


 Cite this: *RSC Adv.*, 2026, **16**, 28200

Efficient preparation of size-controlled sodium alginate microspheres *via* adaptive Bayesian optimization of the spray process

 Shi Feng,^a Jing Liu,^c Zhuxin Li^{*c} and Shengyang Tao^{*,bc}

Preparing sodium alginate (SA) microspheres *via* spray-precipitation challenges precise size control due to complex parameter coupling. We propose an intelligent framework integrating adaptive Bayesian optimization (BO) with microfluidic spraying to maximize the yield of 90–110 μm microspheres. Utilizing a Gaussian process regression model and Latin hypercube sampling, the framework demonstrated exceptional efficiency. Initializing with 10 prior data points achieved convergence in just 12 iterations, reducing iteration cost by 29.4% compared to using 5 priors. Under optimal conditions, the predicted target droplet proportion (18.60%) precisely matched experiments (18.34%), yielding 18.75% target-sized SA gel microspheres post-curing. Additionally, SHAP analysis revealed that gas and liquid pressures dominate size distribution, elucidating the physical mechanism behind multiple local optima *via* multi-feature compensation. This study provides an efficient, low-cost strategy for customizing polymer microspheres, establishing a robust machine-learning paradigm for optimizing complex multiphase flows.

Received 7th April 2026

Accepted 18th May 2026

DOI: 10.1039/d6ra02906j

rsc.li/rsc-advances

1 Introduction

Sodium alginate (SA), a natural polyanionic polysaccharide extracted from marine brown algae,¹ is highly favoured in the field of biomedical materials due to its excellent biocompatibility,² minimal toxicity,³ and favourable biodegradability.⁴ Abundant in characteristic carboxyl groups (–COO[−]) within its molecular structure, SA readily undergoes mild and rapid crosslinking reactions with divalent cations (*e.g.*, Ca²⁺) in aqueous solutions, thereby forming stable three-dimensional hydrogel networks.^{5–7} Owing to this property, SA microspheres have been widely applied in cutting-edge fields such as targeted drug delivery,⁸ tissue engineering scaffolding,⁹ and food science.¹⁰ In recent years, to meet the demands of complex application scenarios, researchers have developed various SA-based composite microsphere systems. For instance, Bi *et al.* fabricated hydroxyapatite/SA/chitosan composite microspheres *via* emulsion crosslinking techniques to serve as pH-responsive drug carriers for bone tissue engineering.¹¹ Additionally, Chen *et al.* synthesized porous SA/cellulose nanofiber hydrogel microspheres using microfluidics, achieving highly efficient adsorption of heavy metal ions.¹² These versatile applications

underscore the immense potential and significance of SA microspheres.

Among various microsphere fabrication technologies, the spray-precipitation technique stands out as a core process for synthesizing gel microspheres owing to its operational simplicity, excellent continuous processing capability, and facile industrial scale-up.^{13–15} However, this method faces a critical challenge in practical applications: the rapid, precise, and tailored fabrication of microspheres with specific target sizes. The final microsphere size in the spray-crosslinking process is directly dictated by the droplet size formed during instantaneous atomization of the precursor solution, as well as by the uniformity of its spatial distribution.^{16–18} The atomization and breakup of droplets are jointly governed by multiple variables, including the physical properties of the precursor (*e.g.*, viscosity and surface tension, which are concentration-dependent),¹⁹ operating parameters (gas and liquid pressures),²⁰ and spray distance.²¹ Given the highly complex, nonlinear coupling among these variables, traditional one-factor-at-a-time (OFAT) methods or orthogonal experimental designs often require extensive, tedious trial-and-error experiments. Such approaches are not only time-consuming, labor-intensive, and costly, but also generally fail to pinpoint the global optimum within a vast parameter space.

To overcome this bottleneck, data-driven machine learning (ML) strategies have been increasingly integrated into the parameter optimization of chemical engineering processes and microfluidic systems in recent years.^{22–25} When addressing high-cost, small-sample experimental optimization involving multiphase flows, Bayesian optimization (BO) demonstrates

^aChina Nuclear Power Engineering Co., Ltd, Beijing, 100142, China

^bState Key Laboratory of Fine Chemicals, Frontier Science Center for Smart Materials, Dalian University of Technology, Dalian, 116024, China. E-mail: taosy@dlut.edu.cn

^cDalian Key Laboratory of Intelligent Chemistry, CR Belt and Road Joint Laboratory on Intelligent Chemistry and Advanced Materials of Liaoning Province, School of Chemistry, Dalian University of Technology, Dalian, 116024, China


unparalleled advantages over traditional grid search or random search.^{26–28} BO employs a probabilistic surrogate model (such as Gaussian process regression, GPR)²⁹ to approximate the objective function and uses an acquisition function to balance exploration of uncharted regions with exploitation of known high-reward areas. This adaptive learn-and-predict mechanism can converge to the global optimum with a minimal number of experimental evaluations, making it exceptionally well-suited to resolving the black-box parameter-regulation challenges inherent in microsphere fabrication.

Based on the above background, this study proposes an intelligent spray process framework integrating adaptive Bayesian optimization (BO) with Gaussian process regression (GPR), aimed at achieving the rapid and precise tailored fabrication of SA microspheres with target sizes (Fig. 1). Utilizing a custom-built micro-spraying system equipped with a real-time laser particle size analyzer, we selected SA concentration, spray distance, gas pressure, and liquid pressure as independent variables. The optimization objective was to maximize the yield proportion of droplets within a specific size range (90–110 μm). During this process, Latin hypercube sampling (LHS) was employed to acquire highly representative prior data, and the impact of varying prior data sizes on the model's convergence efficiency and iteration cost was thoroughly investigated. Furthermore, to address the persistent black-box nature of machine learning models, this study introduced the game-theoretic SHapley Additive exPlanations (SHAP) interpretability model. By extracting the SHAP values of each parameter, we quantitatively elucidated the dominant roles and nonlinear synergistic mechanisms of physical properties and gas/liquid pressures in governing the microsphere size distribution. Consequently, the multiple local optima predicted by the model were scientifically interpreted from the dual perspectives of physics and chemical engineering. This work not only successfully overcomes the chronic limitation of traditional spray-precipitation methods in achieving precise size control, but also establishes a highly interpretable, dual-driven “AI + Experiment” paradigm at the methodological level for the efficient and cost-effective synthesis of smart materials.

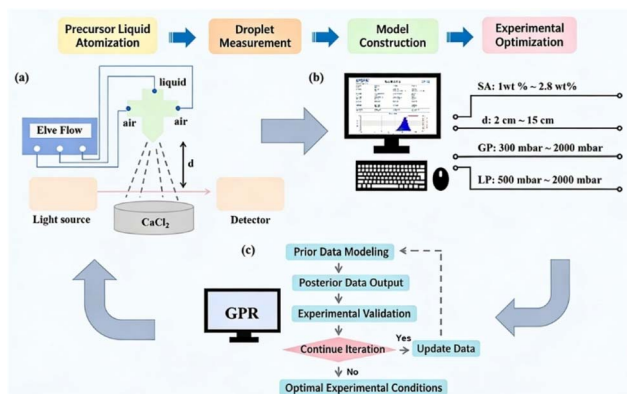


Fig. 1 Experimental optimization of the SA microsphere fabrication process: (a) schematic illustration of the spraying experiment; (b) droplet measurement software and optimization variables; (c) Bayesian optimization workflow.

2 Materials and methods

2.1 Materials and instruments

The primary chemical reagents used in this study included SA (analytical grade) and anhydrous calcium chloride (analytical grade), both purchased from Tianjin Kermel Chemical Reagent Co., Ltd. Deionized (DI) water (resistivity of 18 $\text{M}\Omega\text{ cm}$), produced by a laboratory ultrapure water system, was used throughout the experiments. Fluidic tubing connectors (1/4-28 Luer fittings and ferrules) were acquired from Beijing Yijia Technology Co., Ltd. All reagents were used as received without further purification.

The experimental apparatus included a multi-channel microfluidic pressure pump (OB1 MK3+, Elveflow, France) coupled with an oil-free air compressor (OTS-550, Taizhou Outes Industry and Trade Co., Ltd) to precisely control the flow rates of the gas and liquid phases. A spray laser particle size analyzer (DP-02, Zhuhai OMEC Instruments Co., Ltd) was utilized for the real-time online monitoring of the droplet size distribution. Microscopic morphological characterization was performed using an inverted biological microscope (ECLIPSE Ts2, Nikon, Japan). Furthermore, the solution preparation, processing, lyophilization, and weighing procedures of the samples involved the use of a heating magnetic stirrer (C-MAG HS4, IKA, Germany), an ultrasonic cleaner (KQ2200, Kunshan Ultrasonic Instruments Co., Ltd), a freeze dryer (Scientz-10NIA, Ningbo Scientz Biotechnology Co., Ltd), and a precision electronic balance (ME204E, Mettler Toledo, Switzerland), respectively. Additionally, the internal-mixing micro-spraying nozzle device,³⁰ which constitutes the core of the experimental setup, was custom-designed and fabricated by our research group. The gas and liquid outputs are controlled by a multi-channel ElveFlow pressure controller. One channel is driven by the gas pressure to control the liquid flow, while the other channel serves as the gas source, facilitating the simultaneous control of the material flow and the gas output volume. The spray device uses an internally mixed three-stream air atomization nozzle (the diameter of the liquid outlet of the nozzle is 1 mm, the diameter of the gas outlet is 1.4 mm, and the gas outlet is 0.3 mm higher than the liquid outlet), which can form a uniform shear force compared to traditional two-stream nozzles, resulting in a better shear effect between gas and liquid phases. At the nozzle outlet, there is a significant relative speed between the gas and liquid phases, generating a large shear force, which causes the precursor to be atomized and dispersed into extremely fine droplets. To prevent long-term corrosion by organic solvents, the nozzle is made of stainless steel. The collection device uses a Petri dish made of polytetrafluoroethylene (PTFE) dual-phobic surface material, which reduces issues such as microsphere breakage and agglomeration.

2.2 Spray-precipitation fabrication of SA microspheres

The fabrication of SA gel microspheres was performed using a custom-built micro-spraying system. The specific process workflow was as follows: first, aqueous SA solutions with varying mass fractions (1.00–2.80 wt%) were prepared as precursor solutions, alongside a 2.0 wt% CaCl_2 solution serving as the



receiving crosslinking bath. During the experiments, the gas pressure (GP) and liquid pressure (LP) were precisely regulated by the Elveflow pressure pump. The SA precursor solution was fed into the internal-mixing nozzle *via* the liquid channel, while two streams of compressed air converged through the air inlets. Inside the nozzle, the high-velocity airflow exerted intense shear forces on the highly viscous SA solution, inducing hydrodynamic instability and forcing the liquid to break up and atomize into fine droplets prior to ejection.

The atomized droplets travelled downward over a specific spray distance (d) and traversed the detection optical path of the spray laser particle size analyzer. Based on the Mie scattering theory, the detector captured the variations in laser scattering angles caused by droplets of different sizes (where larger particles yield smaller scattering angles). It enabled the real-time online calculation and output of the droplets' volume median diameter (D_{50}) and size distribution range. Subsequently, the droplets fell into the underlying CaCl_2 -receiving bath, where the carboxyl groups on the SA molecular chains instantaneously underwent ionic crosslinking reactions with Ca^{2+} in solution. The droplets were allowed to cure statically in the receiving bath for 5 min, yielding stable SA gel microspheres. After collection and washing, the actual size distribution of the cured microspheres was further characterized.

2.3 Construction of the Bayesian optimization model

To overcome the low efficiency of traditional trial-and-error methods, this study developed a data-driven Bayesian optimization framework in Python using the Scikit-learn and Scikit-Optimize libraries. Four key parameters that most significantly affect the morphology and size of the microspheres were selected as the independent optimization variables: SA concentration (1.00–2.80 wt% with a step size of 0.10 wt%), spray distance d (2–15 cm with a step size of 1 cm), gas pressure GP (300–2000 mbar with a step size of 50 mbar), and liquid pressure LP (500–2000 mbar with a step size of 50 mbar). This discretization process constructed a multi-dimensional parameter optimization space comprising 371 280 data points ($28 \times 13 \times 34 \times 30$). The optimization objective was to maximize the proportion of droplets in the 90–110 μm size range (y_2), while simultaneously monitoring the volume median diameter (D_{50} ; y_1).

To ensure uniform distribution and diversity of the initial data across the multi-dimensional space, Latin hypercube sampling (LHS) was employed to generate 5 and 10 sets of experimental parameter combinations as prior datasets. Spray experiments were then conducted to obtain the actual droplet proportions, which were subsequently used as starting points for the iterations. Given the characteristics of the small-sample dataset, Gaussian process regression (GPR) was selected as the surrogate model for the Bayesian optimization.

2.4 Adaptive hyperparameter search and model iteration strategy

To ensure optimal generalization, the BayesSearchCV module from the Scikit-Optimize library was used to implement

adaptive hyperparameter optimization. Initially, the program read the data from a CSV file and split it into training and test sets, allocating 20% to testing. During each training iteration, the algorithm automatically traversed four candidate kernel functions: Radial Basis Function (RBF), Matern, Rational Quadratic, and Dot Product. Using the minimization of the negative mean squared error as the evaluation criterion, it dynamically matched the optimal kernel function that best suited the current data characteristics. Simultaneously, to ensure the kernel matrix remains positive definite during computation and enhance the model's adaptability to new data, the optimizer adaptively searched for the optimal noise level parameter (α) within a log-uniform range (1×10^{-6} to 1×10^{-2}). It also optimized the maximum number of optimizer restarts ($n_{\text{restarts_optimizer}}$) within an integer range (5–20) to increase the probability of finding the global optimum in non-convex optimization problems, thereby preventing the model from being trapped in local minima. The entire hyperparameter optimization process used 5-fold cross-validation (4-fold when using 5 prior data sets), with the random seed fixed at 42 to ensure high reproducibility of the optimization results.

Upon completion of the model's self-optimization, the system entered a closed-loop, iterative validation phase. First, the partitioned test set was used to evaluate the model's generalization performance, calculating and reporting key metrics such as mean squared error (MSE), coefficient of determination (R^2), and mean absolute error (MAE). Subsequently, the algorithm generated continuous combinations of independent variables within the predefined parameter boundaries and used the current best model to predict the target variable values along with their standard deviations. By screening globally for variable combinations that satisfied the boundary conditions, the optimal parameter combination that maximized the predicted value of the target variable (*i.e.*, the yield proportion of 90–110 μm droplets) was identified. Its corresponding predicted value, standard deviation, and optimal kernel function were then output. Then, this optimal predicted parameter set was applied in actual spraying experiments for validation. If the error between the actual proportions measured in three consecutive experiments and the predicted value remained within a minimal range, the process optimization was deemed to have converged, and the iteration was terminated. If the convergence criteria were not met, the newly acquired experimental data and their corresponding actual droplet proportions were added to the training dataset, prompting the model to proceed to the next round of active learning. Finally, for the fully optimized model, the game-theoretic SHAP interpretability model was introduced to quantitatively evaluate the marginal contribution of each process feature to the microsphere proportion, thereby revealing the nonlinear coupling mechanisms among the parameters.

2.5 SHAP feature interpretability model

To address the inherent black-box nature of machine learning models such as GPR, this study introduced the game-theoretic SHAP (SHapley Additive exPlanations) model. The SHAP model



provides an in-depth interpretation of prediction results by quantifying the impact of each feature. It calculates the Shapley contribution of each parameter to the model output: a value of zero indicates no impact, a positive value implies an increase in the model's predicted value, and a negative value indicates a decrease. Furthermore, a larger absolute value denotes a more significant impact, enabling a quantitative evaluation of the nonlinear coupling mechanisms among the variables.

3 Results and discussion

3.1 Construction of the optimization parameter space and adaptive search strategy

In the spray-precipitation process for fabricating SA gel microspheres, the physical properties of the precursor solution and the external atomization pressure jointly dictate the droplet size distribution. To achieve the targeted fabrication of microspheres with specific sizes, this study selected SA concentration, spray distance (d), gas pressure (GP), and liquid pressure (LP) as the core independent variables, establishing the volume median diameter D_{50} and the yield proportion of droplets within the 90–110 μm range as the dual-objective function. This specific size range was carefully selected due to its immense practical significance in biomedical applications.³¹ Microspheres within 90–110 μm typically exhibit optimal flowability and injectability, making them highly ideal for targeted drug delivery (e.g., transcatheter arterial embolization) and cell encapsulation. This size window effectively prevents capillary or needle blockage during clinical injection while providing a sufficiently high surface-area-to-volume ratio for payload release. By refining the step sizes,

a vast parameter space comprising 371 280 data points was constructed. To efficiently locate the optimum within this multi-dimensional space, a Bayesian optimization algorithm based on an adaptive strategy was employed, with its operational workflow illustrated in Fig. 2.

This strategy overcomes the limitations of traditional fixed-direction searches. It constructs a surrogate model *via* GPR and adaptively traverses the Radial Basis Function (RBF), Matern, Rational Quadratic, and Dot Product kernels during each iteration. Consequently, it dynamically matches the optimal kernel based on the minimum mean squared error (MSE) criterion. Simultaneously, the algorithm optimizes the noise level parameter over a log-uniform distribution and the maximum number of restarts over an integer range, thereby effectively preventing the model from being trapped in local minima while ensuring the positive definiteness of the kernel matrix. Such an adaptive learning mechanism enables the model to continuously refine its estimation of the hyperparameter space based on prior data, laying a robust algorithmic foundation for subsequent highly efficient parameter optimization.

3.2 Effect of prior data size on the algorithm's convergence and optimization efficiency

In the initial stage of Bayesian optimization, the quality of prior data directly dictates the surrogate model's exploration efficiency within the parameter space.²⁷ In this study, the Latin hypercube sampling (LHS) method was employed to determine the combinations of independent variables for the prior data.

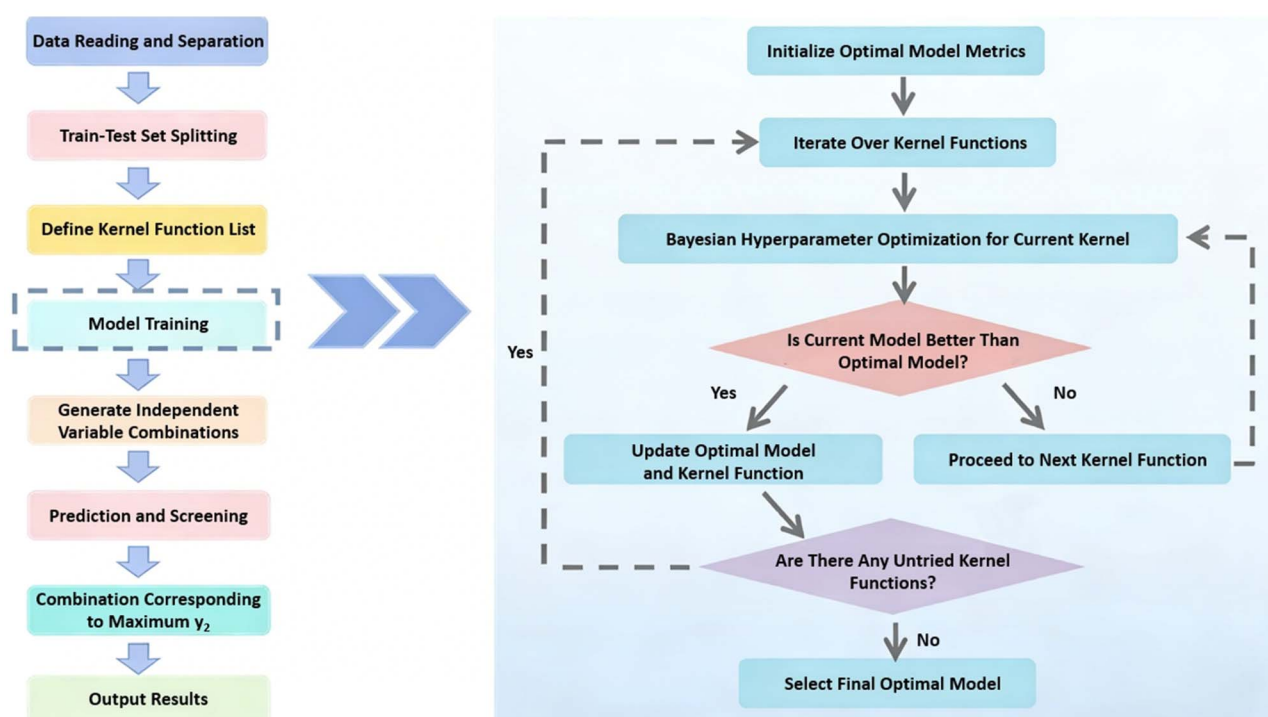


Fig. 2 Workflow of the Bayesian optimization algorithm.



Compared to traditional random sampling or grid search, the LHS method demonstrates irreplaceable advantages. It can uniformly cover the value ranges within the multi-dimensional space consisting of 371 280 independent variable combinations—formed by SA concentration, spray distance (d), gas pressure (GP), and liquid pressure (LP)—thereby effectively avoiding the over-concentration of sample points or the omission of critical regions. Based on this method, 5 and 10 sets of experimental parameter combinations were selected to conduct actual spray measurements, yielding a droplet size distribution of 90–110 μm , which served as prior data to initiate the optimization (Tables 1 and 2).

By comparing the optimization trajectories for the two different prior data sizes, it can be observed that the optimization initiated with 5 prior data points required 17 iterations to complete 22 experiments. Its predicted droplet proportion was 17.47%, exhibiting good agreement with the actual measured value of 18.35%. In contrast, with 10 prior data points, 22 experiments were completed in only 12 iterations, and the predicted value of 18.60% closely matched the measured value of 18.34%. From the perspectives of the number of iterations and experimental cost, the 10 prior data points demonstrated a significant advantage. Although both approaches ultimately completed 22 experiments, the 10 prior data points reduced the number of iterations by 29.4%. Given that a single microfluidic spray experiment and the subsequent particle size measurement are relatively rapid (about 3 min), reducing the number of iterations directly decreases overall optimization time. These results clearly demonstrate that the size of the prior data directly affects the quality of subsequent optimization training. Sufficient prior data not only reduces the number of optimization iterations but also ensures more rational optimal variable combinations. Conversely, insufficient prior data leads to more

Table 2 Experimental results with ten sets of prior data

Entry	SA/wt%	d/cm	GP/mbar	LP/mbar	MV (%)	PV (%)
1	2.38	4	400	2000	10.60	—
2	2.59	13	750	1550	0.00	—
3	2.69	7	1150	1250	18.85	—
4	1.74	15	950	500	15.59	—
5	1.00	11	600	1450	15.32	—
6	2.27	12	1100	1650	0.00	—
7	2.16	6	1500	1850	18.95	—
8	1.11	5	1300	900	14.85	—
9	1.21	14	1400	1100	13.16	—
10	1.42	10	1550	1900	13.48	—
11	1.00	2	1200	650	17.94	23.64
12	1.00	11	600	1450	12.87	15.30
13	1.00	2	750	1250	10.68	27.04
14	1.20	2	2000	500	0.00	26.71
15	1.00	2	550	500	8.76	16.40
16	1.00	2	450	1500	3.28	13.64
17	2.70	15	300	1550	18.71	24.87
18	2.70	15	350	1750	17.59	22.21
19	2.20	6	1500	1850	18.86	18.90
20	2.80	2	800	1650	11.96	17.62
21	2.70	9	800	2000	18.45	18.75
22	2.70	15	350	1750	18.34	18.60

iterations and is prone to local stagnation, where the optimized parameters cluster at boundary values.

Furthermore, a detailed analysis of the algorithm's convergence process is crucial for verifying the model's reliability. The model's convergence process represents the gradual approximation of predicted values to actual values as the number of iterations increases. As illustrated in Fig. 3, the model-predicted droplet yield proportion exhibited anomalous fluctuations, both high and low, during the optimization period. These fluctuations inherently reflect the exploration characteristics of the Bayesian optimization process within the uncharted parameter space.³² Even if a single prediction coincidentally aligns with the actual value, it may merely indicate entrapment in a local extremum. Effective convergence is signified only when the predicted values consistently and stably approximate the actual values across multiple iterations. Regardless of whether 5 or 10 prior data points were used, the model ultimately converged stably as iterations progressed. Furthermore, in conjunction with Fig. 4, it can be observed that as the number of experiments increases, the volume median diameter D_{50} of the micro-sprayed droplets gradually converges toward the expected 90–110 μm range. Concurrently, the yield of particles within the target range steadily increases, which aligns perfectly with the anticipated process optimization objectives.

3.3 Optimal process parameter combinations and experimental validation of microsphere precipitation

Based on the different prior data sizes, the Bayesian optimization model output two distinctly different optimal parameter combinations within the vast parameter space. In the optimization experiment utilizing 5 sets of prior data, the optimal parameter combination obtained after 17 iterations was an SA

Table 1 Experimental results with five sets of prior data

Entry	SA/wt%	d/cm	GP/mbar	LP/mbar	MV (%)	PV (%)
1	1.42	14	1250	1450	10.65	—
2	2.69	10	1400	1550	15.97	—
3	1.74	2	400	1750	9.42	—
4	1.32	13	1050	500	12.38	—
5	2.16	3	600	1900	17.33	—
6	2.60	2	350	600	0.00	20.50
7	2.20	3	600	1900	0.00	17.30
8	2.80	15	1850	2000	0.00	43.47
9	1.00	5	350	1250	9.92	16.00
10	2.80	2	950	1950	20.39	25.54
11	1.00	3	300	1450	1.89	0.00
12	2.80	2	450	500	23.91	43.94
13	2.70	15	1700	700	14.74	11.68
14	1.00	8	1200	800	12.61	14.76
15	1.00	2	2000	500	14.25	24.25
16	1.00	2	2000	500	14.69	14.00
17	2.80	2	450	500	15.45	23.86
18	2.80	2	450	500	15.51	21.17
19	2.80	2	950	1950	20.34	20.34
20	2.80	2	650	500	17.51	18.42
21	2.80	2	950	1950	20.31	20.34
22	2.80	2	650	500	18.35	17.47



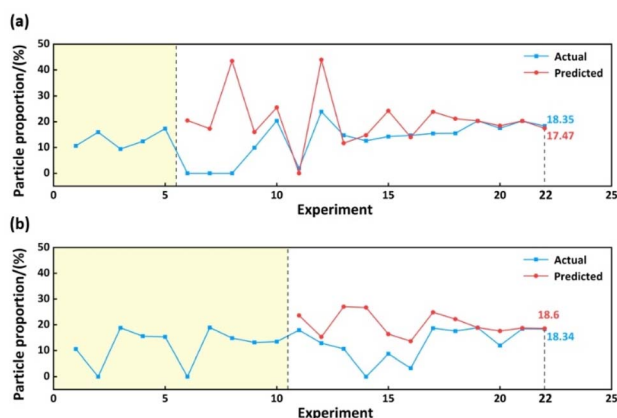


Fig. 3 Actual and GPR-predicted droplet proportions versus experimental iterations for optimizations initiated with (a) five and (b) ten sets of prior data.

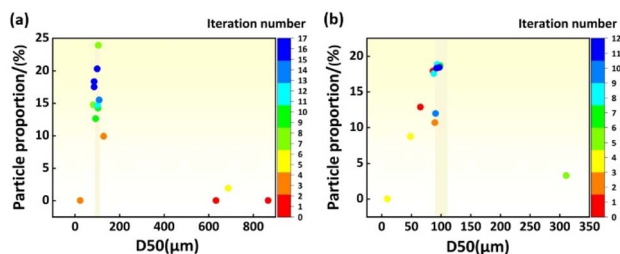


Fig. 4 Scatter plots of actual droplet proportion and D_{50} versus iteration number for optimizations initiated with (a) five and (b) ten sets of prior data.

concentration of 2.80 wt%, a spray distance of 2 cm, a gas pressure of 650 mbar, and a liquid pressure of 500 mbar, with a corresponding measured droplet yield proportion of 18.35%. In contrast, for the optimization experiment based on 10 sets of prior data, the identified optimal parameter combination was an SA concentration of 2.70 wt%, a spray distance of 15 cm, a gas pressure of 350 mbar, and a liquid pressure of 1750 mbar, yielding a measured droplet proportion of 18.34%. Notably, although these two sets of parameters exhibited substantial numerical differences, the ultimately obtained actual yield proportions of 90–110 μm droplets showed remarkable consistency, with a negligible difference of only 0.01%. The occurrence of this phenomenon is primarily attributed to the highly complex nonlinear coupling relationships among various influencing factors during the microfluidic spraying process. For instance, an increase in SA concentration leads to a significant elevation in liquid viscosity, which subsequently hinders the effective breakup of droplets. To counterbalance the increased internal viscous resistance of the liquid and prevent excessive particle sizes, the system must substantially adjust the gas or liquid pressure to re-optimize the spraying kinetic energy. Such mutual constraints and compensation effects among multiple factors result in multiple physically equivalent local optima within the optimization space. Concurrently, the

different distributions of the initial LHS sampling points guided the model to converge to distinct local optima, thereby yielding distinct optimal parameter combinations.

To further validate the model's predictions and their practical applicability, the predicted optimal atomization conditions were applied to the calcium chloride receiving bath to achieve crosslinking and curing of the SA microspheres. As illustrated in Fig. 5, when the experiment was conducted with the parameter combination derived from the 5 prior data points, the yield of cured gel microspheres in the 90–110 μm range reached 19.67%, corresponding to an original droplet proportion of 18.35%. Conversely, when utilizing the parameters derived from the 10 prior data points, this microsphere proportion was 18.75%, corresponding to an original droplet proportion of 18.34%. Compared with the measured size distribution of the original droplets, the yield proportion of target-sized microspheres showed a slight deviation after post-curing. This phenomenon is primarily attributed to two physicochemical mechanisms: first, upon the droplets falling into the receiving bath, the calcium ions in the solution instantaneously undergo intense ionic crosslinking reactions with the carboxyl groups on the SA molecular chains. It induces conformational shrinkage of the polymer chains, thereby decreasing the overall size of the crosslinked gel microspheres relative to the original droplets. Second, during static curing, minor water evaporation from the system also affects the final size of the gel microspheres. Through statistical analysis of the volume median diameter (D_{50}) before and after crosslinking, an average diameter shrinkage of approximately 4.5% to 6.0% was quantified in our experiments. In future practical production, introducing this average shrinkage rate (e.g., $\sim 5\%$) as a calibration coefficient into the objective function of the Bayesian model will further refine the predictive accuracy for the final product, completely bridging the gap between liquid droplet generation and solid microsphere formulation. Therefore, a reasonable discrepancy between the microsphere and droplet

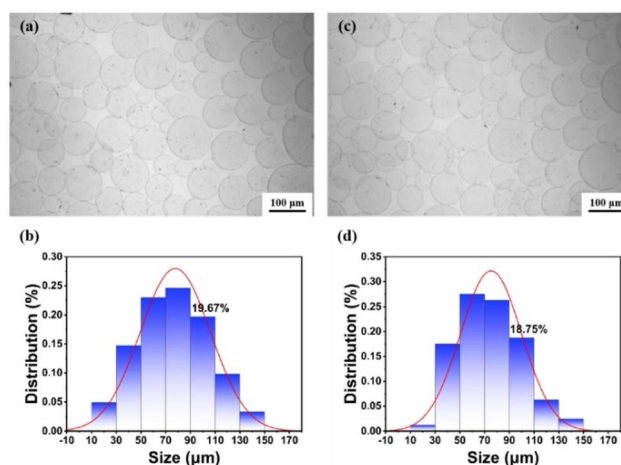


Fig. 5 Optical micrographs and particle size distribution of gel microspheres formed after solidification and precipitation: (a and b) experimental validation results with five sets of prior data; (c and d) experimental validation results with ten sets of prior data.



proportions within the target size range is considered a normal phenomenon. This further substantiates that the constructed Bayesian model can provide a reliable and flexible predictive benchmark for the tailored fabrication of target-sized microspheres.

3.4 Feature parameter analysis

Based on the SHAP model established in Section 2.5, we performed an in-depth interpretation of the prediction results. In Fig. 6, the color of the scatter points intuitively reflects the magnitude of the feature values (red indicates high values, while blue indicates low values).

The analysis reveals significant differences in the impact of varying prior data sizes on the model's training depth. For the 5 sets of prior data, prior to optimization (Fig. 6a and b), the scarcity of experimental data resulted in poor model generalization capability, failing to clearly extract the underlying mechanisms by which the four features influence the droplet

proportion. However, after completing the active learning iterations, the distribution range of SHAP values for the liquid pressure (LP) feature expanded significantly, elevating it to a key dominant factor governing the droplet proportion. This machine learning-derived conclusion perfectly aligns with previously established physical principles that gas and liquid pressures are the most prominent external driving forces for fluid shear atomization. Conversely, in the analysis of the 10 prior data sets (Fig. 6c and d), the spray distance (d) exerted a strong negative inhibitory effect in some experiments; as the spray distance increased (indicated by the red, high-feature values), the predicted droplet proportion decreased significantly. From an aerodynamic perspective, this occurs because the droplet swarm undergoes severe dispersion and spatial secondary evolution over the extended flight distance, inevitably broadening the initially concentrated particle size distribution. In contrast, the SHAP values for the SA concentration feature consistently clustered around zero post-optimization, indicating that its marginal impact on the predicted target proportion is relatively minor within the current parameter window.

A comprehensive analysis of Fig. 6 further reveals a key global pattern: the extremely high values of each feature (represented by red points) tend to cluster at both ends of the SHAP value axis, exerting a tremendous absolute impact on the predicted values. From a data science perspective, this phenomenon perfectly corroborates the physical essence underlying the coexistence of multiple optimal parameter combinations discussed in Section 3.3. Owing to strong coupling among multiple variables, when a specific feature (such as SA concentration or spray distance d) reaches an extreme high value that may adversely affect the target particle size, the model can adaptively identify a compensatory pathway. Specifically, by substantially adjusting the gas pressure (GP) and liquid pressure (LP) to appropriate levels, the system rebalances the kinetic energy of fluid breakup against viscous dissipation. This ensures that the yield proportion of droplets within the 90–110 μm range can still be globally maximized even under extreme conditions. Ultimately, this SHAP-based feature interpretation approach not only endows the optimization model with profound physicochemical interpretability but also provides a solid scientific rationale for the multi-dimensional regulation of complex multiphase flow processes.

4 Conclusions

This study proposed an intelligent optimization framework based on adaptive Bayesian optimization and Gaussian process regression (GPR), successfully overcoming the engineering challenges of strong nonlinear parameter coupling and the difficulty of precisely regulating target-sized (90–110 μm) droplets in the spray-precipitation process for SA microspheres. The results demonstrate that this strategy achieves highly efficient optimization at a minimal experimental cost. Initiating optimization with 10 prior data points required only 12 active learning iterations to reach convergence, reducing the iteration cost by 29.4% compared to 5 prior data points. Under optimal

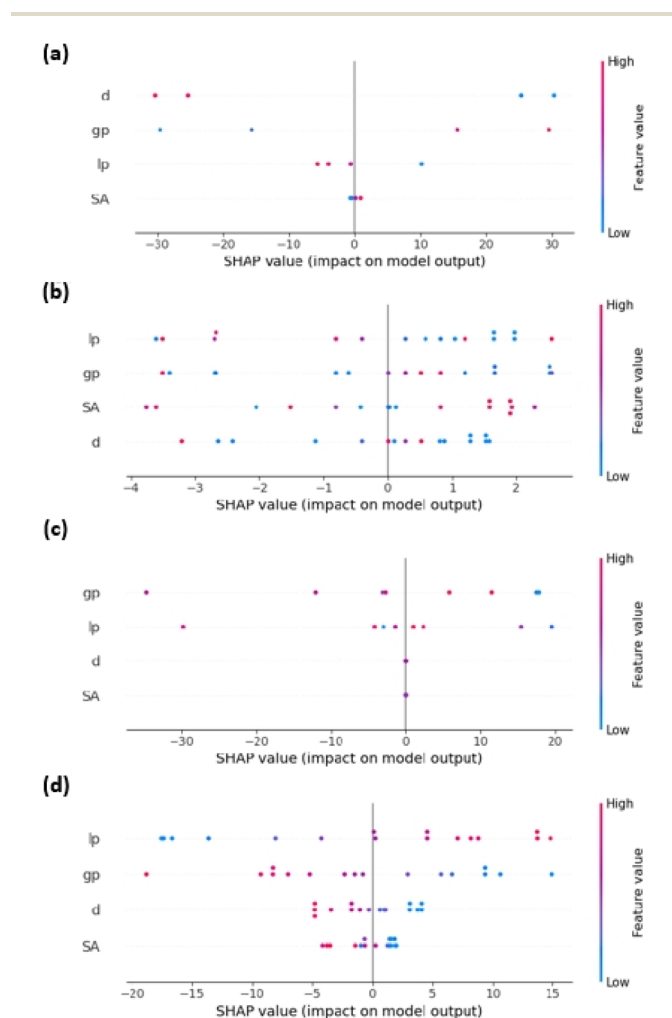


Fig. 6 SHAP value analysis for optimization experiments: (a) SHAP values of features before optimization with five sets of prior data; (b) SHAP values after optimization with five sets of prior data; (c) SHAP values before optimization with ten sets of prior data; (d) SHAP values after optimization with ten sets of prior data.



parameter settings, the model-predicted droplet yield proportion (18.60%) closely matched the measured value (18.34%). Furthermore, the target size proportion of the gel microspheres post-curing and sedimentation remained high at 18.75–19.67%, achieving high-precision closed-loop validation from algorithmic prediction to physical microsphere fabrication. Additionally, the introduction of the SHAP interpretability model demystified the algorithmic black box, precisely revealing the dominant roles of gas and liquid pressures in fluid breakup. It also elucidated, from a physical perspective, the mechanisms of fluid dynamics by which nonlinear compensation among multiple parameters leads to the coexistence of multiple local optima. Furthermore, it is worth emphasizing that this adaptive optimization framework exhibits remarkable flexibility in practical applications. By simply modifying the target size window in the objective function, the system can be easily adjusted to accommodate smaller microspheres (e.g., 30–50 μm) for inhalation therapy, or larger hydrogel beads (e.g., 200–300 μm) for three-dimensional cell culture and tissue engineering. This work not only significantly reduces the trial-and-error costs of tailored fabrication of target-sized microspheres but also establishes a highly efficient research paradigm for smart material fabrication that deeply integrates data-driven optimization with physical mechanism interpretation.

Author contributions

SF: conceptualization, writing – original draft, writing – review & editing. JL: conceptualization, writing – original draft, writing – review & editing. ZL: writing – original draft, writing – review & editing. ST: conceptualization, funding acquisition, writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Data availability

All data supporting the findings of this study are available within the paper and its supplementary information (SI). Supplementary information: the relevant experimental operations described in the text and other results. See DOI: <https://doi.org/10.1039/d6ra02906j>.

Acknowledgements

The authors would like to acknowledge the financial support from the National Natural Science Foundation of China (No. 22372025).

References

- H. A. Smith, J. Zhou and H. L. Buckley, *Green Chem.*, 2026, **28**, 2846–2862.
- F. Zainab, S. Mir, S. W. Khan, N. S. Awwad and H. A. Ibrahim, *RSC Adv.*, 2025, **15**, 19983–20005.
- R. Y. Pastrana-Alta, E. Huarote-Garcia, M. A. Egusquiza-Huamani and A. M. Baena-Moncada, *RSC Adv.*, 2025, **15**, 35807–35843.
- A. E. Ashmar, E. J. Beckman and S. K. Fullerton-Shirey, *Green Chem.*, 2025, **27**, 13480–13488.
- K. Cysewska, L. Schöbel and A. R. Boccaccini, *J. Mater. Chem. B*, 2026, **14**, 2324–2339.
- A. S. Sokolov, A. S. Abdurashitov, P. I. Proshin and G. B. Sukhorukov, *J. Mater. Chem. B*, 2025, **13**, 12166–12171.
- X. Wang, Y. Lin, S. Jiao and X. Liu, *Chem. Commun.*, 2025, **61**, 16572–16575.
- M. E. Astaneh, A. Hashemzadeh and N. Fereydouni, *J. Mater. Chem. B*, 2024, **12**, 10163–10197.
- S. Wang, Y. Fu, Z. Xu, J. Zhang, Y. Niu and J. Wang, *Carbohydr. Polym.*, 2026, **375**, 124772.
- P. Yan, W. Lan and J. Xie, *Trends Food Sci. Technol.*, 2024, **143**, 104217.
- Y. G. Bi, Z. T. Lin and S. T. Deng, *Mater. Sci. Eng., C*, 2019, **100**, 576–583.
- Y. Chen, X. Liu, R. Zhou, J. Qiao, J. Liu, R. Cai, J. Liu, J. Rong and Y. Chen, *Int. J. Biol. Macromol.*, 2024, **278**, 135000.
- S. Y. Yang, J.-S. Park, J. H. Kim, M. Yoon, S. E. Wang, D. S. Jung and Y. C. Kang, *Mater. Today Chem.*, 2024, **35**, 101889.
- V. Prosapio, I. De Marco and E. Reverchon, *Chem. Eng. J.*, 2016, **292**, 264–275.
- Y. Li, J. Mei, X. Guo, B. Zhong, H. Liu, G. Liu and S. Dou, *RSC Adv.*, 2016, **6**, 70091–70098.
- G. D. Park, J. Lee, Y. Piao and Y. C. Kang, *Chem. Eng. J.*, 2018, **335**, 600–611.
- M. Pal, L. Wan, Y. Zhu, Y. Liu, Y. Liu, W. Gao, Y. Li, G. Zheng, A. A. Elzatahry, A. Alghamdi, Y. Deng and D. Zhao, *J. Colloid Interface Sci.*, 2016, **479**, 150–159.
- H. P. Duong, T. Mashiyama, M. Kobayashi, A. Iwase, A. Kudo, Y. Asakura, S. Yin, M. Kakihana and H. Kato, *Appl. Catal., B*, 2019, **252**, 222–229.
- B. Sarma, S. Kumar, A. Dalal, D. N. Basu and D. Bandyopadhyay, *Phys. Rev. Appl.*, 2021, **15**, 014005.
- K. Hanthanan Arachchilage, M. Haghshenas, S. Park, L. Zhou, Y. Sohn, B. McWilliams, K. Cho and R. Kumar, *Adv. Powder Technol.*, 2019, **30**, 2726–2732.
- E. Dalir, A. Dolatabadi and J. Mostaghimi, *Int. J. Heat Mass Transfer*, 2022, **182**, 121969.
- A. Shokry, S. Medina-González, P. Baraldi, E. Zio, E. Moulines and A. Espuña, *Chem. Eng. J.*, 2021, **425**, 131632.
- N. Li, J. Xu, W. Gao, S. Wan, X. Xu, B. Yan, J. Wang and G. Chen, *Chem. Eng. J.*, 2026, **528**, 172413.
- E. Reus, J. Savinsky, S. Wennemaring, J. Käsbaach, F. Kerkhoffs, J. Kehrein, S. B. Rauer, T. Lühmann, A. C. Adams, M. Wessling, J. Magnus and L. Meinel, *J. Controlled Release*, 2025, **388**, 114370.
- M. Liu, H. Hu, Y. Cui, J. Song, L. Ma, Z. Yuan, K. Wang and G. Luo, *Chem. Eng. J.*, 2025, **511**, 161972.
- S. Desimpel, M. Dorbec, K. M. Van Geem and C. V. Stevens, *Chem. Soc. Rev.*, 2026, **55**, 2731–2775.



- 27 B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams and A. G. Doyle, *Nature*, 2021, **590**, 89–96.
- 28 X. Li, Y. Che, L. Chen, T. Liu, K. Wang, L. Liu, H. Yang, E. O. Pyzer-Knapp and A. I. Cooper, *Nat. Chem.*, 2024, **16**, 1286–1294.
- 29 V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, *Chem. Rev.*, 2021, **121**, 10073–10141.
- 30 W. Y. Jing Liu, Y. Lyu and S. Tao, *CIESC J.*, 2024, **75**, 1724–1734.
- 31 Q. Wang, K. Qian, S. Liu, Y. Yang, B. Liang, C. Zheng, X. Yang, H. Xu and A. Q. Shen, *Biomacromolecules*, 2015, **16**, 1240–1246.
- 32 B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick and A. I. Cooper, *Nature*, 2020, **583**, 237–241.

