Electronic Supplementary Information for manuscript

Autonomous Flow Reaction Optimization Guided by NMR Spectroscopy and Advanced Bayesian Optimization Algorithms

Aravind Senthil Vel,[#a] Yuliia Horbenko,[#a] Nour El Sabbagh,[#a] Burkhard Groh,[a,] Daniel Cortés-Borda,[a] Jonathan Farjon,[a] Patrick Giraudeau,[a] Jean-Nicolas Dumez,[†a] Jean-Philippe Krieger,[b] and François-Xavier Felpin*[a]

# Contributed equally

† Passed away on August 18, 2025

[a] Nantes Université, CNRS, CEISAM, UMR 6230, F-44000 Nantes, France.

[b] Syngenta Crop Protection AG, Breitenloh 5, CH-4333 Münchwilen, Switzerland

∗ Corresponding authors.

e-mail: fx.felpin@univ-nantes.fr; ORCID: 0000-0002-8851-246X; Website: http://felpin.univ-nantes.fr/

Table of content

1. Experiment details

The first stock solution was composed of N'-hydroxybenzimidamide (0.7 M) and hexafluorobenzene (0.416 M). The solution was prepared using the 50 mL volumetric flask, in which 4.77 g (0.035 mol) of N'-hydroxybenzimidamide and 2.40 mL (0.021 mol) of hexafluorobenzene were added. The flask was then filled to the mark with ethyl acetate.

The second stock solution was composed of trifluoroacetic anhydride (0.8 M). The solution was prepared using the 50 mL volumetric flask, in which 5.64 mL (0.04 mol) was added. The flask was then filled to the mark with ethyl acetate. The experimental setup consisted of two reagent streams as depicted in Fig. 2. In the first stream, a solution of N-hydroxybenzimidamide (0.7 M) and hexafluorobenzene (0.416 M) in ethyl acetate was continuously pumped and combined in a T-mixer (internal volume: 0.57 μL) with a second stream containing a solution of trifluoroacetic anhydride (0.8 M) in ethyl acetate. The combined stream was then passed through a tubular PEEK reactor (10 mL, 0.76 mm i.d.), immersed in a heated bath to control the reaction temperature. Afterward, the reaction mixture was cooled in a PFA loop (0.6 mL, 0.75 mm i.d.) before being analyzed in real time by quantitative in-line 1D ^{19}F NMR spectroscopy using a benchtop spectrometer, which has been successfully used in the field of flow synthesis. All units equipped with a RS-232 port were autonomously controlled with MATLAB through the use of communication protocols provided by the manufacturer.

2. NMR details

1D ^{19}F NMR experiments were recorded at 41.044 MHz on a ^{1}H/^{19}F benchtop spectrometer (Spinsolve, Magritek) equipped with a flow cell (4 mm id). Experiments were conducted at 28.5°C.

2.1. NMR acquisition

All the instructions about NMR data acquisition were called through the Spinsolve Expert software. The shimming procedure was performed prior to each NMR experiment. 1D ^{19}F spectra were acquired in a stopped-flow mode and in a single scan, using the following parameters: a recovery delay of 3.7 s followed by a 90° pulse and an acquisition time of 1.3 s with 65k data points, and a spectral width of 50 ppm centered at -130 ppm.

2.2. NMR processing

All the instructions about NMR processing were called through the Spinsolve Expert software. The obtained free induction decays were zero-filled up to 128k points, Fourier transformed, apodized using an exponential function (lb = 1 Hz) and finally automatically phased.

The peak area of the product was obtained by integrating the spectral region from -68.5 to -66.4 ppm; for the reference compound (hexafluorobenzene), this was from -166.5 to -164.2 ppm as shown in Fig. S1.

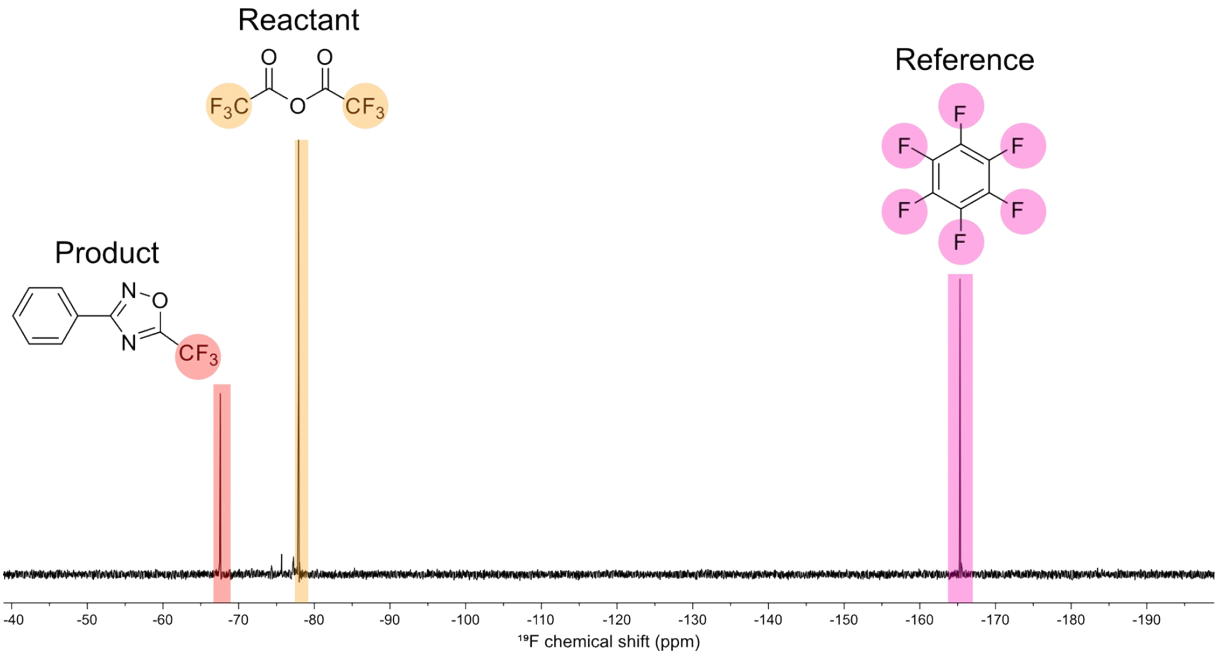


Fig. S1 ^{19}F 1D spectrum of the reaction mixture containing peaks of the product, a reactant and a reference.

3. Optimization Strategy details

3.1. Pareto-oriented approach

We used the MVMOO[1] (Mixed Variable Multi-Objective Optimization) solver for this approach. MVMOO employs a Gaussian Process (GP) as the surrogate model and the Expected Improvement Matrix[2] as the acquisition function. The solver is available at https://github.com/jmanson377/MVMOO?tab=readme-ov-file.

3.2. Adaptive Boundary Constraint in Bayesian Optimization (ABC-BO)

We proposed ABC-BO[3] for Single-Objective Optimization (SOO) problems to avoid futile experiments in reaction optimization. For certain objective functions $f(x)$, such as throughput

or E-factor, it is possible to estimate the maximum achievable objective $f^{\max\, acheivable}$ by assuming a 100% yield for each specific experimental condition. In standard Bayesian Optimization (BO), the algorithm may suggest experiments where the maximum achievable objective is worse than the current best f^*. Such experiments, in our opinion, are futile to perform. Although BO proposes these points to improve the knowledge of the search space, we avoided exploring these conditions since they are already known not to improve the objective. This is achieved by defining this condition as a constraint in the optimization problem:

$$\max_x f(x)$$

$$s.t.\ f^{\max\, acheivable} > f^* - tolerance$$

The tolerance is specified to avoid an overly strict constraint. Since the constraint restricts the search space, if the current best objective is overestimated due to noise, this could result in excluding theoretically valid conditions from the search space. This concept of a futile experiment applies only to certain objective functions whose calculations are influenced by the variables. For example, the throughput depends on the variable residence time. Although this constraint alone is sufficient to avoid futile experiments, in our methodology we further redefine the search space to exclude such regions, which helps reduce the computational burden during optimization. For more details, please refer to the original work.[3]

The constraint specified here falls under a known constraint, meaning any violation can be identified theoretically. Therefore, it can be ensured that the next suggested point does not violate this constraint. In this work, we also defined another constraint — the yield must be greater than or equal to a specified value. Unlike the earlier constraint on the maximum achievable objective, this is an unknown constraint, as its violation cannot be determined theoretically before performing the experiment.[4] Because of this additional unknown constraint (on yield), the intrinsic constraint in ABC-BO had to be adjusted. Restricting the search space based solely on the current best objective value may lead to conditions where the yield constraint cannot be satisfied at all, causing the optimization to fail. To avoid this, we considered a solution as the current best only if it also satisfies the yield constraint. This relaxation allows conditions where the yield requirement is feasible while still ensuring that the improvement in the objective is theoretically valid. We used MATLAB's built-in solver *bayesopt* which can handle these constraints. The problem was solved using the default settings, except that the acquisition function was set to the upper confidence bound (or lower confidence bound).

4. Optimization results

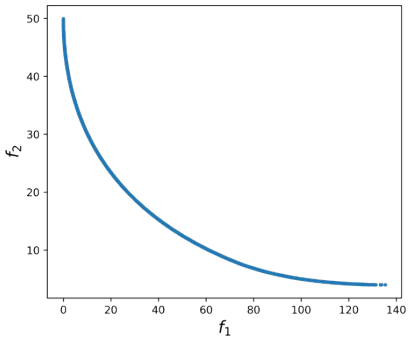
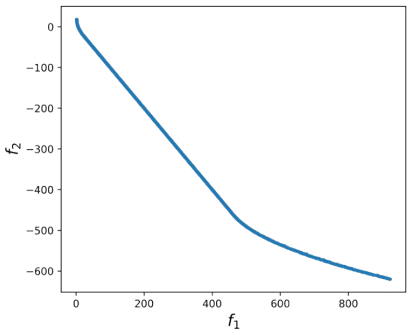
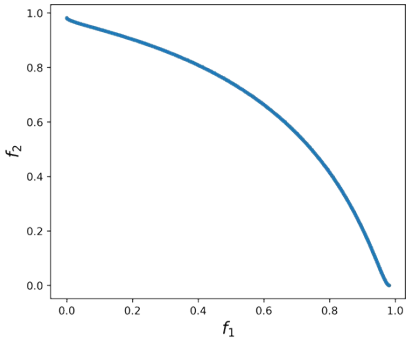
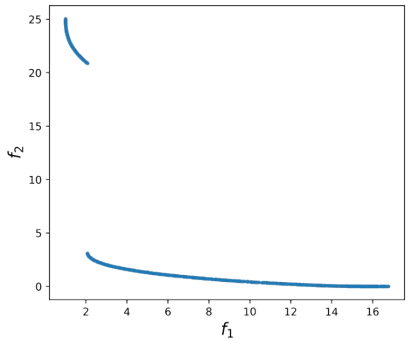
4.1. In silico results

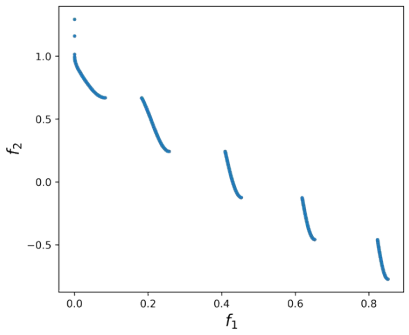
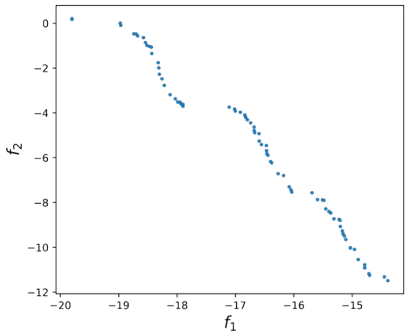
To benchmark the performance between Pareto-oriented and constraint-oriented approaches, the multi-objective problems were constructed from six analytical functions. The optimization problem involves minimising the two objectives (f_1, f_2) defined in Table S1. In the Pareto-oriented approach, the problem is solved to identify the entire Pareto front. In the constraint-oriented approach, we focus on minimising f_1 while constraining f_2 to stay below a chosen threshold. For each function, we set three threshold values based on the f_2 values on the Pareto front at 25%, 50%, and 75% from the minimum of the Pareto front. Each function was then optimised three times using these different constraint levels. Because the limits are defined this way, the feasible region (solutions that satisfy the constraint) changes for each case. The constraint values used for each function, and the corresponding percentage of feasible solutions in the search space, are given in Table 2. Fig. S2 shows the feasible points within the search space for the corresponding constraint level for each function. As the constraint becomes tighter, the percentage of feasible solutions decreases. This helps to assess the performance of the algorithms in more challenging situations, where the feasible region becomes smaller.

Since the concept of ABC-BO applies only to reaction optimization problems, the *in silico* study is solved only with the unknown constraint. The MATLAB's built-in solver bayesopt is used with its default settings. Each problem has been solved with the budget of 50 iterations for 21 runs.

The results are presented in Fig. S3–Fig. S8. The top row compares the distribution of the best objective values obtained (that satisfy the constraint) at different iteration intervals, shown as box-and-whisker plots for the two approaches. It can be seen that the constraint-oriented approach (using bayesopt) identifies better objective values compared to the Pareto approach. In the Pareto approach, the algorithm focuses on capturing the entire Pareto front, so it overlooks the region of interest. In contrast, the constraint-oriented approach focuses on identifying the best solution within the feasible region (satisfying the constraint), so it is able to find better optimal solutions. The bottom row shows the percentage of runs (out of 21) where the algorithm fails to identify a feasible solution at each iteration interval. As the feasible space within the search space becomes smaller, the constraint-oriented approach has a higher chance of identifying feasible solutions than the Pareto-oriented approach.

Table S1. Analytical functions used for the benchmark study. The Pareto front for each function is identified using 10,00,000 random points within the search space.

Function	Objectives (Min)	Bounds	Pareto front
F1 *Binh and Korn function*	$f_1 = 4x^2 + 4y^2$ $f_2 = (x-5)^2 + (y-5)^2$	$x \in [0, 5]$ $y \in [0, 3]$	
F2 *Chankong and Haimes function*	$f_1 = 2 + (x-2)^2 + (y-1)^2$ $f_2 = 9x - (y-1)^2$	$x \in [-20, 20]$ $y \in [-20, 20]$	
F3 *Fonseca-Fleming function*	$f_1 = 1 - exp\left[-\left(\left(x - \frac{1}{\sqrt{2}}\right)^2 + \left(y -\right.\right.\right.$ $f_2 = 1 - exp\left[-\left(\left(x + \frac{1}{\sqrt{2}}\right)^2 + \left(y +\right.\right.\right.$	$x \in [-4, 4]$ $y \in [-4, 4]$	
F4 *Poloni's two objective function*	$f_1 = \left[1 + (A_1 - B_1(x,y))^2 + (A_2 -\right.$ $f_2 = (x+3)^2 + (y+1)^2$ $A_1 = 0.5\sin(1) - 2\cos(1) + \sin(2$ $A_2 = 1.5\sin(1) - \cos(1) + 2\sin(2$ $B_1(x,y) = 0.5\sin(x) - 2\cos(x) +$ $B_1(x,y) = 1.5\sin(x) - \cos(x) + 2$	$x \in [-\pi, \pi]$ $y \in [-\pi, \pi]$	

F5 *Zitzler-DebThiele's function N 3*	$f_1 = x$ $f_2 = g(x)h(f_1(x), g(x))$ $g = 1 + \frac{9}{29}y$ $h(f_1, g) = 1 - \sqrt{\frac{f_1}{g}} - \left(\frac{f_1}{g}\right)sin(10$	$x \in [0, 1]$ $y \in [0, 1]$	
F6 *Kursawe function*	$f_1 = -10e^{-0.2\sqrt{x^2+y^2}} - 10e^{-0.2}$ $f_2 = \lvert x\rvert^{0.8} + 5\sin(x^3) + \lvert y\rvert^{0.8} + 5$	$x \in [-5, 5]$ $y \in [-5, 5]$ $z \in [-5, 5]$	

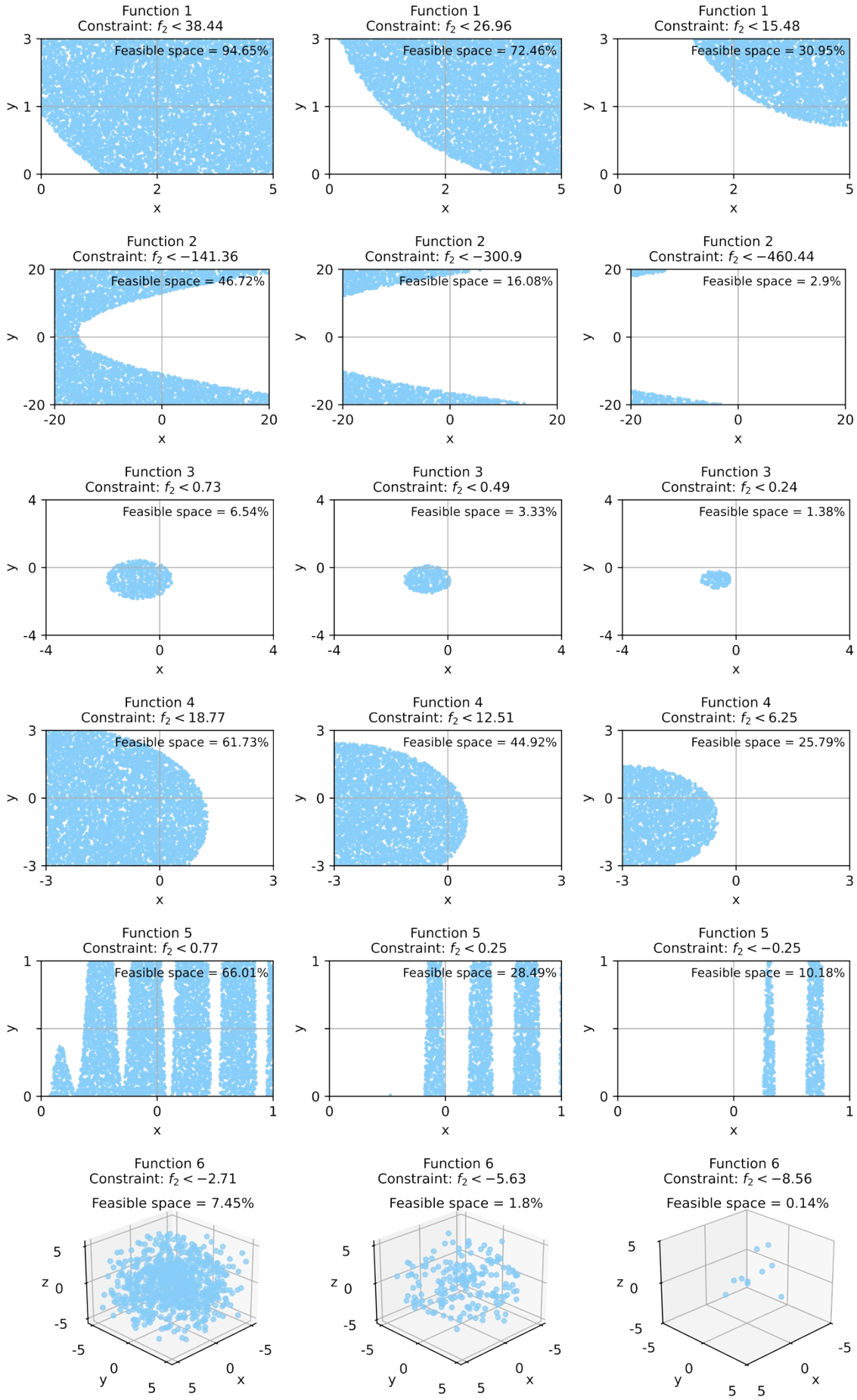

Function 1
Constraint: $f_2 < 38.44$
Feasible space = 94.65%
Function 1
Constraint: $f_2 < 26.96$
Feasible space = 72.46%
Function 1
Constraint: $f_2 < 15.48$
Feasible space = 30.95%
Function 2
Constraint: $f_2 < -141.36$
Feasible space = 46.72%
Function 2
Constraint: $f_2 < -300.9$
Feasible space = 16.08%
Function 2
Constraint: $f_2 < -460.44$
Feasible space = 2.9%
Function 3
Constraint: $f_2 < 0.73$
Feasible space = 6.54%
Function 3
Constraint: $f_2 < 0.49$
Feasible space = 3.33%
Function 3
Constraint: $f_2 < 0.24$
Feasible space = 1.38%
Function 4
Constraint: $f_2 < 18.77$
Feasible space = 61.73%
Function 4
Constraint: $f_2 < 12.51$
Feasible space = 44.92%
Function 4
Constraint: $f_2 < 6.25$
Feasible space = 25.79%
Function 5
Constraint: $f_2 < 0.77$
Feasible space = 66.01%
Function 5
Constraint: $f_2 < 0.25$
Feasible space = 28.49%
Function 5
Constraint: $f_2 < -0.25$
Feasible space = 10.18%
Function 6
Constraint: $f_2 < -2.71$
Feasible space = 7.45%
Function 6
Constraint: $f_2 < -5.63$
Feasible space = 1.8%
Function 6
Constraint: $f_2 < -8.56$
Feasible space = 0.14%

Fig. S2 Feasible space within the search space that satisfies the corresponding constraint value. The blue points represent the conditions that satisfy the constraint out of 10,000 random points sampled within the search space.

Table S2. Constraints applied to f_2 in the constraint-oriented approach while minimising f_1. The percentage of feasible solutions within the search space that satisfy the constraint is shown. The feasible solutions are identified using 100,000 random points (21 runs) within the search space.

Function	Constraint		Feasible space (%)
F1	$f_2 <$	38.44	94.65±0.05
		26.96	72.46±0.13
		15.48	30.95±0.15
F2	$f_2 <$	-141.36	46.72±0.10
		-300.90	16.08±0.09
		-460.44	2.90±0.05
F3	$f_2 <$	0.73	6.54±0.08
		0.49	3.33±0.06
		0.24	1.38±0.04
F4	$f_2 <$	18.77	61.73±0.13
		12.51	44.92±0.16
		6.25	25.79±0.11
F5	$f_2 <$	0.77	66.01±0.10
		0.25	28.49±0.12
		-0.25	10.18±0.07
F6	$f_2 <$	-2.71	7.45±0.08
		-5.63	1.80±0.03
		-8.56	0.14±0.01

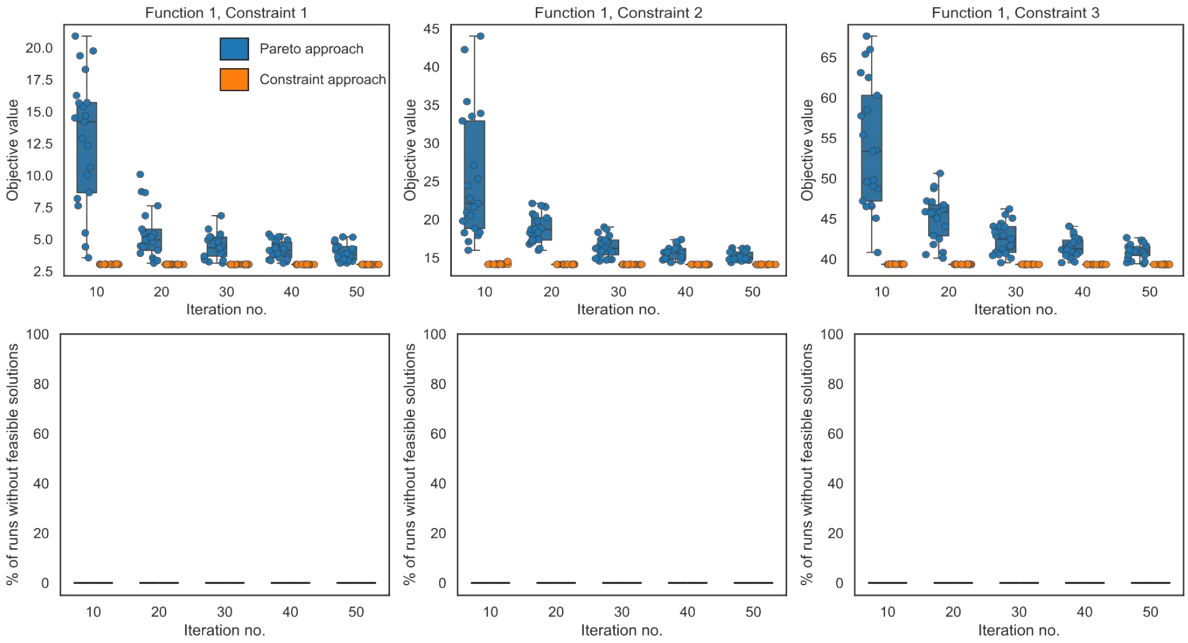


Fig. S3 Results on optimising Function 1 for different constraint values. The top row shows the box-and-whisker plots comparing the distribution of the best objective values (that satisfy the constraint) obtained over 21 runs for the two approaches. The bottom row shows the bar plots representing the percentage of infeasible solutions (solutions that do not satisfy the constraint) out of the 21 runs.

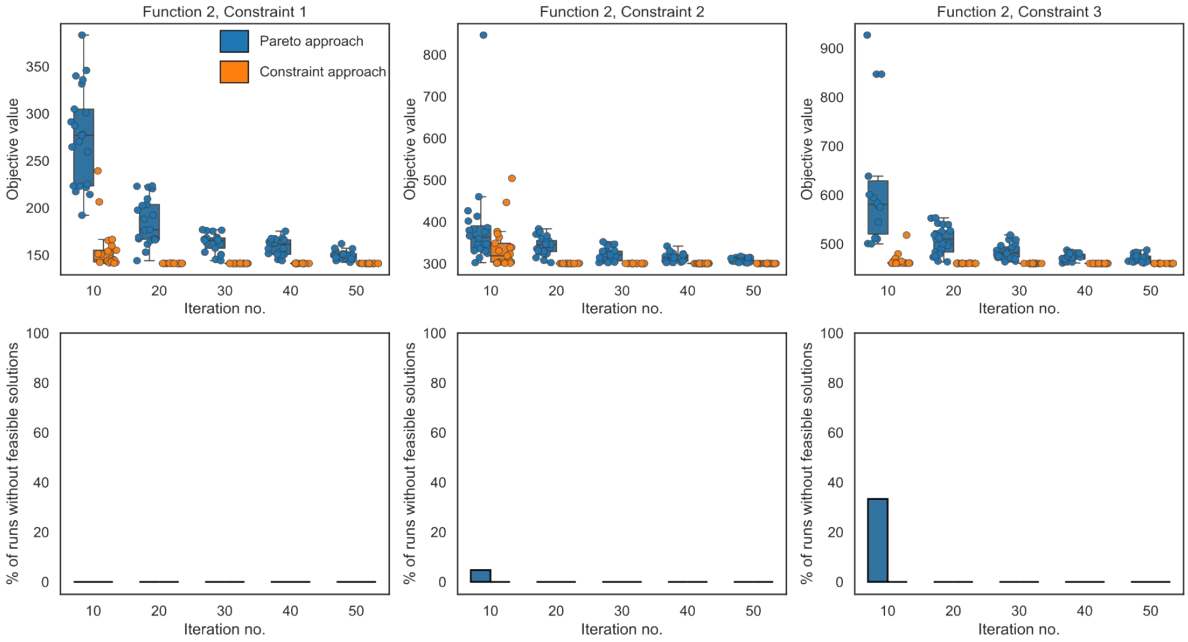


Fig. S4 Results on optimising Function 2 for different constraint values. The top row shows the box-and-whisker plots comparing the distribution of the best objective values (that satisfy the constraint) obtained over 21 runs for the two approaches. The bottom row shows the bar plots representing the percentage of infeasible solutions (solutions that do not satisfy the constraint) out of the 21 runs.

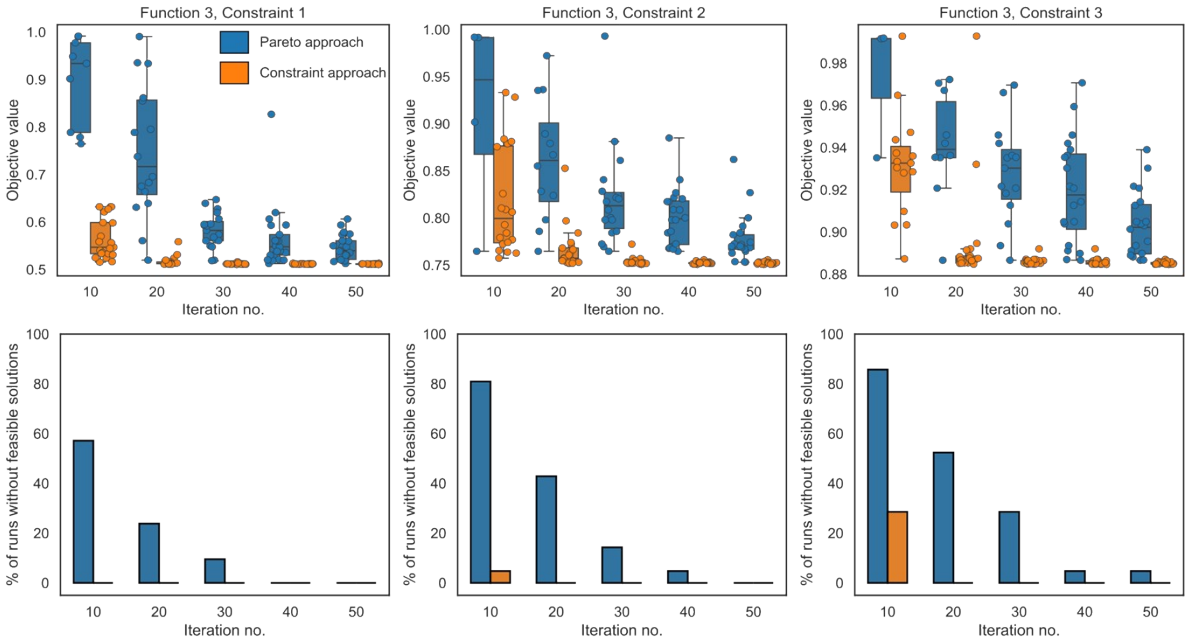


Fig. S5 Results on optimising Function 3 for different constraint values. The top row shows the box-and-whisker plots comparing the distribution of the best objective values (that satisfy the constraint) obtained over 21 runs for the two approaches. The bottom row shows the bar plots representing the percentage of infeasible solutions (solutions that do not satisfy the constraint) out of the 21 runs.

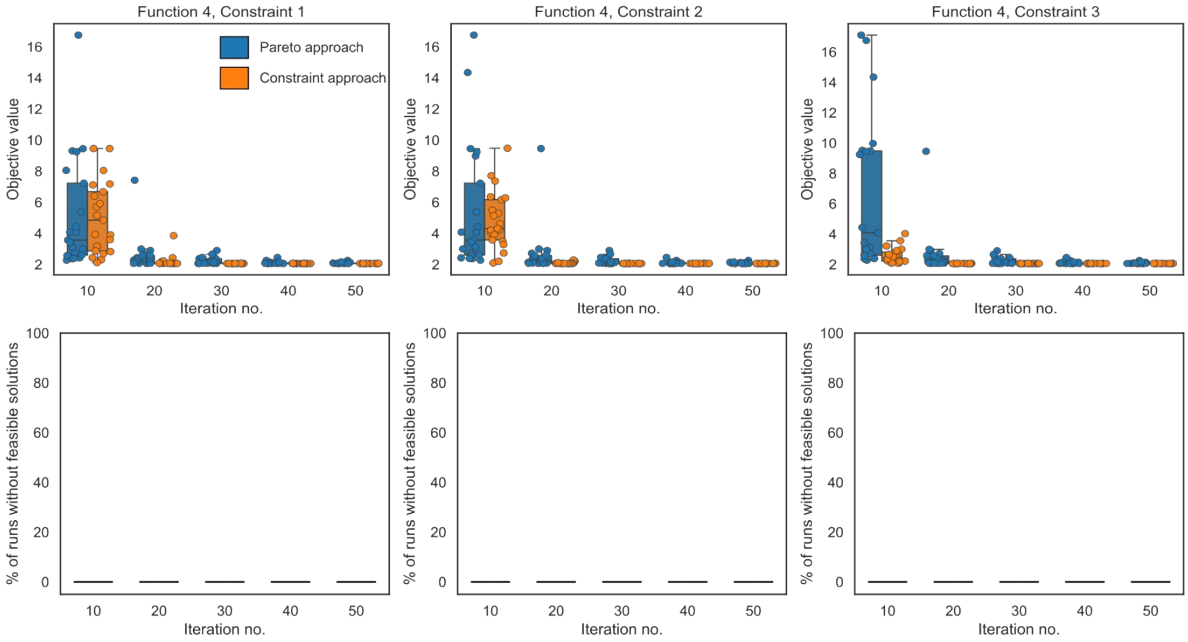


Fig. S6 Results on optimising Function 4 for different constraint values. The top row shows the box-and-whisker plots comparing the distribution of the best objective values (that satisfy the constraint) obtained over 21 runs for the two approaches. The bottom row shows the bar plots representing the percentage of infeasible solutions (solutions that do not satisfy the constraint) out of the 21 runs.

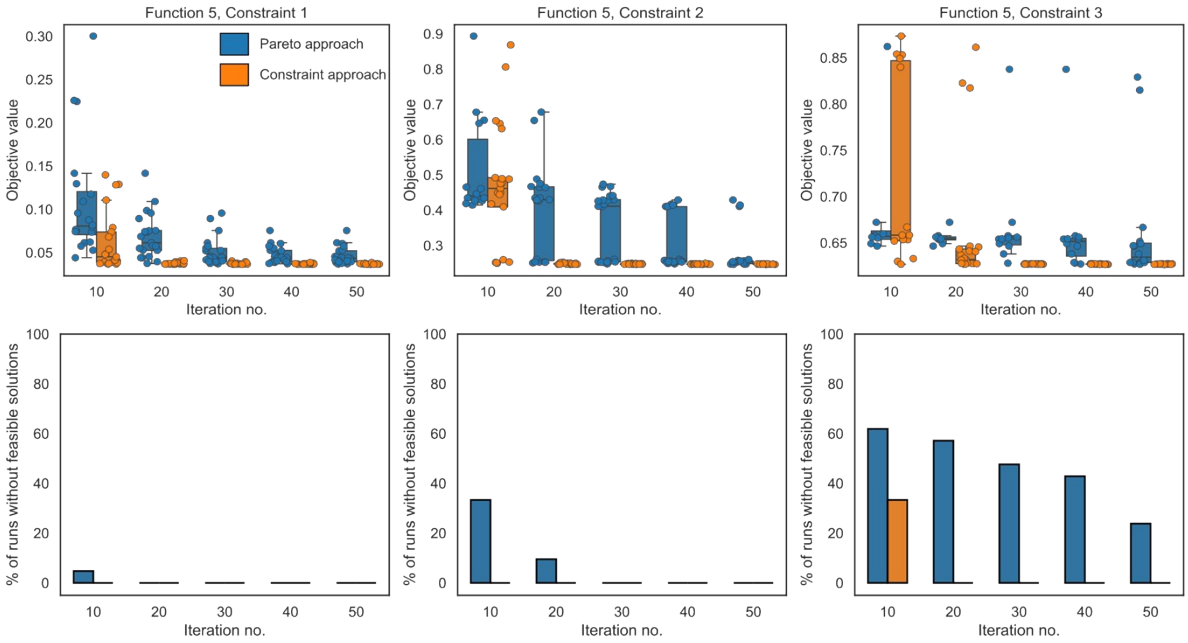


Fig. S7 Results on optimising Function 5 for different constraint values. The top row shows the box-and-whisker plots comparing the distribution of the best objective values (that satisfy the constraint) obtained over 21 runs for the two approaches. The bottom row shows the bar plots representing the percentage of infeasible solutions (solutions that do not satisfy the constraint) out of the 21 runs.

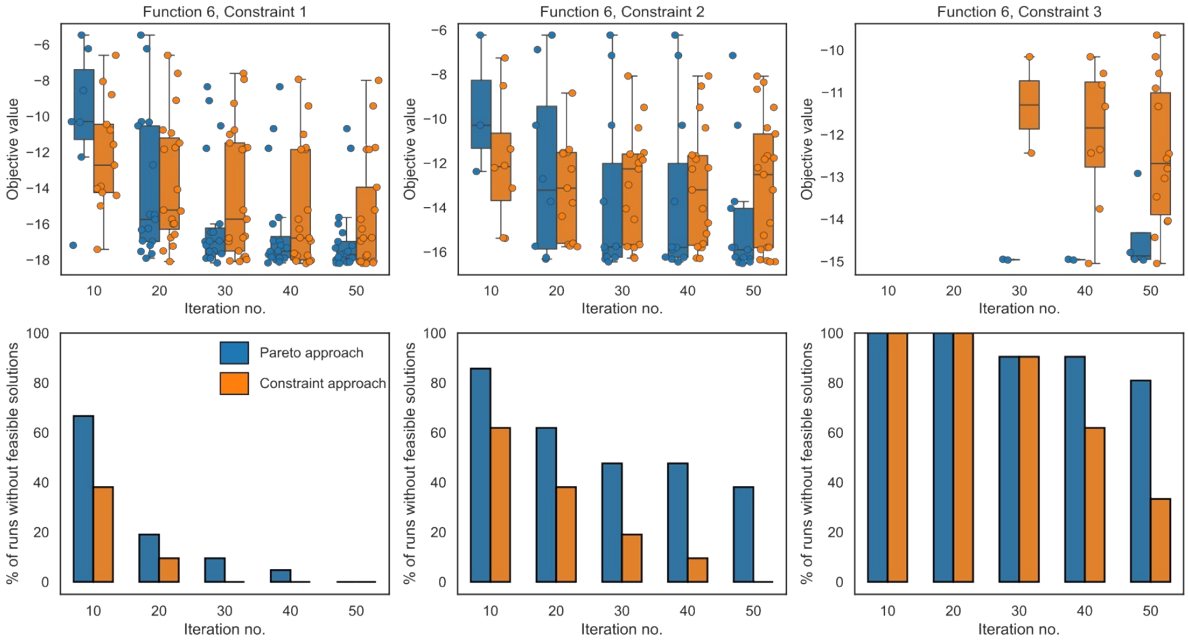


Fig. S8 Results on optimising Function 6 for different constraint values. The top row shows the box-and-whisker plots comparing the distribution of the best objective values (that satisfy the constraint) obtained over 21 runs for the two approaches. The bottom row shows the bar plots representing the percentage of infeasible solutions (solutions that do not satisfy the constraint) out of the 21 runs.

4.2. Experimental results

Table S3 Optimization results for MVMOO

Exp. No.	Reagent eq.	Res. time	Temp.	Yield	Throughput	
	$eq.$	min	^oC	%	g/h	
1	1.94	7.5	31.9	4.5	0.20	Sampling : LHS
2	1.64	58.0	48.1	75.1	0.48	
3	1.02	37.9	23.0	58.9	0.74	
4	1.02	49.3	22.6	62.2	0.60	Optimization : MVMOO
5	1.09	60.0	39.0	85.1	0.65	
6	1.00	52.3	43.1	85.6	0.79	
7	1.00	60.0	50.0	81.0	0.65	
8	1.00	60.0	41.5	85.4	0.68	
9	1.00	43.2	50.0	85.5	0.95	
10	1.00	34.7	50.0	82.5	1.14	
11	1.00	14.7	37.5	69.1	2.25	
12	1.00	29.7	49.3	92.0	1.48	
13	1.00	10.1	50.0	88.3	4.18	
14	1.20	5.7	46.4	62.3	4.81	
15	1.12	5.0	50.0	84.4	7.67	

The constraint used for ABC-BO are defined as follows:

For Phase 1:

$Throughput^{max.\ acheivable} > Throughput^*$

where $Throughput^*$ is the best throughput adjusted for yield noise (5%)

For Phase 2:

$Yield \geq 85\%$

$Throughput^{max.\ acheivable} > Throughput_{85\%\ yield}^*$

Where $Throughput_{85\%\ yield}^*$ is the best throughput that satisfies the yield constraint, adjusted for yield noise (5%)

Table S4. Optimization results for ABC-BO

Exp. No.	Reagent eq.	Res. time	Temp.	Yield	Throughput	
	$eq.$	min	^oC	%	g/h	
1	1.50	32.5	35.0	71.5	0.86	Sampling : Center point
2	1.79	15.6	42.0	49.5	1.11	
3	1.41	12.7	37.1	57.0	1.80	
4	1.66	20.5	22.1	19.9	0.35	
5	1.36	10.7	32.2	35.8	1.38	Optimization:
6	1.00	13.4	48.6	94.0	3.39	ABC-BO
7	1.21	5.0	50.0	76.8	6.68	Phase 1
8	1.97	5.0	49.9	16.9	1.11	
9	1.31	6.1	50.0	73.4	5.03	
10	1.18	6.4	44.6	74.3	5.12	
11	1.18	13.1	50.0	94.9	3.20	
12	1.04	8.5	49.9	79.2	4.41	Optimization:
13	1.13	11.2	49.9	96.7	3.92	ABC-BO
14	1.13	9.1	49.7	86.2	4.30	Phase 2
15	1.15	9.0	49.5	81.1	4.01	

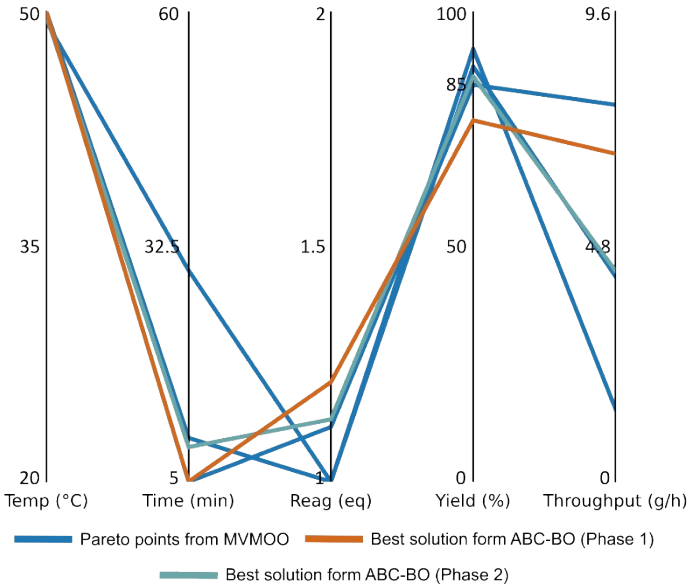


Fig. S9 Optimization results presented as a parallel coordinate plot. Three Pareto solutions were obtained from MVMOO. The best solution from ABC-BO in phase 1 corresponds to the maximum throughput, while the best solution from ABC-BO in phase 2 corresponds to the maximum throughput satisfying the yield constraint (≥85%).

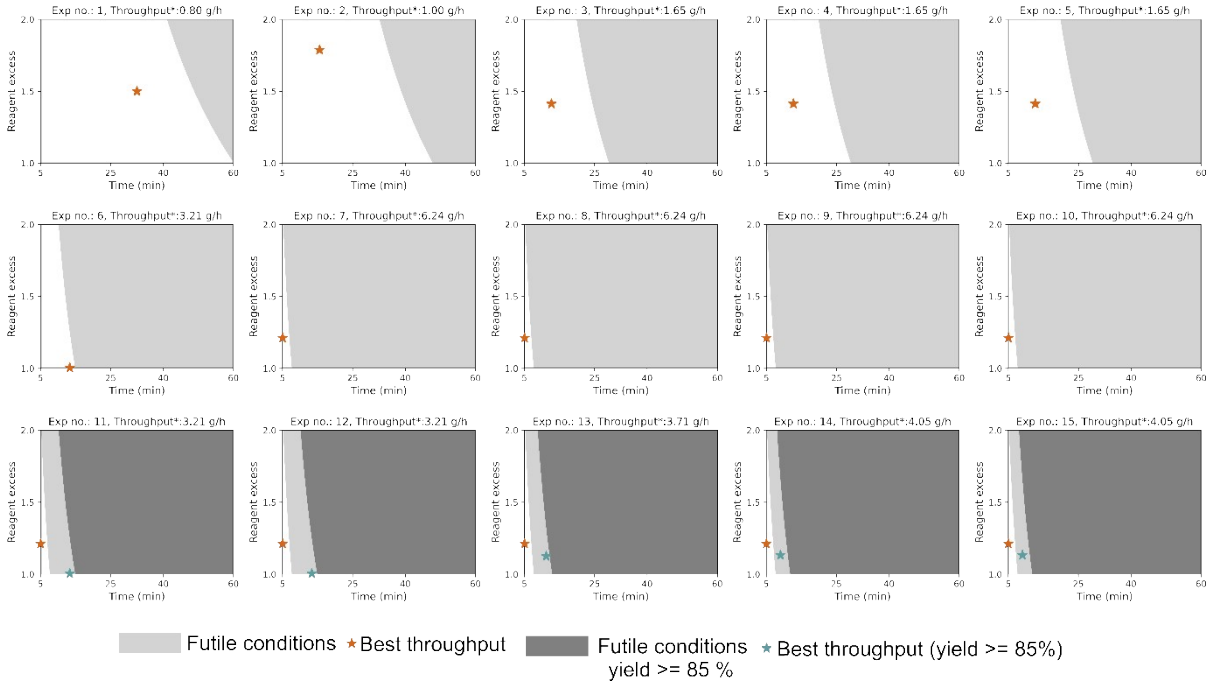


Fig. S10 Evolution of the futile space within the search space, represented in terms of residence

time and reagent equivalent, for the problem solved using ABC-BO. For the first 10 experiments (phase 1), the orange star represents the condition corresponding to the best throughput (adjusted) at that point, and the light grey regions indicate the futile conditions associated with that best solution. For the last 5 experiments (phase 2), the blue star represents the condition corresponding to the best throughput (adjusted) that satisfy the yield constraint

(≥85%), and the dark grey regions indicate the futile conditions for this best solution. The difference between the dark grey and light grey regions illustrates the relaxation in the restricted search space due to the additional constraint applied in this phase of optimization.

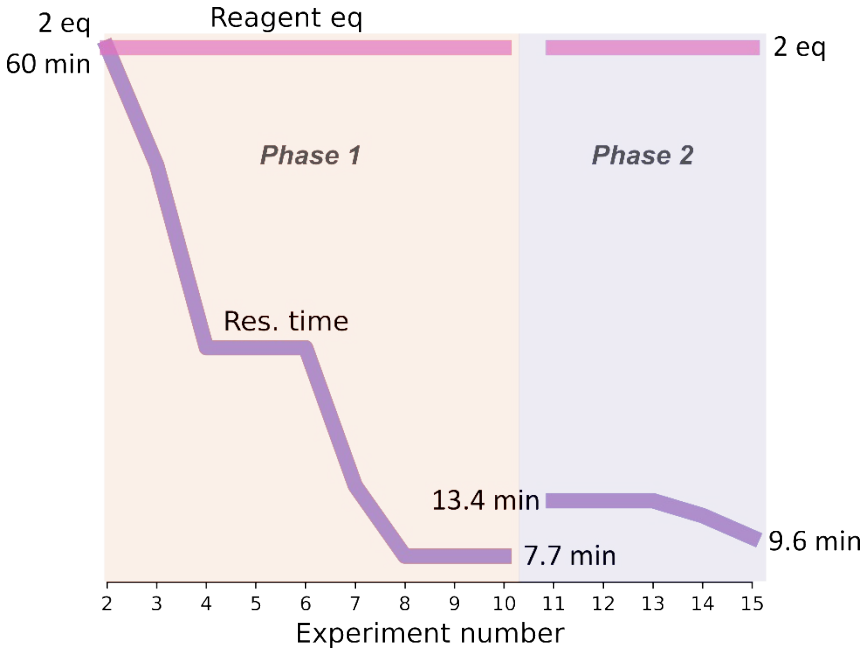


Fig. S11 Boundary reduction of the variables (reagent excess and residence time) during optimization using ABC-BO. The boundaries of these variables are redefined to eliminate futile experiments. For phase 1, the refinement is based on the best adjusted throughput, and for phase 2, it is based on the best throughput that satisfies the yield constraint (≥85%).

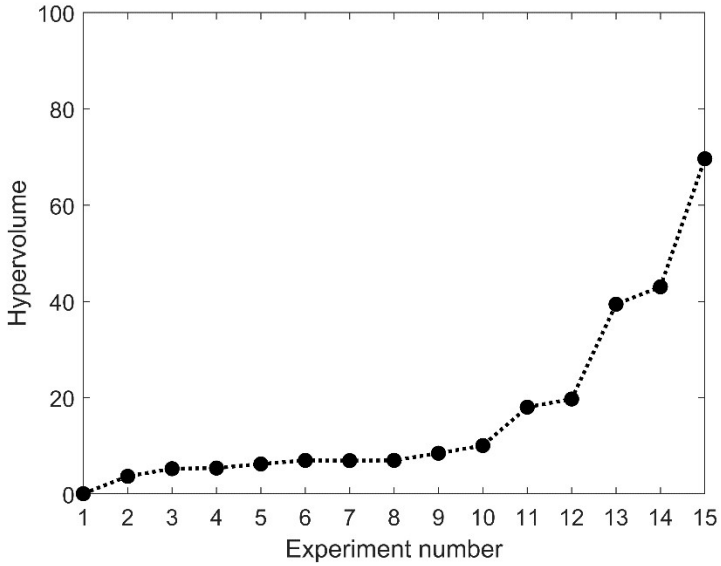


Fig. S12 Hypervolume trend during optimization using MVMOO.

References

1 J. A. Manson, T. W. Chamberlain and R. A. Bourne, *J. Glob. Optim.*, 2021, **80**, 865–886.

2 D. Zhan, Y. Cheng and J. Liu, *IEEE Trans. Evol. Comput.*, 2017, **21**, 956–975.

3 A. Senthil Vel, J. Spils, D. Cortés-Borda and F.-X. Felpin, *React. Chem. Eng.*, 2025, **10**, 2137–2147.

4 M. A. Gelbart, J. Snoek and R. P. Adams, *arXiv*, 2014, preprint, arXiv:1403.5607, DOI: 10.48550/arXiv.1403.5607.