

# Digital Discovery

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: F. Wang, M. Parhizkar, A. Harker and M. Edirisinghe, *Digital Discovery*, 2025, DOI: 10.1039/D5DD00243E.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

# Constrained composite Bayesian optimization for rational synthesis of polymeric particles

Fanjin Wang<sup>1,2,\*</sup>, Maryam Parhizkar<sup>2</sup>, Anthony Harker<sup>3</sup>, Mohan Edirisinghe<sup>1</sup>

<sup>1</sup>Department of Mechanical Engineering, University College London, London, WC1E 7JE, United Kingdom

<sup>2</sup>School of Pharmacy, University College London, London, WC1N 1AX, United Kingdom

<sup>3</sup>Department of Physics and Astronomy, University College London, London, WC1E 6BT, United Kingdom

\*Corresponding author, email: [fanjin.wang.20@ucl.ac.uk](mailto:fanjin.wang.20@ucl.ac.uk)

## Abstract

Polymeric nanoparticles have critical roles in tackling healthcare and energy challenges with miniature characteristics. However, tailoring synthesis processes to meet design targets has traditionally depended on domain expertise and trial-and-errors. Modeling strategies, particularly Bayesian optimization, facilitated materials discovery with maximized/minimized properties. Coming from practical demands, this study integrates constrained composite Bayesian optimization (CCBO) to perform target-value optimization under black-box feasibility constraints for nanoparticle production-by-design. With a synthetic problem that simulates electrospraying, a representative nanomanufacturing process, CCBO avoided infeasible conditions and efficiently optimized towards predefined size targets, surpassing baseline methods and state-of-the-art optimization pipelines. CCBO was also observed to provide comparable decisions as experienced experts in a human vs. BO campaign. Furthermore, laboratory experiments validated CCBO for guided synthesis of poly(lactic-co-glycolic acid) particles with diameters of 300 nm and 3.0  $\mu\text{m}$  *via* electrospraying under minimal initial data. Overall, the CCBO approach presents a versatile and holistic optimization paradigm for next-generation target-driven particle synthesis empowered by artificial intelligence (AI).



## Introduction

View Article Online  
DOI: 10.1039/D5DD00243E

Polymeric micro- and nano-particles have received much attention in pharmaceuticals, catalysis, and energy applications due to their unique properties at a small scale (1,2). Diverse design requirements for particles, under the quality-by-design (QbD) framework, have been put forward by specific usages (3). For example, particles as drug delivery platforms span diverse sizes across hundreds of nanometers for intravenous injection to micrometers for pulmonary administration (4). However, the optimization of synthesis to meet these design requirements, within any manufacturing technology used, has mainly relied on human expertise and extensive trial-and-error experimentation. Modeling strategies could facilitate the optimization of parameters towards design targets (5,6). Traditional design of experiment (DoE) strategies can identify dominating factors in the processing parameters and direct towards optimum, but the methodology becomes less effective in high-dimensional problems or complex relationships (7). For example, orthogonal experiment designs, such as Plackett-Burman and Taguchi methods, can typically accommodate up to three levels for each variable (8). Plus, it is also difficult to incorporate experiment feasibility into DoE optimization frameworks unless analytical descriptions of constraints are available. As a different approach, machine learning (ML) is powerful in modeling complicated relationships (9,10). Using ML models as surrogates, adaptive sampling methods design sequential experiments for laboratory evaluation (11). Bayesian optimization (BO) was developed for efficient optimization of black-box functions, and worked well under small data regime (12–14). It employs a Gaussian process (GP) as a surrogate model, leveraging its ability to provide both mean and variance estimations for candidate selection. A carefully designed acquisition function is then used to score the candidates to explore uncertain points as well as exploit promising optimal points.

More recently, BO has been investigated for materials and drug discovery, assisting the identification of optimal properties (15–17). However, two critical challenges were presented for the application of BO in targeted synthesis of materials. First, conventional BO is developed to seek for a global maximum or minimum, rather than matching a pre-defined target (18,19). The latter has occurred as a common requirement in materials development tasks such as matching physiological mechanical properties for hydrogels and tailoring releasing profiles for drug delivery agents. Despite its relevance, the target-matching problem remains surprisingly underexplored in BO applications for materials discovery. This may be attributed to the prevailing emphasis on discovering materials with extreme or superior properties, rather than those that meet specific design criteria. Another issue is associated with feasibility constraints in experimentation. The majority of current applications of BO within materials discovery and development did not incorporate feasibility (17,20,21). Nevertheless, practical concerns could



emerge in BO recommendations due to a myriad of reasons in laboratory experiments, such as impossible combination of material compositions, incompatible processing parameters, and limitations from apparatus. Shrinking variables' boundaries to a practical region could be a direct solution at the cost of reduced search space. In a special case where known constraints of input variables are available (e.g., as inequality equations), optimization could be performed subjected to these constraints (22,23). For example, Li *et al.* nested an active learning loop for constraint modelling to restrict the candidate space selectable by BO (24). Low *et al.* suggested evolution-guided Bayesian optimization, which impose known constraints to multi-objective optimization problem for nanoparticle synthesis in microfluidics (25). However, these strategies become impossible when these constraints become unknown *a priori* and need to be evaluated through laboratory experiments.

Several prior works on constrained and composite BO respectively have explored applications in hyperparameter tuning. In constraint BO, Gramacy and Lee proposed to weight the expected improvement (EI) acquisition function with a modelled probability to enforce preference for feasible candidates (26). Gardner *et al.* extended this approach to inequality constraints, assuming the feasibility could be derived from a continuous-valued constraint function (27). More recently, Tian *et al.* proposed boundary exploration method that relaxes acquisition function weights to encourage exploration near constraint boundaries (28). For composite BO, Uhrenholt and Jensen investigated target value optimization, specifically minimizing a 2-norm, by warping the GP to a noncentral chi-squared distribution (29). As an improvement, Astudillo and Frazier approached a more general problem of composite BO for arbitrary composite function over the objective function. They transformed the Gaussian posterior in the acquisition function directly with the composited function within the acquisition function (30). Although these strategies have been rigorously tested on synthetic benchmarks and hyperparameter optimization tasks, they have yet to be integrated into a combinatorial framework to facilitate guided laboratory experiments.

Here, we implement a constrained composite Bayesian optimization (CCBO) pipeline showcasing efficient identification of suitable processing parameters in rational synthesis of polymeric particles. Through introducing a variational inference GP component, the black-box experiment feasibility was modeled and incorporated into BO acquisition function. Composite BO, on the other hand, handles the modeling of experimental parameters and targeting particle size through a composite objective function. Amongst various fabrication techniques of particles, electrospraying was selected as the model technique for its simplicity, versatility, and precision as a popular manufacturing method in drug delivery research (31). It utilizes electric fields to deform the meniscus of polymer solution to form fine jets which eventually disintegrate into fine droplets. As these droplets travel towards a collector, they further shrink



and solidify due to solvent evaporation. Processing parameters in electrospraying such as flow rate, voltage, polymer concentration, and the solvent could be adjusted to tailor product characteristics, although the intertwined impact of these factors could lead to prolonged, if not infeasible, trial-and-errors (32). With CCBO, we demonstrate its superior performance in target parameter optimization compared with random baseline and conventional BO strategies through both synthetic data and wet-lab experiments of poly(lactic-co-glycolic acid) (PLGA) particles synthesis at multiple size targets.

## Methods

**Materials.** PLGA (PURASORB PDLG 5004A, 50:50 ratio) was purchased from Corbion (Amsterdam, The Netherlands). Chloroform and N, N-Dimethylacetamide (DMAc) were purchased from Sigma-Aldrich (Gillingham, UK).

**Electrospraying production of particles.** PLGA solutions were prepared by mixing PLGA granules with solvents at ambient temperature with magnetic stirring overnight. The solutions were fed by a syringe pump (Harvard PHD Ultra, Edenbridge, UK) to a 22-gauge needle (outer diameter 0.71 mm) through a capillary. The positive output of a high voltage power supply (Glassman High Voltage Inc., NJ, United States) was connected to the needle through a crocodile clamp and the collection plate was connected to the ground. Before electrospraying, the flow rate and voltage were adjusted to the values recommended by BO. Experiments were conducted at atmospheric pressure. The temperature and humidity in the room were controlled to be 19-22 °C and 40-50%.

**Characterization of particles.** Particles were collected on a glass slide placed on the collection plate for Scanning Electron Microscopy (SEM) analysis. Zeiss Gemini 360 SEM (Germany) was used under an acceleration voltage of 1.0 kV with an SE2 detector. For each sample, three images were taken randomly at different locations. Images were further analyzed using ImageJ (National Institute of Health, USA). To obtain mean particle size, a hundred particles were randomly measured for their diameters. For infeasible experiments, the diameters of splashes from undried droplets on the collecting glass slides were recorded as a measurement of size.

**Constrained composite Bayesian optimization.** Two components were incorporated in the BO pipeline and were developed under the framework of BoTorch (33) and GPyTorch (34). The objective component, which tracked the distance (or particle size in the case of CCBO), followed the classical design of BO (see **Supplementary Note 1** for details of handling categorical inputs) (15). Notably, due to the difficulty in determining the noise level in experiments, we assumed the input data from laboratory experiments, after averaging over triplicates, to be noiseless. In terms of the acquisition function,  $q$ -Expected Improvement ( $qEI$ ,



or batch EI) as a thoroughly investigated strategy was selected to allow consideration of multiple candidates jointly in each iteration (18). In its simplistic form where  $q$  equals 1, EI acquisition function at a single point  $x_0$  can be given by  $\alpha_{\text{EI}}(x_0) = \mathbb{E}[\max(y_o - f^*, 0)]$ , where  $y_o \sim \mathcal{N}(\mu(x_0), \sigma^2(x_0))$  with  $\mu(x_0)$  and  $\sigma^2(x_0)$  being the posterior mean and variance from the Gaussian process at  $x_0$ ,  $f^*$  is the current best observation. As the calculation of expectation requires integrating over the posterior, it becomes analytically intractable under a batched scenario where  $q > 1$ . We followed the strategy in BoTorch where Monte-Carlo sampling was used to approximate the expectation by:

$$\alpha_{\text{qEI}}(\mathbf{X}) \approx \frac{1}{N} \sum_{i=1}^N \max_{j=1, \dots, q} [\max(y_{o,ij} - g^*, 0)], y_{o,ij} \sim \mathbb{P}(GP(\mathbf{X})|\mathcal{D}) \quad (1)$$

where  $N$  was the total number of Monte-Carlo sampling,  $q$  was the number of candidates to be evaluated in parallel, and  $y_{o,ij}$  was sampled through the reparameterization trick from the Gaussian process conditioned on data  $\mathcal{D}$ ,  $g^*$  represents the current closest distance (with respect to the target) achieved. Notably, the data  $\mathcal{D}$  consisted of  $\{(x_i, y_{o,i})\}_{i=1}^n$  where  $y_{o,i} = g(s_i) = -(s_i - s_o)^2$  with  $s_o$  representing the target value. Under such configurations, this vanilla BO pipeline could help identify suitable experiment variables  $\mathbf{X}$  that can maximize this negative distance measure  $y_o$ .

Furthermore, the feasibility component was introduced to learn black-box constraints in the experiment. Here, a variational Gaussian process was implemented for the binary classification of experimental success or failure (34). The details for variational inference for Gaussian classification were described in previous publications (35). Briefly, the latent Gaussian process is further wrapped with a Probit regression to limit the output between 0 and 1, for the purpose of approximating a Bernoulli posterior. For our latent Gaussian process, it followed the same constant mean prior and kernel functions to incorporate mixed inputs. To incorporate feasibility modelling in the Bayesian optimization process, we followed the strategy proposed earlier (26) to extract the posterior probability as a scaling factor in the acquisition function:  $\alpha_{\text{qEIcon}}(\mathbf{X}) = \mathbb{P}(y_c = 1 | \mathbf{X}) * \alpha_{\text{qEI}}(\mathbf{X})$ . Incorporating this factor in the acquisition function allowed the suppression of the value of experiments that are potentially infeasible, creating our constrained BO pipeline.

Both the vanilla and constrained BO pipelines had the Gaussian process modeling  $y_o$  and attempted to minimize this distance. As a different strategy, composite BO used a Gaussian process to directly model the particle size  $s$ . The composite part, namely the negative squared distance function  $g$ , was separated from the input data. Instead, the distance function was directly applied to the Gaussian posterior in the acquisition function (36):





$$\alpha_{\text{qEICF}}(\mathbf{X}) \approx \frac{1}{N} \sum_{i=1}^N \max_{j=1, \dots, q} [\max(g(s_{ij}) - g^*, 0)], s_{ij} \sim \mathbb{P}(GP(\mathbf{X}) | \mathcal{D}') \quad (2)$$

View Article Online  
DOI: 10.1039/D5DD00243E

where  $s_{ij}$  was sampled through the reparameterization trick and  $\mathcal{D}' = \{(x_i, s_i)\}_{i=1}^n$ . When coupling the composite acquisition function  $\alpha_{\text{qEICF}}$  with the constraint probability, we have the acquisition function for CCBO:  $\alpha_{\text{qEICFcon}}(\mathbf{X}) = \mathbb{P}(y_c = 1 | X) * \alpha_{\text{qEICF}}(\mathbf{X})$ .

In the present work, the Monte-Carlo sampling number  $N$  was 512 and  $q$  was fixed to 2 throughout all BO pipelines. All input  $\mathbf{X}$  were normalized to unit cubes, and the flow rate variable was transformed to logarithm before normalization. The outcomes of the objective component, including the distance variable,  $y_o$ , in vanilla BO and constrained BO, as well as the particle size variable,  $s$ , in CCBO, were standardized to zero mean and unit variance. The outcomes of the feasibility component,  $y_c$ , were rescaled to  $\{-1, 1\}$ .

**Synthetic electrospray data generation.** The synthetic data of electrospray was generated through the following functions:

$$s = 2 * \frac{\sqrt{Qc}}{\log(U)} + \alpha + 0.4 \quad (3)$$

$$y_c = \begin{cases} 1, & \text{if } \log(Q) * (\alpha - 0.5) + 1.4 > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

where  $s$  is the particle size ( $\mu\text{m}$ ),  $Q$  is the flow rate ( $\mu\text{L min}^{-1}$ ),  $c$  is the concentration of the polymer solution (% w/v),  $U$  is the applied voltage (kV). The  $\alpha$  is a constant depending on the solvent ( $\text{CHCl}_3$ : 1, DMAc: 0).

**Validating BO with synthetic data.** The targeting particle size  $s_o$  was arbitrarily set to be 0.6, 3.0, 6.0 and 18.0  $\mu\text{m}$  to validate BO performance. In each run, three BO pipelines and the random baseline were performed for 10 iterations with the starting data listed on **Table 1**. The outcomes of experiments were calculated by synthetic **equations (3) and (4)** from the corresponding experimental variables. Each run was repeated 20 times to account for variations. The comparison between human and BO followed similar settings used previously. The starting data (**Table 1**) was first shown to the participants ( $N=14$ ) with varying experience in electrospraying, including advanced users with more than 3 years' experience ( $N=4$ ), intermediate users between 1-3 years' experience ( $N=4$ ), and beginners with less than 1 years' experience ( $N=6$ ). At each iteration, two experiments were recommended by participants to optimize towards a 3.0  $\mu\text{m}$  target, followed by revealing experiment results calculated by the synthetic equations. In total, five iterations were performed for human vs BO campaign considering that CCBO have achieved significant reduction within a few rounds. During the





campaign, the participants were not allowed to access each other's results except for the five initial data provided. They were asked to follow a similar pattern to work out the recommendations as if they were dealing with a new electrospray setup. No strict time constraints were imposed on the participants to provide answers, but a typical time of 2 minutes was observed for participants to suggest experiments for one iteration.

The regret, defined by the closest distance towards the targeting particle size, was plotted at each iteration. The experimental variables proposed in a typical run were visualized on 3D plots with symbols representing solvent and feasibility, and colors encoding the iteration. The area under a curve (AUC) of each strategy and human participant was calculated based on trapezoid rules for quantitative comparison. One-tailed Mann-Whitey U-tests were performed with alternative hypothesis being CCBO had smaller AUC/regret compared to BO baselines or human groups, respectively.

**Table 1.** Boundaries of Experiment Variables for BO and the Starting Data for Synthetic Experiments.

Label	Polymer Concentration (% w/v)	Flow Rate ( $\mu\text{L min}^{-1}$ )	Voltage (kV)	Solvent
Bounds	[0.05-5.00]	[0.01-60.00]	[10.0-18.0]	{CHCl <sub>3</sub> , DMAc}
S-1	0.50	15.00	10.0	DMAc
S-2	0.50	0.10	10.0	CHCl <sub>3</sub>
S-3	3.00	20.00	15.0	DMAc
S-4	1.00	20.00	10.0	CHCl <sub>3</sub>
S-5	0.20	0.02	10.0	CHCl <sub>3</sub>

**Guiding laboratory experiments with CCBO.** The boundaries of experimental variables remained the same as the validation with synthetic data. The starting eight experiments were generated through a Sobol sequence within boundaries for each variable. The targeted particle sizes were 300 nm and 3.0  $\mu\text{m}$  based on domain expertise in drug delivery. The two experiments in each iteration were performed in triplicates. The results were fed back to the BO pipeline to obtain the next recommendations. The stopping criterion was set as achieving  $\pm 10\%$  to the targeting size.

## Results

**Validating CCBO through synthetic data.** Performance of CCBO was first validated with synthetic experimental data. Before introducing the benchmark results, three configurations of BO pipelines tested in this study are presented (**Fig. 1a**). More details of the implementation can be found in the method section. Briefly, the vanilla BO pipeline followed a traditional BO design where the target to be maximized was the negative squared distance  $y_o$ . The feasibility component, which leveraged a variational GP for classification, was added to track



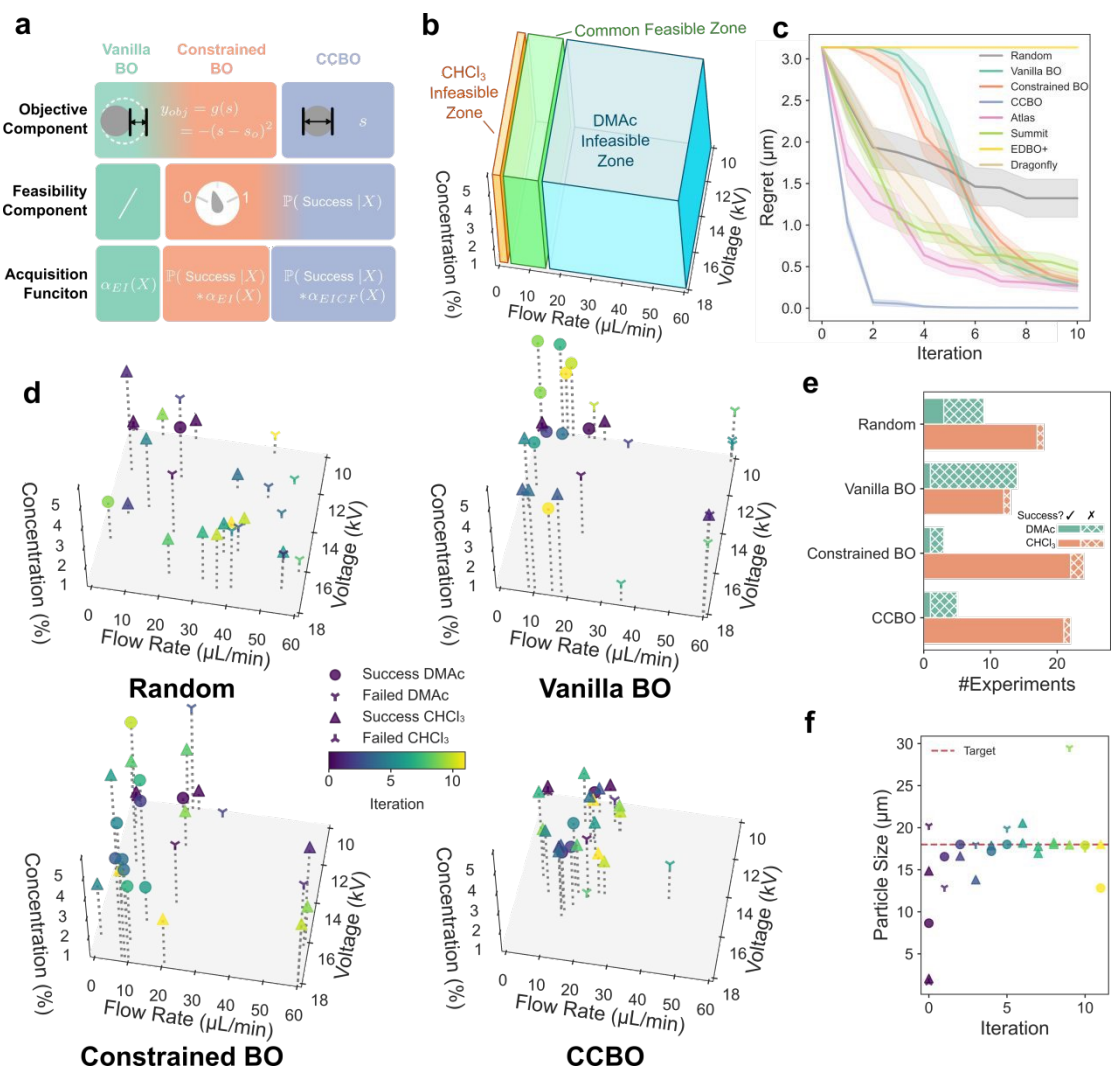
experimental feasibility. Through factoring in a probability term into the acquisition function, the constrained BO pipeline was able to pick candidates with higher chance of success. CCBO adopted the same feasibility modelling, whilst modifying the objective component. It utilized GP to model the fundamental relationship in experiments between the processing variables  $x$  with the size  $s$ . The negative squared distance function was incorporated in the acquisition function to prioritize candidates for minimizing the distance to the pre-set target. In terms of the synthetic problem, the data was produced by equations simulating electrospray processing. Specifically, the function for determining the size of electrosprayed particles (see **equation (3)**) was inspired by scaling laws proposed for electrospray and experimental observations, where flow rate and polymer concentration (through affecting the viscosity) are both positively correlated to the diameter with voltage having a negative impact (32,37). Logarithm and power transformations in the function were intended to add complexity in the modeling process to simulate the nonlinear nature of the electrospraying process. The constant for alpha was added to account for the impact of solvents considered in the process. In addition, the feasibility zone, as visualized in **Fig. 1b**, was set to be highly related to the flow rate and the solvent. The rationale was from practical considerations where chloroform, as a highly volatile solvent, would result in clogged nozzle at lower flow rates. DMAc, at higher flow rates, would lead to insufficient evaporation of the solvent and produce splashes of droplets on the collector instead of solid particles.

As a benchmark, CCBO, together with random baseline, vanilla BO, and constrained BO only, were performed for 10 iterations. Five initial experiments were included, accounting for successful and failed cases for both solvents. The optimization target was set to 18  $\mu\text{m}$ . Results for other target sizes, including 0.6, 3 and 6  $\mu\text{m}$ , can be found in **Supplementary Fig. 1**. In each iteration, two sets of processing parameters were proposed and subjected to simulation functions to retrieve the synthetic experimental result as well as the feasibility. The regret, defined as the difference between the target and the closest candidate, was recorded after each iteration as a measurement of performance (**Fig. 1c**). After 10 iterations, the random baseline reached 0.8  $\mu\text{m}$  regret. Similarly, the vanilla BO and constrained BO both achieved around 0.4  $\mu\text{m}$  regret. By contrast, the CCBO algorithm rapidly converged to the targeted diameter after only two iterations. Moreover, AUC of each strategy, using trapezoidal method, was calculated to quantify the optimization efficiency. CCBO achieved a minimal AUC of  $2.47 \pm 0.85$ , significantly lower than random method ( $19.48 \pm 8.12$ ,  $p < 0.0001$ ), vanilla BO ( $18.35 \pm 3.86$ ,  $p < 0.0001$ ) and constrained BO ( $16.26 \pm 3.73$ ,  $p < 0.0001$ ) under one-tailed Mann-Whitney U-test.

The benchmark on synthetic electrospray was extended to compare CCBO with state-of-the-art optimization methods such as Summit (38), Dragonfly (39), EDBO+ (40), and Atlas (41).



Notably, the implementations in Summit and Dragonfly did not support optimization under unknown constraints. Therefore, their performance was similar to the vanilla BO baseline. In terms of EDBO+, it was developed as a pool-based active learning optimization platform. The EDBO+ algorithm did not provide improvement on minimizing the distance towards design targets potentially due to lack of support for constrained optimization and the limitation of pool-based search space compared to other strategies. Finally, the most recent development of Atlas, a framework library for self-driving libraries, from Hickman *et al.*, utilized a variational GP to model unknown constraints for experiment feasibility (42). The optimization by Altas with *a priori* unknown constraints showed better performance than other existing strategies in the benchmark. However, as all of the state-of-the-art libraries do not natively support target-value optimization, none of these strategies outperform CCBO algorithm proposed here for electrospray optimization. More detailed results can be found in the **Supplementary Fig. 1**. These results have highlighted the importance of incorporating both constrained and composite optimization scheme for target-driven design problems in experiments.



**Fig. 1. Results for CCBO validation with synthetic data.** **a** An illustration of configurations for vanilla BO, constrained BO, and CCBO. **b** Parameter space visualization for the synthetic data with feasibility zone highlighted for each solvent. **c** Benchmark results of target value optimization with random baseline, vanilla BO, constrained BO, Atlas (constrained BO), Summit (single objective BO), EDBO+ (pool-based BO), Dragonfly (bandit BO), and CCBO. The regret is calculated by the closest distance with respect to the design target, achieved at different iterations of BO. Each benchmark experiment was performed for 10 iterations. Shaded areas indicate standard error from 20 times repetition. **d** Visualization of experimental parameters suggested. Each data point represents one synthetic experiment. The corresponding iteration is coded by color. Symbols represent the solvent used and feasibility of experiment. **e** Comparison of total number of successful (filled bars) and failed experiments (hatched bars) in a typical run of 10 iterations with the four strategies. **f** The particle sizes produced with parameters chosen by CCBO. The iteration is color-coded to the data point and the symbols represent the solvent and feasibility. The target (18  $\mu\text{m}$ ) was highlighted as a dashed line.

To understand the recommendation process, the experiments proposed were visualized in **Fig. 1d**. The random baseline sampled uniformly across the experiment space with both solvents, resulting in many failed DMAc experiments due to the flow rate feasibility constraints. Vanilla BO started exploring the boundary conditions in the first few rounds. With an additional model to account for feasibility, the constrained BO algorithm managed to learn the feasible region for DMAc, as reflected by most DMAc experiments being recommended with lower flow rates. This corresponded well to the initial feasible zone visualized in **Fig. 1b**. In addition, the number of failed and successful attempts of each algorithm from the results in **Fig. 1e** were plotted, highlighting the reduction in infeasible experimental conditions with the help of the additional constraint model.

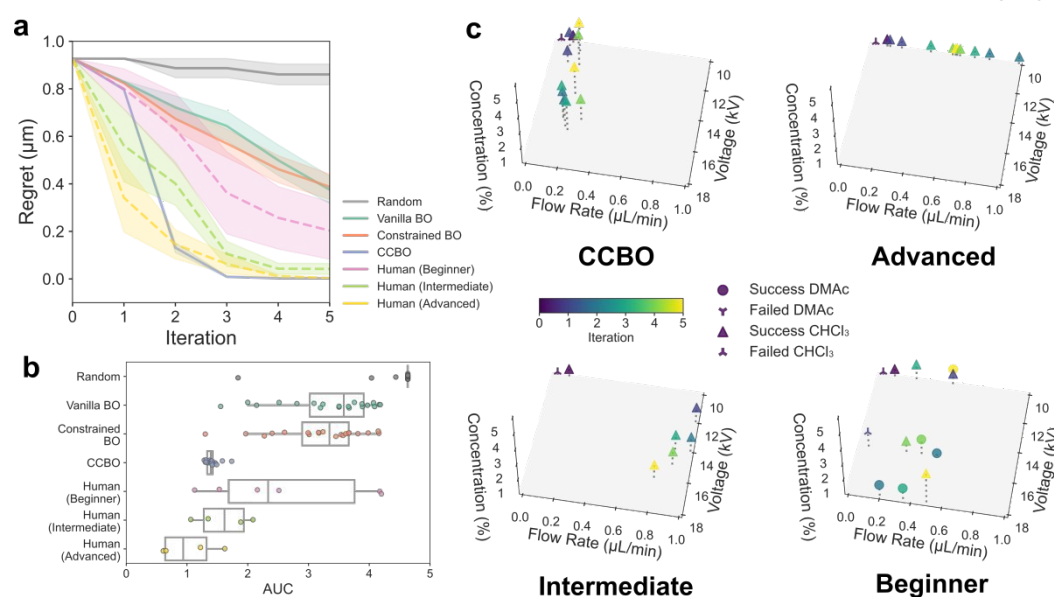
Furthermore, the CCBO strategy was observed to show highly efficient searching in a localized experiment space (**Fig. 1d**). This good performance of CCBO could be explained by its design. The routes taken by vanilla BO and constrained BO were directly minimizing the distance where the surrogate GP was forced to model more complicated results from both the experiment and the superimposed distance function. On the contrary, GP was solely used for modeling the black-box experiment results for CCBO. Our observations with CCBO echoed the findings in composite BO literature: extracting the analytically trackable part from the black-box function can drastically benefit the optimization efficiency (30). In standard BO, the EI acquisition function assumes Gaussian posterior distribution. However, the posterior of the composite function becomes non-Gaussian after the transformation with a non-linear function. To address this, Astudillo and Frazier suggested leaving the GP to model the black-box function. The composite part was instead incorporated into the acquisition function to transform the Gaussian posterior of the black-box function. This allows more efficient optimization through a closer approximation of posterior distribution in a composite scenario(36). In our implementation, the



composite acquisition function was optimized in the CCBO pipeline with Monte Carlo sampling. Through the benchmark validation, we have shown that vanilla BO or constrained BO alone would not be able to efficiently optimize our design problem, highlighting the importance of the integration of CCBO.

Finally, we compared CCBO to human electrospray users with varying levels of expertise on this synthetic campaign. More experienced users were believed to approach the target more efficiently as they've been equipped with prior knowledge of the parameters' influence and the selection of solvent. All participants (N=14) evaluated the same initial experimental data and suggested experiments to achieve a target particle size of 3  $\mu\text{m}$ . The comparative results are plotted in **Fig. 2a**. More detailed human performance results are available in **Supplementary Table 1 and Supplementary Fig. 2**. In the first iteration, the CCBO strategy was behind intermediate (1-3 years' experience, N=4) and advanced users ( $\geq 3$  years' experience, N=4) and performed similarly as beginners ( $<1$  years' experience, N=6). But CCBO soon overtook intermediate users from the second iteration onwards and surpassed advanced user on later iterations. Quantitatively, the AUC was calculated and plotted (**Fig. 2b**) with respect to each strategy or human group, where CCBO ( $1.40 \pm 0.10$ ) owned significantly smaller ( $p=0.01$ ) AUC than beginners ( $2.62 \pm 1.19$ ) under one-tailed Mann-Whitney U-test. There were no significant reductions of AUC with CCBO compared to intermediate ( $1.60 \pm 0.41$ ,  $p=0.34$ ) or advanced ( $1.03 \pm 0.42$ ,  $p=0.95$ ) users. When focusing on overall performance (regret at final iteration), the regret of CCBO strategy was significantly lower than intermediate ( $p=0.02$ ) and beginner ( $p<0.0001$ ) users. Further analysis of parameter selection strategies revealed that advanced users predominantly followed a one-factor-at-a-time (OFAT) approach, resulting in linear adjustment patterns (**Fig. 2c**). Most beginner users and intermediate users attempted to adjust multiple parameters simultaneously. Unlike human participants, CCBO employed more strategic exploration and exploitation, effectively reducing experimental regret by targeting promising regions in the parameter space. Taking together, these findings demonstrated that CCBO could achieve performance comparable to highly experienced participants and navigate complex experimental spaces more effectively than human users. In addition, the performance differences in users with various expertise reflected a successful development of the synthetic problem simulating electrospraying, consolidating our confidence in proceeding to laboratory validation.





**Fig. 2. Comparing Human and BO performance with synthetic data.** **a** Benchmark results with BO pipelines (solid lines) for  $3.0\ \mu\text{m}$  target in comparison with human users (dashed lines) on synthetic data. The experiment was performed for 5 iterations. Shaded areas for random, vanilla BO, constrained BO, and CCBO indicate standard error from 20 times repetition. Shaded areas for human performance, including beginner ( $N=6$ ), intermediate ( $N=4$ ), advanced ( $N=4$ ), come from standard error of participants respectively. **b** Box plot with scatters on area-under-a-curve (AUC), calculated using trapezoid rule, of the benchmark results from each strategy and human groups. **c** Visualization of experiments selected by CCBO and human participants with various experience levels. Each data point represents one ‘synthetic’ experiment. The corresponding iteration is coded by color. Symbols represent the solvent used and feasibility of experiment.

### Guiding laboratory electrospraying with CCBO for targeted particle production.

Following the validation of CCBO with synthetic data, it was applied in real-world experiments to guide electrospraying production of micro- and nanoparticles. The initial experiments, generated through a Sobol sequence, were performed to accumulate starting data for BO pipelines (Table 2).

**Table 2.** Processing parameters generated through a Sobol sequence and the resulting particle sizes and feasibility ( $N=3$ ).

Label	Polymer Concentration (% w/v)	Flow Rate ( $\mu\text{L min}^{-1}$ )	Voltage (kV)	Solvent	Mean Size ( $\mu\text{m}$ )	Feasible?
0-1	2.40	1.73	14.0	DMAc	0.56	1
0-2	4.06	0.44	15.7	CHCl <sub>3</sub>	1.00	0
0-3	2.88	49.11	11.8	DMAc	15.00	0
0-4	0.76	0.01	17.6	CHCl <sub>3</sub>	1.20	0
0-5	0.11	10.43	14.5	CHCl <sub>3</sub>	6.26	1
0-6	3.55	0.06	12.8	DMAc	0.15	1
0-7	4.55	2.39	16.7	CHCl <sub>3</sub>	5.24	1
0-8	1.88	0.21	11.0	DMAc	1.12	1





Two particle sizes, 300 nm and 3.0  $\mu\text{m}$ , were set as the design targets based on pharmaceutical interests as drug carriers for intravenous injection and pulmonary delivery (4). Based on previous reports, the production of PLGA particles with these two particle sizes require distinctive processing parameters involving different solvents and flow rates (32,43). Thus, the setting of these targets could simulate distinct typical experimental scenarios to challenge BO pipelines. The workflow of targeted particle production under CCBO guidance is illustrated in **Fig. 3a**. With the initial data gathered, CCBO pipeline was implemented to propose two experiments in parallel for laboratory investigation. The selection of two experiments was based on the capacity of laboratory work and to avoid wasting materials and preparation time. After collecting samples and characterization, the results from triplicated experiments were evaluated and compared with the target. The next iteration of BO was performed based on the addition of the new data.

The proposed parameters by CCBO can be visualized with heatmaps in **Fig. 3b**. The heatmap of initial experiments reflected the diverse selections of parameters in Sobol sequence. In total, three iterations of BO were performed for the target of 300 nm and four iterations for 3.0  $\mu\text{m}$  target. The selection of solvents was the most obvious difference for these two targets. Indeed, in previous reports of PLGA particle synthesis, DMAc was a popular solvent due to its high boiling point (44). From a mechanistic viewpoint, droplets will experience fission due to the competition between Coulombic repulsion and liquid surface tension in an electrospraying process (45). At the same time, the evaporation of solvents increases the concentration and viscosity of the droplet. As a non-volatile solvent, DMAc allows this fission process to fully develop and thus generates sub-micrometer particles (32). Chloroform, on the contrary, was preferred in literature to produce larger particles within tens of micrometers range (46). These practical considerations, normally accumulated through experiences and trial-and-error, were also picked up by the BO pipeline. The recommendations provided by CCBO clearly showed the trend of adopting DMAc for the 300 nm target and chloroform for the 3.0  $\mu\text{m}$  target.

Linking the recommendations to the experiment results (**Fig. 3c**) could provide a more holistic viewpoint of the selection strategy of CCBO. For 300 nm target, the best candidate in initial experiments (0-8 on **Table 2**) used DMAc with a low polymer concentration, flow rate and voltage to obtain 0.15  $\mu\text{m}$  particles. The recommendations from CCBO pipeline showed exploration of higher concentrations and fine-tuning of the flow rate parameter (**Supplementary Table 2**). Interestingly, the 3-1 and 3-2 experiments both achieved 300 nm particle size with distinct processing parameters, suggesting that the impact from less concentrated polymer solution was compensated by the higher flow rate used for 3-1.

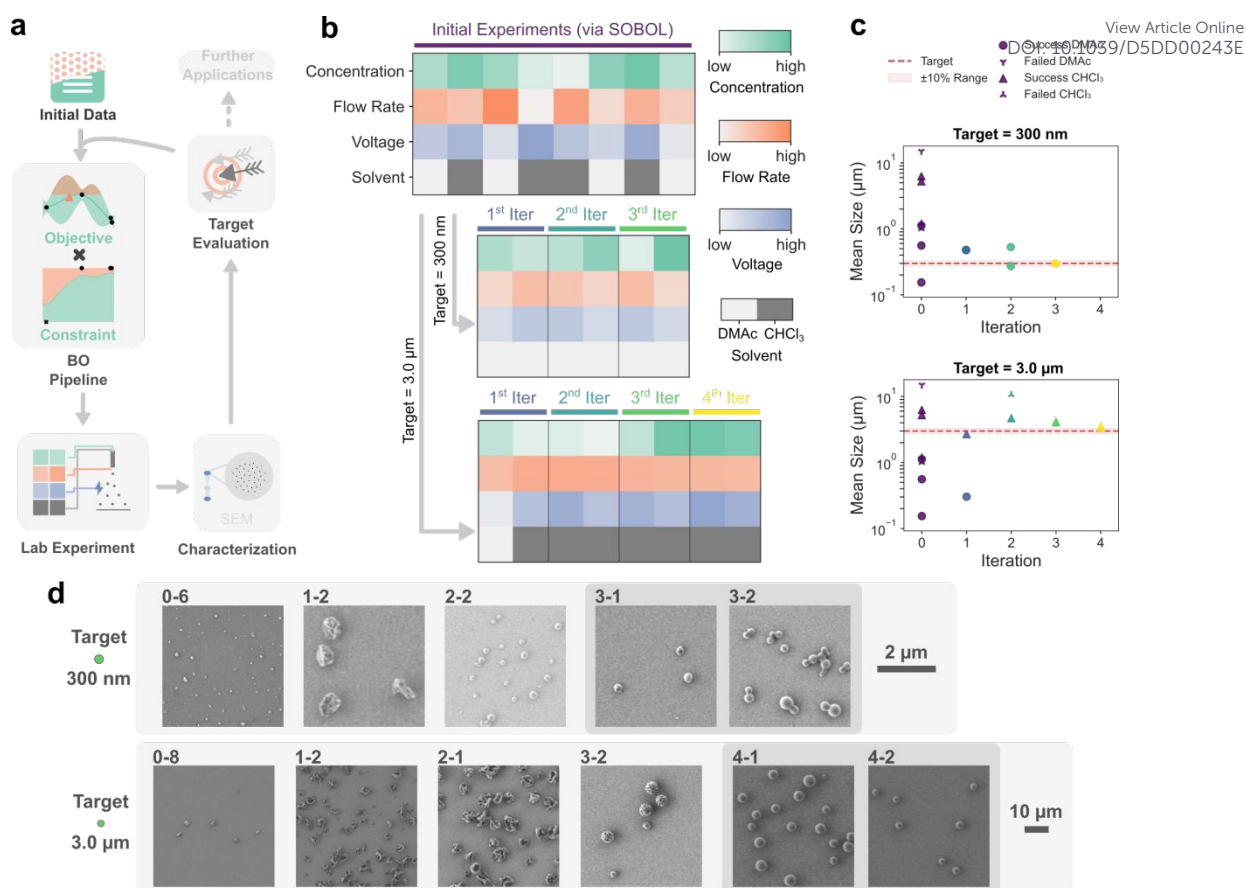




Furthermore, the balance of exploration-exploitation from the EI acquisition function was further demonstrated through the experiment series for 3.0  $\mu\text{m}$  target. In the first iteration, CCBO attempted both DMAc and chloroform as the solvent (**Supplementary Table 3**). The second iteration tested the lowest polymer concentration (0.05% w/v), which was shown as the lightest green in the heatmap (**Fig. 3b**). Finally, the recommendation settled down at higher concentration with reduced flow rates to approach the target with fine-tuning from exploitation. It was also observed from the SEM images (**Fig. 3d**) that the experiment 1-2 for 3.0  $\mu\text{m}$  target managed to produce 2.69  $\mu\text{m}$  particles with rough and polydisperse characteristic from a low polymer concentration (0.36% w/v) sprayed at a high flow rate of 3.65  $\mu\text{L min}^{-1}$ . The final experiments 4-2 suggested 4.02% w/v solution sprayed at 1.08  $\mu\text{L min}^{-1}$  (**Supplementary Table 3**) to obtain 3.29  $\mu\text{m}$  diameter particles. This result again highlighted the ability to achieve similar particle size through balancing polymer concentration and flow rate, together with adjusting other parameters. The SEM images of the final iteration experiments have shown satisfactory particle production at targeting sizes.

Overall, we have verified the performance of CCBO in the automatic identification of the experiment feasibility region and the rapid convergence to design targets through synthetic data validation. The comparison with human experts demonstrated CCBO's competitive performance. The rational exploration of experiment space outperformed the instinct-driven OFAT trial and error by humans. The wet-lab experiments, as a further step, consolidated CCBO's potential in real-world applications for guided particle synthesis within a few iterations.





**Fig. 3. Guiding electrospray experiments with CCBO.** **a** A schematic diagram representing the experiment process with the integration of CCBO. **b** Heatmaps visualizing the processing parameters used for (top) initial experiments, (middle) 300 nm target, and (bottom) 3.0  $\mu\text{m}$  target. The initial experiments were generated with Sobol sequence and the targeted experiment series were suggested by CCBO pipeline. **c** Experiment results of particles generated with electrospraying under parameters proposed for (top) 300 nm and (bottom) 3.0  $\mu\text{m}$  target. Each data point represents the mean of triplicated laboratory experiments. Symbols represent the solvent used and feasibility of experiment. **d** SEM images of particles produced at different iterations for (top) 300 nm and (bottom) 3.0  $\mu\text{m}$  target.

## Discussion

The present work demonstrated the application of efficient CCBO pipeline for target value optimization under black-box constraints. The two components in CCBO worked cohesively to address the need for guiding particle synthesis. For target optimization, the composite BO demonstrated strong capacity in modeling under the composited distance function over the underlying, black-boxed electrospray relationship function. On the other hand, the constraint compartment managed to learn and regulate the suggested experiments with a variational Gaussian process. To deal with unknown feasibility boundaries, many current strategies chose to apply active learning for the identification of unknown feasibility regions, followed by



running BO pipelines under the established boundaries (47,48). As an improvement, CCBO was designed for integrating these two individual processes and focus on identifying the feasibility regions around the design target. This could be seen from the initial experiments where the infeasibility caused by the mismatching of high flow rate with a less volatile solvent DMAc (experiment 0-3 on **Supplementary Table 2**) was not further explored because the target only requires experiments in the lower flow rate region. In comparison, with an active learning pipeline, extra experiments would be needed to determine the possible range for DMAc. Thus, the design of CCBO pipeline allows efficient reduction in the number of experiments to save laboratory resources.

In addition, the innate exploration-exploitation trade-off from BO made possible the identification of multiple possible experimental parameters that can achieve the same design target. This is especially helpful when other design considerations coexist. For example, in the validation with the synthetic problem (**Fig. 1f**), CCBO attempted both DMAc and chloroform and paired them with a wide range of other processing parameters to hit the design target in iterations 6 to 10. From the perspective of production rate, a higher flow rate and polymer concentration might be preferred. Similarly, if the sustainability of the solvent is considered, DMAc would be selected over chloroform as a less harsh solvent. On top of the synthetic data, laboratory experiments also managed to find multiple parameters to produce particles with 300 nm or 3.0  $\mu\text{m}$  diameter. These particles exhibited distinctive morphology and polydispersity, demonstrating varying characteristics for their applications. Although not explicitly coded as a multiple-objective optimization problem, these sets of experimental parameters could be presented to the user as alternative choices. In practice, such flexibility allows the researcher to consider product properties, manufacturing metrics, or other aspects in production, without changing the main design target.

Since only two solvents have been investigated in the present work, categorical representations of the solvent variable were used instead of applying molecular featurization. Many modern BO libraries designed for chemistry and materials research support molecular featurization, such as Atlas and GAUCHE (41,49). Featurizing molecules with their physicochemical properties could incorporate chemistry knowledge in the optimization process and benefit molecular structure optimization and discovery tasks. For example, Griffiths *et al.* managed to leverage BO to optimize molecular design in a latent space generated from variational autoencoders (50). Although optimizing the solvent molecule *per se* was not necessarily a focus in particle synthesis applications, leveraging molecular fingerprints to represent solvents would equip the optimization process with chemically meaningful knowledge (*via* representing similar solvent with close descriptors) (51). In addition, extending the present single-objective optimization paradigm to multiple objectives could benefit more complicated particle design



tasks, including the control of both particle size and size distributions, or morphological features. Our implementation of constrained optimization was through feasibility-weighting of the acquisition function. It should be able to extend to multi-objective optimization seamlessly, considering that the feasibility modelling is irrelevant to the type of acquisition functions. Notably, Li *et al.* recently proposed a new method to balance (unknown) constraint modelling and multi-objective optimization through unifying constraint violation with hypervolume regret (52). They have demonstrated improved efficiency compared to baseline scalarization-based methods such as qParEGO (53). On the other hand, composite optimization should be transferable to multi-objective optimization scenarios, where objectives are scalarized. However, the implementation of these extensions is beyond the scope of the current manuscript and thus left as potential future directions here.

Finally, we highlight that CCBO could potentially be extended to other particle synthesis systems, such as batch methods and microfluidics, to facilitate the guided design and production of particles. In the past, the resource-demanding nature of experimentation and scarcity of data posed significant challenges and prolonged the workflow of particle synthesis. We are expecting CCBO to empower nanotechnology with a smarter and more efficient paradigm for target-driven design.

## Conclusions

Achieving rational synthesis of nanoparticles often relies on extensive domain expertise and trial-and-error experimentation to navigate within the feasibility space with target product specifications. In the present work, we introduced CCBO as a unified framework to address constraint-aware and target-value optimization in nanoparticle production. It was evaluated in a synthetic electrospray problem, presenting superior performance compared to baselines and state-of-the-art strategies. Benchmarking against human electrospray users further demonstrated that CCBO matches experts with at least three years' experience. Laboratory experiments practically validated its ability to guide electrospray synthesis of PLGA particles with biomedically meaningful target diameters of 300 nm and 3.0  $\mu\text{m}$ . These findings highlight CCBO as a powerful and efficient strategy for materials development tasks characterized by complex, black-box constraints and precise design objectives, contributing to the next generation of AI-driven nanomanufacturing.

## Data availability

The code to implement CCBO presented in this study has been made available at <https://github.com/FrankWanger/CCBO.git>. The specific code and data for reproducing CCBO



and baselines is also archived on Zenodo: <https://doi.org/10.5281/zenodo.16614771>. The raw experiment data has been included in the supplementary information.

## Competing interests

The authors declare no competing interests.

## Acknowledgements

The author Fanjin Wang would like to thank the Engineering and Physical Sciences Research Council (EPSRC) for supporting his PhD research (EP/R513143/1 and EP/W524335/1). Dr Jakob Zeitler from Matterhorn Studio is thanked for initial discussions. The participants of the human vs. BO campaign are gratefully acknowledged for their time and expertise.

## Author contributions

F.W. – Conceptualization, Data curation, Methodology and Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing; M.P. – Methodology, Resources, Supervision, Writing – review & editing; A.H. – Methodology, Writing – review & editing; M.E. – Methodology, Resources, Supervision, and Writing – review & editing.

## References

1. Mitchell MJ, Billingsley MM, Haley RM, Wechsler ME, Peppas NA, Langer R. Engineering precision nanoparticles for drug delivery. *Nat Rev Drug Discov.* 2020 Dec 4;20(2):101–24.
2. Stuart MAC, Huck WTS, Genzer J, Müller M, Ober C, Stamm M, et al. Emerging applications of stimuli-responsive polymer materials. *Nat Mater.* 2010 Jan 22;9(2):101–13.
3. Colombo S, Beck-Broichsitter M, Bøtker JP, Malmsten M, Rantanen J, Bohr A. Transforming nanomedicine manufacturing toward Quality by Design and microfluidics. *Adv Drug Deliv Rev.* 2018 Mar;128:115–31.
4. Poon W, Kingston BR, Ouyang B, Ngo W, Chan WCW. A framework for designing delivery systems. *Nat Nanotechnol.* 2020 Sept 7;15(10):819–29.
5. Rao L, Yuan Y, Shen X, Yu G, Chen X. Designing nanotheranostics with machine learning. *Nat Nanotechnol.* 2024 Oct 3;1–13.
6. Tao H, Wu T, Aldeghi M, Wu TC, Aspuru-Guzik A, Kumacheva E. Nanoparticle synthesis assisted by machine learning. *Nat Rev Mater.* 2021 July 13;6(8):701–16.



7. Antony J. Design of experiments for engineers and scientists. 2nd edition. London: Elsevier; 2014. 208 p. (Elsevier insights). View Article Online  
DOI: 10.1039/D5DD00243E
8. Montgomery DC. Design and analysis of experiments. Ninth edition. Hoboken, NJ: John Wiley & Sons, Inc; 2017. 1 p.
9. Batra R, Song L, Ramprasad R. Emerging materials intelligence ecosystems propelled by machine learning. *Nat Rev Mater*. 2020 Nov 9;6(8):655–78.
10. Back S, Aspuru-Guzik A, Ceriotti M, Gryn'ova G, Grzybowski B, Ho Gu G, et al. Accelerated chemical science with AI. *Digit Discov*. 2024;3(1):23–33.
11. Xu P, Ji X, Li M, Lu W. Small data machine learning in materials science. *Npj Comput Mater*. 2023 Mar 25;9(1):1–15.
12. Greenhill S, Rana S, Gupta S, Vellanki P, Venkatesh S. Bayesian Optimization for Adaptive Experimental Design: A Review. *IEEE Access*. 2020;8:13937–48.
13. Rasmussen CE. Gaussian processes for machine learning. 3. print. Cambridge, Mass: MIT Press; 2006. 272 p. (Adaptive computation and machine learning).
14. Wu Y, Walsh A, Ganose AM. Race to the bottom: Bayesian optimisation for chemical problems. *Digit Discov*. 2024 June 12;3(6):1086–100.
15. Shields BJ, Stevens J, Li J, Parasram M, Damani F, Alvarado JIM, et al. Bayesian reaction optimization as a tool for chemical synthesis. *Nature*. 2021 Feb 3;590(7844):89–96.
16. Szymanski NJ, Rendy B, Fei Y, Kumar RE, He T, Milsted D, et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature*. 2023 Nov 29;624(7990):86–91.
17. Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *Npj Comput Mater*. 2019 Feb 18;5(1):1–17.
18. Jones DR, Schonlau M, Welch WJ. Efficient Global Optimization of Expensive Black-Box Functions. *J Glob Optim*. 1998;13(4):455–92.
19. Hoffer JG, Ranftl S, Geiger BC. Robust Bayesian target value optimization. *Comput Ind Eng*. 2023 June;180:109279.
20. Borkowski O, Koch M, Zettor A, Pandi A, Batista AC, Soudier P, et al. Large scale active-learning-guided exploration for in vitro protein production optimization. *Nat Commun*. 2020 Apr 20;11(1):1872.
21. Ortiz-Perez A, van Tilborg D, van der Meel R, Grisoni F, Albertazzi L. Machine learning-guided high throughput nanoparticle design. *Digit Discov*. 2024;3(7):1280–91.
22. Antonio C. Sequential model based optimization of partially defined functions under unknown constraints. *J Glob Optim*. 2021 Feb 1;79(2):281–303.
23. Hickman RJ, Aldeghi M, Häse F, Aspuru-Guzik A. Bayesian optimization with known experimental and design constraints for chemistry applications. *Digit Discov*. 2022;1(5):732–44.





24. Li G, Wang Y, Kar S, Jin X. Bayesian Optimization with Active Constraint Learning for Advanced Manufacturing Process Design. *IISE Trans.* 0(0):1–15. View Article Online  
DOI: 10.1039/D5DD00243E
25. Low AKY, Mekki-Berrada F, Gupta A, Ostudin A, Xie J, Vissol-Gaudin E, et al. Evolution-guided Bayesian optimization for constrained multi-objective optimization in self-driving labs. *Npj Comput Mater.* 2024 May 13;10(1):104.
26. Gramacy RB, Lee HKH. Optimization Under Unknown Constraints [Internet]. arXiv; 2010 [cited 2025 Feb 17]. Available from: <http://arxiv.org/abs/1004.4027>
27. Gardner JR, Kusner MJ, Xu Z, Weinberger KQ, Cunningham JP. Bayesian optimization with inequality constraints. In: *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. Beijing, China: JMLR.org; 2014. p. II-937-II-945. (ICML'14).
28. Tian Y, Zuniga A, Zhang X, Dürholt JP, Das P, Chen J, et al. Boundary Exploration for Bayesian Optimization With Unknown Physical Constraints [Internet]. arXiv; 2024 [cited 2025 Feb 17]. Available from: <http://arxiv.org/abs/2402.07692>
29. Uhrenholt AK, Jensen BS. Efficient Bayesian Optimization for Target Vector Estimation. In: *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* [Internet]. PMLR; 2019 [cited 2024 Apr 16]. p. 2661–70. Available from: <https://proceedings.mlr.press/v89/uhrenholt19a.html>
30. Astudillo R, Frazier PI. Thinking Inside the Box: A Tutorial on Grey-Box Bayesian Optimization. In: *2021 Winter Simulation Conference (WSC)* [Internet]. Phoenix, AZ, USA: IEEE; 2021 [cited 2024 Apr 16]. p. 1–15. Available from: <https://ieeexplore.ieee.org/document/9715343/>
31. Ali A, Zaman A, Sayed E, Evans D, Morgan S, Samwell C, et al. Electrohydrodynamic atomisation driven design and engineering of opportunistic particulate systems for applications in drug delivery, therapeutics and pharmaceuticals. *Adv Drug Deliv Rev.* 2021 Sept;176:113788.
32. Xie J, Jiang J, Davoodi P, Srinivasan MP, Wang CH. Electrohydrodynamic atomization: A two-decade effort to produce and process micro-/nanoparticulate materials. *Chem Eng Sci.* 2015 Mar;125:32–57.
33. Balandat M, Karrer B, Jiang DR, Daulton S, Letham B, Wilson AG, et al. BOTORCH: a framework for efficient monte-carlo Bayesian optimization. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc.; 2020. p. 21524–38. (NIPS '20).
34. Gardner JR, Pleiss G, Bindel D, Weinberger KQ, Wilson AG. GPyTorch: blackbox matrix-matrix Gaussian process inference with GPU acceleration. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc.; 2018. p. 7587–97. (NIPS'18).
35. Hensman J, Matthews A, Ghahramani Z. Scalable Variational Gaussian Process Classification. In: *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics* [Internet]. San Diego, California, USA: PMLR; 2015 [cited 2024 Oct 4]. p. 351–60. Available from: <https://proceedings.mlr.press/v38/hensman15.html>
36. Astudillo R, Frazier P. Bayesian Optimization of Composite Functions. In: *Proceedings of the 36th International Conference on Machine Learning* [Internet]. PMLR; 2019 [cited



2024 Apr 17]. p. 354–63. Available from: <https://proceedings.mlr.press/v97/astudillo19a.html> View Article Online  
DOI: 10.1039/D5DD000243E

37. ZHANG HB, EDIRISINGHE MJ, JAYASINGHE SN. Flow behaviour of dielectric liquids in an electric field. *J Fluid Mech.* 2006 July;558:103.
38. Felton KC, Rittig JG, Lapkin AA. Summit: Benchmarking Machine Learning Methods for Reaction Optimisation. *Chemistry–Methods.* 2021;1(2):116–22.
39. Kandasamy K, Vysyaraju KR, Neiswanger W, Paria B, Collins CR, Schneider J, et al. Tuning hyperparameters without grad students: scalable and robust Bayesian optimisation with dragonfly. *J Mach Learn Res.* 2020 Jan 1;21(1):81:3098-81:3124.
40. Torres JAG, Lau SH, Anchuri P, Stevens JM, Tabora JE, Li J, et al. A Multi-Objective Active Learning Platform and Web App for Reaction Optimization. *J Am Chem Soc.* 2022 Nov 2;144(43):19999–20007.
41. Hickman RJ, Sim M, Pablo-García S, Tom G, Woolhouse I, Hao H, et al. Atlas: a brain for self-driving laboratories. *Digit Discov.* 2025 Apr 9;4(4):1006–29.
42. Hickman RJ, Tom G, Zou Y, Aldeghi M, Aspuru-Guzik A. Anubis: Bayesian optimization with unknown feasibility constraints for scientific experimentation. *Digit Discov* [Internet]. 2025 June 18 [cited 2025 July 14]; Available from: <https://pubs.rsc.org/en/content/articlelanding/2025/dd/d5dd00018a>
43. Wang F, Elbadawi M, Tsilova SL, Gaisford S, Basit AW, Parhizkar M. Machine learning predicts electrospray particle size. *Mater Des.* 2022 July;219:110735.
44. Parhizkar M, Reardon PJT, Knowles JC, Browning RJ, Stride E, Pedley RB, et al. Performance of novel high throughput multi electrospray systems for forming of polymeric micro/nanoparticles. *Mater Des.* 2017 July;126:73–84.
45. Hartman RPA, Brunner DJ, Camelot DMA, Marijnissen JCM, Scarlett B. JET BREAK-UP IN ELECTROHYDRODYNAMIC ATOMIZATION IN THE CONE-JET MODE. *J Aerosol Sci.* 2000 Jan;31(1):65–95.
46. Xu J, Li K, Liu M, Gu X, Li P, Fan Y. Studies on preparation and formation mechanism of poly(lactide-co-glycolide) microrods via one-step electrospray and an application for drug delivery system. *Eur Polym J.* 2021 Apr;148:110372.
47. Khatamsaz D, Vela B, Singh P, Johnson DD, Allaire D, Arróyave R. Bayesian optimization with active learning of design constraints using an entropy-based approach. *Npj Comput Mater.* 2023 Apr 1;9(1):1–14.
48. Arróyave R, Khatamsaz D, Vela B, Couperthwaite R, Molkeri A, Singh P, et al. A perspective on Bayesian methods applied to materials discovery and design. *MRS Commun.* 2022 Oct 26;12(6):1037–49.
49. Griffiths RR, Klarner L, Moss HB, Ravuri A, Truong S, Stanton S, et al. GAUCHE: A Library for Gaussian Processes in Chemistry [Internet]. arXiv; 2023 [cited 2025 July 10]. Available from: <http://arxiv.org/abs/2212.04450>
50. Griffiths RR, Hernández-Lobato JM. Constrained Bayesian optimization for automatic chemical design using variational autoencoders. *Chem Sci.* 2020 Jan 2;11(2):577–86.



51. Wang F, Harker A, Edirisinghe M, Parhizkar M. Tackling Data Scarcity Challenge through Active Learning in Materials Processing with Electrospray. *Adv Intell Syst.* 2024 May 23;6(7):2300798. New Article Online  
DOI: 10.1039/D5DD00243E
52. Li D, Zhang F, Liu C, Chen Y. Constrained Multi-objective Bayesian Optimization through Optimistic Constraints Estimation [Internet]. *arXiv*; 2025 [cited 2025 July 10]. Available from: <http://arxiv.org/abs/2411.03641>
53. Daulton S, Balandat M, Bakshy E. Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization [Internet]. *arXiv*; 2020 [cited 2024 Mar 11]. Available from: <http://arxiv.org/abs/2006.05078>



# Constrained composite Bayesian optimization for rational synthesis of polymeric particles

Fanjin Wang<sup>1,2,\*</sup>, Maryam Parhizkar<sup>2</sup>, Anthony Harker<sup>3</sup>, Mohan Edirisinghe<sup>1</sup>

<sup>1</sup>Department of Mechanical Engineering, University College London, London, WC1E 7JE, United Kingdom

<sup>2</sup>School of Pharmacy, University College London, London, WC1N 1AX, United Kingdom

<sup>3</sup>Department of Physics and Astronomy, University College London, London, WC1E 6BT, United Kingdom

\*Corresponding author, email: [fanjin.wang.20@ucl.ac.uk](mailto:fanjin.wang.20@ucl.ac.uk)

## Data availability

The code to implement CCBO presented in this study has been made available at <https://github.com/FrankWanger/CCBO.git>. The specific code and data for reproducing CCBO and baselines is also archived on Zenodo: <https://doi.org/10.5281/zenodo.16614771>. The raw experiment data has been included in the supplementary information.

