Reaction Chemistry & Engineering

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *React. Chem. Eng.*, 2022, **7**, 2359

Probabilistic framework for optimal experimental campaigns in the presence of operational constraints†

Kennedy Putra Kusumo, [a] Kamal Kuriyan, [a] Shankarraman Vaidyaraman, [b] Salvador García Muñoz, [b] Nilay Shah [a] and Benoît Chachuat *[a]

The predictive capability of any mathematical model is intertwined with the quality of experimental data collected for its calibration. Model-based design of experiments helps compute maximally informative campaigns for model calibration. But in early stages of model development it is crucial to account for model uncertainties to mitigate the risk of uninformative or infeasible experiments. This article presents a new method to design optimal experimental campaigns subject to hard constraints under uncertainty, alongside a tractable computational framework. This computational framework involves two stages, whereby the feasible experimental space is sampled using a probabilistic approach in the first stage, and a continuous-effort optimal experiment design is determined by searching over the sampled feasible space in the second stage. The tractability of this methodology is demonstrated on a case study involving the exothermic esterification of priopionic anhydride with significant risk of thermal runaway during experimentation. An implementation is made freely available based on the Python packages DEUS and Pydex.

Received 19th October 2021,
Accepted 20th July 2022

DOI: 10.1039/d1re00465d

rsc.li/reaction-engineering

1 Introduction

The development of holistic and systematic approaches to tackling modern industrial challenges is nigh synonymous with the construction of high-fidelity mathematical process models. Optimal experiment design (OED) helps accelerate the construction of such models through computing maximally informative experimental campaigns. Two main types of experiment design objectives may be distinguished, each useful in different steps of model construction. When multiple candidate model structures compete, maximally discriminative experiments help determine the most suitable model.[1–6] By contrast, optimal calibration experiments yield maximally precise model parameter estimates.[7–9]

This paper focuses on the latter objective. The term model-based experimental design (MBDoE) was coined to refer to techniques that focus on nonlinear process models and it has found numerous applications in chemical engineering and process systems engineering; see the seminal survey by Franceschini and Macchietto.[10] A particular emphasis in this literature has been on gradient-based approaches to MBDoE and the design of dynamic experiments. By contrast, the concepts of experimental supports and efforts, which are ubiquitous in the statistical experiment design literature,[8,9,11,12] have been somewhat foregone in engineering applications and only recently have effort-based approach to MBDoE been investigated.[13,14]

A primary challenge with nonlinear process models is that optimal campaigns become dependent on the unknown model parameter values that the experiments are meant to help estimate in the first place. The basic technique entails a local design approach, whereby the model is linearized around a nominal model parameter value. The challenge of inappropriate nominal values leading to uninformative experiments can be counteracted to a certain extent with robust experimental designs.[15–22] A second challenge, which remains to be fully addressed, is that of infeasible experiments. To ensure the safety and economic feasibility of experimental runs, many systems must comply with operational constraints. Inappropriate nominal model parameter values exacerbate the risk of designing infeasible experiments under such constraints.

The methodology developed through this paper aims at the construction of experimental campaigns for nonlinear processes subject to hard operational constraints. Both its importance and relevance have been highlighted in the process systems engineering literature. For instance, the

[a] *The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK.*
E-mail: b.chachuat@imperial.ac.uk
[b] *Synthetic Molecule Design and Development, Lilly Research Laboratories, Eli Lilly & Company, Indianapolis, IN 46285, USA*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d1re00465d

position paper by Bonvin *et al.*[23] stresses the need for satisfying safety constraints when applying experimental design techniques to pilot- and full-scale processing plants, or in developing physiological models for personalized health care. To generalize, this problem is of paramount importance whenever data are collected from a constraint-sensitive system. One class of relevant problems is model maintenance experiments, which are run to maintain, update or re-calibrate a model over time; *e.g.*, to capture catalyst degradation,[24,25] fouling,[26,27] aging of heart valves,[28] *etc.* Another class is model addition experiments, which are run to calibrate extra model components introduced to account for previously unaccounted phenomena; *e.g.*, with applications in troubleshooting, retrofitting, or decommissioning.[29–31] These additional data, taken directly from the full-scale processing plant, are required to capture atypical process behaviour before drawing any conclusions.

Precursor work by Galvanin *et al.*[32] involved backing off of the active process constraints in a maximally-informative experiment to ensure its feasibility against process disturbances and modeling uncertainties. Mesbah and Streif[33] formulated a robust OED problem with chance constraints, for which they applied the Cantelli–Chebyshev inequality[34] to obtain a deterministic surrogate of the chance constraints in combination with polynomial chaos expansion for propagating the uncertainty to the process dependent variables. The formulation by Telen *et al.*[35] also maximizes the expected value of the information criterion subject to probabilistic constraints by applying the sigma-point method[36] to approximate the uncertainty propagation. More recently, Petsagkourakis and Galvanin[37] used Gaussian Process (GP) surrogates to establish a safe experimental space and to approximate the expected value and variance of the experimental information criterion.

Due to the computational complexity of considering uncertainty and hard constraints when designing experiments, the aforementioned contributions have focused on the problem of designing a single experimental run, as opposed to experimental campaigns comprising multiple parallel runs. And despite this restricted view, the resulting optimization problems remain computationally challenging, especially for systems with large uncertainty. The standard alphabetic information criteria[38] lead to nonconvex optimization problems for many practically relevant process models, whose solution with gradient-based techniques[39] can result in suboptimality. Numerical failure caused by singular information matrices is also commonplace.

The main objective of this paper is to propose a computational methodology for ensuring feasibility of constrained experimental campaigns under parametric model uncertainty, which relies on the synergistic and tractable combination of sampling and convex optimization techniques, and proceeds in two stages: (i) the discretization of the feasible experimental space using an adaptive sampling method; followed by (ii) the design of feasible experimental campaigns using a convex, continuous-effort OED formulation that searches over the discretized space. Section 2 starts with a formalization of the OED problems subject to process constraints, adopting the Bayesian paradigm to establish the restricted experimental space. An illustrative example is presented that involves a response surface model before the two-stage computational methodology is described in section 3, including numerical results of the illustrative example to highlight key aspects of the solution framework. This is complemented by a simulated case study in section 4 involving the design of a dynamic experimental campaign for the exothermic esterification of propionic anhydride and butan-2-ol, with risks of thermal runaway during experimentation. Another simulated case study for re-calibration of a model of a jacketed stirred tank reactor that is undergoing a doubling in throughput is provided as part of the ESI.†

2 Problem statement and background

Consider a system with n_x experimental controls $\boldsymbol{x} \in \mathscr{X} \subset \mathbb{R}^{n_x}$ and n_y measured responses $\boldsymbol{y} \in \mathbb{R}^{n_y}$,

$$\boldsymbol{y} = \boldsymbol{\eta}(\boldsymbol{\theta}, \boldsymbol{x}) + \varepsilon. \tag{1}$$

The parameters $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{n_\theta}$ in the mathematical model $\boldsymbol{\eta}$ are typically inferred from noisy experimental data using frequentist or Bayesian inference and are thus inherently uncertain. For simplicity, the measurements are assumed to be independent throughout, and the measurement error $\varepsilon \in \mathbb{R}^{n_y}$ is assumed to have zero mean $\mathbb{E}(\varepsilon) = 0$ with uncorrelated and homoscedastic covariance Σ_y. Although it is given in closed form in (1), note that the model $\boldsymbol{\eta}$ may also be defined implicitly *via* a set of nonlinear algebraic or differential equations without the loss of generality.

2.1 Restricted experimental space

Assume that the experiments are subjected to constraints on the controls $\boldsymbol{x}$ and the responses $\boldsymbol{y}$,

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) \leq \boldsymbol{0}. \tag{2}$$

Enforcing these constraints hinges on predictions from the model (1) that is under development. In principle, the model allows for the response y to be abstracted out of the predicted constraints,

$$\boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{x}) := \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\eta}(\boldsymbol{\theta}, \boldsymbol{x})) \leq \boldsymbol{0}. \tag{3}$$

One can define the restricted experimental space in various ways. Ignoring the model uncertainty first, a nominal restricted experimental space is readily defined as

$$\mathscr{R}_0 := \{\boldsymbol{x} \in \mathscr{X} : \boldsymbol{g}(\boldsymbol{\theta}_0, x) \leq \boldsymbol{0}\}, \tag{4}$$

where a nominal value $\boldsymbol{\theta}_0$ needs to be specified *a priori* for the model parameters; for instance, as a maximum likelihood

or a maximum *a posteriori* estimate. But if the uncertainty in $\boldsymbol{\theta}$ is large, such a nominal restricted space could be misleading.[40]

To mitigate the risk of constraint violations, one may adopt the Bayesian paradigm and describe the model parameter uncertainty with a probability distribution $\pi(\boldsymbol{\theta})$; for instance, obtained as the posterior distribution from a Bayesian inference problem with an appropriate likelihood function and prior distribution. In this way, the feasibility probability of the constraints can be predicted as

$$\mathbb{P}[\boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{x}) \leq 0 | \pi(\boldsymbol{\theta})] := \int_{\boldsymbol{\theta}:\boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{x})\leq 0} \pi(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}. \tag{5}$$

This constraint feasibility probability then allows one to define a probabilistic restricted experimental space as

$$\mathscr{R}_\alpha := \{\boldsymbol{x} \in \mathscr{X} : \mathbb{P}[\boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{x}) \leq \boldsymbol{0} | \pi(\boldsymbol{\theta})] \geq \alpha\}. \tag{6}$$

The confidence level $\alpha \in (0,1]$ reflects the level of aversion against the risks of not satisfying the constraints (2), with higher α values corresponding to a higher aversion. A confidence level α close to 1 is typical in problems that may involve industrial process safety or patient safety in health care.

Another definition for the restricted space can adopt a set-based paradigm, which considers a bounded set Θ representing all possible values $\boldsymbol{\theta}$ may take. The set-based restricted experimental space below is the counterpart to the probabilistic set (6),

$$\mathscr{R}_{\boldsymbol{\Theta}} := \{\boldsymbol{x} \in \mathscr{X} : \boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{x}) \leq \boldsymbol{0}, \forall\, \boldsymbol{\theta} \in \Theta\}. \tag{7}$$

In practice, the parameter set Θ may be the result of a set-membership estimation.[41–45] It may as well be chosen as the (linearized) confidence region at a given confidence level in a frequentist estimation, or even as the highest posterior density (HPD) set at probability level α after a Bayesian estimation; see, *e.g.*, Perić *et al.*[44] and Kusumo *et al.*[40] for further details and examples.

Although both provide valid definitions for a restricted experimental space, our focus herein is on the probabilistic restricted space $\mathscr{R}_\alpha$. In particular, the definition of $\mathscr{R}_\alpha$ in (6) assumes that any structural mismatch present in the model $\boldsymbol{\eta}$ is captured through the probability distribution π of the model parameters $\boldsymbol{\theta}$. Note that other methods directly account for the presence of structural model mismatch, for instance by further restricting the experimental space to subdomains where the model is deemed reliable,[46,47] but these are outside the scope of this work.

2.2 Optimal experiment design

An experimental campaign for estimating the model parameters $\boldsymbol{\theta}$ is assumed to comprise a total of $N_\mathrm{t} > 0$ experimental runs. Since such campaigns often consist of repeated runs with identical experimental controls (replicates), it is convenient to denote an experimental design ξ as

$$\xi := \begin{Bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_{N_\mathrm{c}} \\ p_1 & \dots & p_{N_\mathrm{c}} \end{Bmatrix}, \tag{8}$$

where $N_\mathrm{c} \leq N_\mathrm{t}$ is the number of unique runs, and $p_i \in (0, 1]$ refers to the effort (or weight) of the i-th experimental candidate with controls $\boldsymbol{x}_i$ for each $i = 1,\dots, N_\mathrm{c}$. In the context of dynamic experiments, experimental decisions such as sampling times and (parameterized) time-varying controls can be considered as part of $\boldsymbol{x}_i$. The set $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_\mathrm{c}}\}$ is called the support of the experimental design ξ, denoted by $\mathrm{supp}(\xi)$. In an actual campaign a support $\boldsymbol{x}_i$ must not have an effort of zero; that is, a support $\boldsymbol{x}_i \in \mathrm{supp}(\xi)$ is a unique run within ξ with non-zero effort $p_i > 0$ by convention.

If r_i is the number of times $\boldsymbol{x}_i$ is repeated, then the effort $p_i = r_i/N_\mathrm{t}$ is a quantized variable that may only take values of positive multiples of the fraction $1/N_\mathrm{t}$. Designs where the p_i's are quantized are called exact designs and require N_t to be specified. To emphasize this, exact designs are denoted with a subscript ξ_{N_t}. In contrast, a continuous design allows p_i to vary continuously in the standard simplex $\sum_i p_i = 1$ with $p_i > 0$, $\forall i$. Continuous designs are independent of N_t and can be viewed as exact designs with $N_\mathrm{t} \to \infty$. The distinction between continuous and exact thereby vanishes as N_t gets larger.

An optimal design ξ^* is a design that maximizes some scalar information criterion ϕ,

$$\xi^* \in \arg\max_{\xi} \phi(\xi). \tag{9}$$

Searching over an experimental design ξ entails searching over all possible number of supports $N_\mathrm{c} \in \mathbb{Z}^+$, the experimental controls $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_\mathrm{c}} \in \mathbb{R}^{n_x}$, and the corresponding efforts $p_1, \dots, p_{N_\mathrm{c}} \in (0, 1]$. Note that (9) is applicable to both continuous and exact designs.

Herein, the focus is on the problem of computing experimental campaigns which are optimal for model calibration. The information criterion ϕ is often expressed as a function of the Fisher information matrix (FIM), $\mathbf{M} \in \mathbb{R}^{n_\theta \times n_\theta}$ given by[8]

$$\mathbf{M}(\xi, \boldsymbol{\theta}) := \sum_{i=1}^{N_\mathrm{c}} p_i \mathbf{A}(\boldsymbol{x}_i, \boldsymbol{\theta}). \tag{10}$$

Under the assumption of uncorrelated homoscedastic error in (1), the atomic matrix $\mathbf{A}$ of a given experimental support is computed as

$$\mathbf{A}(\boldsymbol{x}_i, \boldsymbol{\theta}) := \mathbf{F}(\boldsymbol{x}_i, \boldsymbol{\theta})^T \Sigma_y^{-1} \mathbf{F}(\boldsymbol{x}_i, \boldsymbol{\theta}) \tag{11}$$

with $\mathbf{F} \in \mathbb{R}^{n_y \times n_\theta}$ the regressor matrix that is specific to the model at hand.

For models that are linear in the parameters $\boldsymbol{\theta}$, the regressor matrix $\mathbf{F}$ becomes independent of $\boldsymbol{\theta}$ and is readily available through rearrangement of the model eqn (1) into its standard form,

$$\boldsymbol{y} = \mathbf{F}(\boldsymbol{x})\boldsymbol{\theta} + \varepsilon. \tag{12}$$

For nonlinear process models the basic local design technique entails a linearization around the nominal parameter value $\boldsymbol{\theta}_0$. The linearized process model is then rearranged into the standard form (12) to recover the regressor matrix,

$$\mathbf{F}(\boldsymbol{x}, \boldsymbol{\theta}_0) := \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{\theta}_0). \tag{13}$$

Designs obtained using the linearized regressor matrix (13) are said to be locally-optimal around $\boldsymbol{\theta}_0$. This local design (LD) approach entails maximizing the following information criterion in problem (9),

$$\phi(\xi) := \Phi(\mathbf{M}(\xi, \boldsymbol{\theta}_0)), \tag{14}$$

with classical choices for the function Φ as reported in Table 1.

Analogous to defining the probabilistic restricted experimental space $\mathscr{R}_\alpha$, the average design (AD) approach makes use of a prior probability distribution $\pi(\boldsymbol{\theta})$ of the uncertain model parameters when designing experiments.[22] It maximizes the expected value of information content over the uncertain scenarios,

$$\phi(\xi) := \int_{\boldsymbol{\theta}\in\Theta} \pi(\boldsymbol{\theta})\Phi(\mathbf{M}(\xi, \boldsymbol{\theta}))\mathrm{d}\boldsymbol{\theta}. \tag{15}$$

This approach, which is also known as the pseudo-Bayesian approach,[48] is our main focus herein.

Notice that robust experimental design techniques such as the AD approach still relies on linearization, as the experimental information at a given $\boldsymbol{\theta}$ is computed through the linearized regressor matrix $\mathbf{F}(\cdot,\boldsymbol{\theta})$ in (13). Other techniques that do not rely on linearization exist, such as the Bayesian experimental design method,[16] but are outside the scope of this work.

2.3 Formal problem statement

Adopting a probabilistic framework, a formal statement of the constrained OED problem is given by

$$\max_{\xi} \ \phi(\xi) \tag{16}$$

$$\text{s.t.} \ \ \mathbb{P}[\boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{x}) \le \mathbf{0} | \pi(\boldsymbol{\theta})] \ge \alpha, \ \forall \boldsymbol{x} \in \mathrm{supp}(\xi). \tag{17}$$

This formulation applies to both continuous and exact designs. Since the N_c constraints (17) are chance constraints, a natural choice for the criterion ϕ in (16) is the AD criterion (15). A local design counterpart to (16) and (17) entails maximizing the LD criterion (14) subject to the nominal constraint $\boldsymbol{g}(\boldsymbol{\theta}_0, \boldsymbol{x}) \le \mathbf{0}$, $\forall \boldsymbol{x} \in \mathrm{supp}(\xi)$; and a maximin design could be formulated similarly.

Depending on the distribution $\pi(\boldsymbol{\theta})$ and the confidence level α, the set $\mathscr{R}_\alpha$ could be empty. This would indicate that the process model is too uncertain, and there is simply not enough information to ascertain a safe experiment at the desired confidence level α using that model. The options available are to lower the confidence level α, modify the experimental setup itself, and/or gather extra information about the system through other means before designing experimental runs.

2.4 Illustrative example

Consider an experimental system with two experimental controls $x_1, x_2 \in \mathscr{X} := [-1, 1]^2$ and a single response y, predicted by the following response-surface model

$$y = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2 + \theta_5 {x_1}^2 + \theta_6 {x_2}^2 + \varepsilon. \tag{18}$$

This model can be written in the standard form (12) with the regressor matrix $\mathbf{F}(\boldsymbol{x}) := [1 \ x_1 \ x_2 \ x_1x_2 \ {x_1}^2 \ {x_2}^2]$ and $\boldsymbol{\theta} \in \mathbb{R}^6$. The measurement error ε is assumed to be Gaussian distributed with zero mean $\mathbb{E}(\varepsilon) = 0$ and homoscedastic variance $\Sigma_y = 1$.

The task at hand is to compute an optimally informative campaign ξ^* to estimate $\boldsymbol{\theta}$ in (18), in the D-optimal sense (Table 1). Since (18) is linear with respect to $\boldsymbol{\theta}$, notice that the LD and AD approaches are equivalent.

A complicating factor is the presence of a safety concern whenever the response y goes outside a known range of $[1.85, 3]$, thereby imposing hard operational constraints. The model (18) is used to predict the response y, allowing for the probabilistic restricted experimental space to be defined as‡

$$\mathscr{R}_\alpha := \{\boldsymbol{x} \in [-1, 1]^2 : \mathbb{P}[1.85 \le \mathbf{F}(\boldsymbol{x})\boldsymbol{\theta} \le 3.00 | \mathcal{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)] \ge \alpha\}, \tag{19}$$

which assumes a normally-distributed model uncertainty with mean $\boldsymbol{\mu}_\theta := [2 \ 1 \ 1 \ 1 \ 2 \ 2]^T$ and covariance $\Sigma_\theta = 0.05 \ \boldsymbol{I}_6$; and where the confidence level is taken as $\alpha = 0.85$.

The resulting constrained OED problem is given by

$$\max_{\xi} \ \log\det\left(\sum_{i=1}^{N_c} p_i \, \mathbf{F}\left(x^i\right)^T \Sigma_y^{-1} \, \mathbf{F}\left(x^i\right)\right) \tag{20}$$

Table 1 Classical information criteria Φ for model calibration. These are given both in standard form and in eigen-form, with λ_k the k-th eigenvalue of the FIM $\mathbf{M}$

Optimality criterion	$\Phi(\mathbf{M})$	
	Standard form	Eigen-form
Determinant	$\log(\det(\mathbf{M}))$	$\log(\Pi_k \lambda_k)$
Average or tr**A**ce	$-\mathrm{Tr}(\mathbf{M}^{-1})$	$-\sum_k \lambda_k^{-1}$
Extreme	$\lambda_{\min}(\mathbf{M})$	$\min_k \lambda_k$

‡ An explicit expression of the constraint feasibility probability is given by $\mathbb{P}[y^L \le \mathbf{F}(\boldsymbol{x})\boldsymbol{\theta} \le y^U | \mathcal{N}(\boldsymbol{\mu}_\theta, \Sigma_\theta)] = \frac{1}{2}\left[\mathrm{erf}\left(\frac{y^U - \mathbf{F}(\boldsymbol{x})\boldsymbol{\mu}_\theta}{\sqrt{2\mathbf{F}(\boldsymbol{x})\Sigma_\theta\mathbf{F}(\boldsymbol{x})^T}}\right) - \mathrm{erf}\left(\frac{y^L - \mathbf{F}(\boldsymbol{x})\boldsymbol{\mu}_\theta}{\sqrt{2\mathbf{F}(\boldsymbol{x})\Sigma_\theta\mathbf{F}(\boldsymbol{x})^T}}\right)\right]$.

$$\text{s.t. erf}\left(\frac{3-\mathbf{F}(\boldsymbol{x}^i)\mu_\theta}{\sqrt{2\,\mathbf{F}(\boldsymbol{x}^i)\Sigma_\theta\,\mathbf{F}(\boldsymbol{x}^i)^{\mathrm{T}}}}\right)-\text{erf}\left(\frac{1.85-\mathbf{F}(\boldsymbol{x}^i)\mu_\theta}{\sqrt{2\,\mathbf{F}(\boldsymbol{x}^i)\Sigma_\theta\,\mathbf{F}(\boldsymbol{x}^i)^{\mathrm{T}}}}\right)\geq 2\alpha,\ \forall i \tag{21}$$

with $\{\boldsymbol{x}^1,\ldots,\boldsymbol{x}^{N_c}\}:=\text{supp}(\xi)$. Although the model is linear in the uncertain parameters, computing a globally optimal solution to the constrained OED problem (20) and (21) is challenging. And this complexity will be exacerbated in the case of more complex, possibly implicit and dynamic models as well as higher-dimensional experimental spaces. In response to this, the following section introduces a computationally-tractable methodology, based on sampling and convex optimization for approximating the solution to the general problem (16) and (17).

3 Computational methodology

We propose a computational methodology that decomposes the constrained OED problem (16) and (17) into two stages. A simplified information flow of these two stages within the overall model calibration cycle is illustrated in Fig. 1. A prerequisite is that the experimenter has determined a suitable process model and selected a probability distribution $\pi(\boldsymbol{\theta})$ that reflects their belief on the unknown parameter values $\boldsymbol{\theta}$.

Leveraging the process model and $\pi(\boldsymbol{\theta})$, the first stage focuses on establishing the probabilistic restricted experimental space $\mathscr{R}_\alpha$ at a given reliability value α. This is achieved through a suitable sampling algorithm that generates a sample of N_s experimental points that satisfy the probabilistic constraints (17),

$$\mathscr{S}_\alpha := \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{N_s}\}\subseteq\mathscr{R}_\alpha, \tag{22}$$

or terminates with an indication that $\mathscr{R}_\alpha$ is indeed empty. One drawback is the loss of continuum within the restricted experimental space, which may lead to sub-optimally informative campaigns. As with any sampling methodology, another drawback is attributed to the curse of dimensionality with the number of experimental controls n_x.

The key component for an efficient stage 1, therefore, is the sampling algorithm that is used to generate the experimental samples $\mathscr{S}_\alpha$. To mitigate the loss of continuum and the chances that large parts of $\mathscr{R}_\alpha$ are missed by the discretization, the N_s samples in $\mathscr{S}_\alpha$ should be as uniformly distributed as possible inside $\mathscr{R}_\alpha$. An efficient sampler will furthermore have a high acceptance rate, in order to minimize the total time spent on drawing samples outside of $\mathscr{R}_\alpha$, and be highly parallelizable to enable reduction in wall-clock time when needed.

The samples $\mathscr{S}_\alpha$ are passed on to the second stage, where the focus is on designing the experimental campaign within $\mathscr{R}_\alpha$, under the crucial assumption that $\mathscr{R}_\alpha$ is sufficiently well described by $\mathscr{S}_\alpha$. This stage comprises 3 steps:

(i) The sensitivity analysis step involves computation of the atomic information matrices $\mathbf{A}(\boldsymbol{x}_i,\cdot)$ of all samples in $\mathscr{S}_\alpha$, and thus the corresponding regressor matrices $\mathbf{F}(\boldsymbol{x}_i,\cdot)$;

(ii) The optimization step solves a convex optimization problem to determine an optimal continuous experimental design ξ^* as a set of optimal efforts p_i^* for each sample $x_i\in\mathscr{S}_\alpha$; and

(iii) The rounding step apportions the fractional efforts p_i^* to non-negative multiples of $1/N_t$, given the total number N_t of experimental runs.

In practice, the rounded efforts would be used to run the experimental campaign. The experimenter would get to choose the frequency of model parameter updates and whether runs are conducted sequentially or in parallel. The optional updates are depicted with the dotted lines connecting the experimental campaign and the model parameter uncertainty in Fig. 1. In principle, as soon as a single measurement is taken, the probability distribution $\pi(\boldsymbol{\theta})$ could be updated, *e.g.*, through Bayesian inference. This would update the restricted space $\mathscr{R}_\alpha$ in turn, and thus the experimental design. Any parameter update in a nonlinear model will affect stage 2 regardless of whether or not stage 1 was affected because the atomic information matrix $\mathbf{A}$ is

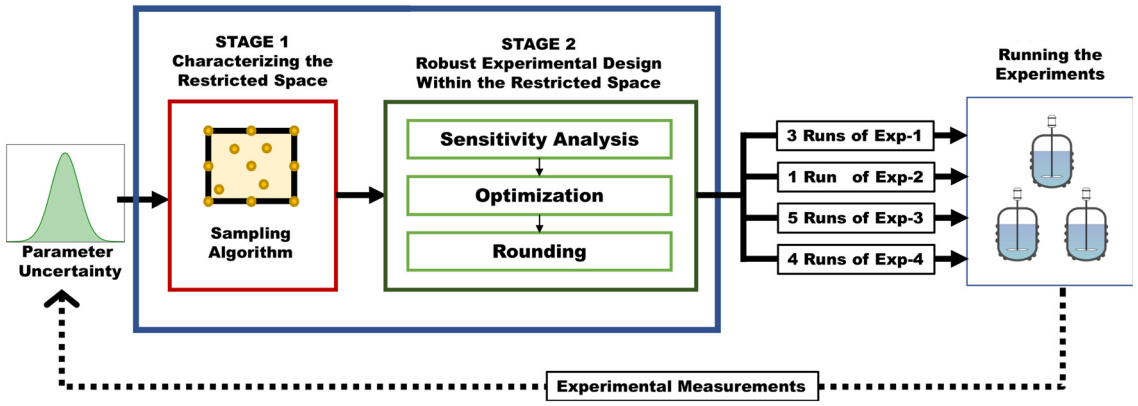

Fig. 1 Illustration of the two-step computational methodology and information flow within the model calibration cycle.

dependent on $\boldsymbol{\theta}$. In other words, a different experimental campaign design may be obtained even when $\mathscr{R}_\alpha$ is unaffected by the parameter update. In the case of a linear model, by contrast, the atomic information matrix $\mathbf{A}$ is independent of $\boldsymbol{\theta}$. A different experimental campaign, therefore, may only be obtained when the experimental samples $\mathscr{S}_\alpha$ have changed.

An ideal model calibration cycle involves conducting parameter updates as frequently as possible thus minimizing the chances of uninformative or infeasible experimental campaigns. However, this frequency is limited by the computational cost for completing an iteration of parameter update and experimental design within the restricted space, on top of actually running the physical experiments. In practice, a termination condition for this iteration could entail checking if the variances in model parameters or predicted responses are within acceptable levels.

Following an outline of the methodology and a discussion on how it fits into the overall model calibration cycle, we describe and discuss both stages in more detail next.

3.1 Characterization of the restricted experimental space

The first stage of the methodology involves drawing N_s samples of the experimental controls to represent the probabilistic restricted experimental space $\mathscr{R}_\alpha$ defined in (6). We consider an extension of nested sampling (NS) for feasible space characterization, as described in Kusumo *et al.*[40] and its improvements in Kusumo *et al.*[49] An excellent feature of NS for our methodology is the ability to specify a minimal number of samples to draw from $\mathscr{R}_\alpha$ as stopping criterion. This allows for easier control over the complexity of the following solution steps. Furthermore, NS is especially designed to uniformly draw samples from a target set with a high acceptance rate, which meets the desiderata of an ideal sampler for stage 1 as discussed before. Only a brief account of NS is given next for the sake of brevity, while interested readers are referred to previous publications[40,49] for further details on this approach.

NS for feasible space characterization[40] is an adaptive sampling technique that evolves a set of live points through regions of increasing feasibility probability, until a stopping criterion is met. The number of live points is chosen as N_s here. The sampling starts with a uniform sample of these N_s live points within the unrestricted experimental space $\mathscr{X}$. The feasibility probability of each live point is approximated by discretizing the probability distribution $\pi(\boldsymbol{\theta})$ into a set $\mathscr{S}_\pi$ comprising N_π parameter scenarios θ_j,

$$\hat{\mathbb{P}}[g(\theta,x) \le 0 \mid \pi(\theta)] \approx \frac{1}{N_\pi} \sum_{\theta \in \mathscr{S}_\pi} \mathbb{1}[g(\theta,x)], \tag{23}$$

where $\mathbb{1}[\cdot]$ is the indicator function with $\mathbb{1}[x] = 1$ when $x_k \le 0, \forall k$ and $\mathbb{1}[x] = 0$ otherwise. Since one can always transform a set of scenarios with unequal weights into an equally weighted one, we shall assume that samples are equally-weighted thereafter. To arrive at a truly Bayesian approach $\mathscr{S}_\pi$ may be chosen as the sampled posterior returned by a Bayesian estimation algorithm on some initial experimental data using an appropriate prior. Nevertheless, NS is applicable regardless of how $\mathscr{S}_\pi$ was generated.

During the NS iterations, the live points are replaced by new points. N_p proposals are generated per iteration, and their feasibility probabilities are again estimated using (23). Each proposal replaces the current live point with the lowest feasibility probability if they have a higher feasibility probability, otherwise they are rejected. Rejected replacements are called ghost points, and replaced live points are called dead points. The effectiveness of NS hinges on the chosen strategy to propose these replacements. The strategy used here samples within an enlarged ellipsoid around the current live points, as first proposed by Mukherjee *et al.*[50] The enlargement factor of the ellipsoids decreases over the iterations according to two user-defined tuning parameters, the initial enlargement factor and the shrinking rate of this factor. Replacements are done one at a time within a single iteration, following the order that the points are generated. It is possible, therefore, that a new live point is replaced by another one within the same iteration. At the end of an iteration, two stopping criteria are checked: the primary criterion checks if all live points have feasibility probabilities higher than the target reliability value α; a secondary criterion checks the maximum number of iterations, without which the sampling technique could never terminate when $\mathscr{R}_\alpha$ is empty. When the primary stopping criterion is met, the final collection of live points are (at least) N_s samples from $\mathscr{R}_\alpha$. The follow-up work by Kusumo *et al.*[49] implements a vectorized function evaluation and two enhancements of this algorithm: (i) increasing the number of live points as the algorithm progresses closer to the target reliability; and (ii) warm-starting the initial live points by conducting a nominal search first.

3.2 Robust experimental design within the restricted space

The second stage of the methodology aims to conduct an optimal experimental design within the sampled restricted space $\mathscr{S}_\alpha$ obtained from stage 1 and proceeds in three steps, which are further discussed below. This dependence of stage 2 on stage 1 highlights the importance of an accurate characterization of the restricted space, and in turn the need for selecting an accurate probability distribution π. It is worth re-emphasizing the benefits of adopting a Bayesian approach in this context, where π corresponds to the posterior distribution from an inference problem.

3.2.1 Sensitivity analysis. This step computes the atomic information matrices $\mathbf{A}$ in (11) for each experiment candidate sample in $\mathscr{S}_\alpha$ passed on from stage 1, and possibly for each uncertainty realization in $\mathscr{S}_\pi$ in the AD approach (15). These information matrices are computed out of the sensitivities of the response model $\boldsymbol{\eta}$ with respect to the model parameters $\boldsymbol{\theta}$, as given in (13). In principle, the analysis can start as soon as a first sample is drawn from the restricted space $\mathscr{R}_\alpha$.

However, for ease of conveying computational complexity and run-times, we do not use this enhancement herein. The atomic information matrices are furthermore completely independent from one another, allowing them to be computed entirely in parallel.

For process models that are linear with respect to the model parameters and given in the standard form (12), computation of each atomic information matrix (11) is cheap, requiring a single matrix multiplication. For nonlinear models available in explicit form (1), one can apply either symbolic differentiation or automatic differentiation[51,52] to compute the linearized regressor matrices (13). For models given in implicit form, such as systems of differential-algebraic equations (DAEs), a finite-difference scheme may be used; for instance, simple forward or central differentiation or Richardson extrapolation[53,54] for higher accuracy. Although readily implemented, this approach can become computationally prohibitive for large process models. Instead, methods and software for sensitivity analysis of DAE models may be used. They have reached a stage of maturity where both forward and adjoint sensitivities can be computed reliably and efficiently.[55–58] They rely on numerical integration techniques tailored to the special structure of the sensitivity system, such as staggered or simultaneous corrector methods for forward-in-time sensitivity integration and state interpolation and check-pointing for backward-in-time adjoint integration, and can be conveniently combined with automatic differentiation tools.[52]

3.2.2 Effort-based optimization. The second step determines a maximally-informative experimental campaign out of the sample set $\mathscr{S}_\alpha$ passed on from stage 1. The optimization problem searches over the experimental efforts p_i associated with each experimental point $\boldsymbol{x}_i \in \mathscr{S}_\alpha$ and its construction relies on the atomic matrices computed in step 2.1. For instance, the local design (LD) problem can be defined as

$$\left(p_1^*, \ldots, p_{N_s}^*\right) \in \arg\max_p \Phi\left(\sum_{i=1}^{N_s} p_i \mathbf{A}(\boldsymbol{x}_i, \boldsymbol{\theta}_0)\right) \tag{24}$$

$$\text{s.t.} \quad \sum_{i=1}^{N_s} p_i = 1, \;\; p_i \geq 0 \,\forall i, \tag{25}$$

for a nominal parameter value $\boldsymbol{\theta}_0$. Likewise, the average design (AD) problem is given by

$$\left(p_1^*, \ldots p_{N_s}^*\right) \in \arg\max_p \frac{1}{N_\pi} \sum_{\boldsymbol{\theta} \in \mathscr{S}_\pi} \Phi\left(\sum_{i=1}^{N_s} p_i \mathbf{A}(\boldsymbol{x}_i, \boldsymbol{\theta})\right) \tag{26}$$

$$\text{s.t.} \quad \sum_{i=1}^{N_s} p_i = 1, \quad p_i \geq 0 \,\forall i, \tag{27}$$

where the AD criterion (15) is discretized over the sampled uncertainty set $\mathscr{S}_\pi$.

Both formulations take advantage of the available discretization $\mathscr{S}_\alpha$ of the restricted experimental space to reduce the constrained, typically nonconvex, optimization problem (16) and (17) into a more tractable optimization problem. In particular, the chance constraints (17) are always satisfied by construction (up to the error introduced by the uncertainty discretization $\mathscr{S}_\pi$). Both cost functions (24) and (26) are furthermore convex whenever the selected criterion Φ is concave—a property shared by the classical A, D and E criteria listed in Table 1. Convexity eliminates potential issues with local optima during optimization and implies quicker solutions when modern convex optimization solvers are used. Another feature is that, while (24)–(27) scale linearly with the number N_s of samples in the restricted experimental space $\mathscr{R}_\alpha$, they are independent of the total number of experimental runs N_t, thereby making the formulation tractable for designing large experimental campaigns.

An important difference with the general OED problem (16) and (17) is that the optimal efforts $\left(p_1^*, \ldots, p_{N_s}^*\right)$ obtained from (24) and (25) or (26) and (27) are allowed to take a value of zero. But the support of the optimal design ξ^* is readily recovered after solving the latter, as the collection of experimental samples in $\mathscr{S}_\alpha$ with a nonzero effort,

$$\text{supp}(\xi^*) := \left\{\boldsymbol{x}_j \in \mathscr{S}_\alpha : p_j^* > 0\right\}. \tag{28}$$

Regarding initialization, an effective and generic initial guess that seeks to avoid a singular FIM consists of allocating equal efforts to all candidates, $p_i = 1/N_s$, $\forall_i$. Note that this approach is akin to conducting a practical identifiability (or estimability) analysis[59] with exactly one repetition of all N_s experimental candidates: obtaining a regular FIM with these initial efforts provides a confirmation that the model is practically identifiable, while a singular FIM suggests a lack of practical identifiability with the current discretization of the experimental space. In this latter scenario, either fixing the values of certain model parameters or adding more experiment candidates may help recover practical identifiability.

3.2.3 Effort rounding. In general, the optimal efforts computed by solving the continuous design problem (26) and (27) will not be exact fractional repetitions of the experimental candidates for the desired number N_t of experimental runs. A rounding procedure must, therefore, be applied to convert the continuous design into an exact design ξ_{N_t} for the desired N_t. A characteristic of many such rounding procedures is that the rounded designs keep the support points of the continuous designs which may lead to sub-optimal campaigns. It is also clear that any rounding causes a loss of information compared to the optimal continuous-effort design.

The focus herein is on the efficient rounding method, defined in terms of the following maximum likelihood ratio.[60] Given two experimental designs (whether exact or continuous) ξ^1 and ξ^2 such that $\text{supp}(\xi^1) \subseteq \text{supp}(\xi^2)$ and with respective efforts p^1 and p^2, the maximum likelihood ratio of ξ^1 relative to ξ^2 is computed as

Table 2 D-Optimal campaign for the response-surface model (18). The experimental supports together with their continuous efforts are reported in the left section. The right section reports the rounded experimental efforts using the efficient and greatest effort rounding procedures for varying number of runs in the experimental campaign. The bottom two rows report the minimum likelihood ratio of the rounded designs and the actual efficiency to indicate rounding quality for different experimental campaigns

			Total number of runs, N_t								
x_1	x_2	p^{*a} (%)	6	7	8	9	10	12	14	20	30
−0.485	−0.437	1.1	—	—	—	—	—	—	1	1	1
0.129	−0.524	1.4	—	—	—	—	—	1	1	1	1
0.138	−0.841	8.5	1	1	1	1	1	1	1	2	2
−0.113	−0.624	0.6	—	—	—	—	—	—	1	1	1
−0.282	−0.823	8.3	—	1	1	1	1	1	1	1	2
−0.764	0.311	12.9	1	1	1	1	1	1	1	2	3
−0.849	−0.192	5.4	—	—	—	—	1	1	1	1	2
0.481	−0.186	12.4	1	1	1	1	1	1	1	2	3
−0.719	−0.486	9.8	1	1	1	1	1	1	1	2	3
−0.296	0.497	6.0	—	—	—	1	1	1	1	1	2
−0.401	0.188	4.7	—	—	—	—	—	1	1	1	2
0.414	−0.645	6.8	—	—	1	1	1	1	1	1	2
0.092	−0.029	11.1	1	1	1	1	1	1	1	2	3
0.042	−0.388	11.0	1	1	1	1	1	1	1	2	3
Efficiency bound (%)			—	—	—	—	—	—	50.24	70.34	81.72
Actual efficiency (%)			98.6	95.7	95.6	95.6	93.5	95.2	93.3	98.3	99.8

[a] Experiment candidates with efforts p_i^* below a minimum threshold of 10^{-4} are reassigned a zero effort and the remaining non-zero efforts are re-normalized so they add up to one.

$$\varepsilon_{\xi^1/\xi^2} := \min_{i \in \mathrm{supp}(\xi^2)} \left(\frac{p_i^1}{p_i^2} \right). \tag{29}$$

Under mild assumptions,[60] it can be shown that the ratio $\varepsilon_{\xi^1/\xi^2}$ provides the following bound,

$$\phi(\xi^1) \geq \varepsilon_{\xi^1/\xi^2} \phi(\xi^2), \tag{30}$$

for a convex information criterion ϕ. A complementary metric is the actual rounding efficiency κ given by

$$\kappa_{\xi^1/\xi^2} := \frac{\phi(\xi^1)}{\phi(\xi^2)}, \tag{31}$$

where an efficiency of 0.5 means that running the experimental campaign ξ^1 twice would yield an equivalent amount of information as running ξ^2 a single time.

When used to indicate rounding quality, ξ^1 and ξ^2 take the role of the rounded design ξ_{N_t} and the continuous design ξ^*, respectively. The efficient rounding method guarantees that the rounded design ξ_{N_t} has a criterion value that is at least $\varepsilon_{\xi_{N_t}/\xi^*}$ times the criterion value of the optimal continuous design ξ^*. Another property of this method is that an efficiently rounded design for $N_t + 1$ runs corresponds to the rounded design ξ_{N_t} for N_t runs with one extra run assigned to one of the support points of ξ^*. In that sense, ξ_{N_t} is necessarily a subset of ξ_{N_t+1}. Regarding the actual rounding efficiency, the closer $\kappa_{\xi_{N_t}/\xi^*}$ to 1 the closer the information content of the rounded design ξ_{N_t+1} to that of the optimal continuous design ξ^*. But the actual rounding efficiency remains a relative measure of the information content, not an absolute one. Although a rounded campaign ξ_{N_t+1} will always contain more information than ξ_{N_t} when using the efficient rounding method, nothing can be said as to the relative magnitude of $\kappa_{\xi_{N_t}/\xi^*}$ compared to $\kappa_{\xi_{N_t+1}/\xi^*}$.

A limitation of the efficient rounding method is that it may only be meaningful when the continuous design ξ^* has a number of supports $N_c \leq N_t$; otherwise, $\varepsilon_{\xi_{N_t}/\xi^*}$ takes a value of zero. When the number of supports $N_c > N_t$, we may apply the greatest effort rounding procedure in place of the efficient rounding, which involves assigning an equal effort $1/N_t$ to the N_t number of supports in ξ^* with the largest efforts and an effort of zero to the rest.

3.3 Illustrative example (continued)

We revisit the illustrative example introduced in section 2.4 and apply the two stage solution framework to approximate the solution of (20) and (21). Two key parameters that dictate the overall computational complexity of the methodology as well as the fidelity of the computed experimental design are: (i) the number of point samples N_s to be drawn from the probabilistic restricted experimental space $\mathscr{R}_{0.85}$; and (ii) the number of Monte Carlo scenarios N_π to discretize the model uncertainty $\mathcal{N}(\mu_\theta, \Sigma_\theta)$. A simple rule-of-thumb is to start small for both parameters and gradually increase them. A discussion of this process for the illustrative example is reported in the ESI.† The computational statistics with the settings described below are reported in Table 3.

The first stage generates a set of experimental control points from within $\mathscr{R}_{0.85}$. Running DEUS with $N_\pi = 1000$ uncertainty scenarios and drawing $N_p = 10$ proposals at each nested sampling iteration leads to a sampled restricted space $\mathscr{S}_\alpha$ with $N_s = 3885$ points from $\mathscr{R}_{0.85}$, illustrated as the blue

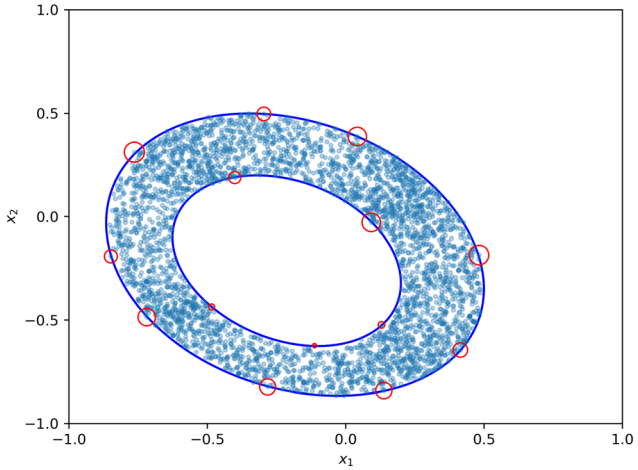

Fig. 2 Continuous D-optimal design for the response-surface model (18) using 3885 experimental candidates drawn from the restricted experimental space $1.85 \leq y \leq 3$ at reliability level 0.85. Translucent blue circles represent the drawn candidates. The red hollow circles represent the effort allocated to the experimental candidates, their radii being proportional to the corresponding optimal efforts. The solid blue line illustrates the boundaries of the analytical feasible region at reliability level 0.85.

circles in Fig. 2. Believing in the model (18), therefore, suggests that $\mathscr{R}_{0.85}$ takes the shape of a 2-dimensional torus. This is a consequence of the model (18) describing a paraboloid and the process constraints $1.85 \leq y \leq 3$ cutting this paraboloid at two different levels. It is also worth re-emphasizing that although the model is linear in the parameters $\boldsymbol{\theta}$, the corresponding parametric uncertainty still affects the restricted space *via* the chance constraints (21).

The second stage computes the optimal experimental design within the restricted space $\mathscr{R}_{0.85}$, further divided into three steps. The first step computes the 6-by-6 atomic information matrices $\mathbf{A}$ for all N_s experimental samples generated in stage 1, according to (11). Given the simplicity of the model, the regressor matrices $\mathbf{F}$ were provided analytically.

The second step of stage 2 determines the optimal efforts of each experimental candidate in order to maximize the D-optimal criterion, by solving the optimization problem (26) and (27). The resulting D-optimal efforts are reported in Table 2 and represented as red hollow circles in Fig. 2. It is worth reemphasizing that no explicit constraint is put on the efforts to limit the number of non-zero efforts in the optimal solution that is recovered as the support set (28). This is a consequence of Caratheodory's theorem, which when applied to experimental designs states that any optimal information matrix $\mathbf{M}$ can always be written as a convex combination of at most $d = n_\theta(n_\theta + 1)/2 = 21$ atomic information matrices; see theorems 2.1 and 2.2 of Fedorov and Leonov.[9] The constrained D-optimal campaign comprises 14 support points and shows a general pattern where all support points lie on the inner and outer boundaries defined by the constraints $y \leq 3$ and $y \geq 1.85$, respectively. A larger share of the total experimental effort is allocated to the outer boundary with 9 supports, while the remaining 5 supports lie on the inner boundary.

The third step of stage 2 apportions the continuous design to the desired number of experimental runs N_t. Because the response-surface model (18) has 6 parameters, the theoretical minimum for an invertible FIM is also $N_t = 6$ experimental runs. Moreover, the continuous D-optimal campaign (Table 2) has 14 support points so the efficient rounding method[60] cannot be applied if $N_t < 14$, in which case the greatest effort rounding method is applied. Table 2 summarizes the rounded D-optimal designs for between 6–30 runs. The bottom two rows of Table 2 report the efficiency bound and the actual efficiency. Overall, the actual efficiencies $\kappa_{\xi_{N_t}/\xi^*}$ are consistently above 90%, even for a low number of runs N_t. This indicates that the exact design ξ_{N_t} retains a majority of the information of the optimal continuous campaign ξ^*, despite the low values of corresponding efficiency bounds $\varepsilon_{\xi_{N_t}/\xi^*}$. Notice also that $\kappa_{\xi_{12}/\xi^*} < \kappa_{\xi_6/\xi^*}$ and $\kappa_{\xi_{14}/\xi^*} < \kappa_{\xi_7/\xi^*}$, an indication that the rounding scheme applied is suboptimal for $N_t = 12$ or 14. Indeed, more

Table 3 Computational statistics for the illustrative example (section 3.3) and case study (section 4). Sampling of the probabilistic restricted experimental space $\mathscr{R}_\alpha$ in stage 1 was conducted using the Python code DEUS.[40,49] Optimal experimental campaigns within the sampled restricted space $\mathscr{S}_\alpha$ as part of stage 2 were computed using the Python code Pydex. The solver MOSEK[61] interfaced through CVXPY[62] was used to solve the convex continuous-effort optimization problems. For the illustrative example, the atomic matrices needed to construct these optimization problems relied on analytical sensitivities passed by the user. For the case study, the DAE model was implemented using Pyomo[63] and Pyomo.DAE[64] and their sensitivity analysis was automated with the solver IDAS part of the SUNDIALS suite[65] coupled with the automatic differentiation tool CasADi.[52] The effort rounding procedure in Pydex implements the efficient rounding method[60] and defaults to the greatest effort rounding procedure when the continuous design has more supports than the total number of experimental runs in the campaign. Links to both Python code DEUS and Pydex can be found in the ESI†

	Sampling of experimental space		Sensitivity analysis			Optimization	
Case	CPU model	Time (hh:mm:ss)	CPU model	# cores	Time (hh:mm:ss)	CPU model	Time (hh:mm:ss)
Illustrative example	AMD Ryzen 2600x	00:00:53	AMD Ryzen 2600x	1	00:00:11	AMD Ryzen 2600x	00:00:15
Restricted, average	Intel Xeon E5-2667 v4	01:34:36	Intel Xeon E5-2697 v2	24	00:08:31	Intel Xeon E5-2667 v4	03:42:23
Restricted, local	Intel Xeon E5-2667 v4	01:34:36	Intel Xeon E5-2667 v4	1	00:00:46	Intel Xeon E5-2667 v4	00:03:44
Unrestricted, local	AMD Ryzen 2600x	00:00:01	Intel Xeon E5-2667 v4	1	00:00:44	Intel Xeon E5-2667 v4	00:02:52

informative exact designs for ξ_{12} and ξ_{14} could be obtained by doubling the efforts in ξ_6 and ξ_7, respectively. The same does not occur for either N_t = 20 or 30 since $\kappa_{\xi_{20}/\xi^*} > \kappa_{\xi_{10}/\xi^*}$ and $\kappa_{\xi_{30}/\xi^*} > \kappa_{\xi_{10}/\xi^*}$.

4 Case study: esterification of propionic anhydride

Consider the esterification reaction between butan-2-ol and propionic anhydride, following the reaction

$$\underset{\text{A:propionic anhydride}}{(CH_3CH_2CO)_2O} + \underset{\text{B:butan-2-ol}}{CH_3CHOHC_2H_5} \xrightarrow{H_2SO_4} \underset{\text{C:butyl propionate}}{C_2H_5CO_2(CH_3)CHC_2H_5} + \underset{\text{D:propionic acid}}{C_2H_5CO_2H}. \quad (32)$$

Since this reaction is known to be (moderately) exothermic,[66,67] experiments are conducted on a bench-scale calorimeter comprised of a coil-jacketed semi-batch reactor fitted with an internal cooling coil as part of a temperature controller system. The temperature control system is capable of maintaining isothermal reaction conditions under normal operation, but there is risk of thermal runaway during experimentation if this control system were to fail, that needs to be carefully managed.[68]

Experiments start with an initial charge of the anhydride (A), containing N_A^0 moles with volume V_A^0 at the given reaction conditions. The alcohol (B) is fed into the reactor at a time-varying volumetric feed rate u (L s^{-1}) and concentration C_B^{in} (mol L^{-1}). The anhydride is acidified prior to the initial charge, containing 0.8% of sulfuric acid (relative to the mass of A). Both the initial charge of A and the feed of B are pre-heated to the set reaction temperature before starting the experiments. The esterification reaction is initiated by introducing the alcohol into the reactor and runs for a given batch time τ.

The candidate mechanistic model of the calorimeter posits that the reaction is first-order in each reactant,[69]

$$\frac{d\xi_A(t)}{dt} = \frac{r(t)V(t)}{N_A^0} \quad \text{with } \xi_A(0) = 0 \quad (33)$$

$$\frac{dV(t)}{dt} = u(t) \quad \text{with } V(0) = V_A^0 \quad (34)$$

$$r(t) = k^\circ \exp\left(\frac{-E_a}{RT}\right)\left[\frac{N_A(t)}{V(t)}\right]^\alpha\left[\frac{N_B(t)}{V(t)}\right]^\beta \quad (35)$$

$$N_A(t) = N_A^0[1 - \xi_A(t)] \quad (36)$$

$$N_B(t) = C_B^{in}[V(t) - V_A^0] - N_A^0\xi_A(t) \quad (37)$$

$$\dot{Q}_r(t) = (-\Delta H_r)r(t)V(t), \quad (38)$$

where t (s) denotes time, ξ_A (–) the molar conversion of A, V (L) the volume of reaction mixture, N_A and N_B (mol) the respective amounts of A and B, r (mol L^{-1} s^{-1}) the reaction rate, $\dot{Q}_r$ (J s^{-1}) the thermal power converted from the chemical reaction, T (K) the reaction temperature, ΔH_r (J mol^{-1}) the molar enthalpy of reaction, k° (L, mol, s) the pre-exponential factor, E_a (J mol^{-1}) the activation energy, α and β the reaction orders for A and B, respectively, and R = 8.314 J K^{-1} mol^{-1} the universal gas constant.

To quantify the risk of a thermal runaway should the temperature control system fail, a safety constraint is imposed on the cooling failure temperature T_{cf}—also known as the maximum temperature of synthesis reaction (MTSR). This constraint is expressed in terms of the amount of accumulated non-converted reactant at failure of the cooling system, in order to ensure that a maximum cooling failure temperature T^{max} = 405 K is never exceeded even under adiabatic conditions,

$$T_{cf}(\tau) := T + \min\{N_A(\tau), N_B(\tau)\}\frac{-\Delta H_r}{\rho c_p V(\tau)} \le T^{max} \quad (39)$$

with ΔH_r = 62.5 × 10^3 J mol^{-1} the heat of reaction, ρ = 900 g L^{-1} the density of the reaction mixture, and c_p = 2 J g^{-1} K^{-1} the specific heat capacity of the reaction mixture.

The goal of the case study is to design maximally-informative experiments using the bench-scale calorimeter to precisely estimate the thermokinetic parameters E_a, k°, α, and β while guaranteeing a safe operation. The prior knowledge on the thermokinetic parameters assumes independent and Gaussian distributed uncertainties, given by $E_a \sim \mathcal{N}(8.25 \pm 0.41) \times 10^4$ J mol^{-1}, and $k^\circ \sim \mathcal{N}(9.72 \pm 2.92) \times 10^7$ (L, mol, s). These values are based on the model calibration results reported in Ubrich *et al.*,[69] herein with a 2- and 4-fold increase in the uncertainty of E_a and k°, respectively, to add to the conservatism. Additionally, discrete uncertainties are introduced on both the reaction orders α and β, which assume a 75% probability of taking a value of 1 and a 25% probability of taking a value of 2.

The calorimeter and the installed spectroscopic probes provide measurements, at regular 50 minute intervals, of the thermal power $\dot{Q}_r$ and the molar conversion ξ_A.[67] The measurements corresponding to each output are assumed to be independently and identically distributed and follow Gaussian distributions with a standard deviation of 0.01 for the conversion ξ_A, and 1 J s^{-1} for the thermal power $\dot{Q}_r$.

The reaction temperature T and feedrate $u(t)$ are selected as the experimental controls. In particular, the time-varying feedrate $u(t)$ is parameterized into a piecewise-constant profile on 4 stages,

$$u(t) = \begin{cases} u_1, & 0 \le t < 0.25\tau \\ u_2, & 0.25\tau \le t < 0.50\tau \\ u_3, & 0.50\tau \le t < 0.75\tau \\ u_4, & 0.75\tau \le t < \tau. \end{cases} \quad (40)$$

After consideration of the physical limits of the experimental setup, the experimental controls are bounded as $T \in [338, 348]$ (K) and $u_1, u_2, u_3, u_4 \in [0.00, 2.80] \times 10^{-5}$ (L s^{-1}). Other

potential experimental factors are fixed: initial amount of A, N_A^0 = 2.50 mol; initial volume of A, V_A^0 = 0.32 L; feed concentration of B, c_B^{in} = 10.87 mol L^{-1}; and overall batch time, τ = 30 000 s (8 h 20 min).

4.1 Restricted D-optimal average design

We apply the computational methodology of section 3 to design an optimal experimental campaign, subject to meeting the safety constraint with 95% confidence. In other words, we only permit running experiments for which there is less than a 5% chance that temperature within the reactor will exceed safe levels in the event of a cooling system failure.

Stage 1. We apply the nested sampling algorithm (section 3.1) to draw $N_s \geq 1000$ samples from the restricted experimental space at α = 0.95 reliability level, using the default tuning parameters. The probability distributions of E_a, k°, α, and β are jointly discretized using N_π = 100 Monte Carlo scenarios, a threshold obtained by gradually increasing N_π to a point where no significant difference can be observed between the experimental designs.

The resulting samples are presented in a corner plot in Fig. 3. Of the 1793 candidate experiments drawn by nested sampling, 1080 experiments belong to the target restricted experimental space at 95% reliability (red points), another 599 experiments have a feasibility probability in the range 80–95% (yellow points), 91 a feasibility probability in the range 70–80% (blue points), and the remaining 23 a feasibility probability below 70% (magenta points), with the lowest feasibility probability being 44%. Clearly, a much

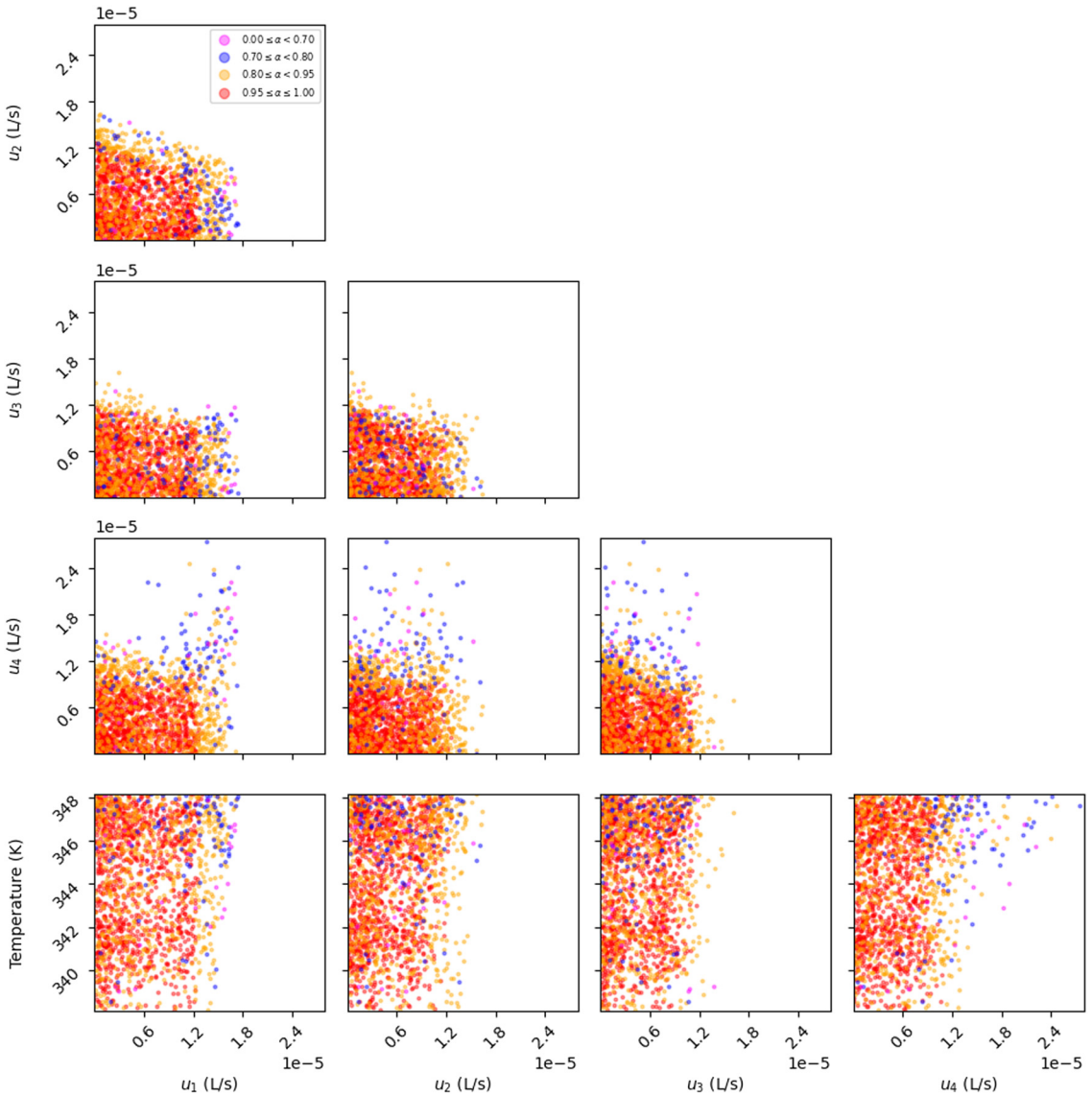

Fig. 3 Corner plot showing all possible 2D projections of the 1793 experiment candidates drawn from the 5-dimensional restricted experimental space. Red points: ≥95% feasibility probability. Yellow points: 80–95% feasibility probability. Blue points: 70–80% feasibility probability. Magenta points: <70% feasibility probability. The overlap between samples from different groups is a visual artifact due to the projection.

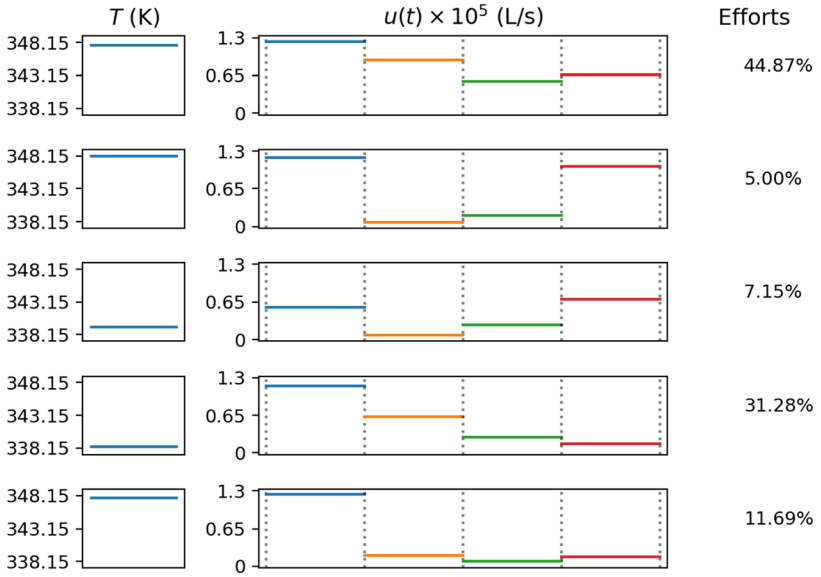

Fig. 4 Experimental supports and efforts in the D-optimal average design for the esterification case study. The time-invariant temperature T is shown as the left-most column. The time-varying alcohol feedrate $u(t)$ is shown in the middle column as a piecewise-constant profile with colours denoting the time periods. Dotted gray lines indicate the switching times in the piecewise-constant feedrate $u(t)$. The optimal efforts are reported in the right-most column.

larger number of samples are drawn inside or near the target level (c. 60% acceptance rate), thereby confirming the suitability of nested sampling in the proposed methodology.

Despite the significant overlap between various coloured samples, meaningful insights can be drawn from Fig. 3. First of all, notice that a large part of the experiment design space is empty (top 3 rows and columns), particularly areas of high alcohol feed rate throughout the batch where it would be overly risky to operate in the event of a cooling system failure. This aligns with the intuition that a lower alcohol feedrate lowers the risk of thermal runaway. Consequently, the red samples occupy a smaller subspace in the lower feedrate range, $u(t) \leq 1.2 \times 10^{-5}$ L s^{-1}. Regarding the effect of temperature (bottom row), the cluster of red samples are relatively unaffected by the reaction temperature. This indicates that within the relatively narrow temperature range considered, the reaction temperature plays a smaller role than the alcohol feedrate on the risk of thermal runaway. Notice also that a majority of the red samples are found within the higher temperature range, which may seem counter-intuitive: a higher reaction temperature means a faster rate of reaction and a lower amount N_A throughout the batch, thus decreasing the second term of the MTSR to an extend that overpowers the effect of the increase in reaction temperature T, and ultimately permitting a wider range of feed rates to satisfy the safety constraint (39).

The yellow samples occupy a similar, albeit larger, subspace than the red samples. Intuitively, if a lower reliability level were selected experiments with a larger alcohol feedrate would be accepted, leading to more

Table 4 D-Optimal average design for the esterification case study. The experimental supports together with their continuous efforts are reported in the left section. The right section reports the rounded experimental efforts using the efficient and greatest effort rounding procedures for varying number of runs in the experimental campaign. The bottom two rows report the maximum likelihood ratio of the rounded designs and the actual efficiency to indicate rounding quality for different experimental campaigns

Temperature	Feed rate $u(t) \times 10^5$ (L s^{-1})				Effort	Total number of runs, N_t				
T (K)	u_1	u_2	u_3	u_4	p^{*a} (%)	2	3	4	5	10
347.67	1.23	0.91	0.54	0.66	44.9	1	1	1	1	4
348.14	1.19	0.07	0.19	1.04	5.0	—	—	—	1	1
339.33	0.55	0.07	0.26	0.69	7.1	—	—	1	1	1
338.34	1.16	0.63	0.27	0.16	31.3	1	1	1	1	3
347.85	1.24	0.19	0.09	0.17	11.7	—	1	1	1	1
					Efficiency bound (%)	—	—	—	44.6	85.5
					Actual efficiency (%)	89.2	98.1	91.1	87.8	99.5

[a] Experiment candidates with efforts p_i^* below a minimum threshold of 10^{-4} are reassigned a zero effort and the remaining non-zero efforts are re-normalized so they add up to one.

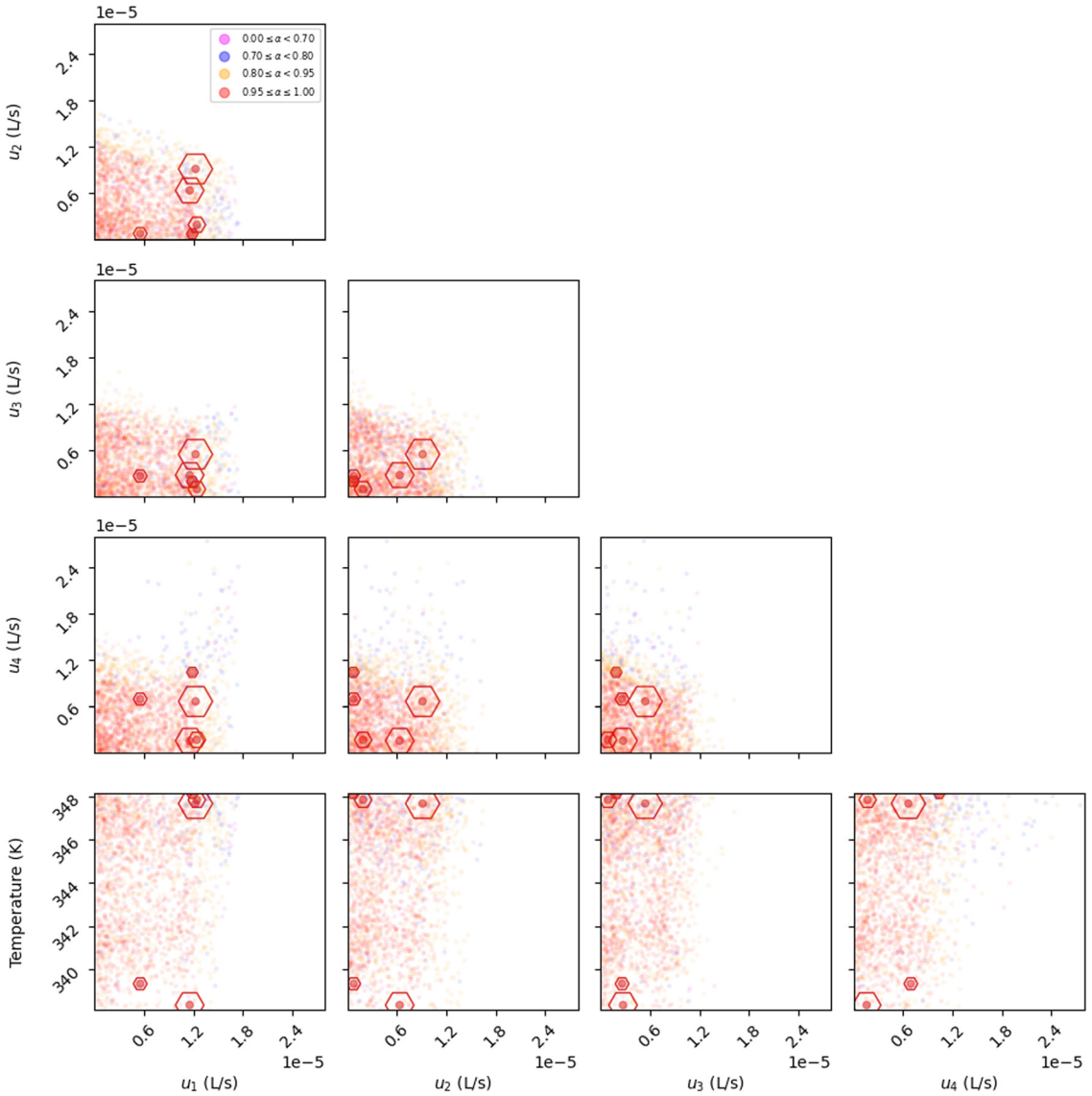

Fig. 5 Corner plot showing the restricted D-optimal average design in 2D projections. The 1793 experiment candidates drawn from the 5-dimensional restricted experimental space are made translucent to indicate boundaries of the restricted space. The opaque red points are the experimental supports, with hollow hexagonal markers around them indicating the corresponding amount of experimental effort.

informative, yet also riskier, experimental campaigns. Despite their small numbers, the blue and magenta samples indicate even riskier combinations of feedrates and temperatures. Located away from the main cluster of red points, they depict a clear separation between safe combinations of low alcohol feedrates and unsafe ones with overly large feedrates.

Stage 2. Following discretization of the restricted experimental space, we compute a D-optimal average design within that space. The optimal continuous-effort campaign is tabulated in Table 4, alongside the efficiency bounds and actual efficiencies after rounding the efforts for different numbers of runs N_t, which is necessary before implementing a campaign in practice. To make it more convenient to interpret and communicate Fig. 4 illustrates this optimal continuous-effort campaign graphically, whereby rows correspond to the supports of the campaign and columns to the different experimental controls. On top of these, Fig. 5 presents a corner plot showing the same restricted space as in Fig. 3, now with translucent points except for the ones that make up the supports of the maximal average design. Around these support points are hollow hexagonal markers that denote the corresponding optimal efforts, using larger hexagons to indicate larger efforts.

Overall, the D-optimal average design comprises five supports. The temperature subplots (bottom row of Fig. 5) show that the supports correspond to low and high temperature experiments only, that is, no experiments with mid-range temperatures. The optimal experimental effort is split amongst the high and low temperature experiments, with slightly larger total (c. 62%) for the high temperature experiments. The u_1 subplots (first column) show that four of the supports have high u_1

values among the restricted experimental space. The u_2 subplots (second column) show that three supports have low u_2 values, and medium to high u_2 values for the other two. By contrast, the u_3 and u_4 subplots (third and fourth columns) show a wider range of values, with a less discernible pattern in the maximal average D-optimal design as compared to u_1 and u_2. Four of the five supports are visually at, or near the boundary of the restricted space from the projections in Fig. 5. This is confirmed by Fig. S7,† which presents the cooling failure temperature T_{cf} over time for each support.

From Table 4, the actual efficiencies obtained with the rounding scheme (section 3.2.3) are found to be acceptable, even though the efficiency bounds do not provide any guarantee when the total number of runs N_t is lower than the number supports $N_s = 5$. It is worth re-emphasizing that we apply the greatest effort rounding for N_t that are less than N_s, and the efficient rounding otherwise. As expected, the actual efficiencies approach 100% when a large number of runs N_t are considered. Interestingly, a relatively high actual efficiency is obtained for $N_t = 3$, as compared to the others except $N_t = 10$, with a rounded design that has 3 supports and equal effort distribution. However, one should keep in mind that the actual efficiencies do not provide any information on the relative amount of predicted information between the rounded design and the optimal exact design.

4.2 Discussions

The computational statistics reported in Table 3 show that the effort-based optimization using MOSEK (interfaced through CVXPY) as part of stage 2 is the most time-consuming step, at nearly 4 h of wall-time. It is worth reiterating that since the formulation (26) is convex, MOSEK can provide a numerical certificate that the computed experimental design is indeed globally optimal within the given discretized space. The time needed to sample the restricted experimental space (stage 1), although 2.5-fold smaller than the optimization time, is also significant. The sensitivity analysis (stage 2) is must faster in comparison, but this is largely due to parallelization on a 24-core workstation —running this analysis in serial on a single core would take between 1–1.5 h of wall-time, which is commensurate with the experimental space sampling step. These statistics suggest that computing a restricted maximal-average design could become prohibitively costly for problems with more experimental controls, more uncertainty scenarios, or more complex models. Moreover, memory requirement could also become limiting. In the present case study, it is indeed the large memory requirement for solving the optimization problem using MOSEK that prevented the solution of larger problems.

In experimental design problems where the optimization becomes intractable, one way of lowering both computational time and memory requirement is to consider a local design instead of an average design in stage 2. While this ignores the effect of uncertainty on the predicted information content, the crucial effect of uncertainty on the restricted experimental space is still accounted for as part of stage 1. As reported in Table 3, opting for a local design cuts the optimization wall-time by 60 fold, down from about 4 h to 4 min. Notice that the cut in sensitivity analysis wall-time (from 9 min to 1 min) is significantly less because this step was parallelized for the average design using 24 cores while conducted in serial for the local design.

The consequence of a local design is of course a higher risk for designing a sub-optimal or uninformative experiment campaign. The results of the restricted D-optimal local design are shown in Fig. S8 and S9 and Table S1 of the ESI.† They present large similarities with the restricted D-optimal average design (Fig. 4 and 5, Table 4), implying that a local design may not cause significant information loss in the present case study under the prior probability distribution used for the uncertain model parameters.

A rather radical simplification to further reduce the computational time would be to ignore the operational constraints altogether, and conduct the experiment design in the unrestricted space instead. The hope is that the resulting optimal experiments would still satisfy the operational constraints with a high enough confidence—the feasibility probability of the experiment being computed *a posteriori*—but there is naturally no way that this can be guaranteed. In the present case study, this approach entails designing the experiments within the whole range of experimental factors values permitted, $T \in [338, 348]$ (K) and $u_1, u_2, u_3, u_4 \in [0.80, 2.80] \times 10^{-5}$ (L s^{-1}). Doing so essentially eliminates the time required to sample the experimental space, so an unrestricted local design can now be computed within minutes of wall-time. However, opting for an unrestricted local design leads to all the experimental supports having feasibility probabilities lower than 70%, which is considered to be unsafe—the results of this unrestricted D-optimal local design are shown in Fig. S10 and S11 and Table S2 of the ESI.† A possible compromise could be to consider a nominal restricted space in stage 1, as defined in (4), where the effect of the model uncertainty is ignored. To increase the likelihood of satisfying the desired reliability target, one could then choose the nominal parameter values $\boldsymbol{\theta}_0$ conservatively. For instance, in the present case study, one could recognize that higher reaction rates are more likely to cause the system to violate the MTSR constraint and adjust the chosen $\boldsymbol{\theta}_0$ accordingly.

5 Conclusions

Modeling uncertainty poses a unique challenge in optimal experiment design, especially in early stages of process development. This paper highlights the importance of accounting for model uncertainty during optimal experimental design in the presence of operational

constraints. The main contribution of the paper is a computational methodology that decomposes such constrained experimental design problems under uncertainty into a sequence of two tractable stages, whereby the first stage generates a finite sample of the restricted experimental space at a given reliability level and the second stage determines a maximally-informative experimental campaign within this sampled space so as to certify feasibility. Key advantages include a convex optimization problem in the second stage and the ability to handle non-convex chance constraints seamlessly at the cost of discretizing the experimental space in the first stage.

Applications of the methodology in a simple response-surface model, then in a more challenging dynamic reaction system described by differential-algebraic equations, demonstrate its suitability and tractability in designing robust experimental campaigns. In particular, the latter depicts a realistic experimental design problem with safety constraints and model uncertainty, that we believe to be currently out of reach for state-of-the-art model-based design of experiment techniques. Although discretization brings advantages, some key challenges with this solution approach also seem to stem from it. In problems with a large number of experimental degrees of freedom, a prohibitively large number of samples may be required to retain a faithful enough representation of the experimental space. Of course, a silver lining is that discretization allows one to start with a small number of samples and gradually refine the covering of the experimental space whenever needed and possible. Herein, we utilize a tailored sampling approach adapted from nested sampling in Bayesian inference to efficiently and uniformly sample the feasible experimental space at a desired reliability level. This sampling method furthermore enables control over the number of feasible experimental candidates generated, thereby controlling the complexity of the subsequent steps as well. Parallel computation of the atomic information matrices is also demonstrated to be effective in significantly cutting computational times and enabling the solution of larger problems. Although not shown, both stages of the methodology could furthermore be conducted in parallel rather than sequentially in order to further expedite the solution.

We envision a few important follow-ups to further improve the methodology. Investigations to enhance computational tractability need to be explored. For instance, the use of the sigma points could help propagate uncertainty more efficiently than through standard Monte Carlo simulation. Deployment of surrogate modeling techniques is another promising approach to mitigate computational intractability with expensive process models, while for tackling problems with more experimental controls recursive schemes that either adapt or augment the set of sampling points recursively could be devised. Lastly, analyzing the trade-offs between experiments conducted sequentially or in parallel, especially in the presence of operational constraints, delineates a promising research area.

Author contributions

Kennedy Putra Kusumo: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing – original draft, writing – review & editing. Kamal Kuriyan: data curation, formal analysis, investigation, software, validation, writing – review & editing. Shankarraman Vaidyaraman: conceptualization, validation, writing – review & editing. Salvador García Muñoz: conceptualization, funding acquisition, supervision, validation, writing – review & editing. Nilay Shah: conceptualization, funding acquisition, resources, supervision, validation, writing – review & editing. Benoît Chachuat: conceptualization, formal analysis, funding acquisition, project administration, resources, supervision, validation, writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work is co-funded by Eli Lilly & Company through the Pharmaceutical Systems Engineering Lab (PharmaSEL) and by the Engineering and Physical Sciences Research Council (EPSRC) as part of its Prosperity Partnership Programme under grant EP/T518207/1.

Notes and references

1 W. G. Hunter and A. M. Reiner, *Technometrics*, 1965, **7**, 307–323.
2 G. E. P. Box and W. J. Hill, *Technometrics*, 1967, **9**, 57–71.
3 P. D. H. Hill, *Technometrics*, 1978, **20**, 15–21.
4 D. Espie and S. Macchietto, *AIChE J.*, 1989, **35**, 223–229.
5 M. Schwaab, F. M. Silva, C. A. Queipo, A. G. Barreto, M. Nele and J. C. Pinto, *Chem. Eng. Sci.*, 2006, **61**, 5791–5806.
6 F. Galvanin, E. Cao, N. Al-Rifai, A. Gavriilidis and V. Dua, *Comput. Chem. Eng.*, 2016, **95**, 202–215.
7 F. Pukelsheim, *Optimal Design of Experiments*, Society for Industrial & Applied Mathematics, USA, 2006, vol. 50.
8 A. C. Atkinson, A. N. Donev and R. Tobias, *Optimum experimental designs, with SAS*, Oxford University Press, 2007.
9 V. Fedorov and S. L. Leonov, *Optimal Design for Nonlinear Response Models*, CRC Press, 2013, p. 2014.
10 G. Franceschini and S. Macchietto, *Chem. Eng. Sci.*, 2008, **63**, 4846–4872.
11 R. Harman and L. Pronzato, *Stat. Probab. Lett.*, 2007, **77**, 90–94.
12 Y. De Castro, F. Gamboa, D. Henrion, R. Hess and J. B. Lasserre, *Ann. Stat.*, 2019, **47**, 127–155.
13 C. Vanaret, P. Seufert, J. Schwientek, G. Karpov, G. Ryzhakov, I. Oseledets, N. Asprion and M. Bortz, *Comput. Chem. Eng.*, 2021, **146**, 107218.
14 K. P. Kusumo, K. Kuriyan, S. García-Muñoz, N. Shah and B. Chachuat, *Comput.-Aided Chem. Eng.*, 2021, **50**, 867–873.
15 L. Pronzato and E. Walter, *Math. Biosci.*, 1988, **89**, 161–176.

16 K. Chaloner and I. Verdinelli, *Stat. Sci.*, 1995, **10**, 273–304.
17 S. Asprey, S. Macchietto and C. Pantelides, *IFAC Proceedings Volumes*, 2000, **33**, 845–850.
18 S. Asprey and S. Macchietto, *J. Process Control*, 2002, **12**, 545–556.
19 S. Körkel, E. Kostina, H. G. Bock and J. P. Schlöder, *Optim. Methods Softw.*, 2004, **19**, 327–338.
20 C. R. Rojas, J. S. Welsh, G. C. Goodwin and A. Feuer, *Automatica*, 2007, **43**, 993–1008.
21 D. Telen, F. Logist, E. V. Derlinden and J. F. V. Impe, *IFAC Proceedings Volumes*, 2012, **45**, 689–694.
22 K. P. Kusumo, K. Kuriyan, S. Vaidyaraman, S. García-Muñoz, N. Shah and B. Chachuat, *Comput. Chem. Eng.*, 2022, **159**, 107680.
23 D. Bonvin, C. Georgakis, C. C. Pantelides, M. Barolo, M. A. Grover, D. Rodrigues, R. Schneider and D. Dochain, *Ind. Eng. Chem. Res.*, 2016, 6891–6903.
24 T. Jahnke, G. A. Futter, A. Baricci, C. Rabissi and A. Casalegno, *J. Electrochem. Soc.*, 2019, **167**, 013523.
25 E. Quiroga, J. Moltó, J. A. Conesa, M. F. Valero and M. Cobo, *Catalysts*, 2020, **10**, 508.
26 E. Diaz-Bejarano, F. Coletti and S. Macchietto, *Heat Transfer Eng.*, 2017, **38**, 681–693.
27 M. Indumathy, S. Sobana and R. C. Panda, *Appl. Therm. Eng.*, 2021, **189**, 116674.
28 E. J. Weinberg, F. J. Schoen and M. R. K. Mofrad, *PLoS One*, 2009, **4**, 1–10.
29 A. A. Alsanousie, O. A. Elsamni, A. E. Attia and M. Elhelw, *Energy*, 2021, **223**, 120079.
30 Y. Wang, R. Smith and J.-K. Kim, *Appl. Therm. Eng.*, 2012, **43**, 7–13.
31 M. J. Kaiser and S. Narra, *Energy*, 2018, **163**, 1150–1177.
32 F. Galvanin, M. Barolo, F. Bezzo and S. Macchietto, *AIChE J.*, 2010, **56**, 2088–2102.
33 A. Mesbah and S. Streif, *IFAC-PapersOnLine*, 2015, **48**, 100–105.
34 A. W. Marshall, I. Olkin and B. C. Arnold, *Inequalities: theory of majorization and its applications*, Springer, 1979, vol. 143.
35 D. Telen, D. Vercammen, F. Logist and J. Van Impe, *Comput. Chem. Eng.*, 2014, **71**, 415–425.
36 S. Julier and J. K. Uhlmann, *A General Method for Approximating Nonlinear Transformations of Probability Distributions, technical report*, 1996.
37 P. Petsagkourakis and F. Galvanin, *Comput. Chem. Eng.*, 2021, **151**, 107339.
38 J. Kiefer, *Ann. Stat.*, 1974, **2**, 849–879.
39 L. T. Biegler, *Nonlinear Programming – Concepts, Algorithms, and Applications to Chemical Processes*, MOS-SIAM Series on Optimization, 2010.
40 K. P. Kusumo, L. Gomoescu, R. Paulen, S. Garciá Munõz, C. C. Pantelides, N. Shah and B. Chachuat, *Ind. Eng. Chem. Res.*, 2020, **59**, 2396–2408.
41 L. Jaulin and E. Walter, *Automatica*, 1993, **29**, 1053–1064.
42 A. R. Gottu Mukkula and R. Paulen, *Comput. Chem. Eng.*, 2017, **99**, 198–213.
43 A. Pankajakshan, M. Quaglio and F. Galvanin, *Comput.-Aided Chem. Eng.*, 2018, 355–360.
44 N. D. Perić, R. Paulen, M. E. Villanueva and B. Chachuat, *J. Process Control*, 2018, **70**, 80–95.
45 R. Paulen, L. Gomoescu and B. Chachuat, *IFAC-PapersOnLine*, 2020, **53**, 7228–7233.
46 M. Quaglio, E. S. Fraga and F. Galvanin, *Chem. Eng. Res. Des.*, 2018, **136**, 129–143.
47 M. Quaglio, E. S. Fraga and F. Galvanin, *IFAC-PapersOnLine*, 2018, **51**, 515–520.
48 E. G. Ryan, C. C. Drovandi and A. N. Pettitt, *Entropy*, 2015, **17**, 1063–1089.
49 K. P. Kusumo, L. Gomoescu, R. Paulen, S. García-Muñoz, C. C. Pantelides, N. Shah and B. Chachuat, *Comput.-Aided Chem. Eng.*, 2020, **48**, 1957–1962.
50 P. Mukherjee, D. Parkinson and A. R. Liddle, *Astrophys. J., Lett.*, 2006, **638**, L51.
51 A. Griewank and A. Walther, *Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation*, SIAM, Philadelphia, 2nd edn, 2008.
52 J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings and M. Diehl, *Math. Program. Comput.*, 2019, **11**, 1–36.
53 L. F. Richardson and R. T. Glazebrook, *Philos. Trans. R. Soc., A*, 1911, **210**, 307–357.
54 L. F. Richardson and J. A. Gaunt, *Philos. Trans. R. Soc., A*, 1927, **226**, 299–361.
55 T. Maly and L. R. Petzold, *Appl. Numer. Math.*, 1996, **20**, 57–79.
56 W. F. Feehery, J. E. Tolsma and P. I. Barton, *Appl. Numer. Math.*, 1997, **25**, 41–54.
57 Y. Cao, S. Li and L. R. Petzold, *J. Comput. Appl. Math.*, 2002, **149**, 171–191.
58 A. C. Hindmarsh, P. N. Brown, S. L. Grant, K. E. Lee, R. Serban, D. E. Shumaker and C. S. Woodward, *ACM Trans. Math. Softw.*, 2005, **31**, 363–396.
59 K. Z. Yao, B. M. Shaw, B. Kou, K. B. McAuley and D. W. Bacon, *Polym. React. Eng.*, 2003, **11**, 563–588.
60 F. Pukelsheim and S. Rieder, *Biometrika*, 1992, **79**, 763–770.
61 E. D. Andersen and K. D. Andersen, *High performance optimization*, Springer, 2000, pp. 197–232.
62 S. Diamond and S. Boyd, *J. Mach. Learn. Res.*, 2016, **17**, 1–5.
63 W. E. Hart, J. P. Watson and D. L. Woodruff, *Math. Program. Comput.*, 2011, **3**, 219–260.
64 B. Nicholson, J. D. Siirola, J.-P. Watson, V. M. Zavala and L. T. Biegler, *Math. Program. Comput.*, 2018, **10**, 187–223.
65 A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker and C. S. Woodward, *ACM Trans. Math. Softw.*, 2005, **31**, 363–396.
66 H. Hernández, J. M. Zaldívar and C. Barcons, *Comput. Chem. Eng.*, 1993, **17**, S45–S50.
67 F. Strozzi, *PhD thesis*, Universiteit Twente, 1997.
68 T. J. Snee, C. Barcons, H. Hernández and J. M. Zaldívar, *J. Therm. Anal.*, 1992, **38**, 2729–2747.
69 O. Ubrich, B. Srinivasan, P. Lerena, D. Bonvin and F. Stoessel, *J. Loss Prev. Process Ind.*, 1999, **12**, 485–493.