Electronic Supplementary Material (ESI) for Reaction Chemistry & Engineering.

ELECTRONIC SUPPLEMENTARY INFORMATION

Probabilistic Framework for Optimal Experimental Campaigns in the Presence of Operational Constraints

Kennedy Putra Kusumo[1], Kamal Kuriyan[1], Shankarraman Vaidyaraman[2], Salvador García Muñoz[2], Nilay Shah[1], and Benoît Chachuat*[1]

[1]The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK
[2]Synthetic Molecule Design and Development, Lilly Research Laboratories, Eli Lilly & Company, Indianapolis, IN 46285, USA

*Corresponding Author: b.chachuat@imperial.ac.uk

1 Pydex

The Python package Pydex is available from: `https://github.com/omega-icl/pydex`. Files related to the case studies covered in this article can also be retrieved from this link.

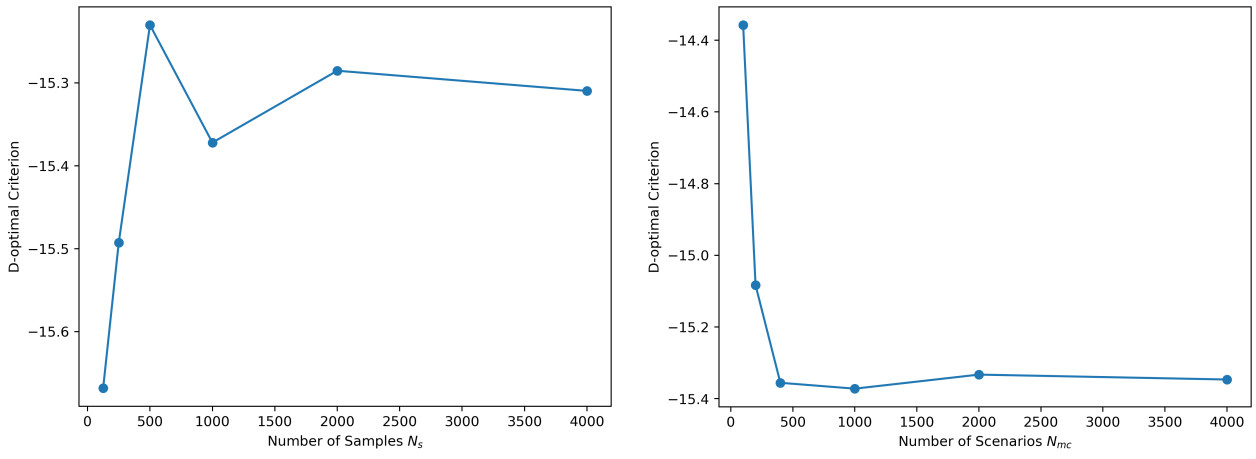
2 DEUS

The Python package DEUS is available from: `https://github.com/omega-icl/deus`. Files related to the case studies covered in this article can also be retrieved from this link.

3 More on the Response Surface Model Example

We analyze the effect of varying number of samples N_s and number of scenarios N_π using the response-surface model (RSM) example (cf. Section 3.4). The idea is to start with small numbers of feasible experiments N_s and model uncertainty realizations N_π and increasing them gradually and separately. We begin by fixing the number of uncertainty realizations $N_\pi = 1000$ and investigate the effects of changing N_s. The RSM example is solved six times using varying number of samples $N_s \in \{125, 250, 500, 1000, 2000, 4000\}$, recording the D-criterion value each time. Following this, the number of samples is fixed at $N_s = 1000$ and the RSM example is solved six times using varying number of scenarios $N_\pi \in \{100, 200, 400, 1000, 2000, 4000\}$. The recorded criterion values are presented in Figure S1, and the control plots showing experimental samples drawn by DEUS and their D-optimal experimental efforts are given in Figure S2.

It is expected that, as the number of samples N_s is increased, the fidelity of the experimental space improves and so does the criterion values. Rather unexpectedly, Figure S1a shows that increasing the number of drawn samples does not guarantee an improvement in the criterion value. We identified that this is because drawing samples using the nested sampling (NS) algorithm in DEUS separately for each case does not yield the following desirable property: When two sets of experimental samples $\mathscr{S}_\alpha^1$ and $\mathscr{S}_\alpha^2$ are drawn independently, with $N_s^1 < N_s^2$ corresponding number of samples, an ideal sampling method shall ensure that $\mathscr{S}_x^1 \subset \mathscr{S}_x^2$. The effect of not having this recursive property can be seen by comparing any two subplots in Figure S2, for instance in Figures S2a & S2b. Notice how certain samples are present in Figure S2a but not in Figure S2b. This implies that fidelity of the experimental space is not guaranteed to improve as the number of samples is increased. Because the RSM example features a linear model in the parameters θ, increasing the number of Monte Carlo scenarios N_π is expected to affect the analytical restricted space $\mathscr{R}_{0.85}$, thereby affecting the drawn samples, and ultimately the criterion value. Figure S1b shows the recorded criterion values as a function of the number of scenarios. As the number of scenarios N_π increases, the optimal criterion values seem to stabilize, but still fluctuate. Upon inspection of the control plots in Figure S3, it is apparent that the issue with lacking the progressive property is still in place, and is likely to affect the conclusion significantly.

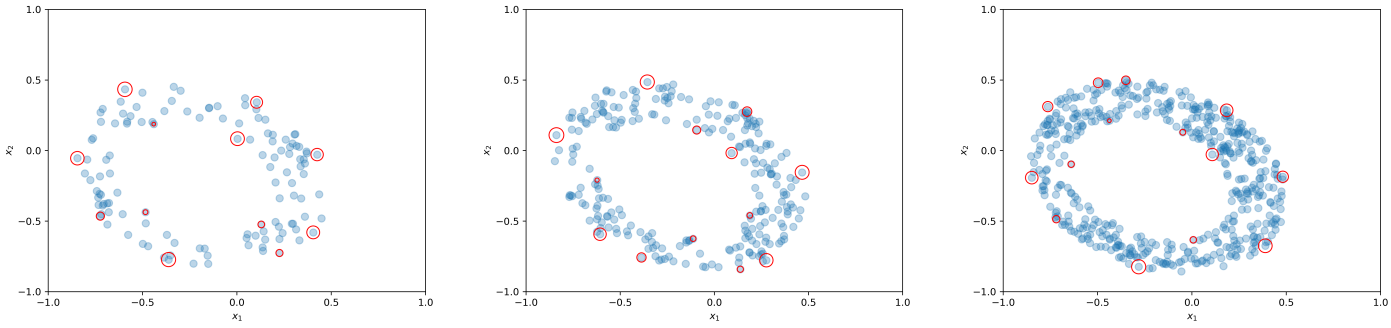


(a) Varying number of samples N_s

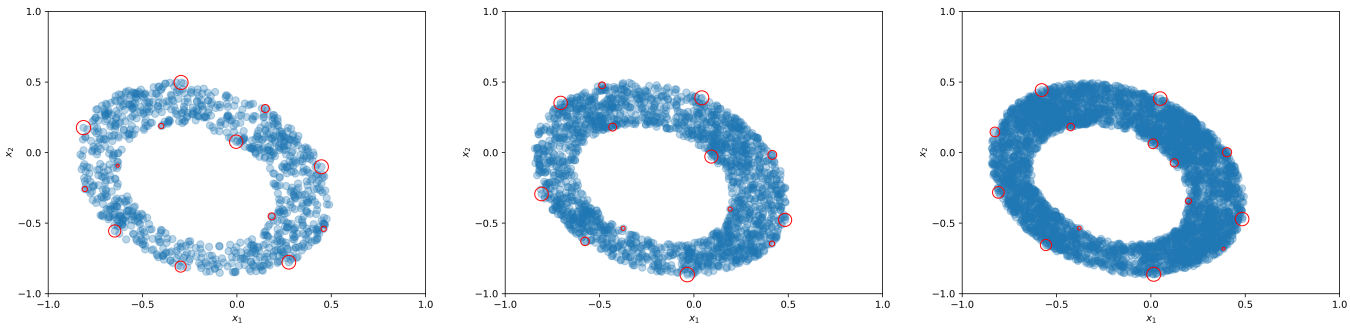
(b) Varying number of scenarios N_π

Figure S1 Without merging: D-optimal criterion value for the RSM example as a function of the number of samples N_s and number of scenarios N_π.

There is a simple remedy that gives the NS algorithm the desired recursive property. Instead of comparing the samples directly, the samples from the different cases are combined. To investigate the effect of varying number of samples N_s, the set of samples presented in Figure S2a and S2b were combined to form the results in Figure S5b; the samples of

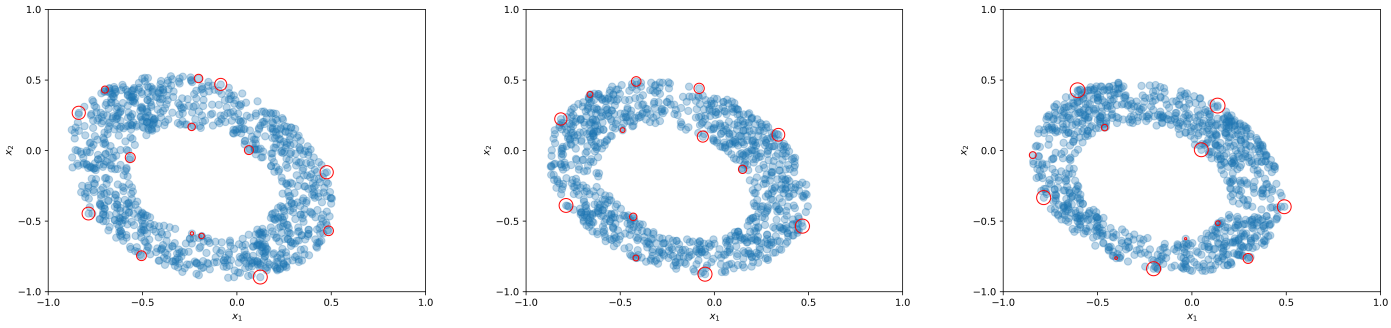


(a) $N_s = 125$ samples **(b)** $N_s = 250$ samples **(c)** $N_s = 500$ samples

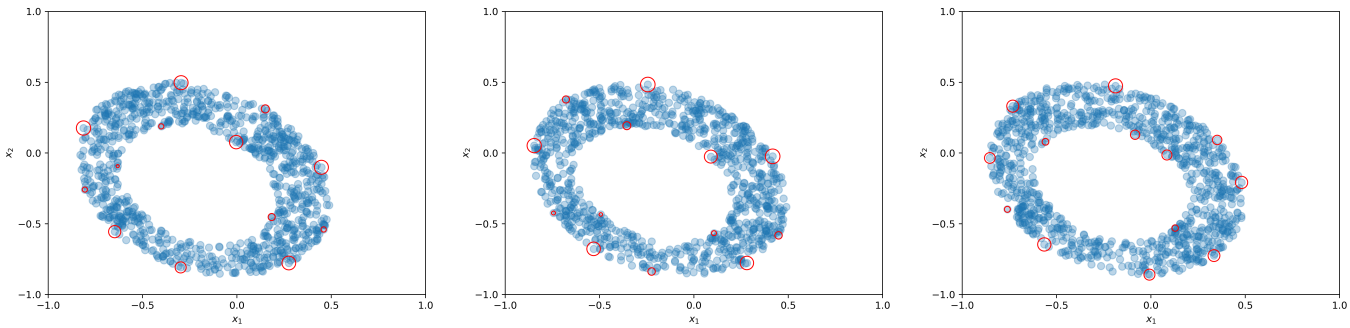


(d) $N_s = 1000$ samples **(e)** $N_s = 2000$ samples **(f)** $N_s = 4000$ samples

Figure S2 Without merging: the RSM example's experimental samples and the D-optimal efforts at varying number of samples N_s for a fixed number of Monte Carlo scenarios $N_\pi = 1000$.

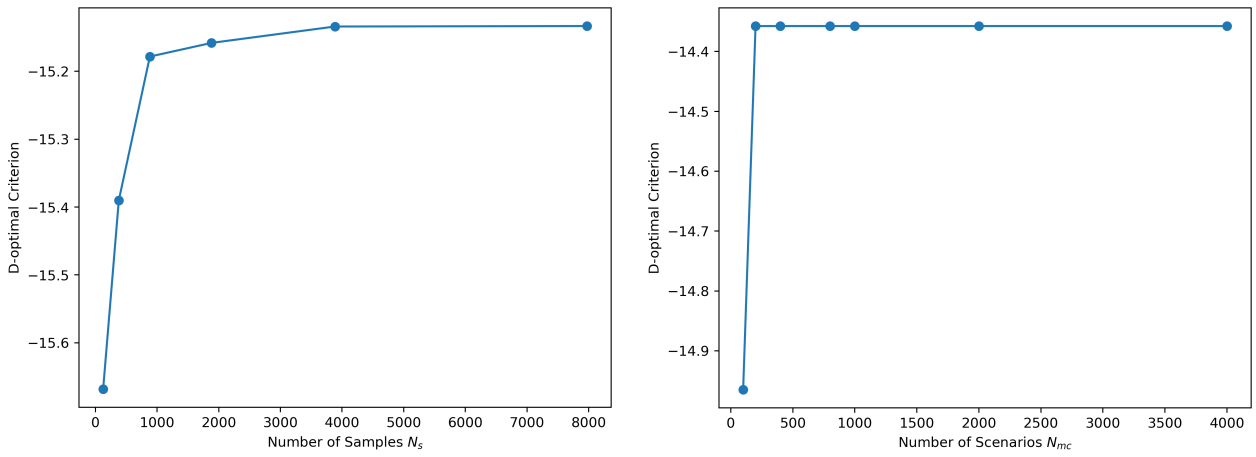


(a) $N_\pi = 100$ scenarios **(b)** $N_\pi = 200$ scenarios **(c)** $N_\pi = 400$ scenarios



(d) $N_\pi = 1000$ scenarios **(e)** $N_\pi = 2000$ scenarios **(f)** $N_\pi = 4000$ scenarios

Figure S3 Without merging: the RSM example's experimental samples and the D-optimal efforts at varying number of Monte Carlo scenarios N_π for a fixed number of samples $N_s = 1000$.



(a) Varying number of samples N_s

(b) Varying number of scenarios N_π

Figure S4 With merging: D-optimal criterion value for the RSM example as a function of the number of samples N_s and number of scenarios N_π.

Figure S2a, S2b, and S2c were combined to form the results of Figure S5c. To investigate the effect of varying number of scenarios N_π, samples from all scenarios (Figures S3a–S3f) are merged together and re-filtered for the given number of scenario N_π. The criterion values and control plots once the merging is done is given in Figures S4, S5, and S6 respectively.

Combining sample sets affects the conclusions significantly. The criterion values consistently improve as the number of samples N_s is increased, see Figure S4a. Their values also stabilize quite quickly as the number of uncertainty realizations N_π is increased, see Figure S4b. The criterion values are consistent after just 200 scenarios, unlike in Figure S1 where criterion values still fluctuate even with $N_\pi = 4000$. Thus we only present the control plots of the first three cases investigated in Figure S6 because the other three are equivalent to Figures S6b and S6c.

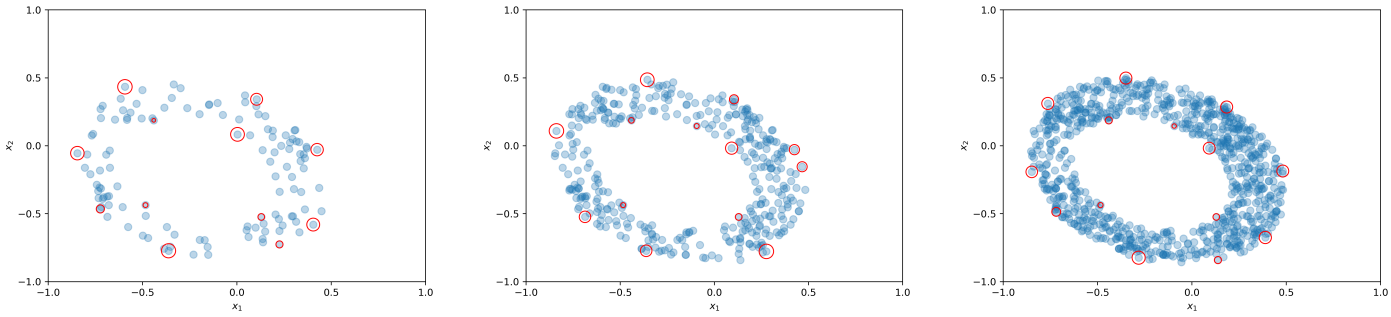
Based on the findings, we recommend to start by determining the appropriate number of scenarios N_π for a constant small number of samples N_s. Like other Monte Carlo techniques a large N_π may be required for consistency, potentially restricting the applicability of the methodology to larger problems. Nevertheless the experimenter must also consider the accuracy of the probability distribution π that is used to describe their beliefs on the model parameter values to guide them. When π is constructed based on ample and informative experimental data, extra care should be taken to ensure that the chosen N_π is sufficient to capture π. In contrast, when π is largely constructed out of one's prior belief (that in itself may be highly subjective), an excessive N_π to capture the fidelity of π may not be needed. Once an appropriate N_π has been determined the experimenter has the option to gradually improve the experimental campaign by increasing the number of experimental factor samples N_s. With such a flexibility, the only parameter that the experimenter needs to carefully decide on is N_π because one can easily tailor N_s to each individual case depending on the trade-off between the time needed to compute the designs and the fidelity of capturing the restricted space, thereby maximizing the information content of the final campaign.

4 Additional Results: Esterification of Propionic Acid

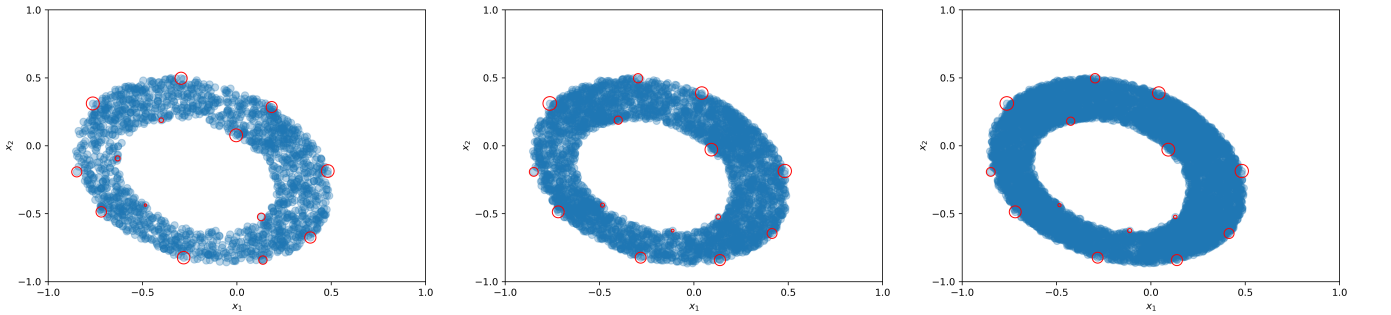
In this extra subsection of the esterification case study, we present additional figures and tables alongside the other designs referenced in the main text. We also provide further discussions on a few notable features of the figures.

4.1 Feasibility of the Restricted D-optimal Average Design

In this subsection, the trajectory of the cooling failure temperature T_{cf} over time is presented in Figure S7 to demonstrate that the restricted, average design's supports satisfy the MTSR constraint. The predicted T_{cf} trajectories involve *groups* due to the discrete probability distribution put on α, and β. Since each can take a value of 1 or 2, there are a total of four trajectory groups corresponding to the different permutations of these values. The groups have varying thickness as well, with thicker lines indicating a higher probability of occurring. The thickest group trajectory corresponds to the scenario when both $\alpha = 1$, and $\beta = 1$, with the highest probability 0.75^2. The thinnest group trajectory corresponds to when both $\alpha = 2$, and $\beta = 2$, with probability 0.25^2. The other two groups with similar thickness correspond to $\alpha = 1$ and $\beta = 2$, and $\alpha = 2$, $\beta = 1$, both with the same probability $0.75 \times 0.25 = 0.1875$. The group trajectories coincide with each other, in support 1 for instance, the thinnest group coincides with the thickest group closely between approximately the 120 minute, and 250 minute mark. Another interesting observation is that four the of five supports "touch" the MTSR

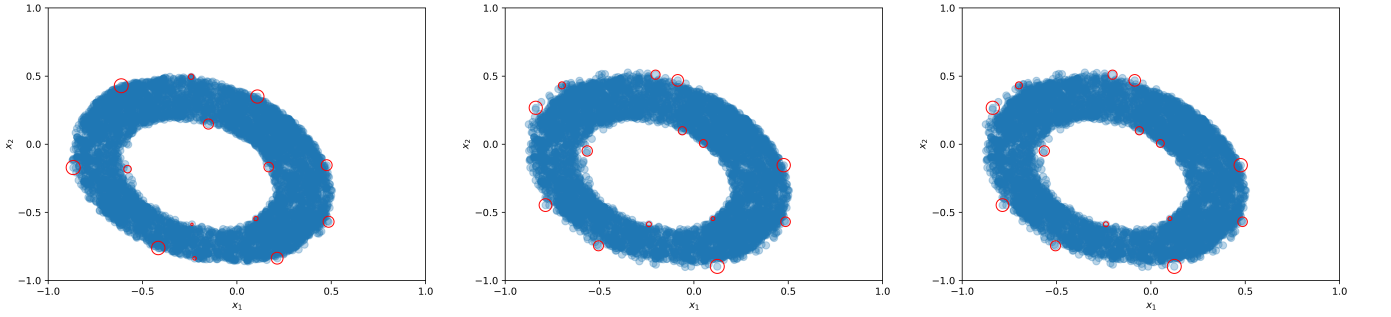


(a) $N_s = 125$ samples **(b)** $N_s = 375$ samples **(c)** $N_s = 875$ samples



(d) $N_s = 1875$ samples **(e)** $N_s = 3875$ samples **(f)** $N_s = 7875$ samples

Figure S5 With merging: the RSM example's experimental samples and the D-optimal efforts at varying number of samples N_s for a fixed number of Monte Carlo scenarios $N_\pi = 1000$.



(a) $N_\pi = 100$ scenarios **(b)** $N_\pi = 200$ scenarios **(c)** $N_\pi = 400$ scenarios

Figure S6 With merging: the RSM example's experimental samples and the D-optimal efforts at varying number of Monte Carlo scenarios N_π for a fixed number of samples $N_s = 1000$.

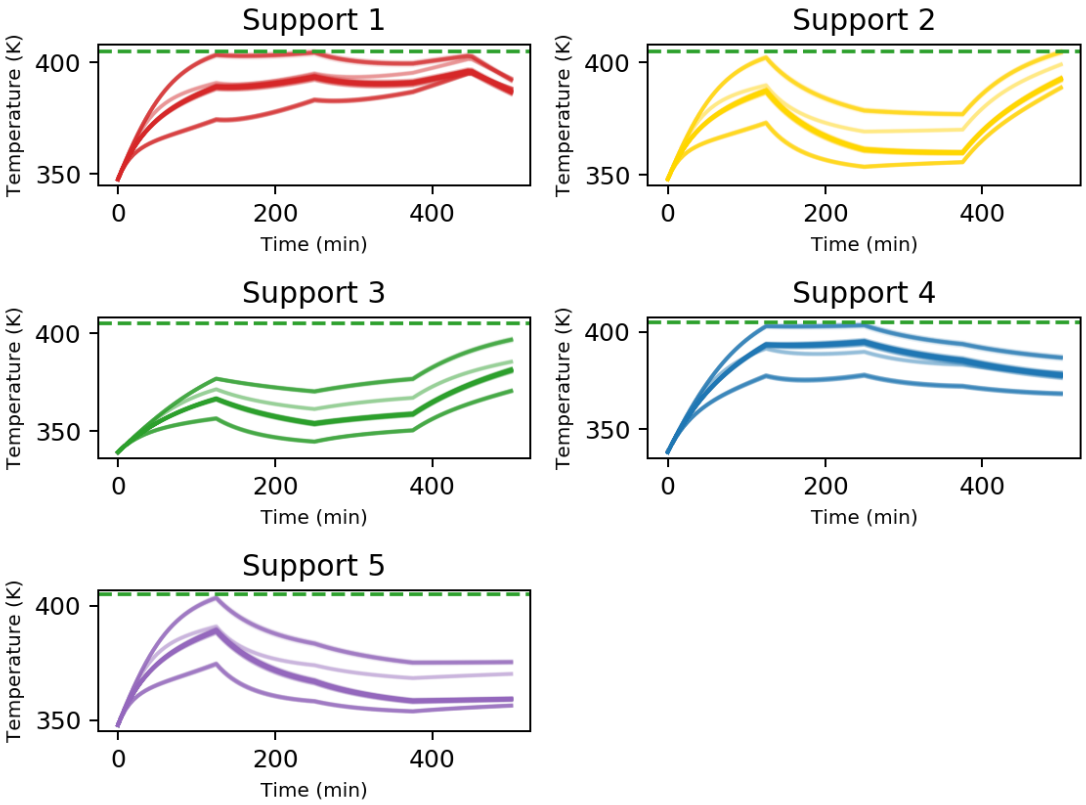


Figure S7 Trajectory of the cooling failure temperature T_{cf} over time of the restricted, average designs supports.

constraint. Indicating that the MTSR constraint does actively constraint the amount of information that the experimental campaign could bring. Thus, as one might expect, many of the supports does activate (or at least are close to activating) the MTSR constraint.

4.2 Restricted D-Optimal Local Design

With the same setup, but now ignoring the effect of uncertainty on the predictions of the information content, we compute a locally D-optimal design within the restricted space. We kept all solution parameters consistent with the restricted, average design. The restricted locally D-optimal design is presented in Figures S8, S9, and Table S1.

Figure S8 presents a corner plot visualization of the locally D-optimal design. It contains several similarities with the average design (AD) of Figure 4, implying that under the prior p.d.f. used in this study, opting for a local design (LD) does not lead to an overly different experimental campaign. This gives some re-assurance that opting for a LD does not lead to a large increase in risk that an uninformative experiment is computed. To provide a measure of this risk, we compute the average D-criterion values for both designs under the given prior: the AD is predicted to have an average D-criterion value of 2.823, whilst the LD 2.806. Both have 5 supports, all of them red, indicating predicted feasibility probability of at least 95%. The temperature subplots again show that both designs are made up of high, and low temperature experiments with no mid-range temperatures. The effort split between the high, and low temperature experiments is similar, with a total of 65.03% allocated to the high-temperature experiments. The u_1 subplots are visually similar, showing four supports at high values of u_1, and one support with lower value of u_1. The u_4 versus u_1 subplot shows a noticeable difference, where the value of u_4 for one the supports is significantly lower than in the AD. These similarities are more easily confirmed when comparing Figures 5, and S9. In Figure 5, support number 1, 2, 4, and 5 (top-to-bottom) of the AD are shared by the LD, appearing as support number 1, 2, 3, and 5 in Figure S9, respectively. All shared supports, furthermore, are allocated similar efforts except for number 5, that is allocated 11.69% in the AD, and 2.74% in the LD. This implies that the difference between the AD, and the LD will be more pronounced in larger experimental campaigns. The supports that are not shared do share similarities to each other. They are both low-temperature experiments, with relatively low alcohol feed rates u for the first three time periods. The main difference between two is that the AD recommends a high alcohol feed rate in the final time period whilst the LD recommends a low alcohol feed rate.

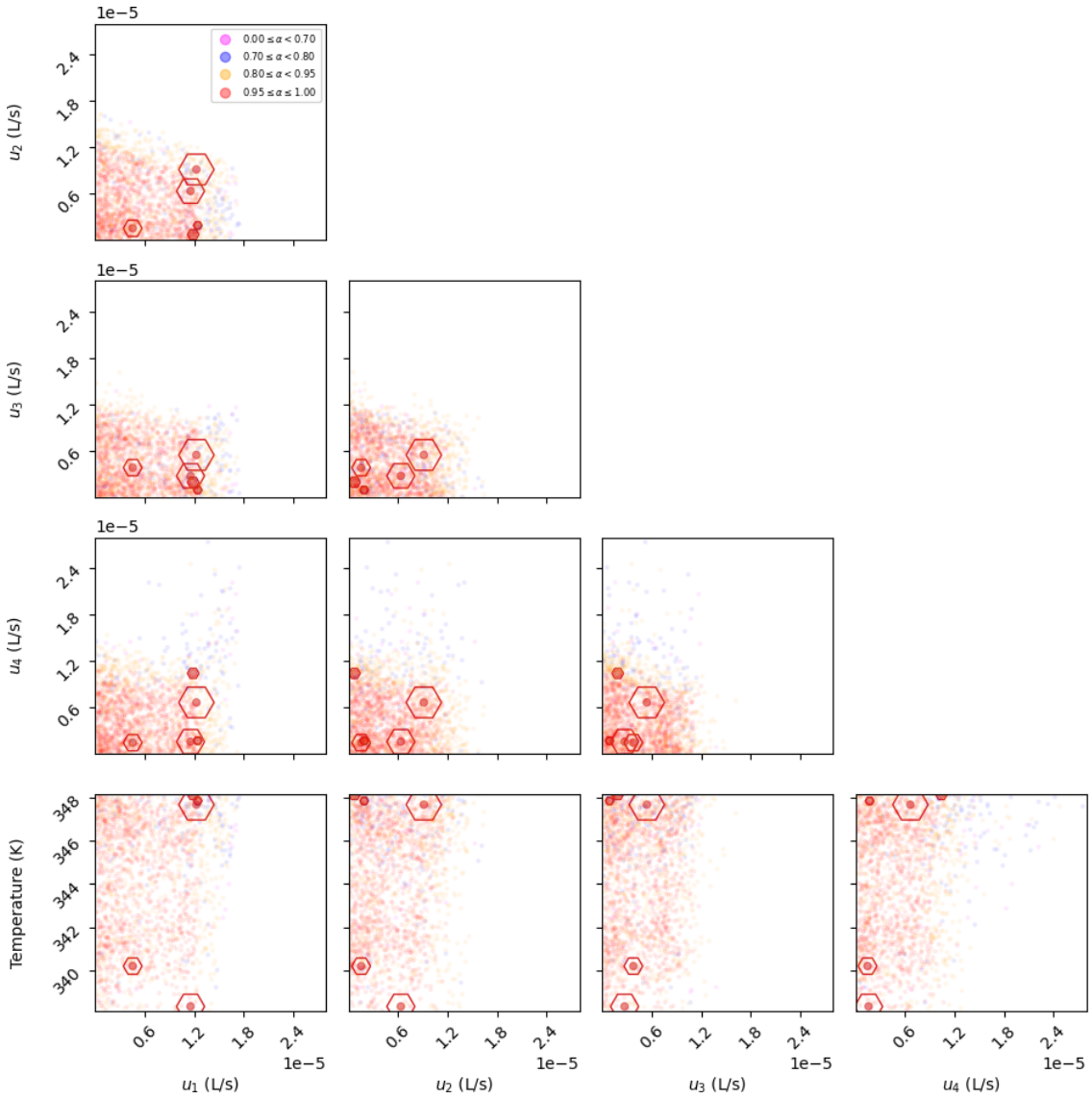


Figure S8 The corner plot of the restricted locally D-optimal design. Despite ignoring the effects of uncertainty on the predicted information content, similarities to the maximal average design are observed. Because of the similarities, it proved to be a viable option when the model proved prohibitively expensive under the project's constraints. Do note that the risk that locally D-optimal designs become uninformative rises as the uncertainty in model parameter increases.

Table S1 The locally D-optimal campaign satisfying the MTSR constraint

Temperature	Feed Rate				Effort	Total number of runs, N_t				
T (K)	$u(t) \times 10^5$ (L s^{-1})				p^* (%)	2	3	4	5	10
	u_1	u_2	u_3	u_4						
347.67	1.23	0.91	0.54	0.66	48.27	1	1	1	1	4
348.14	1.19	0.07	0.19	1.04	5.07	0	0	1	1	1
338.34	1.16	0.63	0.27	0.16	30.35	1	1	1	1	3
340.21	0.45	0.15	0.38	0.14	13.56	0	1	1	1	1
347.85	1.24	0.19	0.09	0.17	2.74	0	0	0	1	1
					Efficiency Bound (%)	0.00	0.00	0.00	41.43	73.73
					Actual Efficiency (%)	87.52	93.54	91.22	86.58	99.19

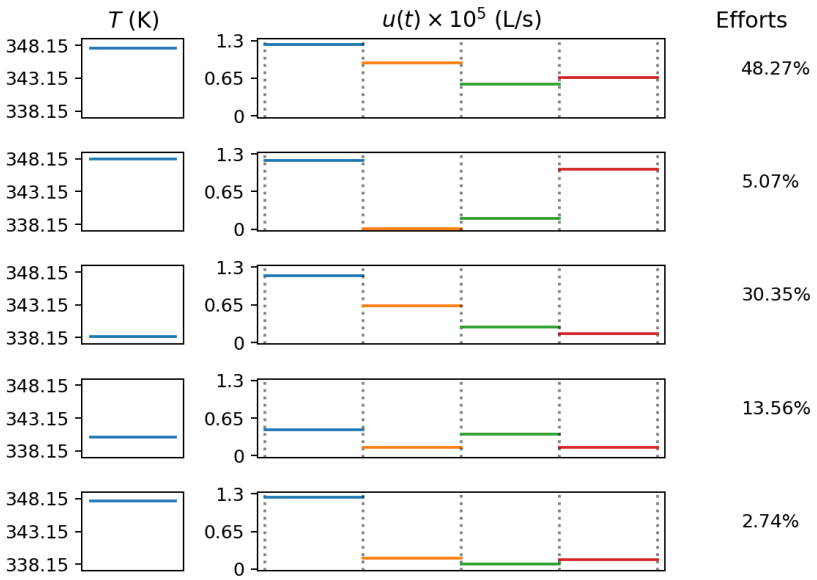


Figure S9 The dynamic experiment plot of the supports of the restricted D-optimal local design. Four supports are shared between the local, and average designs. The similarities between the two designs imply that local design is a viable option when computational cost becomes prohibitive.

4.3 Unrestricted D-Optimal Local Design

The unrestricted experimental space considers the whole range of experimental factors values permitted, i.e., $T \in [338, 348]$ K, and $u_1, u_2, u_3, u_4 \in [0.80, 2.80] \times 10^{-5}$ L s^{-1}. This 5-dimensional space is discretized into $4 \times 4 \times 4 \times 4 \times 5 = 1280$ mesh-centered grid of experimental candidates, with 5 temperature levels. We aim to maintain a comparable, but larger number of candidates with the restricted designs to not under-represent the effectiveness of the unrestricted design with an overly coarse discretization relative to the restricted space.

In the same formats as before, we report the unrestricted, local design in Figures S10, S11, and Table S2. Since the MTSR constraint no longer restricts the experimental space, the unrestricted, local design is expected to provide more information. As comparison, we report the average information criterion under the given prior. The unrestricted, locally D-optimal design is predicted to provide 8.350, compared to 2.823 of the restricted, average design, and 2.806 of the restricted, local design. This increased information content, however, comes at a hefty, and unacceptable safety risks.

The axes limits of Figure S10 are enlarged to clearly illustrated the supports of the unrestricted, local design. It has 4 experimental supports. As the corner plot illustrates, many supports are positioned at the boundaries of the unrestricted space. Unlike the restricted designs, the supports are magenta, indicating the they have predicted feasibility probabilities below 70%. Although some of the supports appear to be within the main cluster of res samples in some subplots, they are located away from it in the other subplots. The T versus u_1 subplot, for instance, shows a clear separation between the experimental supports and runs with high feasibility probabilities.

Although significantly different to the restricted designs, there are two similarities we were able to discern. Like others, the supports of the unrestricted design comprise experiments at low, and high temperatures without mid-range temperatures. Only this time, only a total of 49.04% of the effort is allocated to high-temperature experiments. The alcohol feed rate in the first quarter u_1 of all the supports take relatively high values, indicating that introducing as much alcohol early is informative. Notice, though, that the u_1 values of the unrestricted, local design are significantly higher than those of the restricted ones.

5 Additional Case Study: Increasing Throughput of a Jacketed Stirred-Tank Reactor

Consider the production of a chemical species B in a helical-finned jacketed stirred-tank reactor (for enhanced heat transfer) hosting an exothermic reaction $\mathrm{A} + \mathrm{C} \longrightarrow \nu\mathrm{B}$ and designed to operate in continuous mode. The following mechanistic model posits that species C is always in excess and makes other basic assumptions like perfect mixing and

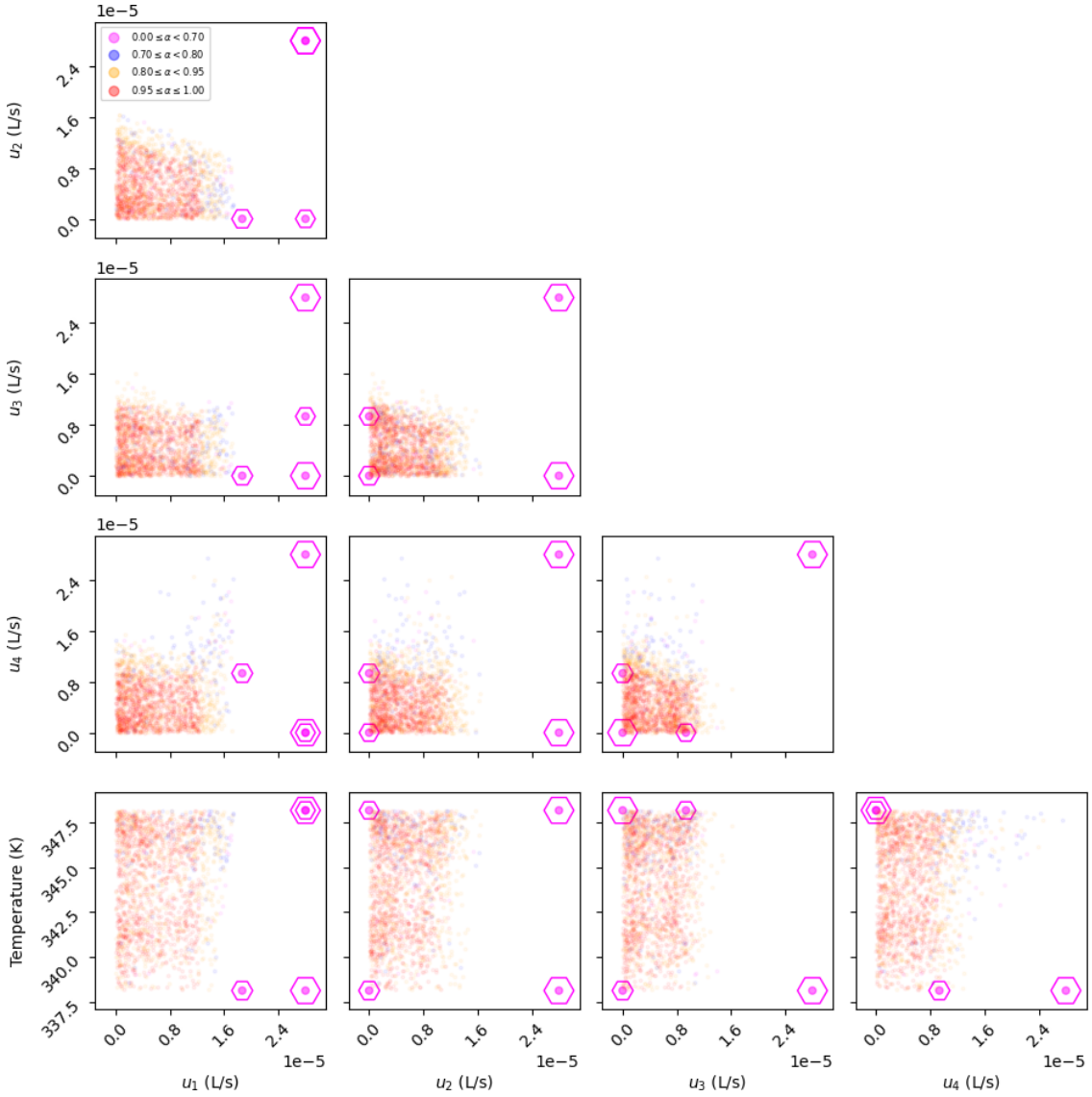


Figure S10 Corner plot of the unrestricted locally D-optimal design. Unlike previous designs, the supports are magenta coloured, having feasibility probabilities lower than 70%. Although some supports appear to be located within the main cluster of red samples, note that they are located outside in the other subplots/projections. The axes limits are slightly enlarged from previous corner plots for clarity as many supports lie on the boundaries of the subplots.

Table S2 The locally D-optimal campaign ignoring the MTSR constraint

Temperature	Feed Rate				Effort	Total number of runs, N_t				
T (K)	$u(t) \times 10^5$ (L s^{-1})				p^* (%)	2	3	4	5	10
	u_1	u_2	u_3	u_4						
338.15	18.70	0.00	0.00	9.30	16.79	0	1	1	1	2
338.15	28.00	28.00	28.00	28.00	34.17	1	1	1	2	3
348.15	28.00	0.00	9.30	0.00	14.98	0	0	1	1	2
348.15	28.00	28.00	0.00	0.00	34.06	1	1	1	1	3
					Efficiency Bound (%)	0.00	0.00	73.16	58.72	87.79
					Actual Efficiency (%)	86.75	95.05	97.51	98.56	99.51

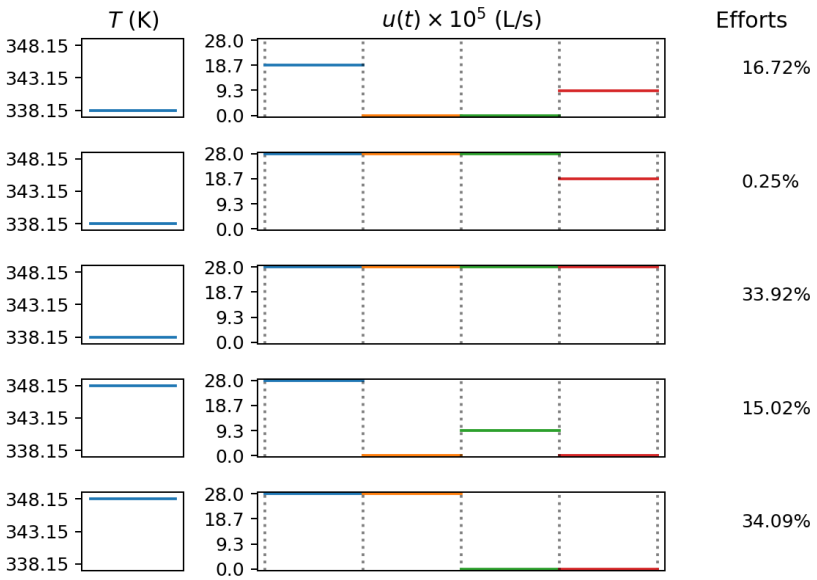


Figure S11 The dynamic experiment plot of the supports of the unrestricted D-optimal local design.

constant reaction mixture volume,

$$\frac{dc_{\mathrm{A}}}{dt} = \frac{q}{V}\left(c_{\mathrm{A},in} - c_{\mathrm{A}}\right) - kc_{\mathrm{A}}^{\gamma} \tag{S1}$$

$$\frac{dc_{\mathrm{B}}}{dt} = -\frac{q}{V}\left(c_{\mathrm{B}}\right) + \nu kc_{\mathrm{A}}^{\gamma} \tag{S2}$$

$$\frac{dc_{\mathrm{C}}}{dt} = \frac{q}{V}\left(c_{\mathrm{C},in} - c_{\mathrm{C}}\right) - kc_{\mathrm{A}}^{\gamma} \tag{S3}$$

$$k = \exp\left(\theta_0 + \theta_1\left(\frac{T - T_{ref}}{T}\right)\right) \tag{S4}$$

$$\frac{dH}{dt} = q(h_{in} - h_{out}) - Q_j \tag{S5}$$

$$H = M_t h_{out} \tag{S6}$$

$$M_t = \sum_{i\in\{\mathrm{A,B,C}\}} c_i V \tag{S7}$$

$$h_{in} = \sum_{i\in\{\mathrm{A,B,C}\}} c_{i,in}\left(h_{F,i} + c_{P,i}(T_{in} - T_{ref})\right) \tag{S8}$$

$$h_{out} = \sum_{i\in\{\mathrm{A,B,C}\}} c_i\left(h_{F,i} + c_{P,i}(T - T_{ref})\right) \tag{S9}$$

$$\frac{dH_j}{dt} = q_w(h_{w,in} - h_{w,out}) + Q_j \tag{S10}$$

$$H_j = M_w h_{w,out} \tag{S11}$$

$$M_w = \rho_w V_j \tag{S12}$$

$$h_{w,in} = h_{F,w} + c_{P,w}\left(T_{w,in} - T_{ref}\right) \tag{S13}$$

$$h_{w,out} = h_{F,w} + c_{P,w}\left(T_j - T_{ref}\right) \tag{S14}$$

$$Q_j = UA(T - T_j) \tag{S15}$$

where c_i ($\mathrm{mol\,L^{-1}}$) is the concentration of species i, t (min) the time, V (L) the volume of reaction mixture, k ($\mathrm{min^{-1}}$) the reaction constant, H (J) the enthalpy of the reaction mixture, h_{in} ($\mathrm{J\,mol^{-1}}$) the molar enthalpy of the inlet, h_{out} ($\mathrm{J\,mol^{-1}}$) the molar enthalpy of the outlet, T (K) the reactor temperature, Q_j ($\mathrm{J\,min^{-1}}$) the heat transferred between the reaction mixture and the jacket, M_T (mol) the total molar holdup of reaction mixture, H_j (J) the jacket enthalpy, $h_{w,in}$ ($\mathrm{J\,mol^{-1}}$) the molar enthalpy of cooling water into the jacket, $h_{w,out}$ ($\mathrm{J\,mol^{-1}}$) the molar enthalpy of cooling water out of the jacket,

Table S3 Model parameters, experimental factors, and default start-up strategy for the jacketed stirred-tank reactor case study.

Model Parameters

Symbol		Name	Value
θ_0	(–)	Pre-exponential constant	-3.896
θ_1	(–)	Activation energy	44.03
ν	($\mathrm{mol\,mol^{-1}}$)	Stoichiometric ratio	1.0
γ	(–)	Order of reaction	1.0
T_{ref}	(K)	Reference temperature	273.15
UA	($\mathrm{W\,K^{-1}}$)	Thermal conductance	2000
V_j	(L)	Jacket volume	2.00
V	(L)	Reaction volume	2.00
ρ_w	($\mathrm{mol\,L^{-1}}$)	Molar density of water	55.56
$c_{P,\mathrm{A}}$	($\mathrm{J\,mol^{-1}\,K^{-1}}$)	Heat capacity of A	112.4
$c_{P,\mathrm{B}}$	($\mathrm{J\,mol^{-1}\,K^{-1}}$)	Heat capacity of B	120.0
$c_{P,\mathrm{C}}$	($\mathrm{J\,mol^{-1}\,K^{-1}}$)	Heat capacity of C	130.0
$h_{F,\mathrm{A}}$	($\mathrm{J\,L^{-1}}$)	Heat of formation of A	-80,000
$h_{F,\mathrm{B}}$	($\mathrm{J\,L^{-1}}$)	Heat of formation of B	-180,000
$h_{F,\mathrm{C}}$	($\mathrm{J\,L^{-1}}$)	Heat of formation of C	0

Experimental Controls

Symbol		Name	Range
q	($\mathrm{L\,min^{-1}}$)	Throughput	$[0.50, 2.50]$
τ_d	(min)	Duration	$[1.0, 20.0]$
q_w	($\mathrm{L\,min^{-1}}$)	Cooling flow	$[0.0, 5.0]$

Default Start-up

Symbol		Name	Value
q	($\mathrm{L\,min^{-1}}$)	Throughput	1.00
q_w	($\mathrm{L\,min^{-1}}$)	Cooling flow	1.1
$c_{\mathrm{A},0}$	($\mathrm{mol\,L^{-1}}$)	Initial A concentration	0.03
$c_{\mathrm{B},0}$	($\mathrm{mol\,L^{-1}}$)	Initial B concentration	0.00
$c_{\mathrm{C},0}$	($\mathrm{mol\,L^{-1}}$)	Initial C concentration	24.66
$c_{\mathrm{A},in}$	($\mathrm{mol\,L^{-1}}$)	Inlet A concentration	0.50
$c_{\mathrm{B},in}$	($\mathrm{mol\,L^{-1}}$)	Inlet B concentration	0.00
$c_{\mathrm{C},in}$	($\mathrm{mol\,L^{-1}}$)	Inlet C concentration	24.66
T_{in}	(K)	Inlet temperature	298.15
T_0	(K)	Initial reaction temperature	298.15
$T_{j,0}$	(K)	Initial jacket temperature	273.15
$T_{w,in}$	(K)	Inlet water temperature	273.15

and M_w (mol) the total molar holdup in the jacket. The rest of the model parameters are collected in the left section of Table S3. The experimental controls are specified in the right section.

The default initial conditions and start-up procedure to achieve steady-state operation is also provided in the right section of Table S3. As shown in Figure S12, the model with nominal parameter estimates predicts that steady-state can be established within 30 minutes. The green dashed lines in the left and right plots of Figure S12 indicate the minimum product concentration $c_{\mathrm{B}}^{\min} = 0.24$ to be suitable for downstream separations and the maximum temperature $T^{\max} = 30\,^{\circ}\mathrm{C}$ for a safe operation, respectively.

To address changing market demand, the throughput needs to be doubled to $q = 2.00\ \mathrm{L\,min^{-1}}$. Although the process was designed with the flexibility to accommodate throughput increases, there is a concern that the actual process behavior has drifted due to aging. This drift could either be attributed to a decrease in UA due to fouling on the reactor-jacket interface, or to a change in θ_0 because of a significant buildup of undetected inhibitor(s) inside the reactor. The process, therefore, may no longer be able to accommodate the throughput doubling without leading to dangerously high reaction temperatures exceeding $T^{\max}$ or off-spec product concentrations below $c_{\mathrm{B}}^{\min}$.

It was estimated that $\theta_0 \in [-4.101, -3.896]$ and $UA \in [1750, 2000]$. With such a belief, the range of predictions under a doubling of the throughput flow from 100 minutes onward is presented in Figure S13. A total of $N_\pi = 100$ Monte Carlo scenarios are drawn from the uniform distribution of the two uncertain parameters UA and θ_0, chosen from gradual increase of N_π until there is no significant differences in the shaded curves in Figure S13. A significant portion (56 out of 100 scenarios) of the c_{B} trajectories on the left plot fall below the green dashed line representing $c_{\mathrm{B}}^{\min}$ after the set-point change. By contrast, all of the reaction temperature profiles on the right plot are predicted to remain below the green dashed line representing $T^{\max}$ after the same change.

To mitigate this large uncertainty, an experimental campaign is designed to support the recalibration of UA and θ_0. The experiments are planned with smaller and revertible set-point changes to test the reactor's capacity to cope with the target throughput increase. Since the experiments are conducted on the process plant itself, the feasibility of the process during the experiment is critical. Three control variables are chosen as experimental controls (Table S3): the modified throughput flow q, imposed 100 minutes after startup time to ensure that the process is initially at steady state and for a duration τ_d before reverting to the default throughput; and the coolant flowrate q_w taken constant throughout the experiment.

The process allows monitoring of c_{A} and c_{B} offline through manual sampling of the reaction mixture. Both the reaction mixture temperature T and the jacket temperature T_w are also recorded at the sampling time. A small quantity of an inhibiting agent is added to the samples for quenching any ongoing reactions; small enough to have negligible effect on temperature of the drawn sample. The sample is then sent to the lab to measure the concentrations c_{A} and c_{B}. As long as the feasibility constraints are met, sampling the reactor outlet directly is considered safe. Nevertheless, because sampling is done manually, only one sample can be drawn per experimental run. Thus, the time at which to sample is another important aspect for experimental design.

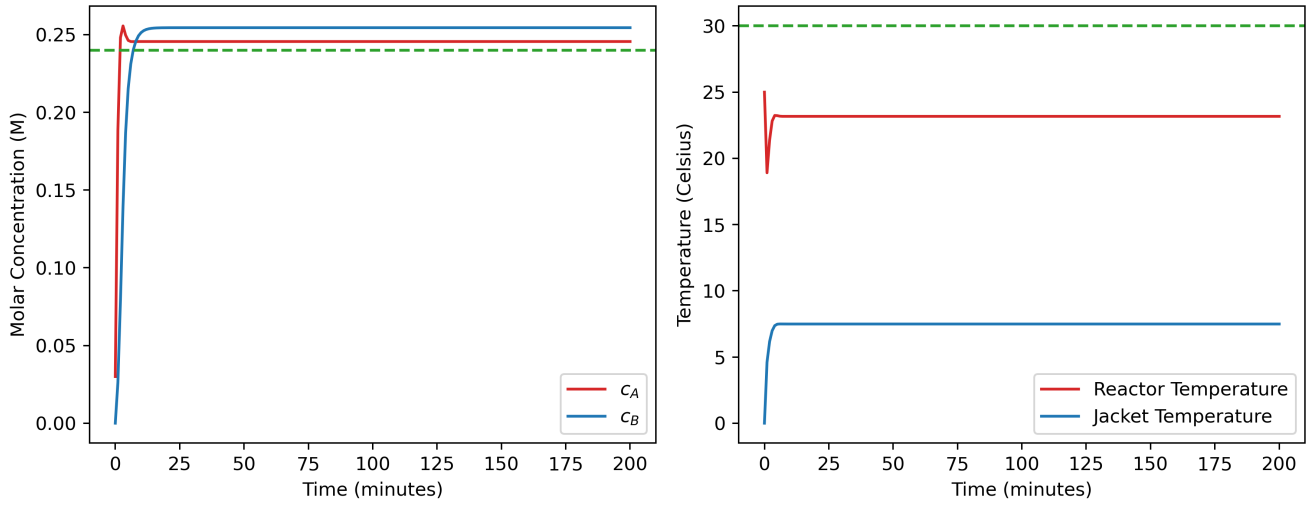


Figure S12 Response trajectories predicted by the process model with nominal parameter values during a default start-up. The green dashed lines in the left and right plots illustrate the minimum concentration $c_{\mathrm{B}}^{\min}$ and maximum reaction temperature $T^{\max}$, respectively.

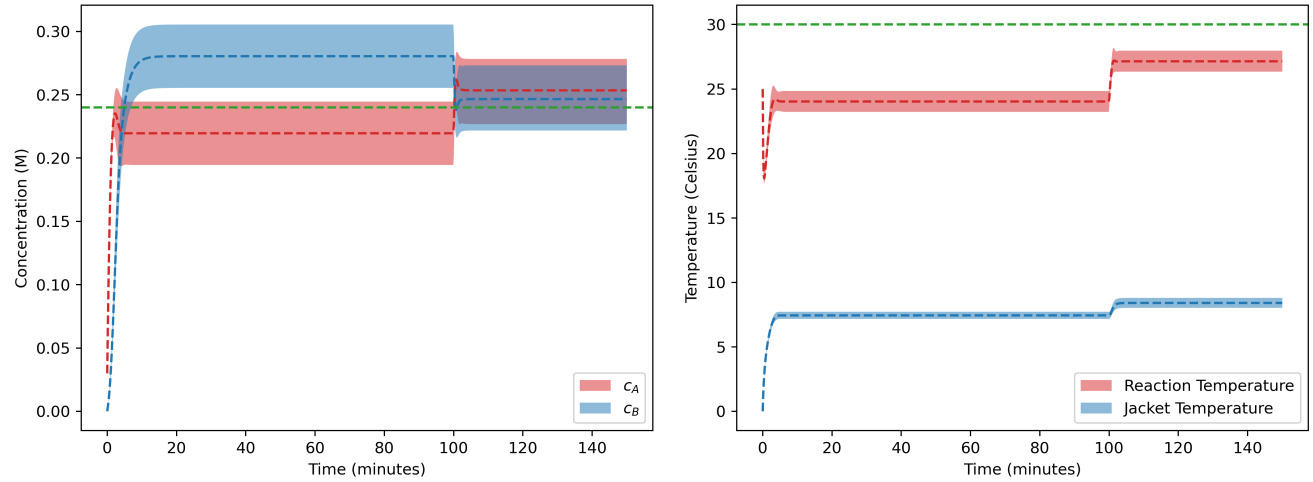


Figure S13 Range of response trajectories predicted by the process model during a default start-up followed by a doubling of the throughput after 100 minutes. The filled regions enclose the minimal and maximal trajectories among all sampled scenarios of the uncertain parameters $(UA, \theta_0) \in [-4.101, -3.896] \times [350, 400]$ with the rest of the parameters at their nominal values, and the dotted curves are the mean profiles. The green dashed lines in the left and right plots illustrate the minimum concentration $c_{\mathrm{B}}^{\min}$ and maximum reaction temperature $T^{\max}$, respectively.

Table S4 Unrestricted, locally D-optimal campaign for the the jacketed stirred-tank reactor. 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.

						Number of Experimental Runs N_t						
UID	q	q_w	τ_d	t_{sp}	p (%)	5	6	7	8	9	10	20
295	2.06	0.00	1.0	102.0	7.3	–	1	1	1	1	1	2
325	2.06	0.00	16.8	117.0	14.3	1	1	1	1	1	2	3
344	2.28	1.00	4.2	105.0	14.1	1	1	1	1	1	1	3
392	2.50	1.00	7.3	108.0	18.3	1	1	1	1	2	2	3
398	2.50	1.00	10.5	111.0	18.4	1	1	1	2	2	2	4
409	2.50	0.00	16.8	117.0	27.6	1	1	2	2	2	2	5
				Efficiency Bound (%)		0.00	60.33	77.49	68.47	77.95	70.87	82.17
				Actual Efficiency (%)		98.01	96.33	98.52	98.37	99.08	98.49	99.62

5.1 Unrestricted Locally D-optimal Design

Although the task is to recalibrate UA and θ_0 only, measuring c_A, c_B, T, and T_w provides information for other model parameters to be reestimated as well. The experiments, therefore, are designed to be maximally informative for the precise estimation of five model parameters, namely UA, θ_0, θ_1, ν and γ. The problem addressed in this subsection is a classical local D-optimal experiment using the current estimates (Table S3) and ignoring the feasibility constraints on $c_B^{\min}$ and $T^{\max}$ initially. The ranges of the controls q, τ_d and q_w (Table S3) are discretized on a $10 \times 7 \times 6$ grid, in combination with 31 sampling time choices—one every minute following the set-point change (excluded) up to the 30 minutes mark (included). This is considering that 20 minutes is the upper bound for τ_d, thus allowing an extra 10 minutes to take measurements when the process is reverting to the original set point.

For this unconstrained locally-optimal design, sensitivity analysis (Step 2.1) and optimization (Step 2.2) was completed using a single core of Intel Xeon E5-2667 V4 @3.2GHz and 128GB of RAM within 6 minutes and 12 minutes of computational time in Pydex, respectively. The maximal D-information criterion is 39.2. The corresponding campaign is reported in Table S4, where each experiment is assigned a unique identification number (UID). The experimental supports all have a sampling time after $100 + \tau_d$, the time at which the system starts reverting. The locally-optimal campaign, therefore, exclusively takes measurements during the reverting phase. Moreover, none of the experimental supports involve a reduction in throughput q, and that the experimental effort is split about equally between cooling flows $q_w = 0$ (49.2%) and $q_w = 1$ (50.8%). Notice finally that the effort apportionment based on the efficient rounding method is effective at preserving a majority of the D-information content—96% and above—for campaigns having between 5–20 runs.

In the rest of the case study, we shall assume a reliability value of $\alpha = 0.95$, such that any experiment predicted to violate either of the constraints in more than 5 realizations of the (UA, θ_0) uncertainty is considered infeasible. Remarkably, none of the experimental supports of the unrestricted local design are feasible at this reliability level, with a ratio of uncertainty realizations violating one or both constraints ranging between 40% to 100%. Another observation is that, although the supports 295 and 325 only differ in their duration τ_d and sampling time t_{sp}, the former obeys $T^{\max}$ under all uncertainty realizations, while the latter violates $T^{\max}$ in every realization. This confirms that imposing a shorter set-point change can effectively prevent violation of $T^{\max}$.

5.2 Restricted Locally D-optimal Design

Since the experimental supports of the unrestricted design are infeasible, the experimental space is now reassessed to ensure that the support points will meet the safety constraints. The number of scenarios is set to $N_\pi = 100$ as earlier (cf. Section 5). A total of $N_s = 424$ samples are drawn from the probabilistic restricted feasible space $\mathscr{R}_{0.95}$ using the nested sampling algorithm implemented in DEUS, which takes about 790 minutes of computational time. The sampled probabilistic restricted space for (q, q_w, τ_d) is illustrated in Figure S14 as a 2D scatter plot with the marker sizes proportional to the duration τ_d; an alternative 3D scatter plot is presented in the ESI (Figure S7).

The general pattern is that a large fraction of the restricted space at 95% reliability level lies within the bottom left corner ($q \in [0.5, 1.75]$, $q_w \in [0, 1.5]$). The abrupt decrease in the density of feasible samples (red markers) for either large throughput $q \geq 1.75$ or reduced cooling capacity $q_w \leq 0.5$ is due to a rise in the reaction mixture temperature T, which quickly exceeds $T^{\max}$ but for the shortest of set-point change τ_d. A similar decrease is observed at large cooling capacity $q_w \geq 1.5$ due to over-cooling in the reactor, which reduces the reaction rate drastically and prevents the concentration c_B from reaching $c_B^{\min}$. The presence of feasible samples outside of the main population that correspond to aggressive changes

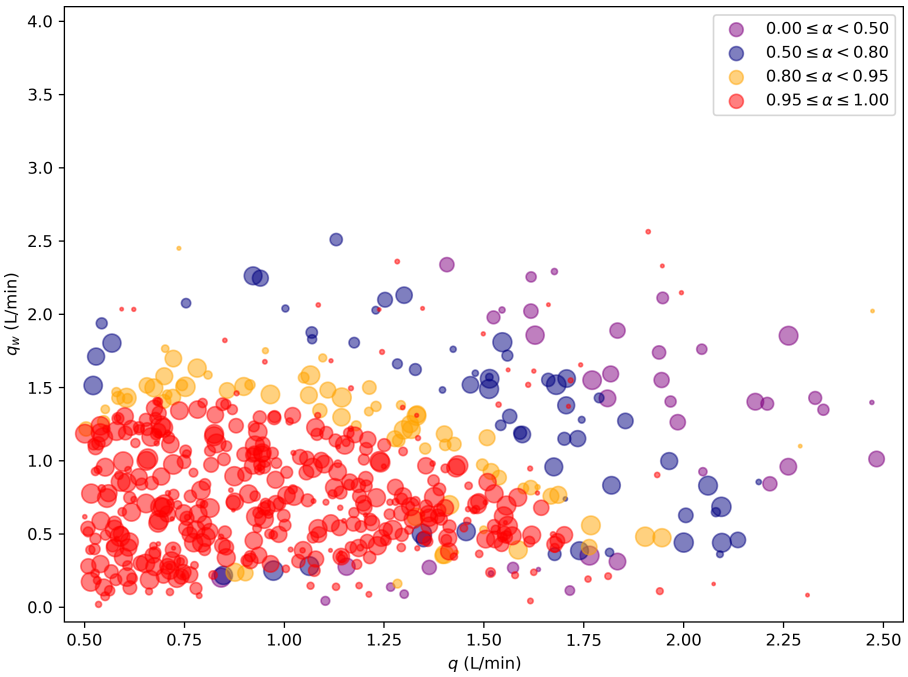


Figure S14 Probabilistic restricted feasible space for the the jacketed stirred-tank reactor. The 3-dimensional data points are scattered in a 2-dimensional space with the marker sizes proportional to the duration τ_d. Red samples satisfy the 95% reliability threshold. The yellow, blue, and purple colored samples are byproducts with feasibility probabilities between $[0.8, 0.95)$, $[0.5, 0.8)$, and $[0, 0.5)$, respectively.

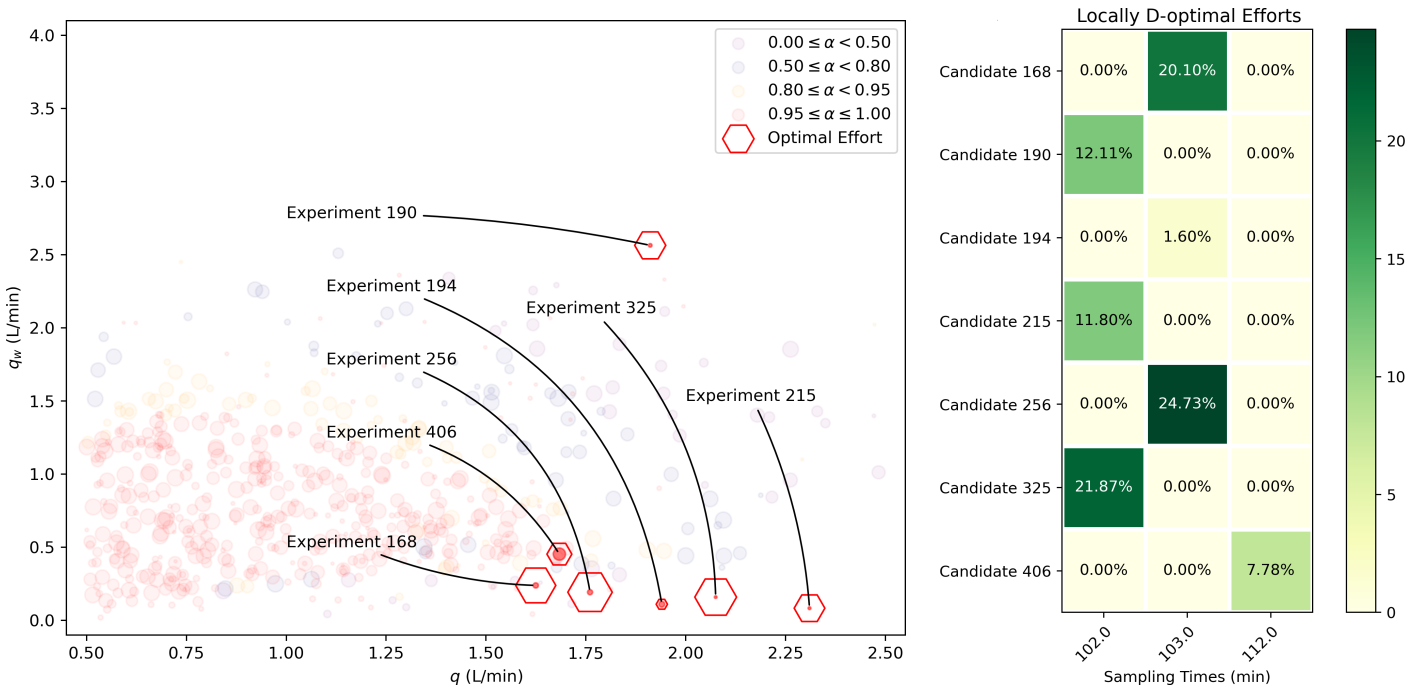
in q and q_w may seem surprising. But these are all runs with short set-point change τ_d, thereby mitigating violations of $T^{\max}$ (cf. Section 5.1).

Next, the locally D-optimal design is computed within the restricted space. Using Pydex, the sensitivity analysis (step 2.1) and optimization (step 2.2) on a single core of Intel Xeon E5-2697 V2@2.7GHz and 48GB of RAM take about 4 minutes and 8 minutes of computational time, respectively. Because the experimental space is now restricted, the predicted D-information criterion drops to 29.6 from 39.2 without restrictions (cf. Section 5.1). The corresponding campaign is illustrated on the plots and table embedded in Figure S15. All support points are located at or near the boundary of the main population of feasible samples. Unlike the unrestricted local design, all experimental supports furthermore increase the throughput q significantly, between 1.6–2.3. Other than the support 406, all the experimental supports involve short set-point changes, with $\tau_d < 3$. All the measurements are furthermore taken during the reverting phase. It is also noticeable that only the support 190 increases the cooling flow q_w to above 2.5, while the others decrease q_w to below 0.5. Lastly, the effort apportionment based on the efficient rounding method is again effective at retaining above 95% of the D-information content for campaigns comprising between 5–20 runs.

5.3 Restricted Average D-Optimal Design

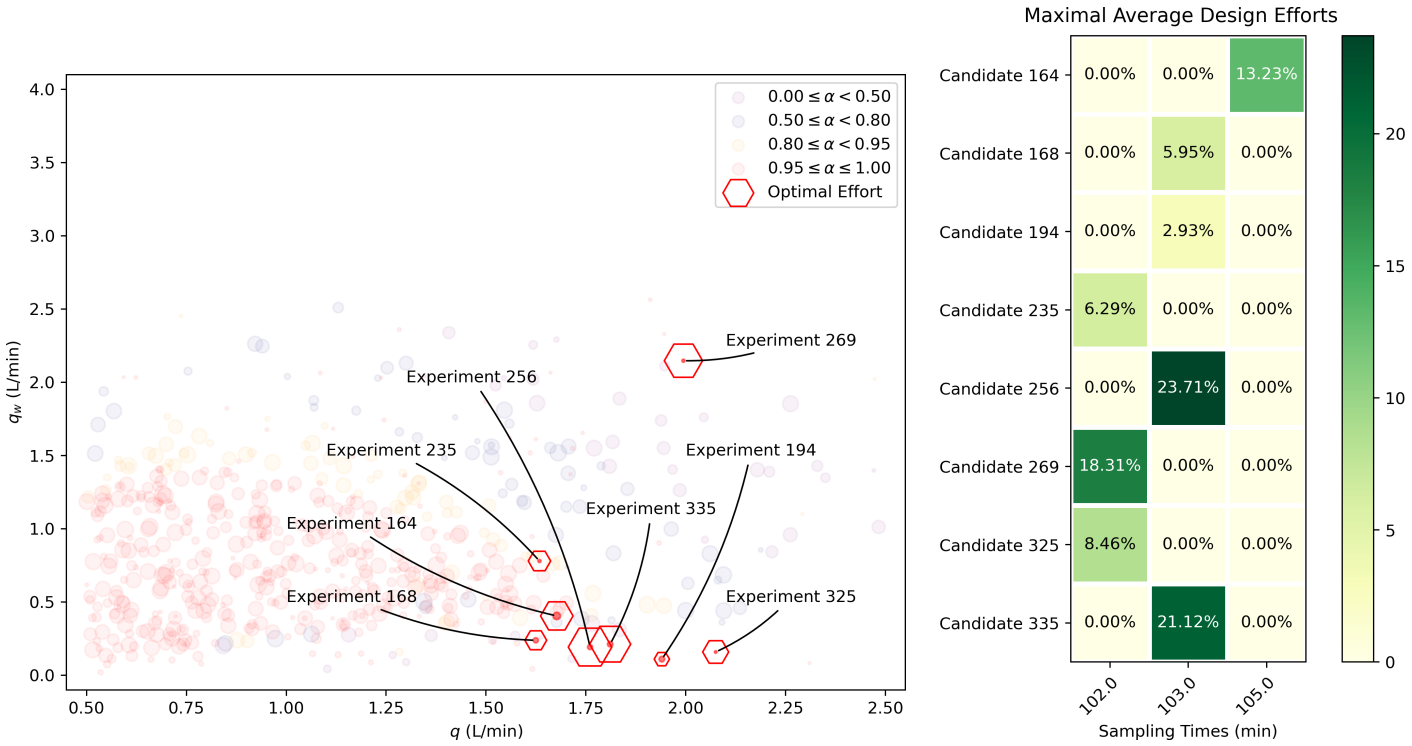
Because the information content may be inaccurately predicted through a local design, the experimental campaign is now designed to maximize the average information content by directly taking the uncertainty of UA and θ_0 into account. The same samples from the restricted space are reused for the average D-optimal design (Figure S14), and the same 100 realizations of UA and θ_0 as for the restricted space sampling are also used to compute the average D-optimal criterion. Since the number of atomic matrices is now 100 times larger, these computations were conducted in parallel using Python's multiprocessing library to utilize 4 CPU cores of a workstation with Intel Xeon E5-2697 V2 @2.7GHz and 48GB of RAM, taking 115 minutes of computational time. After that, the optimization was run on a single core of Intel Xeon E5-2667 V4 @3.2GHz and 128GB of RAM, taking a total of 690 minutes of computational time. Note that CVXPY spends 15% of the total time (over 100 minutes) compiling the optimization problem and the solver MOSEK uses the rest, indicating that problem compilation and solution could be significant impediments to computational tractability in practice.

The average D-optimal campaign is represented on the plots and table in Figure S16. There are notable similarities between the local and average designs. Like the local design, all the experimental supports involve increasing the throughput q. Both designs share 4 common supports (168, 194, 256, 325). Furthermore, the supports 269 and 164 of the maximal average design resemble the supports 190 and 406 of the local design (Figure S15), respectively. The former pair (269 from local and 190 from average) corresponds to a short set-point change τ_d along with a drastic change in both throughput q and cooling flow q_w; the latter pair (164 from local and 406 from average) corresponds to a long



UID	q	q_w	τ_d	t_{sp}	p (%)	Number of Experimental Runs N_t 5	6	7	8	9	10	20
168	1.63	0.24	2.5	103.0	20.1	1	1	1	1	1	2	4
190	1.91	2.56	1.4	102.0	12.1	1	1	1	1	1	1	2
194	1.94	0.11	2.8	103.0	1.6	–	–	1	1	1	1	1
215	2.31	0.08	1.1	102.0	11.8	1	1	1	1	1	1	2
256	1.76	0.19	2.4	103.0	24.7	1	1	1	2	2	2	5
325	2.08	0.16	1.0	102.0	21.9	1	1	1	1	2	2	4
406	1.68	0.45	11.2	112.0	7.8	–	1	1	1	1	1	2
					Efficiency Bound (%)	–	–	57.77	57.16	55.28	80.88	82.54
					Actual Efficiency (%)	98.33	96.78	94.97	96.54	97.36	98.69	99.68

Figure S15 Locally D-optimal campaign within the restricted experimental space at 95% reliability for the the jacketed stirred-tank reactor. Left plot: Translucent markers corresponding to the samples drawn for the restricted experimental space, with experimental supports shown as opaque markers. The size of the red octagons around the experimental supports are proportional to the optimal efforts. Right plot: Heatmap representing the optimal efforts associated with the experimental supports, only showing the optimal time choices for clarity. Table: The left section shows the experimental supports and their associated experimental efforts. 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.



						Number of Experimental Runs N_t						
UID	q	q_w	τ_d	t_{sp}	p (%)	5	6	7	8	9	10	20
164	1.68	0.41	4.4	105.0	13.2	1	1	1	1	1	1	3
168	1.63	0.24	2.5	103.0	6.0	–	–	1	1	1	1	1
194	1.94	0.11	2.8	103.0	2.9	–	–	–	1	1	1	1
235	1.63	0.78	1.3	102.0	6.3	–	1	1	1	1	1	2
256	1.76	0.19	2.4	103.0	23.7	1	1	1	1	2	2	4
269	1.99	2.15	1.3	102.0	18.3	1	1	1	1	1	1	3
325	2.08	0.16	1.0	102.0	8.5	1	1	1	1	1	1	2
335	1.81	0.21	2.7	103.0	21.1	1	1	1	1	1	2	4
					Efficiency Bound (%)	–	–	0.00	52.73	52.61	54.61	81.92
					Actual Efficiency (%)	96.57	95.46	93.61	93.96	96.68	97.16	99.23

Figure S16 Average D-optimal campaign within the restricted experimental space at 95% reliability for the the jacketed stirred-tank reactor. Left plot: Translucent markers corresponding to the samples drawn for the restricted experimental space, with experimental supports shown as opaque markers. The size of the red octagons around the experimental supports are proportional to the optimal efforts. Right plot: Heatmap representing the optimal efforts associated with the experimental supports, only showing the optimal time choices for clarity. Table: The left section shows the experimental supports and their associated experimental efforts. 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.

set-point change with rather mild changes in q and q_w. The main difference between the two designs is that the maximal average design comprises one additional support point that has no resemblance with the local design. The extra support 235 is located around the main population of yellow samples but with a short duration, making it feasible under the given uncertainty. The experimental efforts are reported on the table and illustrated on the left plot in Figure S16. Both the local and maximal average designs assign about two-third of the total experimental effort to three supports only. And similarly to the local designs (cf. Table S4 & Figure S15), the applied rounding scheme is effective at maintaining a majority of the information content in various experimental campaigns having from 5 up to 20 runs.

5.4 Alternative Plots

An alternative 3D plot of the drawn samples of the restricted experimental space is presented in Figure S17 of this document. The 3D scatter plot shows the uniformity of the drawn samples over the restricted space which may not be obvious in the 2D scatter plot of Figure 5 of the main article.

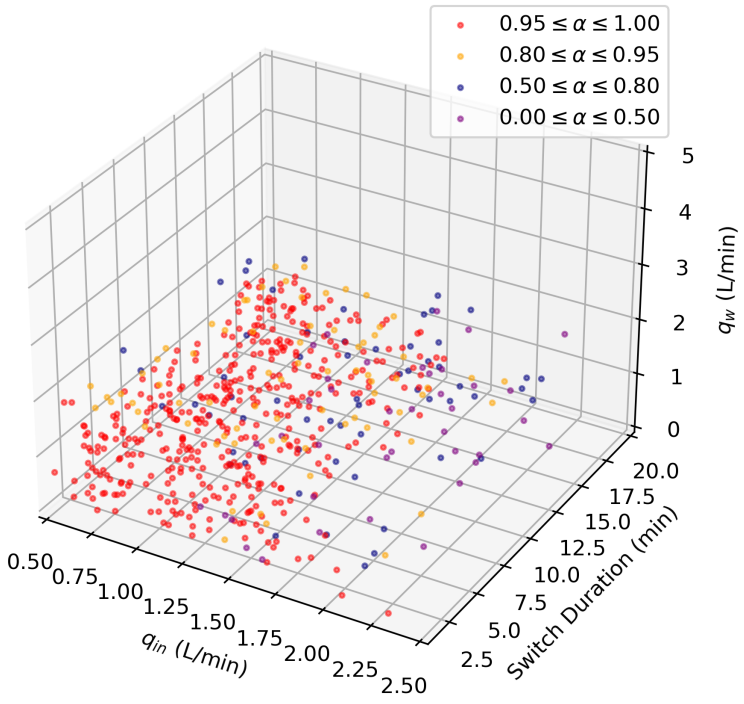


Figure S17 3D scatter plot of samples from the restricted experimental space at 95% reliability drawn using nested sampling.