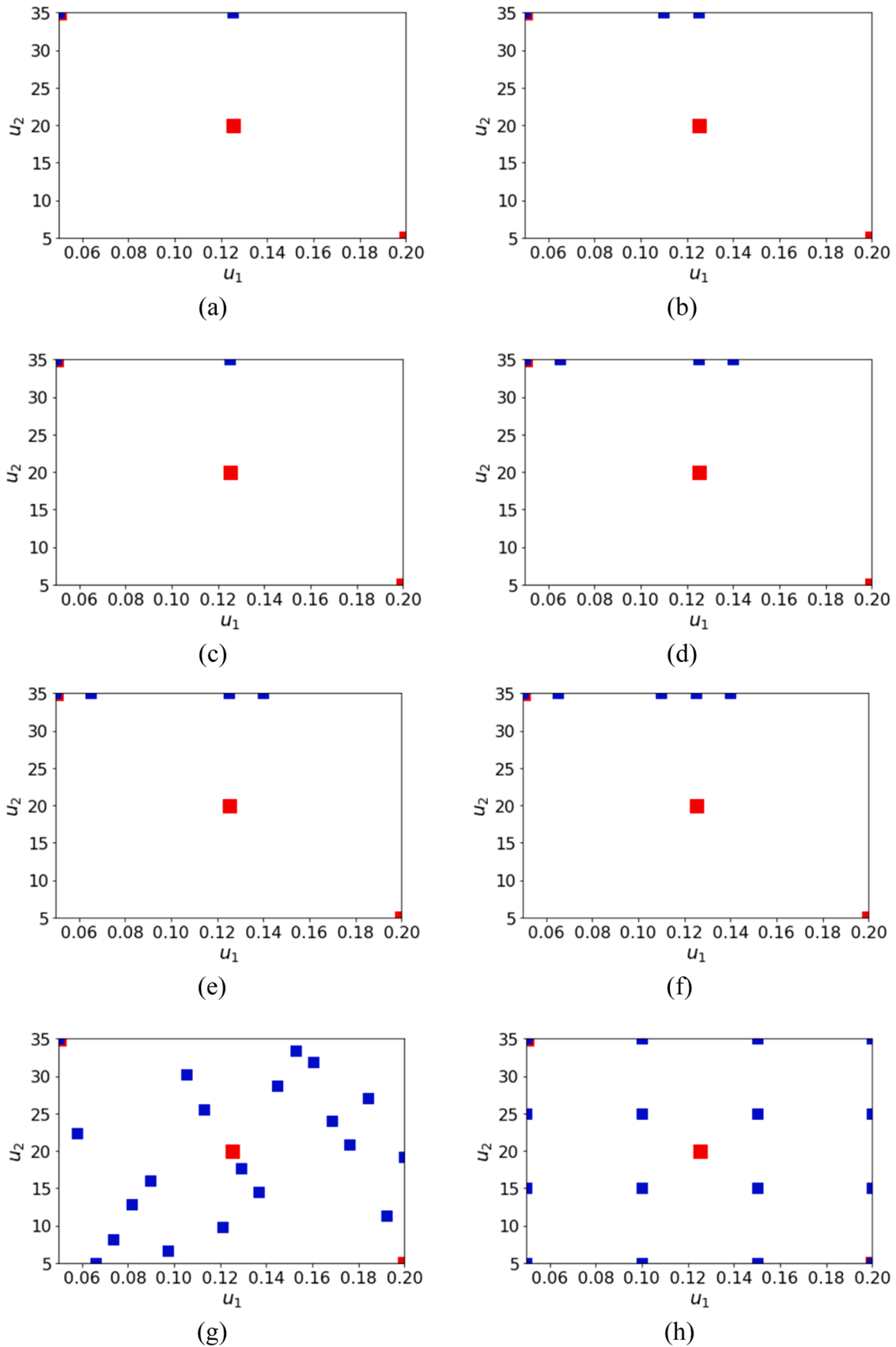


Fig. 11. Design space with the experiments selected by: (a) MBDoE; (b) G-map eMBDoE $J_{G,thr} = 0.25$; (c) G-map eMBDoE $J_{G,thr} = 0.50$; (d) G-map eMBDoE $J_{G,thr} = 0.65$; (e) G-map eMBDoE $J_{G,thr} = 0.75$; (f) G-map eMBDoE $J_{G,thr} = 0.85$; (g) Latin Hypercube; (h) 4^2 full factorial DoE. Red squares indicate preliminary experiments, blue squares indicate experiments designed with the method considered.

Table 4

Number of distinct design points for every scenario compared in the study.

Scenario	No. distinct design points
MBDoE	2
eMBDoE, thr:0.25	3
eMBDoE, thr:0.50	2
eMBDoE, thr:0.65	4
eMBDoE, thr:0.75	4
eMBDoE, thr:0.85	5
LH	16
DoE	16

2.2.4. Implementation of G-maps and H-maps

The two case studies of Section 3.1 and 3.2 are implemented in Python 3.9 (Spyder) and simulated in an Intel® Core™ i7–10875H CPU, @ 2.30 GHz processor with 64.0 GB RAM. The grid of points used to select candidate design points based on their G-optimality value is obtained by discretising the ranges of control variables into equal intervals. Maps of information content across the design space (*H-maps*) are also built using the same discretization of the design space used with G-maps. The procedures of the two optimization-based methods, namely MBDoE and G-map eMBDoE, differ mainly for the additional step of selection of candidate design points based on their J_G values. To compare the methods in terms of required computational time, the time to design one experiment in each case (MBDoE or G-map eMBDoE) are reported in

Section 3.1 and 3.2.

3. Results and discussion

Two simulated case studies are used to compare the performance of the following design of experiments techniques in terms of space exploration, precise parameter estimation and minimization of model prediction uncertainty:

- MBDoE: this is an exploitative design, i.e. only optimal experiments based on the iterative solution of Eq. (5) are selected;
- full-factorial DoE and LH: these designs are exploratory designs, i.e. they guarantee an exploration of the whole experimental design space;
- G-map eMBDoE: this design seeks a trade-off between information maximization and space exploration through the definition of a threshold $J_{G, \text{thr}}$. Different values of $J_{G, \text{thr}}$ are considered.

The space-filling LH design is generated with the doepy package for Python (<https://doepy.readthedocs.io/en/latest/>). Preliminary simulations revealed that the results in terms of parameters precision and model prediction variance are similar regardless of random variations of the selected LH samples.

In both case studies, in silico data is generated according to the following procedure:

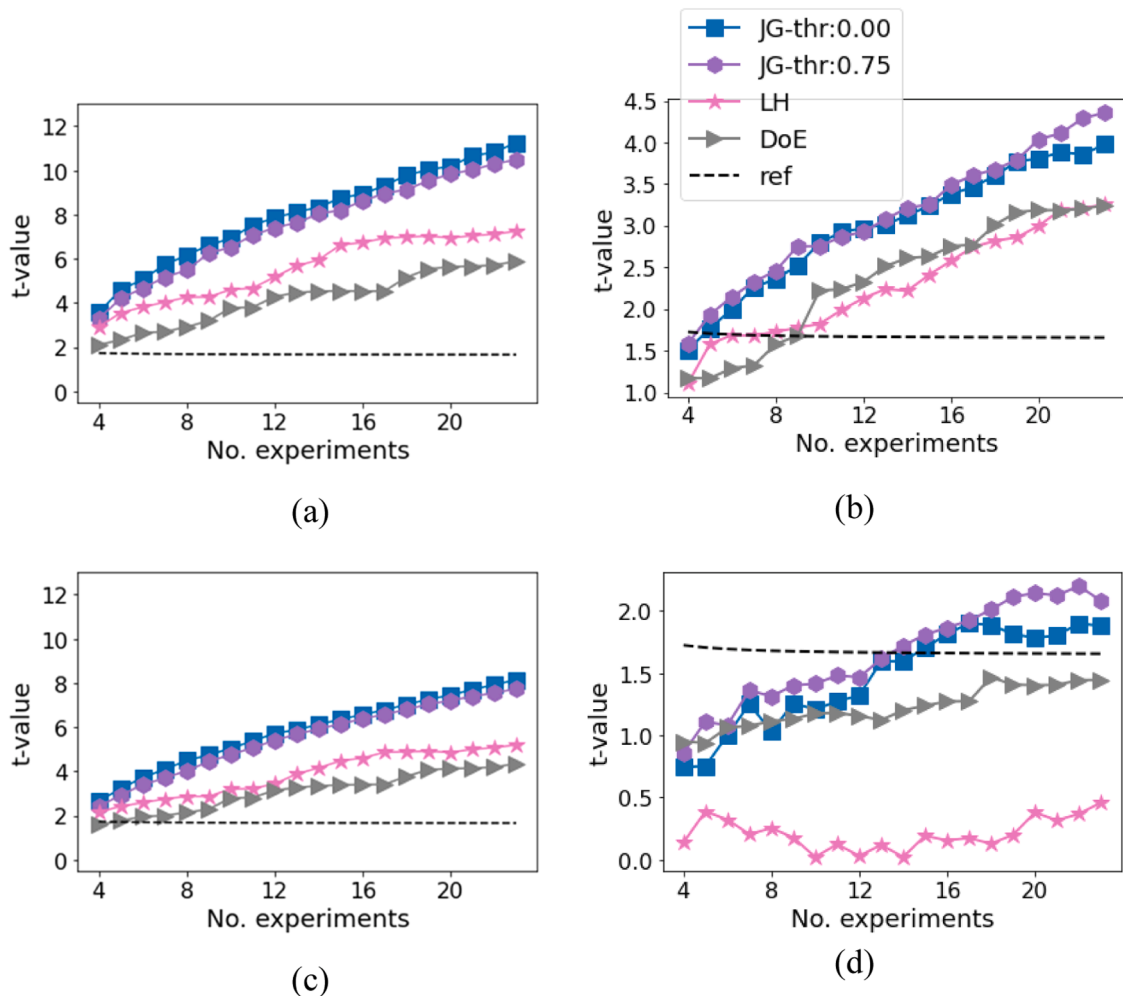


Fig. 12. Profiles of t -values calculated with: MBDoE ($J_{G, \text{thr}} = 0.00$); G-map eMBDoE ($J_{G, \text{thr}} = 0.75$); Latin Hypercube (LH); 4^2 full factorial DoE. The t -values are compared against the reference t -value ('ref' in the legend) for parameters 1–4 in figures (a)–(d), respectively.

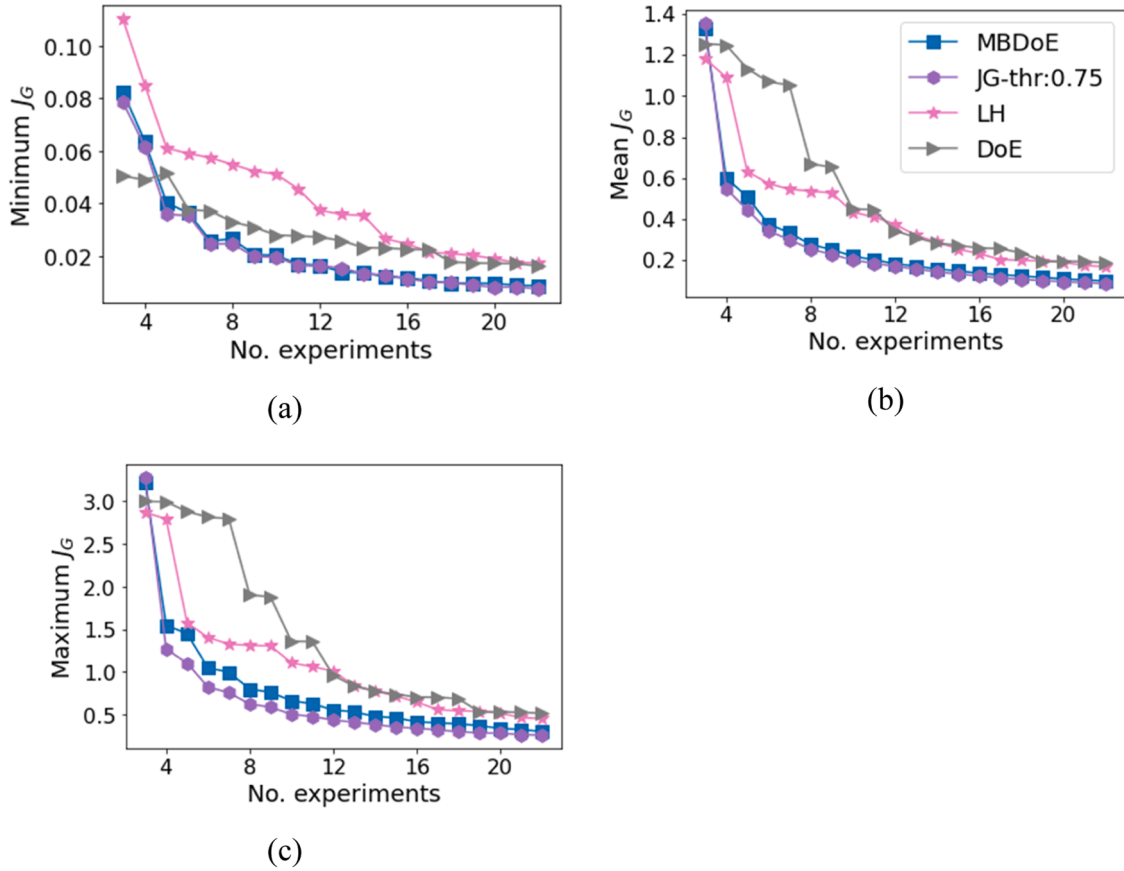


Fig. 13. Profiles of scalar indices of G-optimality calculated with: MBDoe ($J_{G, \text{thr}}=0.00$); G-map eMBDoE ($J_{G, \text{thr}}=0.75$); Latin Hypercube (LH); 4^2 full factorial DoE. Results are obtained by summing up the variance calculated for every response at every time point. Three different scalar measures are considered: (a) minimum G-optimality; (b) mean G-optimality; (c) maximum G-optimality.

- Model equations and true parameter vector θ_{true} (see Section 3, Table 1 and 3) are used to generate the exact value of the model responses y_{exact} at the selected experimental condition;
- A gaussian error with zero mean and a user-defined standard deviation σ_y is then added to y_{exact} to obtain a “noisy” measurement y_{noisy} .

The user-defined standard deviation σ_y is chosen by the user to mimic the precision of the measurements in the physical system and can be typically evaluated from a set of preliminary replicated experiments.

To make the results comparable, the same initial settings are used for all scenarios of the same case study: true model parameter vector (θ_{true}) for the in silico data generation; initial parameters values and lower and upper bounds (θ_{LB} , θ_{UB} respectively) for parameter estimation; standard deviation of the response measurement error (σ_y); ranges for the control variables; set of preliminary experiments and total (maximum) number of experiments (N_e) to be executed. Finally, the selection of the proper threshold for G-optimality is case-dependent, therefore different sets of $J_{G, \text{thr}}$ may be considered in case study 1 and 2.

Finally, robustness of results to the selection of sampling points for case study 2 and different realisations of random noise for the response variables of case study 1 and 2 are shown in Section 1 and 2, respectively, of the Supplementary Material.

3.1. Case study 1

The following two-inputs, single response algebraic model is considered in case study 1.

$$y = \theta_1 u_1 + \theta_2 u_1 u_2 + \theta_3 u_1^2 + \theta_4 u_2^2 + \theta_5 \sin(u_1) \quad (17)$$

Preliminary analysis showed that this model is structurally

identifiable (by using the structural identifiability technique of Asprey and Macchietto, 2000), therefore MBDoe for parameters precision can be applied. Moreover, E-optimality criterion is used to optimize information content, since Asprey and Macchietto (2000) showed its efficacy for the model of case study 2 (in Section 3.2) and the same optimality criterion is used with both case studies for comparison purposes. Initial settings on variables and parameters are provided in Table 1. In this case, the design vector ϕ is equal to $\mathbf{u} = [u_1, u_2]$.

Different thresholds of G-optimality are employed for the explorative MBDoe: $J_{G, \text{thr}} \in \{0, 0.25, 0.50, 0.65, 0.75, 0.85\}$ to evaluate the impact of threshold choice on design performance. Notice that $J_{G, \text{thr}} = 0$ corresponds to a state-of-the-art E-optimal MBDoe, since the constraint in Eq. (11) becomes $J_G \geq 0$ and J_G is always non-negative. All eMBDoE scenarios employ the E-optimal criterion (Eq. (6)).

Even though LH and 4^2 full factorial DoE experiments are designed at once and should be evaluated at the end of the experimental campaign, in this paper intermediate results are included in order to understand and compare the evolution of different design methods.

For all the scenarios, the same preliminary dataset is used: 5 experiments selected through a LH sampling. These preliminary experiments are used to achieve a first parameter estimation and to initialize \mathbf{H}_0 calculations to avoid potential singularity issues in the information matrix. Moreover, parameter estimates from each iteration become initial parameters values for the Maximum Likelihood estimation in the subsequent iteration. A maximum budget of 16 designed experiments is considered in each scenario; therefore, $N_e = 21$ experiments are obtained at the end of the experimental campaign.

All the scenarios are compared in terms of space exploration: Fig. 4 shows the location of the $N_u = 2$ control variables within the entire design space, while Table 2 shows the number of distinct experimental

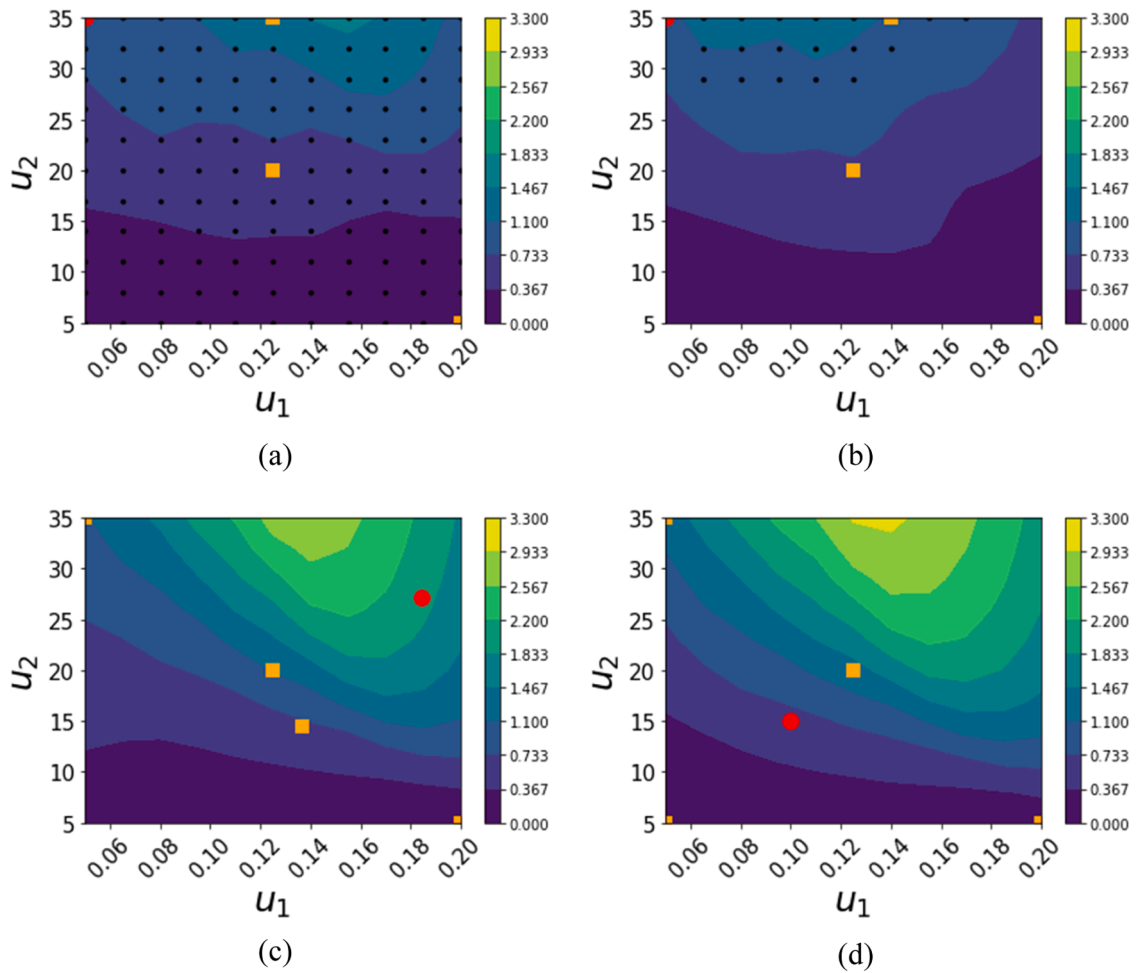


Fig. 14. G-maps generated after 1 calibration experiment. Four methods are compared: (a) MBDoE; (b) G-map eMBDoE with $J_{G,thr} = 0.75$; (c) LH; (d) 4^2 full factorial DoE. Orange squares indicate already measured data (namely, data used to calibrate the model); black dots indicate candidate design points; the red point indicates the experiment designed at the current iteration.

conditions (i.e., different ϕ s) selected by different methods. Results show that the novel explorative MBDoE has the best performance in terms of space exploration when a threshold of 0.65–0.75 is selected (Fig. 4d-e; 16 different design points indicated in Table 2). Similarly, LH (Fig. 4g) and 4^2 full factorial DoE (Fig. 4h), which are inherently explorative, select 16 distinct points that cover all regions of the design space. Moreover, the smaller the threshold $J_{G,thr}$, the less eMBDoE experiments are spread in the design space (see Figs. 4b-d): with $J_{G,thr} = 0.50$, 15 different optimal experiments are selected; with $J_{G,thr} = 0.25$, 12 different are selected; with $J_{G,thr} = 0$, namely a conventional E-optimal MBDoE, only 7 distinct points are selected (see Table 2).

Although G-map eMBDoE with $J_{G,thr} = 0.85$ is one of the most explorative methods (Fig. 4f) and has the greatest reduction of G-optimality throughout the entire design space (results are provided in Appendix A, section A.1), it is the eMBDoE scenario that requires more experiments to precisely estimate parameter θ_5 (Appendix A, section A.1). Hence, to find a better trade-off between space exploration and information maximization, eMBDoE with $J_{G,thr} = 0.75$ is considered in the analysis of precise parameter estimation. For this purpose, the t -values calculated at every iteration for the full set of model parameters are shown in Fig. 5. As shown in Fig. 5, the most critical parameter which requires a higher number of calibration experiments to be precisely estimated is still θ_5 (Fig. 5e). MBDoE and eMBDoE with $J_{G,thr} = 0.75$ require 10 experiments (5 preliminary and 5 optimal) to pass the t -test, LH requires 17 experiments, while DoE is not able to pass the t -test. Finally, details on parameters accuracy (i.e. distance from the assumed

true value) can be found in Appendix A, section A.2; moreover, reproducibility of the LH results despite random variations of different LH designs is shown in Appendix A.3.

In Fig. 6 the different scenarios are compared in terms of model prediction variance across the whole experimental design space. Results are shown for MBDoE, G-map eMBDoE with $J_{G,thr} = 0.75$ threshold, factorial DoE and LH. The smaller the scalar index of J_G , the better the performance in terms of reduction of model prediction uncertainty. Scalar indices are calculated for the G-maps generated during experiment design step: therefore, the iterative generation of G-maps starts with a calibration dataset of 5 preliminary experiments and terminates with a calibration dataset of 20 experiment, which is the map used to calculate the last optimal experiment since $N_e = 21$. The following ranking is obtained in terms of scalar measures of model prediction variances of Eqs. (14)–(16):

- mean G-optimality $J_{G,mean}$, from the 11th experiment onwards (Fig. 6a):
MBDoE > DoE > LH > eMBDoE ($J_{G,thr} = 0.75$)
- maximum G-optimality $J_{G,max}$, from the 14th experiment onwards (Fig. 6b):
MBDoE > DoE > LH > eMBDoE ($J_{G,thr} = 0.75$)

Instead, the minimum G-optimality is equal to zero in all scenarios throughout the experimental campaign, therefore it is omitted here for sake of conciseness.

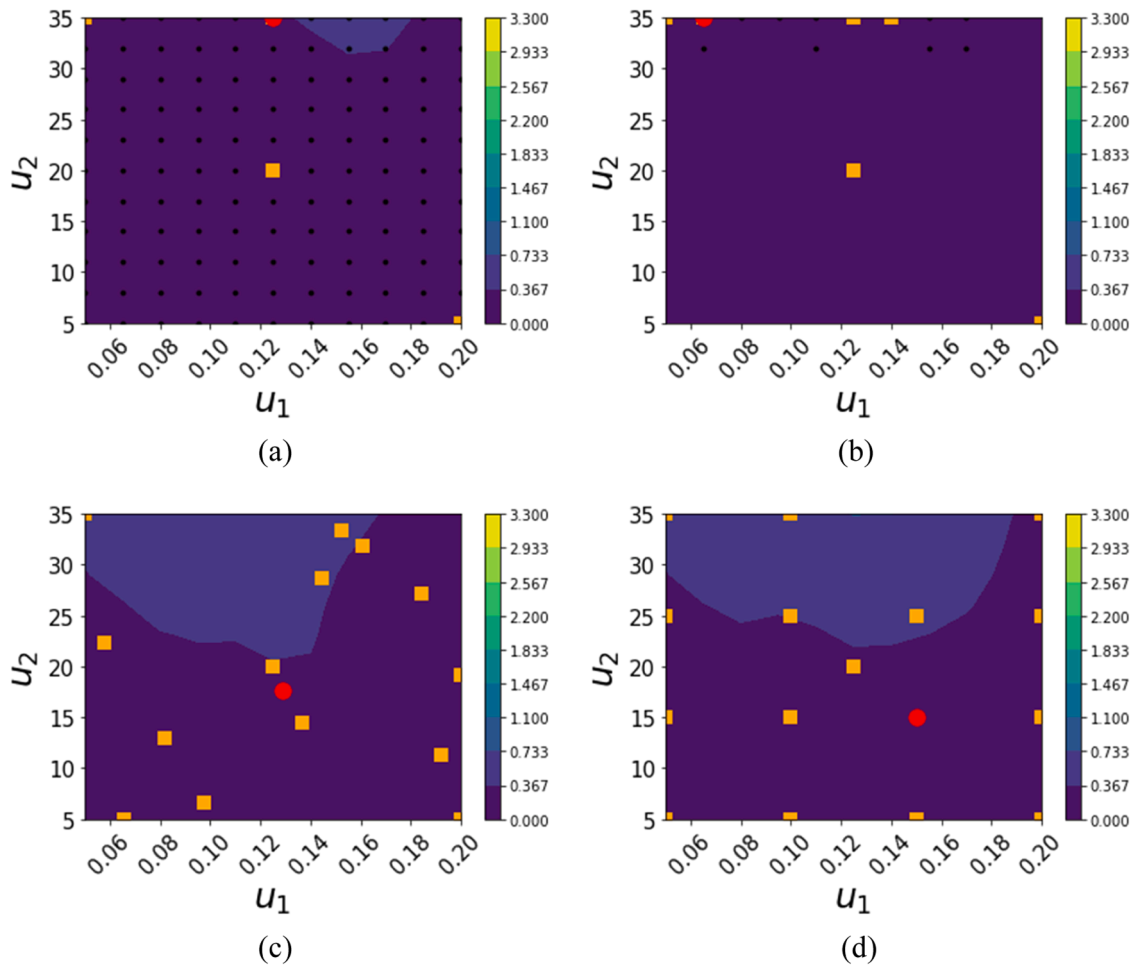


Fig. 15. G-maps generated after 15 calibration experiments. Four methods are compared: (a) MBDoE; (b) G-map eMBDoE with $J_{G,thr} = 0.75$; (c) LH; (d) 4^2 full factorial DoE. Orange squares indicate data already used to calibrate the model; black dots indicate candidate design points; the red point indicates the experiment designed at the current iteration.

To conclude, scalar indices of G-optimality prove that eMBDoE with $J_{G,thr} = 0.75$ has the best performance in terms of reduction of model prediction variance. Moreover, compared to MBDoE, both mean and maximum values of G-optimality are smaller in explorative design methods such as DoE and LH. This suggests that space exploration promotes the reduction of prediction uncertainty, but the best overall result (i.e. minimum prediction variance and maximum parameter precision) is only achieved when a trade-off between space exploration and information maximization is realised.

G-maps are shown in Figs. 7 and 8 to visualize the regions of higher prediction variance within the design space for MBDoE, G-map eMBDoE with $J_{G,thr} = 0.75$, LH and DoE. This result is compared to the maps of information content (H-maps, Figs. 9, 10) for every scenario. For sake of conciseness, only a subset of G-maps and H-maps are included:

- Calibration dataset of 6 experiments (5 preliminary and 1 optimal): maps obtained after the first iteration;
- Calibration dataset of 20 experiments (5 preliminary and 15 optimal): maps generated in the last iteration.

In these maps, different experimental conditions are highlighted: data from experiments already performed, which are used to calibrate the model (orange squares); candidate design points based on the G-optimality threshold (black dots); experiment selected at the current iteration (red dot). Notice that the discretization of the design space and the selection of candidates (black dots) are not performed in LH and

DoE, therefore black points are not present in these figures (see Figs. 7c-d and 8c-d).

Finally, it must be noticed that:

- G-maps represent the model prediction variance, which must be minimized; therefore, the best performance is found in ‘blue’ regions of Figs. 7 and 8 (i.e., the smallest G-optimality values) and the worst one is found in yellow regions (i.e., the highest G-optimality values);
- H-maps represent the amount of information, based on the E-optimal criterion, which must be maximized; in this case, the best performance is found in dark green regions of Figs. 9 and 10 (i.e., high information values) and the worst performance is found at the red regions (small information values).

After measuring the first optimal experiment (Fig. 7a-7d), the G-maps of MBDoE, eMBDoE and LH are quite similar in terms extension of regions with small model prediction variance (blue regions). Moreover, in all of them the G-optimality is smaller in central regions, while it increases towards extreme values of u_1 and u_2 . The G-map of DoE (Fig. 7d) is similar, but slightly worse due to a larger extension of regions with high model prediction variance (i.e., yellow regions). The differences in the distribution of G-optimality becomes more evident after measuring the 20th experiment: indeed, the best performance is achieved with eMBDoE using $J_{G,thr} = 0.75$ (Fig. 8b) since it has the largest extension of the blue region; moreover, the yellow regions at extreme values of the two control variables disappear. The second-best

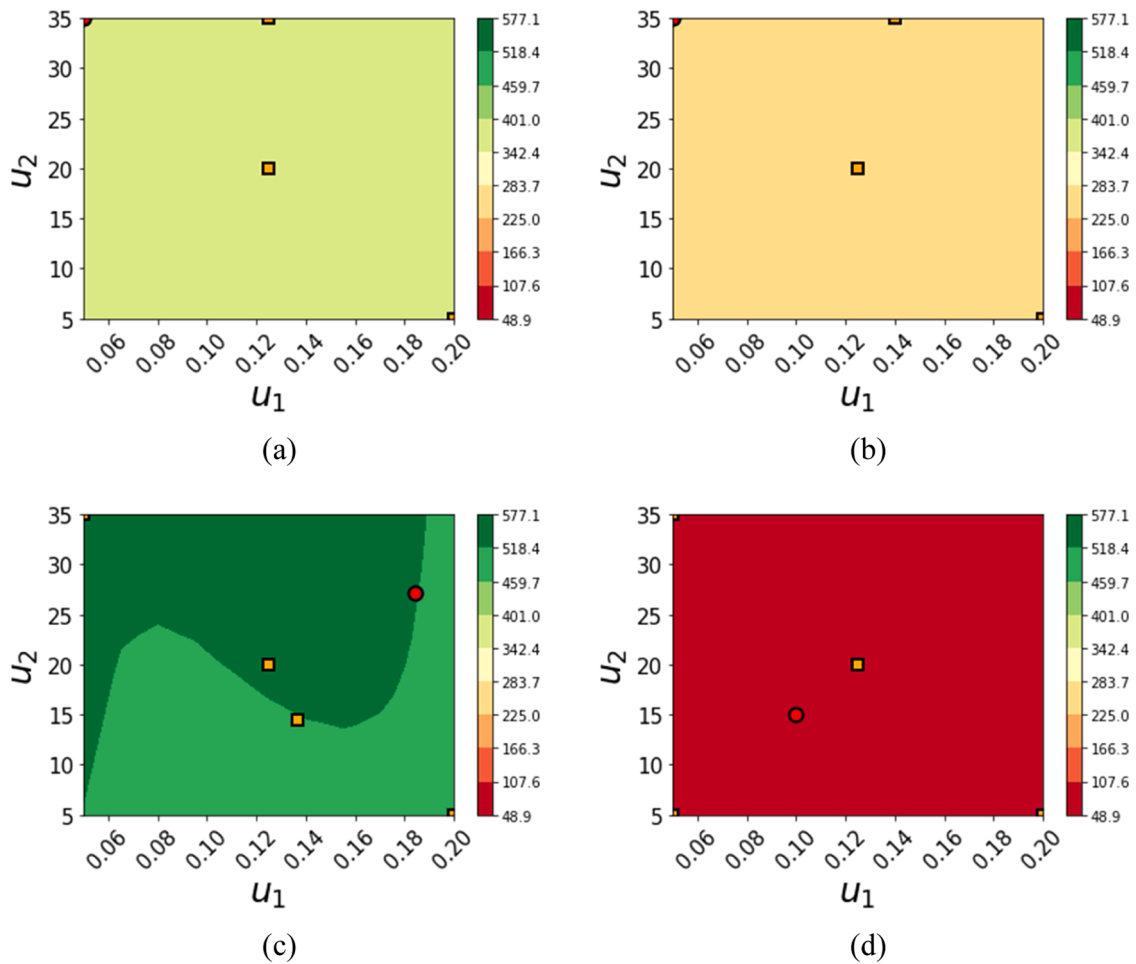


Fig. 16. H-maps generated with 1 calibration experiment. The compared methods are: (a) MBDoE; (b) G-map eMBDoE with $J_{G, \text{thr}} = 0.75$; (c) LH; (d) 4^2 full factorial DoE. Orange squares indicate already measured data (namely, data used to calibrate the model); the red point indicates the experiment designed at the current iteration.

performance in terms of reduction of model prediction variance is found with LH (Fig. 8c); instead, MBDoE and DoE (Fig. 8a and 8d, respectively) still have regions with larger model prediction variance (yellow regions). This suggests that an explorative strategy as LH can improve the prediction precision with respect to an optimal design, but the best solution is found when a good trade-off between space exploration and information maximization is achieved.

By looking at the H-maps generated with 6 measured experiment (Fig. 9) and with 20 measured experiments (Fig. 10), it is clear that MBDoE and G-map eMBDoE outperforms LH and DoE: indeed, their H-maps are red in the first iteration (Figs. 9a-b) and become green in the subsequent iterations (Figs. 10a-b), while LH and DoE end up with a light green (Figs. 10c) and a red maps (Figs. 10d), respectively. This further confirms that the enhancement of space exploration of the proposed explorative MBDoE does not entail a significant loss of information content with respect to a completely optimal design.

The computational times to build G-maps and H-maps (with E-optimal criterion) and to design one experiment with Python 3.9 in an Intel® Core™ i7–10875H CPU, @ 2.30 GHz processor with 64.0 GB RAM are:

- 0.12 s with G-map eMBDoE and $J_{G, \text{thr}} = 0.75$;
- 0.12 s with MBDoE.

3.2. Case study 2

The G-map eMBDoE is applied to the Canoid-type kinetic model describing the material balances of the fermentation of baker's yeast in a fed-batch reactor. Original model and proof of structural identifiability can be found in Asprey and Macchietto (2000).

The two-response dynamic model is represented by the following set of differential and algebraic equations:

$$r = \frac{\theta_1 x_2}{\theta_2 x_1 + x_2} \quad (18)$$

$$\frac{dx_1}{dt} = (r - u_1 - \theta_4) x_1, \quad (19)$$

$$\frac{dx_2}{dt} = -\frac{rx_1}{\theta_3} + u_1(u_2 - x_1). \quad (20)$$

Some simplifying assumptions are made: the inputs u_1 and u_2 are constant in time, and the number of sampling points is fixed ($N_{\text{sp}} = 3$). Therefore, the design vector becomes: $\boldsymbol{\phi} = \mathbf{u} = [u_1, u_2]$; u_1 is the dilution factor with range 0.05–0.20 h^{-1} , while u_2 is the substrate concentration in the feed with range 5.0–35.0 g/L . The two measured concentrations are the biomass concentration x_1 [g/L] and the substrate concentration x_2 [g/L].

As in Section 3.1, model calibration data are obtained by using a simulated process: 'true' parameter vector $\boldsymbol{\theta}_{\text{true}}$ is used in the model Eqs. (18)–(20) to generate noise-free simulated responses for x_1 and x_2 ; then,

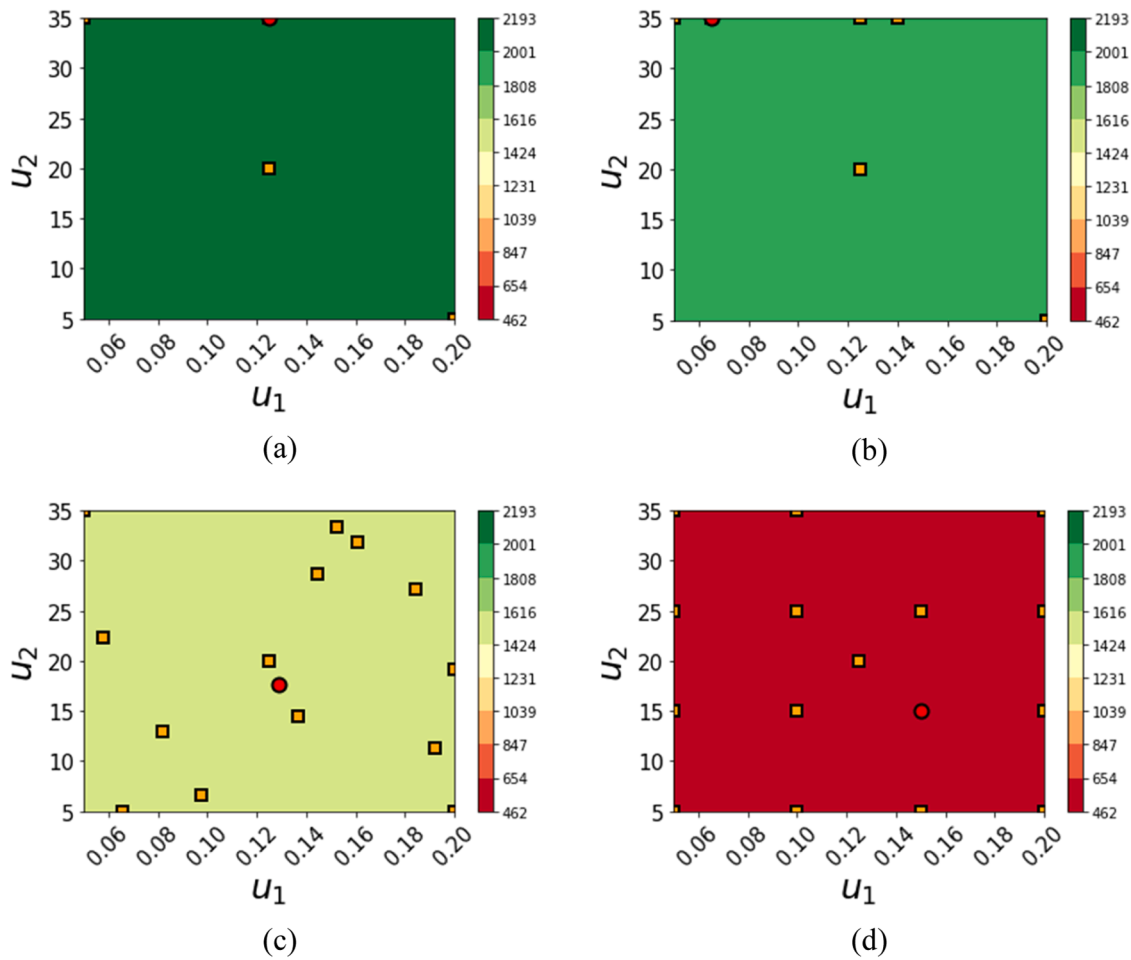


Fig. 17. H-maps generated with 15 calibration experiments. The compared methods are: (a) MBDoE; (b) G-map eMBDoE with $J_{G, \text{thr}}=0.75$; (c) LH; (d) 4^2 full factorial DoE. Orange squares indicate already measured data (namely, data used to calibrate the model); the red point indicates the experiment designed at the current iteration.

gaussian noise with zero mean and a user-defined standard deviation σ_y ($\sigma_y = [1.0, 1.0] \text{gL}^{-1}$, see Table 3) is added in order to mimic measurement errors. Model parameters can be estimated using the in-silico calibration data starting from a set of initial parameters values θ_0 within the ranges $\theta_{LB} - \theta_{UB}$.

Details on settings and parameters and variables ranges are reported in Table 3.

The same G-map based eMBDoE method applied to the algebraic model of Section 3.1 can be applied to this model, but a higher level of complexity is introduced here. In case study 1, the algebraic model was simulated to get the value of a single response variable that corresponds to the single measurement in an experiment at a particular time instance, such as at steady state. However, in this second case study, the dynamic model is simulated to obtain output responses at different time points, which corresponds to a typical fed-batch experiment with multiple sampling points in time. Since we set $N_{sp} = 3$, 6 values of model prediction variance can be calculated: V_y of x_1 at $t_{sp} = [7, 14, 21]$ h and V_y of x_2 at $t_{sp} = [7, 14, 21]$ h. To summarize these results, the sum J_G of all contributions V_y is calculated as in Eq. (9), which is used in the G-map eMBDoE method to design the optimal and explorative experiments. More details on the single contributions V_y can be found in Appendix B, section B.4.

As in Case Study 1, 4 different design of experiments techniques are compared:

- An E-optimal MBDoE, which is Gmap eMBDoE with $J_{G, \text{thr}} = 0.00$;

- Two explorative designs, namely LH and 4^2 full factorial DoE;
- A G-map eMBDoE with E-optimal criterion and different threshold values $J_{G, \text{thr}} \in \{0.25, 0.50, 0.65, 0.75, 0.85\}$.

Three preliminary experiments are designed with LH in order to initialize all the four design of experiments methods. Then, the maximum number of experiments designed in each scenario is fixed to 20, providing a total number of experiments $N_e = 23$.

The extent of space exploration realized by each method can be deduced from Fig. 11 and Table 4. After 20 iterations of experiments design, MBDoE (Fig. 11a) has selected replicates at 2 different experimental conditions and design space exploration is very limited. Instead, eMBDoE (Fig. 11b-f) is able to increase space exploration in such a way as the number of replicated points is reduced, i.e., more distinct experimental conditions are obtained. This is evident with $J_{G, \text{thr}} \in \{0.65, 0.75, 0.85\}$ (Fig. 11d-f). In these cases, the distinct points increase to 3 or 4 (Table 4) instead of the 2 selected by the conventional E-optimal MBDoE. The control variable which is affected by the $J_{G, \text{thr}}$ is u_1 , whereas the different optimal designs select the same value for u_2 for all optimal experiments.

Parameters precision is assessed through t -tests; Fig. 12 shows the results of MBDoE (i.e., $J_{G, \text{thr}} = 0$), eMBDoE with $J_{G, \text{thr}} = 0.75$ (the others are omitted for sake of conciseness), LH and factorial DoE. The G-optimality threshold $J_{G, \text{thr}} = 0.75$ is selected since it has the best performance in terms of precise parameters estimation and in reduction of model prediction variance (for more details, see Appendix B, section B.1).

Parameters $\hat{\theta}_1$ (Fig. 12a) and $\hat{\theta}_3$ (Fig. 12c) pass the t -test with few experiments in every scenario, while the most critical parameters are $\hat{\theta}_2$ (Fig. 12b) and, especially, $\hat{\theta}_4$ (Fig. 12d). Conventional E-optimal MBDoE requires 2 optimal experiments to estimate parameter $\hat{\theta}_2$ (Fig. 12b) and 12 experiments to estimate $\hat{\theta}_4$ (Fig. 12d). The performance is improved by enhancing space exploration through G-map eMBDoE with $J_{G, \text{thr}} = 0.75$, requiring 2 experiments to pass the t -test for $\hat{\theta}_2$ (Fig. 12b) and 11 experiments to pass the t -test for $\hat{\theta}_4$ (Fig. 12d). However, explorative designs such as factorial DoE and LH do not allow to precisely estimate $\hat{\theta}_4$ (Fig. 12d) within the experimental budget of $N_e = 23$ experiments. This suggests that a trade-off between space exploration and information maximization provides the best results in terms of parameters precision.

Additional details on parameter estimation accuracy (i.e. distance from the assumed true value) can be found in Appendix B, section B.2; reproducibility of the LH results is shown in Appendix B.3.

As indicated in Eq. (9) Section 2), the G-map for the selection of eMBDoE experiment is built by summing the prediction variance contributions from the two model responses. More details of the contributions to the overall J_G can be found in Appendix B, section B.4. To compare quantitatively the grids of G-optimality values obtained at every iteration of MBDoE, eMBDoE, factorial DoE and LH, scalar measures of G-optimality are calculated: minimum, mean and maximum values of the J_G calculated for each point in the grid (see Eqs. (14)–(16) in Section 2.2.3).

Considering the minimum, mean and maximum values of G-optimality (Figs. 13a–c), explorative designs such as LH and factorial DoE provide a slower reduction in model prediction variance at the beginning of the experimental campaign and they stabilize at higher values when the maximum experimental budget is reached. When $J_{G, \text{mean}}$ and $J_{G, \text{max}}$ are considered (Fig. 13b–c), the explorative MBDoE has a better performance than conventional MBDoE and it is able to reduce the model prediction variance with the lowest number of experiments. This further suggests that the trade-off between space exploration and information maximization realized by eMBDoE leads to the best performance also in terms of prediction variance.

The different experimental design methods are compared in terms of G-maps and H-maps:

- E-optimal MBDoE (Fig. 14a, 15a, 16a and 17a);
- eMBDoE, with E-optimal criterion and a threshold of $J_{G, \text{thr}} = 0.75$ (Fig. 14b, 15b, 16b and 17b);
- LH (Fig. 14c, 15c, 16c and 17c);
- 4^2 full-factorial DoE (Fig. 14d, 15d, 16d and 17d).

Both maps are built at every iteration in order to calculate the optimal and explorative experiments as in Eq. (11). Here, only the results obtained in two iterations are shown for the sake of conciseness. This includes results after 4 calibration experiments (Figs. 14–15), to assess model prediction variance reduction and information gain after adding only one designed experiment besides the 3 preliminary ones; and results after 15 calibration experiments (Fig. 15 and 17), being the first iteration where G-map eMBDoE is able to reduce completely model prediction variance in the entire design space (darkest blue in the whole design space, as shown in Fig. 15).

After calibrating the model with the data obtained from the fourth experiment (Fig. 14), the G-map is characterized by high G-optimality values with u_2 between 30 and 35 g/L for MBDoE (Fig. 14a) and eMBDoE (Fig. 14b) and with u_2 between 15 and 35 g/L in case of LH (Fig. 14c) and factorial DoE (Fig. 14d). Therefore, designs that take into account information content, such as MBDoE and eMBDoE, provide a better performance at the first iteration with respect to completely explorative designs, such as LH and factorial DoE. When the number of experiments used in calibration increases to 15, G-map eMBDoE (Fig. 15b) is the only scenario able to reduce completely model

prediction variance in the entire design space, confirming the results obtained with scalar indices of G-optimality (Fig. 13).

By analysing the distribution of information content throughout the design space with four calibration experiments, factorial DoE (Fig. 16d) provides the smallest amount of information, while eMBDoE (Fig. 16b) and MBDoE (Fig. 16a) guarantee higher information levels. Unexpectedly, LH generates the highest values of information (Fig. 16c); this may be caused by the initialization of the MBDoE and eMBDoE procedure with parameters estimates that are still quite far from the true values. Indeed, with 15 experiments, the following rank of information content is found: MBDoE (Fig. 17a) > eMBDoE (Fig. 17b) > LH (Fig. 17c) > factorial DoE (Fig. 17d). Therefore, conventional MBDoE generates the maximum amount of information content as expected, while eMBDoE provides an intermediate result between optimal (MBDoE) and explorative designs (LH, factorial DoE). Maps generated in the last experiment design iteration are shown in Appendix B, section B.5.

The computational time required to discretize the design space, characterize it in terms of information content and model prediction variance (i.e., to build H-maps and G-maps) and to calculate the optimal and/or explorative experiment is approximately 0.65 s for both MBDoE and eMBDoE.

This suggests that building of both G-maps and H-maps does not lead to an excessive computational burden even though the model complexity is increased significantly with respect to case study 1.

4. Conclusions and future work

A novel exploratory model-based design of experiments method (eMBDoE) has been proposed in this paper with the objective of precisely estimating model parameters and minimizing model prediction uncertainty in the whole domain of model utilization with the minimum experimental effort. The proposed method is based on a mapping of model prediction variance evaluated across the entire design space through a scalar measure of G-optimality (G-map), which is calculated based on the evaluation of Fisher information matrix and requires knowledge on model structure (set of equations) and estimated parameter values. Experimental conditions within the design space are then selected from the candidate design points, which are associated to a G-optimality value higher than a user-defined threshold. Therefore, the exploration of the space is pushed towards regions that are still not well-described by the model. The G-map based explorative MBDoE is compared against purely explorative methods, i.e. full factorial DoE and LH, and against a purely information-based exploitative method, i.e. MBDoE.

These experimental design techniques were applied to two simulated case studies of increasing complexity: an algebraic model characterized by one response and two control variables and a nonlinear differential equation model of baker's yeast fermentation in a fed-batch reactor with two response variables measured at three sampling points and two control variables. In both cases, the constraint on G-optimality enables an increase of space exploration: the larger the threshold on G-optimality, the more the designed experiments depart from the ones selected by MBDoE.

Results from both case studies suggest that the trade-off between space exploration and information maximization achieved by G-map eMBDoE allows to minimize the number of experiments required to precisely estimate model parameters and to minimize model prediction variance in the whole design space. In case study 1, 14 calibration experiments designed through G-map eMBDoE with $J_{G, \text{thr}} = 0.75$ allows to estimate all model parameters with statistical precision and to reduce G-optimality to the minimum values among all considered scenarios. As regards case study 2, the experimental burden is minimized by eMBDoE with $J_{G, \text{thr}} = 0.75$: 15 calibration experiments are enough to precisely estimate model parameters and reduce G-optimality to a minimum value throughout the design space. Different simulations of the systems under study suggest that a good trade-off is found with a G-optimality threshold of 0.65–0.85.

Ongoing work is focused on further validation of G-map eMBDoE through data generated by a physical system and on the development of a systematic method to determine the best G-optimality threshold.

Finally, the additional step of candidate design points selection required by G-map eMBDoE leads to a negligible increase in computational time with respect to the state of the art MBDoE. Moreover, the time required to design a single experiment increases with model complexity, namely with the increase of the number of control variables, response variables and/or sampling points, but it is still negligible in the case studies analysed in this work: 0.12 s for case study 1 and 0.65 s for case study 2. Thanks to the satisfactory performance of the G-map eMBDoE with the two simulated processes and the limited computational burden, we envisage the future implementation of this technique in automated platforms for online model identification, to integrate and test the proposed experimental design method in an actual experimental system.

CRedit authorship contribution statement

Francesca Cenci: Conceptualization, Software, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Arun Pankajakshan:** Software, Writing – review & editing. **Pierantonio Facco:** Conceptualization, Writing – review & editing, Project administration, Supervision. **Federico Galvanin:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.compchemeng.2023.108353](https://doi.org/10.1016/j.compchemeng.2023.108353).

References

- Asprey, S.P., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. *Comput. Chem. Eng.* 24 (2–7), 1261–1267. [https://doi.org/10.1016/S0098-1354\(00\)00328-8](https://doi.org/10.1016/S0098-1354(00)00328-8).
- Bard, Y., 1974. *Nonlinear Parameter Estimation*. Academic Press, New York, NY.
- Box, M.J., 1968. The occurrence of replications in optimal designs of experiments to estimate parameters in non-linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 30 (2), 290–302. <https://doi.org/10.1111/j.2517-6161.1968.tb00728.x>.
- Buzzi-Ferraris, G., Forzatti, P., 1983. A new sequential experimental design procedure for discriminating among rival models. *Chem. Eng. Sci.* 38 (2), 225–232. [https://doi.org/10.1016/0009-2509\(83\)85004-0](https://doi.org/10.1016/0009-2509(83)85004-0).
- Czitrom, V., 1999. One-factor-at-a-time versus designed experiments. *Am. Stat.* 53 (2), 126–131. <https://doi.org/10.1080/00031305.1999.10474445>.
- Dasgupta, S., Mukhopadhyay, S., Keith, J., 2021. G-optimal grid designs for kriging models. *arXiv*, 2111.06632. [10.48550/arXiv.2111.06632](https://arxiv.org/abs/2111.06632).
- de Prada, C., Pantelides, C.C., Pitarch, J.L., 2019. Special Issue on “Process Modelling and Simulation. *Processes* 7 (8), 511. <https://doi.org/10.3390/pr7080511>.
- Duarte, B., Saraiva, P., Pantelides, C., 2004. Combined Mechanistic and Empirical Modelling. *Int. J. Chem. React. Eng.* 2 (1) <https://doi.org/10.2202/1542-6580.1128>.
- Echtermeyer, A., Amar, Y., Zakrzewski, J.L., 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. <https://doi.org/10.3762/bjoc.13.18>.
- Espie, D., Macchietto, S., 1989. The optimal design of dynamic experiments. *AIChE J.* 35 (2), 223–229. <https://doi.org/10.1002/aic.690350206>.

- Fisher, R.A., 1950. *Contributions to Mathematical Statistics*. John Wiley and Sons, New York.
- Franceschini, G., Macchietto, S., 2008. Novel anticorrelation criteria for model-based experiment design: theory and formulations. *AIChE J.* 54 (4), 1009–1024. <https://doi.org/10.1002/aic.11429>.
- Galvanin, F., Ballan, C.C., Barolo, M., Bezzo, F., 2013. A general model-based design of experiments approach to achieve practical identifiability of pharmacokinetic and pharmacodynamic models. *J. Pharmacokinet. Pharmacodyn.* 40 (4), 451–467. <https://doi.org/10.1007/s10928-013-9321-5>.
- Galvanin, F., Cao, E., Al-Rifai, N., Gavrilidis, A., Dua, V., 2016. A joint model-based experimental design approach for the identification of kinetic models in continuous flow laboratory reactors. *Comput. Chem. Eng.* 95, 202–215. <https://doi.org/10.1016/j.compchemeng.2016.05.009>.
- Galvanin, F., Macchietto, S., Bezzo, F., 2007. Model-based design of parallel experiments. *Ind. Eng. Chem. Res.* 46, 871–882. <https://doi.org/10.1021/ie0611406>.
- Garud, S.S., Karimi, I.A., Kraft, M., 2017. Design of computer experiments: a review. *Comput. Chem. Eng.* 106, 71–95. <https://doi.org/10.1016/j.compchemeng.2017.05.010>.
- Granato, D., de Araújo Calado, V.M., Granato, D., 2013. The use and importance of design of experiments (DOE) in process modelling in food science and technology. D. and Ares, G. (Eds). *Mathematical and Statistical Methods in Food Science and Technology*. John Wiley & Sons, Ltd, pp. 1–18. <https://doi.org/10.1002/9781118434635.ch1>.
- Hunter, W.G., Reiner, A.M., 1965. Designs for discriminating between two rival models. *Technometrics* 7, 307–323. <https://doi.org/10.2307/1266591>.
- Kiefer, J., Wolfowitz, J., 1960. The equivalence of two extremum problems. *Can. J. Math.* 12, 363–366. <https://doi.org/10.4153/CJM-1960-030-4>.
- Kiefer, J., Wolfowitz, J., 1959. Optimum designs in regression problems. *Ann. Math. Stat.* 30, 271–294. <https://doi.org/10.1214/AOMS/1177706252>.
- Killeen, P.R., 2005. An Alternative to Null-Hypothesis Significance Tests. *Psychol. Sci.* 16 (5), 345–353. <https://doi.org/10.1111/j.0956-7976.2005.01538.x>.
- Mihaluta, M., Martin, P., Dantan, J.Y., 2008. Manufacturing process modeling and simulation. ICME, Italy. 7p. fihal-00999479f. <https://hal.archives-ouvertes.fr/hal-00999479>.
- Montgomery, D., 2013. *Design and Analysis of Experiments*. John Wiley & Sons, Hoboken, NJ.
- Prus, M., 2019. Various optimality criteria for the prediction of individual response curves. *Stat. Probab. Lett.* 146, 36–41. <https://doi.org/10.1016/j.spl.2018.10.022>.
- Pukelsheim, F., 1993. *Optimal Design of Experiments*. SIAM, Philadelphia.
- Quaglio, M., Fraga, E.S., Cao, E., Gavrilidis, A., Galvanin, F., 2018. A model-based data mining approach for determining the domain of validity of approximated models. *Chemom. Intell. Lab. Syst.* 172, 58–67. <https://doi.org/10.1016/j.chemolab.2017.11.010>.
- Reichert, I., Olney, P., Lahmer, T., 2021. Combined approach for optimal sensor placement and experimental verification in the context of tower-like structures. *J. Civ. Struct. Health Monit.* 11 (1), 223–234. <https://doi.org/10.1007/s13349-020-00448-7>.
- Shahmohammadi, A., McAuley, K.B., 2019. Sequential model-based A- and V-optimal design of experiments for building fundamental models of pharmaceutical production processes. *Comput. Chem. Eng.* 129, 106504. <https://doi.org/10.1016/j.compchemeng.2019.06.029>.
- Singh, B., Kumar, R., Ahuja, N., 2005. Optimizing drug delivery systems using systematic “design of experiments.” Part I: fundamental aspects. *Crit. Rev. Ther. Drug Carrier Syst.* 22 (1), 27–105. <https://doi.org/10.1615/CritRevTherDrugCarrierSyst.v22.i1.20>.
- Singh, B., Dahiya, M., Saharan, V., Ahuja, N., 2005. Optimizing drug delivery systems using systematic “design of experiments.” Part II: retrospect and prospects. *Crit. Rev. Ther. Drug Carrier Syst.* 22 (3), 215–293. <https://doi.org/10.1615/CritRevTherDrugCarrierSyst.v22.i3.10>.
- Smith, K., 1918. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika* 12 (1/2), 1–85. <https://doi.org/10.2307/2331929>.
- Stigler, S.M., 1971. Optimal experimental design for polynomial regression. *J. Am. Stat. Assoc.* 66 (334), 311–318. <https://doi.org/10.2307/2283928>.
- Vo, A.D.D., Shahmohammadi, A., McAuley, K.B., 2021. Model-based design of experiments for polyether production from bio-based 1,3-propanediol. *AIChE J.* 67, e17394. <https://doi.org/10.1002/aic.17394>.
- Waldron, C., Pankajakshan, A., Quaglio, M., Cao, E., Galvanin, F., Gavrilidis, A., 2019. Closed-loop model-based design of experiments for kinetic model discrimination and parameter estimation: benzoic acid esterification on a heterogeneous catalyst. *Ind. Eng. Chem. Res.* 22165–22177. <https://doi.org/10.1021/acs.iecr.9b04089>.
- 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. *React. Chem. Eng.* 5 (1), 112–123. <https://doi.org/10.1039/c9re00342h>.
- Wong, W.K., 1995. On the equivalence of D and G-optimal designs in heteroscedastic models. *Stat. Probab. Lett.* 25, 317–321. [https://doi.org/10.1016/0167-7152\(94\)00236-1](https://doi.org/10.1016/0167-7152(94)00236-1).
- Zullo, L., 1991. *Computer Aided Design of Experiments. An Engineering Approach*. Ph.D. Thesis, The University of London, U.K.
- Websites <https://doep.readthedocs.io/en/latest/>.