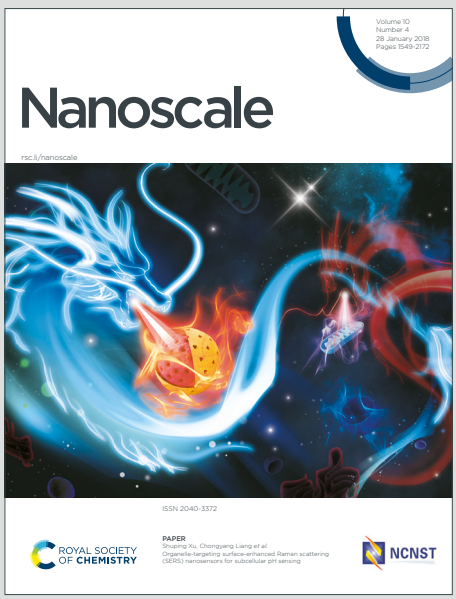
Nanoscale

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: S. Zahedi Asl, S. Abaei, H. Alimardani, M. Beydaghdari, F. Hooriabad Saboor and M. Asgari, *Nanoscale*, 2026, DOI: 10.1039/D5NR04817F.



ROYAL SOCIETY OF CHEMISTRY

rsc.li/nanoscale

From Surface Area to Functionality: Data-Driven Insights into MIL-100(Fe) Synthesis for Enhanced Dye Removal Efficiency

View Article Online
DOI: 10.1039/D5NR04817F

Saeid Zahedi Asl[1, †], *Shayan Abaei*[2, †], *Hosein Alimardani*[2], *Mohammadreza Beydaghdari*[1], *Fahimeh Hooriabad Saboor*[1,*], *Mehrdad Asgari*[3,4,*]

[1]Department of Chemical Engineering, Faculty of Engineering, University of Mohaghegh Ardabili, Ardabil 56199-11367, Iran

[2]School of Chemical Engineering, College of Engineering, University of Tehran, Tehran, Iran

[3]Department of Chemical Engineering & Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, UK

[4] Lucy Cavendish College, University of Cambridge, Cambridge CB3 0BU, United Kingdom

† These authors have contributed equally to this work.

* Corresponding authors: Fahimeh Hooriabad Saboor: f.saboor@uma.ac.ir, Mehrdad Asgari: ma2000@cam.ac.uk

View Article Online
DOI: 10.1039/D5NR04817F

Abstract

Traditional MOF design often maximizes generic metrics such as BET surface area and crystallinity, assuming they universally predict performance. We present a machine-learning-guided optimization framework for MIL-100(Fe) from experimentally synthesized samples. All materials were synthesized via an acid-free, water-based hydrothermal route. We trained small-data ML models to link synthesis parameters, including temperature, time, metal-to-ligand molar ratio, and ion concentration, to key properties comprising surface area, total pore volume, average crystallite size, crystallinity, yield, and methylene blue (MB) removal. SHAP analysis showed that time and metal-to-ligand molar ratio dominated dye removal, whereas surface area was more sensitive to temperature and time. The most accurate model, Gaussian process regression was coupled with a genetic algorithm (GA) to optimize synthesis for property-specific targets. Through optimization, the BET-optimized sample increased surface area from the highest baseline value in the initial experimental dataset, 1748 to 1841.9 m²/g, corresponding to a 5.37% relative increase. The MB-optimized sample increased MB removal from the highest baseline value in the initial experimental dataset, 88.6% to 98.3%, corresponding to a 9.7 percentage-point improvement and a 10.9% relative increase. The MB removal optimized sample with a surface area of 1274.3 m²/g and 17.4% crystallinity, delivers the highest MB uptake of 98.3%, corresponding to the highest adsorption capacity under the benchmark test conditions. While, optimized sample for surface area, reaches 1841.9 m²/g (about 44% higher) with 34.9% crystallinity but achieves only 85.1% removal, about 13% lower than optimized MB sample; indicating that even substantial increases in surface area do not govern adsorption performance. A qualitative t-SNE embedding of the descriptor space shows that the optimized samples

occupy distinct neighborhoods, elucidating that MOF synthesis should be tailored to the target application rather than a single metric such as surface area.

View Article Online
DOI: 10.1039/D5NR04817F

1. Introduction

Metal-organic frameworks (MOFs) are a class of porous, crystalline materials constructed from metal ions or clusters coordinated to organic ligands, forming highly ordered, multidimensional networks [1]. Since the groundbreaking synthesis of the first microporous MOFs by Omar Yaghi and colleagues in the late 1990s%[2, 3], research in this area has accelerated rapidly. Today, MOFs represent one of the most extensively explored families of materials in chemistry and materials science. Over 120,000 unique MOF structures have been experimentally synthesized to date, while high-throughput computational screenings predict the existence of millions more with diverse structural and functional characteristics%[4]. The remarkable appeal of MOFs lies in their structural diversity, synthetic tunability, and ultrahigh surface areas, which make them ideal candidates for a broad range of applications from gas storage and separation to catalysis, sensing, and environmental remediation. However, the real-world deployment of MOFs, particularly in aqueous and harsh environments, is often hindered by stability issues. While frameworks constructed from high-valent metals such as zirconium, hafnium, or chromium typically exhibit excellent hydrolytic and thermal stability, those based on divalent metals like zinc and copper are sensitive to degradation under humid or elevated-temperature conditions%[5, 6]. The intrinsic stability of a MOF is governed by a combination of factors, including the strength of the metal-ligand bonds, the rigidity of the framework topology, and the specific environmental conditions in which the material is used%[5-7]. For instance, Zr-based UiO-66, iron-based MIL-100 and Cr-based MIL-101 have demonstrated exceptional hydrolytic stability in aqueous

View Article Online
DOI: 10.1039/D5NR04817F

environments, while Zn-based MOFs like ZIF-8 can be more susceptible to water-induced structural degradation depending on the exposure conditions [5-7].

Among the many potential applications of MOFs, water purification has emerged as a up-and-coming field due to the urgent need for effective and selective pollutant removal technologies. MOFs offer several advantages in this context, including high porosity, large surface area, and the ability to tailor pore environments through functional group modifications. Notably, materials such as ZIF-8, UiO-66, and MIL-100(Fe) have exhibited strong adsorption capabilities toward a wide range of water contaminants, including heavy metals (e.g., Pb^{2+}, Hg^{2+}, Cr^{6+}), synthetic dyes, and pharmaceutical compounds%[8, 9]. Their selective adsorption behavior is primarily driven by host-guest interactions, including electrostatic attraction, hydrogen bonding, π-π stacking, and molecular sieving effects. For example, UiO-66-NH_2 has shown effective removal of Pb^{2+} ions through strong coordination interactions between the metal centres and amine-functionalized ligands [10]. Similarly, MIL-100(Fe) demonstrates high affinity for methylene blue (MB), attributed to its large pore volume, extensive π-π interactions, and the presence of Lewis acidic sites that enhance guest molecule binding%[11, 12]. MB was selected as a representative cationic dye pollutant and benchmark adsorbate for evaluating water-treatment materials. It is widely used in adsorption studies because its cationic charge, aromatic structure, strong visible absorbance, and aqueous stability make it a convenient model molecule for probing electrostatic attraction, π–π interactions, and pore-accessibility effects in porous adsorbents [13-15].

For water-treatment applications, adsorption capacity alone is insufficient to establish practical relevance; the adsorbent must also retain structural and chemical integrity in aqueous media, with minimal release of framework components into the treated water. This is particularly important for MOFs, where metal-node

View Article Online
DOI: 10.1039/D5NR04817E

or linker leaching can compromise both long-term performance and secondary water quality. MIL-100(Fe)-based adsorbents are attractive in this respect because several aqueous adsorption studies have reported good water stability, negligible iron leaching, and successful reuse over multiple cycles, supporting their suitability as iron-based MOF platforms for pollutant removal [16, 17].

Although it is now well recognized that BET surface area alone does not universally determine MOF performance, many synthesis studies still use surface area and crystallinity as convenient proxy metrics for judging material quality. High-throughput computational screenings have repeatedly shown that adsorption, separation, and transport performance depend on multiple descriptors, including pore size, void fraction, accessible pore volume, electrostatic environment, and adsorbate–framework interactions, rather than surface area alone. The remaining challenge is therefore not simply to restate that surface area is insufficient, but to determine how synthesis conditions within a fixed MOF family shift the balance between textural properties and application-specific performance%[18]. This distinction is central to the present study, where MIL-100(Fe) samples synthesized under different hydrothermal conditions show that the route maximizing BET surface area is not the same as the route maximizing MB removal [19, 20].

It is important to distinguish between framework-level selection and synthesis-condition optimization within a fixed framework family. Across different MOFs, an optimally chosen crystalline framework with suitable pore dimensions, topology, chemical functionality, and stability may indeed provide the best performance for a given adsorbate. In that broader materials-selection context, high crystallinity and well-defined pores are often desirable. However, within a single MOF system synthesized under different conditions, such as MIL-100(Fe) in the present work, variations in crystallinity, defect density, pore accessibility, and surface chemistry can shift the relationship between BET surface area and adsorption performance. Thus, our argument is not that defects or lower crystallinity are universally superior, but rather

View Article Online
DOI: 10.1039/D5NR04817F

that within a fixed MOF family, maximizing BET surface area alone may not identify the synthesis condition that gives the best application-specific performance. This distinction is consistent with recent discussions of pore engineering and defect engineering in MOFs, where performance depends on the balance between accessible pore architecture, chemical functionality, active sites, and framework integrity [20, 21].

Recent research has illuminated cases where MOFs with moderate or even lower surface areas exhibit enhanced adsorption capabilities, challenging the notion that "more surface" always means "better function." For instance, nanoscale MOFs, here referring to MOF nanocrystals or nanoparticle assemblies with reduced crystal/particle dimensions relative to bulk crystals, can exhibit improved adsorption kinetics despite lower apparent surface areas. This behavior is commonly attributed to shorter diffusion pathways, larger external surface-to-volume ratios, and improved accessibility of adsorption sites, rather than to a change in framework topology itself% [22, 23]. Sikdar et al. demonstrated that such nanoscale structures facilitate faster guest-host interactions, a key advantage in time-sensitive separation or remediation processes% [24]. While high surface area remains an asset in applications like gas storage and catalysis; for instance HKUST-1, with a surface area exceeding 2000%m²/g have enabled exceptional hydrogen and methane uptake% [25, 26], conforming that in these cases, surface area must be viewed through the lens of usability. Computational studies, such as the one by Moosavi et al., reinforce this view by identifying surface area as a leading performance descriptor in gas adsorption, yet also acknowledging the contribution of other structural variables% [27]. More precisely, BET surface area should be regarded as a textural descriptor related mainly to the potential number of accessible adsorption sites and equilibrium uptake, whereas adsorption kinetics depend on mass-transfer and accessibility factors such as pore-window size, diffusion path length, particle/crystallite size, temperature, concentration gradients, and agitation.

View Article Online
DOI: 10.1039/D5NR04817F

Therefore, the present discussion distinguishes between surface-area-related capacity effects and transport/accessibility-controlled kinetic effects [28, 29].

Indeed, parameters such as pore geometry, defect density, framework flexibility, and surface chemistry play pivotal roles in determining the actual utility of accessible surface area%[30]. Experimental results further substantiate this more nuanced perspective. Ba Luan Tran et al. reported that a ZIF-8 sample with a lower surface area of 891%m²/g exhibited greater adsorption capacity than a higher surface area counterpart (1285%m²/g), attributed to enhanced pore connectivity and diffusion dynamics%[31]. Similarly, Piscopo et al. observed that fluorine-functionalized UiO-66 MOFs exhibited superior oxygen storage capacity despite reduced surface area, due to altered electronic environments that increased adsorption site affinity [32]. These findings collectively highlight the importance of balancing physical parameters with tailored chemical functionality. The path to optimal MOF performance lies not in unquestioningly maximizing a single property, but in tuning multiple features such as structural and chemical properties according to the demands of the target application%[33].

To address this multifaceted challenge, data-driven methodologies, particularly machine learning (ML) and heuristic optimization strategies, are revolutionizing how MOFs are designed and synthesized. Recent reviews further highlight that ML is becoming a central tool for MOF discovery, structure and property mapping, synthesis optimization, and performance prediction across diverse application domains [34]. These approaches offer the ability to uncover complex, nonlinear correlations between synthesis parameters and performance metrics, significantly reducing reliance on trial-and-error experimentation%[35, 36]. Genetic algorithms (GAs) have shown particular promise in this context, as their population-based search mechanisms efficiently explore high-dimensional, multimodal synthesis spaces where trade-offs like surface area versus defect density must be strategically balanced% [37]. Moreover, advancements in surrogate

View Article Online
DOI: 10.1039/D5NR04817F

modeling, notably Gaussian process regression (GPR), are enabling the development of accurate predictive models even with limited experimental data, which is particularly valuable in fields where data acquisition is costly and labour-intensive%[38]. Collectively, these innovations are leading in a new era of intelligent MOF design, where computational foresight and experimental insight converge to enable application-specific, performance-optimized materials.

In this study, we present a comprehensive and data-driven approach to systematically evaluate and optimize the hydrothermal synthesis of MIL-100(Fe) for MB dye adsorption. Moving beyond traditional emphasis on BET surface area and crystallinity, we investigate the nuanced effects of synthesis parameters including temperature, metal-to-ligand molar ratio, reaction time, and ion concentration on a suite of functional properties including surface area, total pore volume, crystallinity, average crystallite size, material yield and MB removal efficiency. To efficiently explore the high-dimensional synthesis space with minimal experimental burden, we employ a Taguchi Design of Experiments (DoE) strategy, generating a compact yet robust dataset of only 11 synthesis trials. This intentionally limited dataset serves to demonstrate the feasibility of applying machine learning and optimization workflows in low-data regimes, a critical capability for experimental domains where data generation is costly or time-consuming. Structural and performance characterizations are conducted using BET, XRD, FTIR, and TG analyses. To decode the complex relationships between synthesis conditions and material performance, we integrate a machine learning framework based on GPR and GA optimization. GPR's strength in modeling small datasets enables accurate predictions of key response variables, including surface area (m^2/g), total pore volume (cm^3/g), crystallinity (%), crystallite size (nm), yield (%), and MB removal efficiency (%). The GA module then efficiently navigates the multidimensional design space to identify optimal synthesis pathways tailored to enhance specific properties. This hybrid ML-GA framework alongside SHAP (SHapley Additive

View Article Online
DOI: 10.1039/D5NR04817F

exPlanations) feature importance analysis and t-SNE technique, represents a decisive paradigm shift from empirical trial-and-error toward rational, application-driven MOF design which offering a versatile blueprint for next-generation materials engineering in dye removal, wastewater remediation, and broader environmental applications.

2. Experimental

2.1. Materials and Synthesis of MIL-100(Fe)

The synthesis of MIL-100(Fe) was conducted using key reagents: deionized water (DW), trimesic acid (H_3BTC) from Sigma-Aldrich, iron(III) chloride hexahydrate ($FeCl_3 \cdot 6H_2O$) and ethanol from Merck, as well as methylene blue from Merck. **Figure 1A** illustrates the acid-free hydrothermal synthesis of MIL-100(Fe) using $FeCl_3 \cdot 6H_2O$ and BTC ligand under moderate conditions. A hydrothermal method was applied, with deionized water used as a green solvent. **Figure 1B** shows the UV-Vis absorbance spectrum of MB, highlighting a characteristic peak at 664%nm used to monitor adsorption performance. Solutions of 0.1 M hydrochloric acid (HCl) and sodium hydroxide (NaOH) were utilized to adjust the pH.

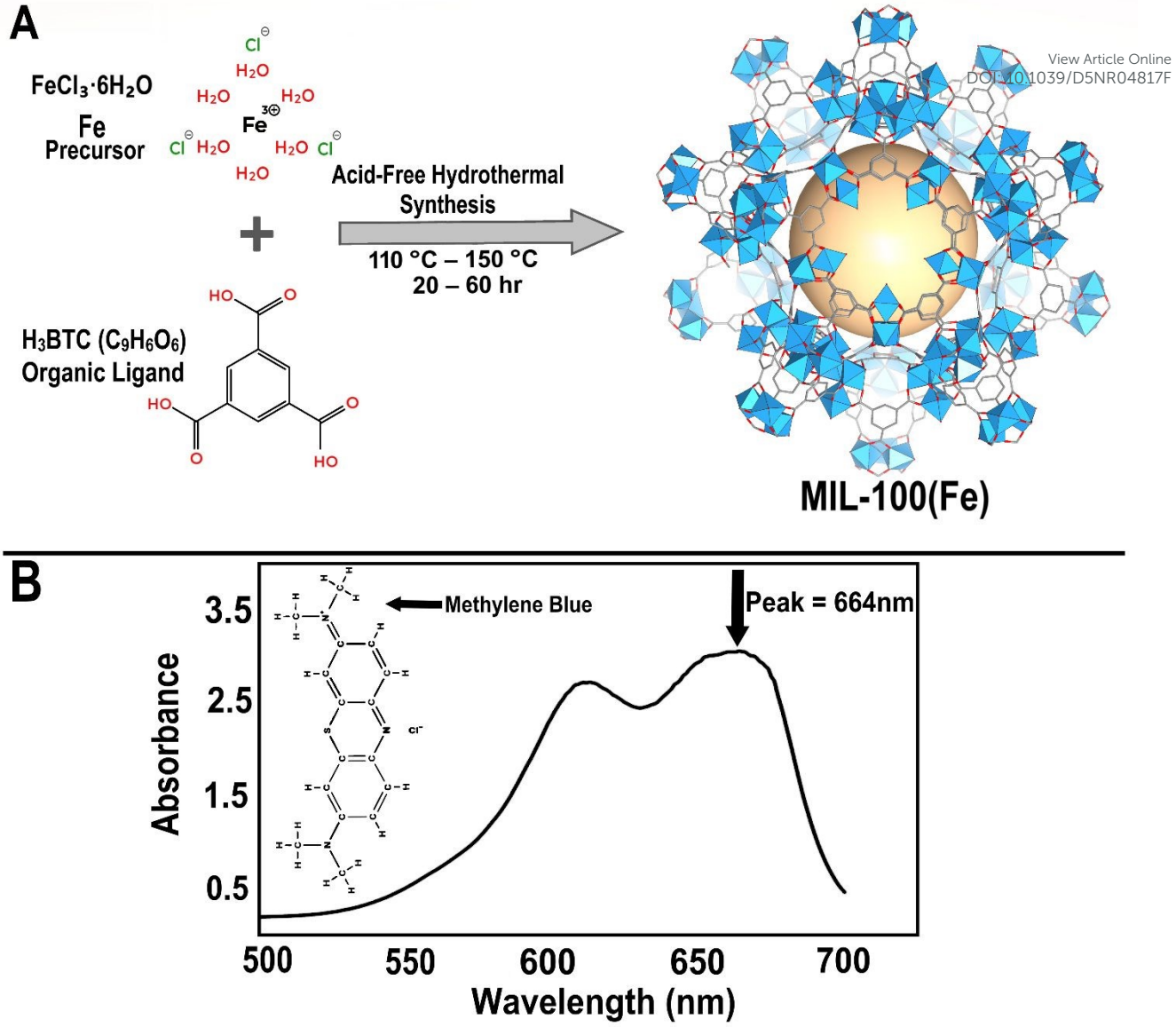


Figure 1. (A) Schematic representation of the green synthesis of MIL-100(Fe) using $FeCl_3 \cdot 6H_2O$ as the metal source and benzene- $C_9H_6O_6$ (BTC) as the organic ligand under mild, acid-free hydrothermal conditions (110%˚C - 150%˚C, 20 - 60 hours). (B) UV-Vis absorbance spectrum of MB, showing a maximum peak at 664%nm.

In a typical synthesis, $FeCl_3 \cdot 6H_2O$ and H_3BTC were dissolved in 40 mL of DW, stirred at room temperature for 30 minutes until a homogeneous orange solution was formed, and then transferred to a Teflon-lined autoclave. The synthesis was carried out in an oven at a given temperature and time. The resulting powder was separated using centrifugation at 10,000 rpm for 10 minutes, washed with DW and ethanol several times, then dried and activated in a vacuum oven at 60 ˚C for 24 hours.

View Article Online
DOI: 10.1039/D5NR04817F

The synthesis parameters, including hydrothermal temperature (110–150 ˚C), hydrothermal time (20–60 h), Fe:H3BTC molar ratio (1:1 to 1:3.5), and Fe ion concentration (0.15–0.45 M), were systematically varied to study their influence on material properties and dye-removal performance. The investigated parameter ranges are summarized in **Table 1**, while the detailed synthesis conditions for the 11 experimental runs are provided in **Table 2**. **Figure 2** provides a schematic of the synthesis process, showing each stage from initial solution preparation to drying and activation. The yield for each synthesis run was determined as the ratio of the actual mass of the dried MIL-100(Fe) powder to the theoretical mass expected from stoichiometric calculations, enabling consistent comparison of material recovery across synthesis batches. The experimental framework was constructed using a modified Taguchi-based design. Four synthesis variables were considered hydrothermal time, hydrothermal temperature, Fe:H_3BTC molar ratio, and Fe ion concentration; each at three levels. A full factorial design for these variables would require $3^4 = 81$ experiments. To reduce the experimental burden while maintaining systematic coverage of the synthesis space, a Taguchi L9 orthogonal array was used to define the first nine synthesis conditions. In this array, the levels of each factor are distributed in a balanced manner across the experimental runs, enabling efficient screening of the main synthesis effects. Two supplementary non-orthogonal design points were then added to broaden coverage of the synthesis space and improve model robustness. The final 11-run design is shown in **Table 2**, and a pairwise visualization of the design-space coverage is provided in the Supporting Information (**Figure S32**). Also, A brief summary of Taguchi different levels has been brought in the **Table S5.**

View Article Online
DOI: 10.1039/D5NR04817F

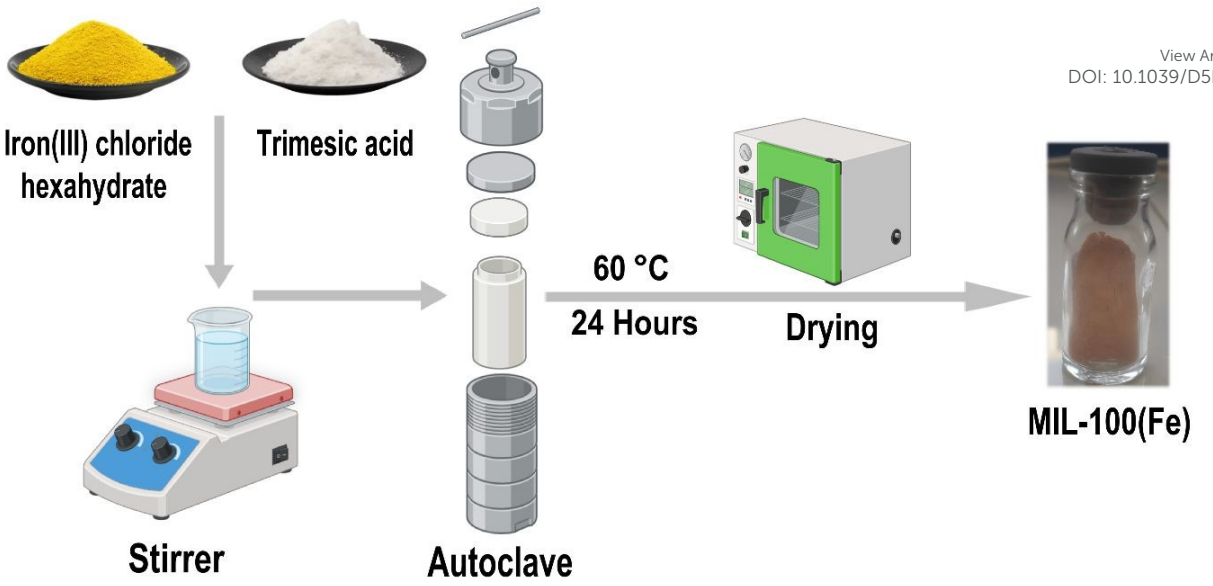


Figure 2. Schematic representation of the synthesis steps for MIL-100(Fe).

Table 1. Summary of synthesis parameters and their investigated ranges for the hydrothermal synthesis of MIL-100(Fe). The parameters include temperature, reaction time, metal-to-ligand molar ratio, and metal concentration.

Parameter	Range
Hydrothermal Temperature (˚C)	110 - 150
Hydrothermal Time (hours)	20 - 60
Fe: H3BTC Molar Ratio	1:1 - 1:3.5
Fe ion Concentration (M)	0.15 - 0.45

2.2. Characterization techniques

BET analysis through N_2 adsorption-desorption (77 K), (BELSORP-mini II surface area and porosity analyzer (BEL Japan)), X-ray diffraction (XRD; Rigaku Ultima IV with Cu Kα radiation (1.5406 Å, 0.04˚ steps)), thermogravimetric analysis (TGA; Linseis STA PT-1000), Fourier-transform infrared spectroscopy (FTIR; PerkinElmer LX185256 (4000 to 400 cm^{-1})), and scanning electron microscopy (SEM; Zeiss Ultra-

View Article Online
DOI: 10.1039/D5NR04817E

Plus) techniques were employed to characterize the synthesized MIL-100(Fe). The material's chemical structure and functional groups essential for adsorption were analyzed via FT-IR, directly linking functional groups to the interaction with methylene blue molecules. This was performed using the KBr pellet method, where MIL-100(Fe) powder was mixed with spectroscopic-grade KBr in a 1:10 mass ratio (sample: KBr). The mixture was finely ground and pressed into a transparent pellet for analysis. X-ray powder diffraction (XRD) was used to verify MIL-100(Fe)'s crystalline structure, crystallinity, average crystallite size, and any iron oxide impurities.

The crystallinity of the MIL-100(Fe) samples was assessed using two complementary methods based on XRD data. First, the relative crystallinity was estimated by integrating the area under the characteristic diffraction peaks in the XRD pattern. This approach involves calculating the ratio of the total integrated peak area to the entire area under the diffraction curve, providing a quantitative measure of crystallinity that accounts for both peak sharpness and overall intensity. This method captures the overall crystalline order and framework regularity of the samples, aligning with the surface area and porosity data obtained from BET analysis, as larger surface areas often correspond to more disordered structures. Additionally, the average crystallite size (D) was determined using the Debye-Scherrer equation, which relates the size of coherent scattering domains to the full width at half maximum (FWHM) of the dominant diffraction peak, as shown below:

$$D = \frac{K\lambda}{\beta \cos\theta} \quad (1)$$

Where D is the crystallite size (nm), K is the shape factor (as usual 0.9), λ is the X-ray wavelength (0.15406 nm for Cu Kα radiation), β is the FWHM of the selected peak (in radians), θ is the Bragg angle (in degrees). This method captures the average size of coherent scattering domains, where broader peaks correspond to smaller crystallite sizes. Together, these approaches provide a comprehensive assessment of

View Article Online
DOI: 10.1039/D5NR04817F

both the crystallinity and crystallite size, aligning with the guidance to incorporate both intensity-based and peak width-based analyses for a more accurate evaluation of the structural properties of MIL-100(Fe). Surface area and porosity measurements were done through BET and BJH methods after degassing the samples at 150˚C for 12 hours under vacuum. TGA was performed with sample heating from room temperature to 600˚C at 10˚C/min in an air environment. Morphological characteristics, such as particle size and shape, were examined using SEM. To ensure accurate characterization, all samples were pre-treated at 60˚C for 24 hours under vacuum.

2.3. Adsorption experiments

The synthesized MIL-100(Fe)'s efficacy in removing MB from water was evaluated under controlled conditions. In a typical adsorption experiment, 10 mg of MIL-100(Fe) was added to 10 mL of dye solution of 100 ppm at pH 9, shaken at 200 rpm for 180 minutes at room temperature. After this period, the material was separated via centrifugation at 10,000 rpm for 10 minutes. Dye removal efficiency was quantified as removal percentage (R%) calculated as follows:

$$R[\%] = \frac{C_0 - C_t}{C_0} \times 100 \qquad (2)$$

where $\boldsymbol{C_0}$ is the initial MB concentration in the adsorbent-free solution before adding MIL-100(Fe), and $\boldsymbol{C_t}$ is the residual MB concentration in the solution phase after contact time $\boldsymbol{t}$, measured after separating the MOF particles from the suspension. MB concentrations were determined using a UV-Vis spectrometer (Nanospec2 nanolytik) through a calibration curve (**Figure S1**). In addition to removal percentage, the adsorption capacity q (mg g^{-1}) was calculated using

$$q = \frac{(C_0 - C_e)V}{m} \times 100 \qquad (3)$$

View Article Online
DOI: 10.1039/D5NR04817F

where C_0 and C_e are the initial and equilibrium methylene blue concentrations (mg L^{-1}), V is the solution volume (L), and mmm is the adsorbent mass (g). This metric was used to provide a quantitative comparison of dye uptake across the synthesized MIL-100(Fe) samples under the fixed benchmark adsorption conditions employed in this study. The adsorption experiments reported in the main study were conducted at **pH 9**, which was selected based on preliminary pH-screening experiments performed over multiple samples and contact times. As shown in the Supporting Information (**Figure S27,**), MB removal consistently increased from **pH 5** to **pH 9** across the sample set, and **pH 9** gave the highest removal under the screening conditions. Therefore, pH 9 was chosen as the benchmark condition for the comparative adsorption tests. This choice is further supported by the measured point of zero charge of the optimized MB-removal sample,$pH_{PZC} = 3.36$ (**Figure S26**), indicating that the adsorbent surface is negatively charged under the final adsorption condition. All adsorption experiments were performed using an incubator shaker under fixed temperature and agitation settings to minimize variability arising from ambient conditions and differences in shaking intensity.

2.4. Machine Learning Strategies for Limited Dataset Prediction

In working with a modest-sized experimental dataset (11 synthesis trials), challenges were faced in extracting reliable insights without overfitting. Limited data often leads to high sensitivity to noise and insufficient representation of variable interactions, which can compromise model generalization [39]; therefore, the resulting models should be interpreted as low-data, system-specific surrogates rather than broadly generalizable predictors. To address this, we applied a ten-fold cross-validation strategy, where the data was divided into ten subsets. We used this technique during model training to maximize the utility of each data point. Each subset was used once as the validation set, while the remaining data was used for training [40, 41]. This approach provided a reliable performance assessment, reducing overfitting and

View Article Online
DOI: 10.1039/D5NR04817E

strengthening the model's ability to generalize[25, 39-41]. Additionally, hyperparameter tuning was conducted systematically for each model using grid search technique. These practices ensured the reliability and generalizability of the predictive models despite data limitations.

A range of ML models was applied, including decision trees, XGBoost, random forest, support vector regressor (SVR), K-nearest neighbors (KNN), and GPR, with Dummy regression, and Linear regression. The Dummy Regressor was used as a constant-property baseline, while Linear Regression was retained as a simple parametric baseline. To further address the limitations of evaluating models with only 11 samples, we additionally performed a leave-one-out cross-validation (LOOCV) audit for the two primary optimization targets: BET surface area and MB removal. In each LOOCV fold, one sample was held out as the test point, not as a validation point, while the remaining ten samples were used for model fitting. This process was repeated until each sample had served once as the held-out test point. No separate validation split was introduced within the LOOCV audit because the remaining training set contains only ten samples. A training-mean Dummy Regressor was evaluated under the same LOOCV protocol to provide a direct mean-property baseline. Model performance in this conservative audit was assessed using RMSE, MSE, and MAE. The detailed LOOCV predictions, parity plots, and baseline comparisons are provided in the Supporting Information, **Table S6** and **Figure S33**. Model performance in this conservative audit was assessed using RMSE, MSE, MAE, and R^2. R^2 is reported for completeness because it is commonly used in regression analysis; however, in very small held-out datasets it can be sensitive to individual samples and to the variance of the target values [42]. Therefore, RMSE, MSE, and MAE are emphasized in the discussion because they provide more direct and physically interpretable measures of prediction error [42]. To further improve reproducibility and clarify the GPR prior/model specification, the final GPR settings used for the two primary targets are reported in the Supporting Information (**Table S7**). These include the

View Article Online
DOI: 10.1039/D5NR04817F

kernel form, length-scale/periodicity parameters, regularization/noise parameter α, target normalization setting, optimizer setting, and random state. This information is provided to make the model assumptions explicit and to distinguish the initial cross-validated model-screening stage from the subsequent LOOCV audit and GA-guided candidate generation.

This diversity allowed models capable of capturing complex relationships between MIL-100(Fe) synthesis conditions and material structural/adsorption performance properties to be identified. To complement predictive modeling, t-distributed stochastic neighbor embedding (t-SNE) was employed to visualize high-dimensional property-parameter relationships along the principal component analysis (PCA), enabling cluster detection and latent trend analysis within the limited dataset. Notably, GPR excelled due to its non-linear capabilities and flexibility, making it particularly effective in handling the intricate, non-linear interactions typical in material science. The probabilistic nature of GPR also helped account for experimental noise, yielding smoother and more interpretable predictions [41]. In comparison, random forest and XGBoost, while effective in large datasets, were found to be less suited for the smaller dataset. Therefore, GPR emerged as the preferred model for further analysis, due to its balance of accuracy and interpretability[41]. To further enhance model performance, hyperparameters were optimized through Grid Search Cross-Validation, with settings like kernel type in GPR and tree count in random forest adjusted to minimize prediction error. Finally, GPR was selected as the primary surrogate because it is suitable for low-data regression, can capture nonlinear response trends, and provides a probabilistic framework for modeling experimental uncertainty. However, in view of the small dataset size, the GPR model is used here primarily as a candidate-generation surrogate within the experimentally explored MIL-100(Fe) synthesis window. The final assessment of the optimized candidates is therefore based on prospective experimental synthesis and characterization, rather than on the surrogate model alone.

View Article Online
DOI: 10.1039/D5NR04817F

For tree-based ensemble models, the number of trees was treated primarily as a stabilization parameter rather than as the main source of model complexity. Increasing the number of trees generally reduces ensemble variance and improves prediction stability until performance reaches a plateau, after which additional trees mainly increase computational cost. Given the small dataset size, overfitting was controlled more directly by limiting tree complexity through hyperparameters such as maximum tree depth, minimum samples per split, and minimum samples per leaf. Thus, ensemble hyperparameter tuning was performed to balance cross-validated predictive accuracy, model stability, and parsimony rather than to simply maximize the number of trees. Combined with ten-fold cross-validation [39, 41], this approach consistently delivered low root mean square error (RMSE), underscoring the models' robustness in predicting material performance.

2.5. Optimizations Via Genetic Algorithm

To optimize the synthesis conditions of MIL-100(Fe) for maximizing the specific surface area and enhancing MB removal efficiency independently, a GA was applied. Given the limited dataset of 11 samples, GA was particularly valuable in efficiently navigating the design space while complementing the predictive strength of the ML models. This nature-inspired technique is highly effective for solving complex, nonlinear problems and offers a robust approach to finding global optima in parameter-rich spaces [43]. GA operates on principles of natural selection, with a population of solutions evolving over generations. In each generation, solutions are evaluated by a fitness function, and the top-performing individuals are selected to form the next generation [44]. Through crossover, in which genetic information between solutions is combined, and mutation, in which slight variations are introduced, GA explores a broad solution space while gradually refining toward optimal synthesis conditions [45]. In this study, GA's dynamic search capabilities were used to optimize surface area and MB removal independently, accounting for their distinct

View Article Online
DOI: 10.1039/D5NR04817F

roles in performance. Unlike traditional methods that risk becoming trapped in local optima, the broader search coverage provided by GA was crucial in this non-linear system for exploring diverse synthesis parameters [43].

The fitness function for GA evaluated two main goals: maximizing surface area and enhancing dye adsorption efficiency. These goals often involve trade-offs, as higher surface area does not automatically yield improved adsorption due to factors like pore volume and crystallinity. The independent optimization of these targets by GA offered a nuanced balance between the competing demands of surface area and adsorption percentage. The crossover mechanism of GA played a key role by allowing synthesis parameters from different solutions to combine in novel ways, promoting innovative approaches not initially considered [43, 44]. Although crossover occasionally disrupted beneficial structures, it also allowed the algorithm to adapt and refine solutions over time, yielding promising synthesis conditions [44].

The equation used in GA for calculating the dynamic parameter K is given below [44]:

$$K = \frac{G + 2\sqrt{g}}{3G} \quad (3)$$

Where G is the total number of generations, and g is the number of generations completed. This adaptability prevented premature convergence, ensuring a balanced approach between exploration and exploitation [43, 44]. GA was crucial in efficiently exploring the synthesis conditions for MIL-100(Fe), achieving a balance between high surface area and optimal adsorption. By circumventing the limitations of traditional one-variable-at-a-time optimization, the global search capabilities of GA allowed for a comprehensive, efficient discovery of conditions that significantly improved both surface area and dye adsorption. This GA approach demonstrates its potential for optimizing other material systems where complex parameter interactions demand sophisticated strategies [43, 44, 46]. In this study, we conducted two separate optimizations: one targeting BET surface area ("OPT BET") and another targeting MB removal

View Article Online
DOI: 10.1039/D5NR04817F

("OPT MB"). After optimization, OPT BET and OPT MB were characterized by FTIR, XRD, N_2 adsorption (BET/BJH), TGA, and SEM; full acquisition parameters and datasets are provided in the SI (**Figures S19-S24**). The predictive models were trained and evaluated using the original 11 experimental synthesis runs, with ten-fold cross-validation used during model development. In this workflow, the GA did not optimize the physical experiment directly. Instead, it optimized the trained GPR surrogate model, using the GPR-predicted property value as the fitness function. The GA searched within the experimentally defined synthesis-variable bounds and proposed candidate synthesis conditions predicted to maximize the selected target property. These proposed optima were then synthesized experimentally as validation samples. The two optimized samples (OPT BET and OPT MB) were generated only after model construction, as prospective experimental validations of the GPR-guided GA framework, and were not included in model training.

Constraint Handling in Prediction Outputs:

During GA optimization, the synthesis variables (reaction time, temperature, metal-to-ligand molar ratio, and ion concentration) were strictly constrained within the experimentally defined design ranges. No explicit upper clipping was initially imposed on the surrogate-predicted MB removal response, in order to avoid introducing an artificial truncation into the regression surface near the upper physical limit. As a result, the GPR model could occasionally return saturation-level predictions slightly above 100%, which are not physically interpretable as actual removal efficiencies. These values were therefore treated only as indicators that the solution had reached the upper-performance regime, and the reported prediction was capped at 100% for physical interpretability. Importantly, this did not expand the GA search domain beyond the experimentally defined synthesis space. The experimentally validated OPT MB sample showed 98.3% removal, confirming that the selected optimum remained within a physically realistic regime.

View Article Online
DOI: 10.1039/D5NR04817F

3. Results and discussion

3.1 Characterization Results and Adsorption Evaluation

Table 2 summarizes the synthesis parameters for 11 MIL-100(Fe) samples, designed using a Taguchi-based DoE framework with two additional random points intentionally added to enhance variability. This deliberately modest dataset was chosen to demonstrate that reliable modeling and optimization, via machine learning and genetic algorithms, can be achieved even with a limited number of experiments. The synthesis parameters, including reaction time, temperature, metal-to-ligand molar ratio, and metal ion concentration, were systematically varied to explore their influence on key material properties such as BET surface area, crystallinity, total pore volume, average crystallite size, yield and MB adsorption efficiency. To visualize the coverage of the experimental design space, a lower-triangle pairwise scatter-matrix of the four synthesis variables (time, temperature, metal-to-ligand molar ratio, and metal concentration) is provided in the Supporting Information (**Figure S32**).

Table 2. Synthesis Conditions for MIL-100(Fe) Samples.

Sample No.	Time [h]	Temp. [˚C]	Metal-to-Ligand Molar Ratio [Fe: Ligand]	Metal Conc. [M]
1	20	110	1:1	0.15
2	40	110	2.25:1	0.3
3	60	110	3.5:1	0.45
4	20	130	2.25:1	0.45
5	40	130	3.5:1	0.15
6	60	130	1:1	0.3

7	20	150	3.5:1	0.3
8	40	150	1:1	0.45
9	60	150	2.25:1	0.15
10	40	130	1:1	0.15
11	40	110	3.5:1	0.3

View Article Online
DOI: 10.1039/D5NR04817F

*Note: Samples S10 and S11 are supplementary non-orthogonal runs added outside the core Taguchi array to extend coverage of the synthesis-variable space.

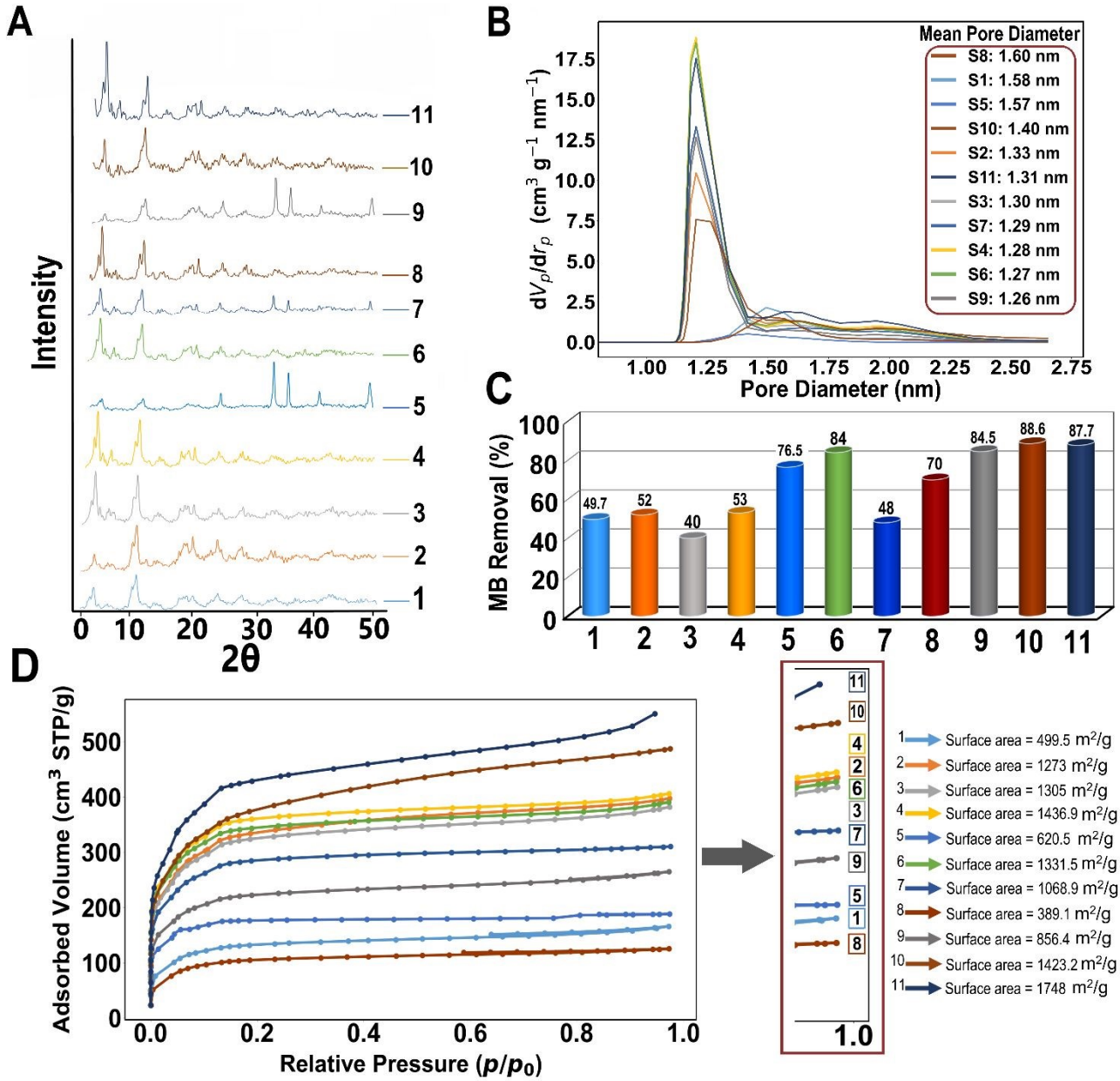


Figure 3. Characterization and adsorption performance of MIL-100(Fe) samples. (A) XRD patterns of eleven MIL-100(Fe) samples synthesized under varied conditions, showing phase differences and minor crystalline iron-oxide impurity peaks tentatively assigned to α-Fe_2O_3 in samples S5, S7, and S9. (B) Micropore size distribution curves, derived from N_2 adsorption data, indicating pore sizes between 1.2-1.8%nm and sample-dependent shifts in peak location and intensity. (C) MB removal efficiencies for the samples, highlighting the strong correlation between synthesis conditions and adsorption performance. (D) N_2 adsorption isotherms and corresponding BET surface areas, confirming differences in microporosity and structural quality across samples. Higher surface areas are associated with stronger MB removal performance.

3.1.1. Characterization Results

The structural and compositional properties of the 11 synthesized MIL-100(Fe) samples were characterized using XRD for crystallinity, FTIR for chemical bonding, TGA for thermal stability, and N_2 adsorption-desorption isotherms for BET surface area and pore structure analysis. These techniques provide a comprehensive profile of each sample's structural integrity, porosity, and functional group distribution, forming the foundation for subsequent performance assessments.

Figure 3A displays the XRD patterns of MIL-100(Fe) synthesized under varied conditions. The prominent diffraction peaks at 3.5˚, 4˚, 10.4˚, and 11.2˚ confirm the formation of the MIL-100(Fe) crystalline phase. However, additional weak reflections at 24.3˚, 33.2˚, 35.8˚, 41.4˚, and 49.5˚ observed in samples S5, S7, and S9 indicating the presence of a minor crystalline iron-oxide impurity phase (Based on peak-position agreement, this impurity is tentatively assigned to α-Fe_2O_3, as summarized in **Table S4 (SI)**) and suggesting the emergence of iron oxide impurities, likely a result of incomplete coordination or missing linker defects at high metal-to-ligand molar ratios%[47]. Based on the position of these reflections, the impurity is more consistent with α-Fe_2O_3 (hematite) than γ-Fe_2O_3. This assignment is supported by the close match between the observed peaks and the characteristic hematite pattern, whereas the more

View Article Online
DOI: 10.1039/D5NR04817F

diagnostic reflections of γ-Fe_2O_3 are not evident in the present data. The appearance of these oxide peaks likely reflects incomplete coordination and/or partial framework degradation under harsher synthesis conditions. Because the impurity produces distinct diffraction peaks, it is unlikely to be purely amorphous; however, XRD alone does not resolve whether the oxide exists as discrete particles, intergrown domains, or partial surface deposits on MIL-100(Fe) [48, 49]. Furthermore, prolonged synthesis at elevated temperatures might have promoted partial degradation of the MIL-100(Fe) framework, facilitating partial iron oxide formation%[50]. Crystallite size and crystallinity values, calculated via the Scherrer equation (Eq. 1), are reported in **Table 3**, indicating a broad distribution attributable to synthesis variability.

Figure 3B presents the micropore size distribution curves for the 11 MIL-100(Fe) samples derived from N_2 adsorption data. All samples exhibit a dominant micropore network with peak sizes between 1.2 and 1.8%nm, consistent with expected pore dimensions of MIL-100(Fe). These values align with literature, where MIL-100(Fe) has been shown to feature micropore peaks around 0.9, 1.4, and 1.7%nm, supporting successful formation of the intended framework [51]. This narrow distribution indicates a highly microporous structure with minimal mesoporosity, consistent with the high surface areas measured by BET (ranging from 389.1 m²/g to 1748 m²/g). The variability in micropore volume reflects the influence of synthesis conditions, including temperature, time, and precursor ratios, on the formation and stability of the microporous frameworks. The total pore volumes are listed in **Table 3**, highlighting the structural heterogeneity across samples.

Figure 3D shows the N_2 adsorption-desorption isotherms for the 11 MIL-100(Fe) samples exhibit characteristic Type I/IV hybrid behavior, marked by a sharp micropore-filling step below p/p_0 ≈%0.1 and a gradual uptake at intermediate pressures attributed to microporosity and interparticle voids [52]. Sample S11 (1748%m²/g) demonstrated the highest N_2 uptake throughout the p/p_0 range, closely followed by S4

View Article Online
DOI: 10.1039/D5NR04817F

(1436.9%m²/g) and S10 (1423.3%m²/g), consistent with their larger micropore volumes. In contrast, samples like S1 (499.5%m²/g) and S8 (389.1%m²/g) showed limited porosity and uptake, likely due to less developed pore networks stemming from suboptimal crystallization or precursor imbalances [52]. Detailed adsorption-desorption isotherms for each of the 11 samples are provided in **Figure%S6** in the supplementary information, offering a complete comparison of textural properties across all synthesized materials.

Fourier-transform infrared (FT-IR) spectra (**Figure S3, SI**) confirmed characteristic MIL-100(Fe) functional groups, including OH, C=O, C-O, C=C, and Fe-O vibrations. While these validate the framework composition, their relative intensities showed no strong correlation with MB removal, indicating limited predictive power for adsorption capacity. TGA (**Figure S4, SI**) demonstrated thermal stability up to about 330%˚C, with water desorption below 110%˚C, major linker decomposition between 300-400%˚C, and a residual plateau beyond 400%˚C corresponding to iron oxide formation. SEM imaging (**Figure S5, SI**) revealed octahedral microstructure, with sample S9 showing pronounced particle agglomeration (×20,000 magnification) likely caused by Van der Waals interactions

3.1.2. Methylene Blue Adsorption Performance Evaluation

Figure 3C illustrates the MB removal efficiencies of the synthesized MIL-100(Fe) samples, revealing substantial variation in adsorption performance. Samples with moderate surface areas but optimized pore architectures such as samples S9 and S10 achieved dye removal levels comparable to S11, which had the highest BET surface area. Conversely, Samples S4 and S10, despite similar surface areas, showed distinct adsorption behaviors, confirming that surface area alone does not dictate performance. Structural factors such as crystallinity, pore connectivity, and impurities are key contributors. Notably, the impurity-containing samples do not behave uniformly: while S5 and S9 show relatively strong MB removal, S7 does not. This indicates that the presence of iron-oxide impurity alone is not sufficient to explain adsorption

View Article Online
DOI: 10.1039/D5NR04817F

performance. Rather, limited oxide formation may contribute in combination with other factors such as defect generation, altered crystallinity, and modified pore accessibility. In this sense, the oxide phase may influence adsorption both directly, by introducing additional iron-oxide/hydroxyl surface sites, and indirectly, by perturbing the formation of the primary MIL-100(Fe) framework and thereby affecting the accessibility of adsorption-relevant regions. Overall **Table 3** summarizes key material properties for MIL-100(Fe) samples including BET surface area, crystallinity, pore volume, crystallite size, yield and MB removal.

Table 3. Summary of Key Structural and Performance Metrics for MIL-100(Fe) Samples. Summary of BET surface area, crystallinity, total pore volume, average crystallite size (D_{avg}), yield, and MB removal efficiency for the 11 MIL-100(Fe) samples synthesized under varying hydrothermal conditions. These metrics provide a comprehensive overview of the textural and adsorption properties, highlighting the influence of synthesis parameters on material performance.

Sample No.	BET [m^2/gr]	Crystallinity [%]	Total pore volume [cm^3/gr]	Avg. crystallite size [D_{avg}, nm]	Yield [%]	Methylene blue removal [%]
1	499.4	20.9	0.257	17.5	75.4	49.7
2	1273	12.9	0.615	33.1	96.8	52
3	1305	29.0	0.591	22.7	81.4	40
4	1436.9	28.5	0.628	21.8	72	53
5	620.5	10.5	0.691	34.1	69.4	76.5
6	1331.5	23.6	0.604	27.2	82.8	84
7	1068.9	21.3	0.479	20.6	73.6	48
8	389.1	24.1	0.194	33.4	64	70
9	856.4	7.6	0.401	31.6	72.3	84.5
10	1423.3	19.0	0.753	27.5	93.2	88.6
11	1748	36.2	0.827	24.7	98.6	87.7

View Article Online
DOI: 10.1039/D5NR04817F

In addition to removal percentage, the corresponding adsorption capacities were calculated and are now reported in **Figure S25 Supplementary information**, providing a direct mass-normalized comparison of dye uptake across the sample set. Under the fixed adsorption conditions used here, the corresponding adsorption capacities span 40.0–88.6 mg g^{-1} across the 11 synthesized samples, while the optimized MB sample reaches 98.3 mg g^{-1}. Because the present work focuses on synthesis–property optimization under a standardized adsorption-screening protocol, detailed dye-adsorption isotherm and kinetic-model fitting were not included here and will require a dedicated mechanistic study. To better rationalize these trends, the likely adsorption mechanism is discussed separately in Section 3.1.3.

3.1.3. Adsorption Mechanism Analysis

The adsorption of methylene blue (MB) by MIL-100(Fe) is most reasonably interpreted as a synergistic process involving electrostatic attraction, π–π interactions, hydrogen bonding/polar interactions, and pore-accessibility effects, rather than being governed by surface area alone. First, MB is a cationic aromatic dye, and the point of zero charge determined for the optimized MB-removal sample was $pH_{PZC} = 3.36$ (**Figure S26, SI**). Because the adsorption experiments were conducted at pH 9, the MIL-100(Fe) surface is expected to be negatively charged under the test conditions, favoring electrostatic attraction with positively charged MB molecules. This interpretation is consistent with prior reports on MOF-based dye adsorption, where solution pH relative to pH_{PZC}strongly influences the uptake of ionic dyes [12, 53-55].

Second, π–π interactions are also likely to contribute. MB contains conjugated aromatic rings, while MIL-100(Fe) is constructed from benzene-1,3,5-tricarboxylate linkers, providing aromatic domains that

View Article Online
DOI: 10.1039/D5NR04817F

can promote dye–framework stacking interactions. In addition, FTIR comparison of the optimized samples (**Figures S20, SI**) suggests that the MB-optimized material retains accessible oxygen-containing surface functionalities, including hydroxyl- and carboxylate-containing groups, which may further assist adsorption through hydrogen bonding or other polar interactions with MB [54-56].

Third, the superior performance of the MB-optimized sample appears to be linked to structural accessibility rather than maximum BET surface area. The SI characterization shows that the OPT MB sample exhibits a pore-size distribution and microstructure consistent with improved accessibility, including open mesopores, and broadened pore features relative to the BET-optimized material. Likewise, its lower crystallinity and broader XRD features suggest a less perfectly ordered framework, which may expose additional adsorption-relevant sites and shorten diffusion pathways. Taken together, these observations explain why the MB-optimized sample achieves higher dye removal despite a lower BET surface area than OPT BET [53, 56]. This interpretation is further supported by preliminary pH-dependent adsorption screening (**Figure S27, SI**), which showed stronger MB uptake under alkaline conditions, and by the pH_{PZC} value of 3.36 for the optimized MB-removal sample, indicating a negatively charged surface at pH 9 and therefore favorable electrostatic attraction toward cationic MB.

Therefore, the present results support a mechanism in which MB removal by MIL-100(Fe) arises from the combined action of surface charge effects, aromatic dye–linker interactions, accessible polar surface groups, and favorable pore/diffusion characteristics. This interpretation is consistent with the broader MOF-dye adsorption literature and reinforces the central conclusion of this work: maximizing textural surface area alone is insufficient for application-specific adsorption optimization [54, 55].

View Article Online
DOI: 10.1039/D5NR04817F

3.1.4. Exploratory Correlation Analysis of Synthesis Variables and Material Responses

To provide a descriptive first-pass view of how synthesis variables relate to the measured material responses, we examined Pearson correlation coefficients (**Figure 4**) and Spearman rank correlations (**Figure S2**) [57]. Because the present dataset contains only 11 samples, these correlations are interpreted exploratorily rather than inferentially and are used only to highlight broad trends that are later examined more rigorously through machine-learning modeling. In addition, the synthesis variables were assigned by the experimental design rather than sampled from a natural population; therefore, correlations among the input variables themselves are not interpreted mechanistically. The analysis instead focuses on the relationships between synthesis conditions and output properties. Within this scope, synthesis time shows a moderate positive association with MB removal and average crystallite size, temperature shows a negative association with yield and crystallinity, and ion concentration shows a positive association with crystallinity. These trends are consistent with the broader behavior captured later by the predictive models, but should be regarded only as qualitative guidance given the limited sample size.

View Article Online
DOI: 10.1039/D5NR04817F

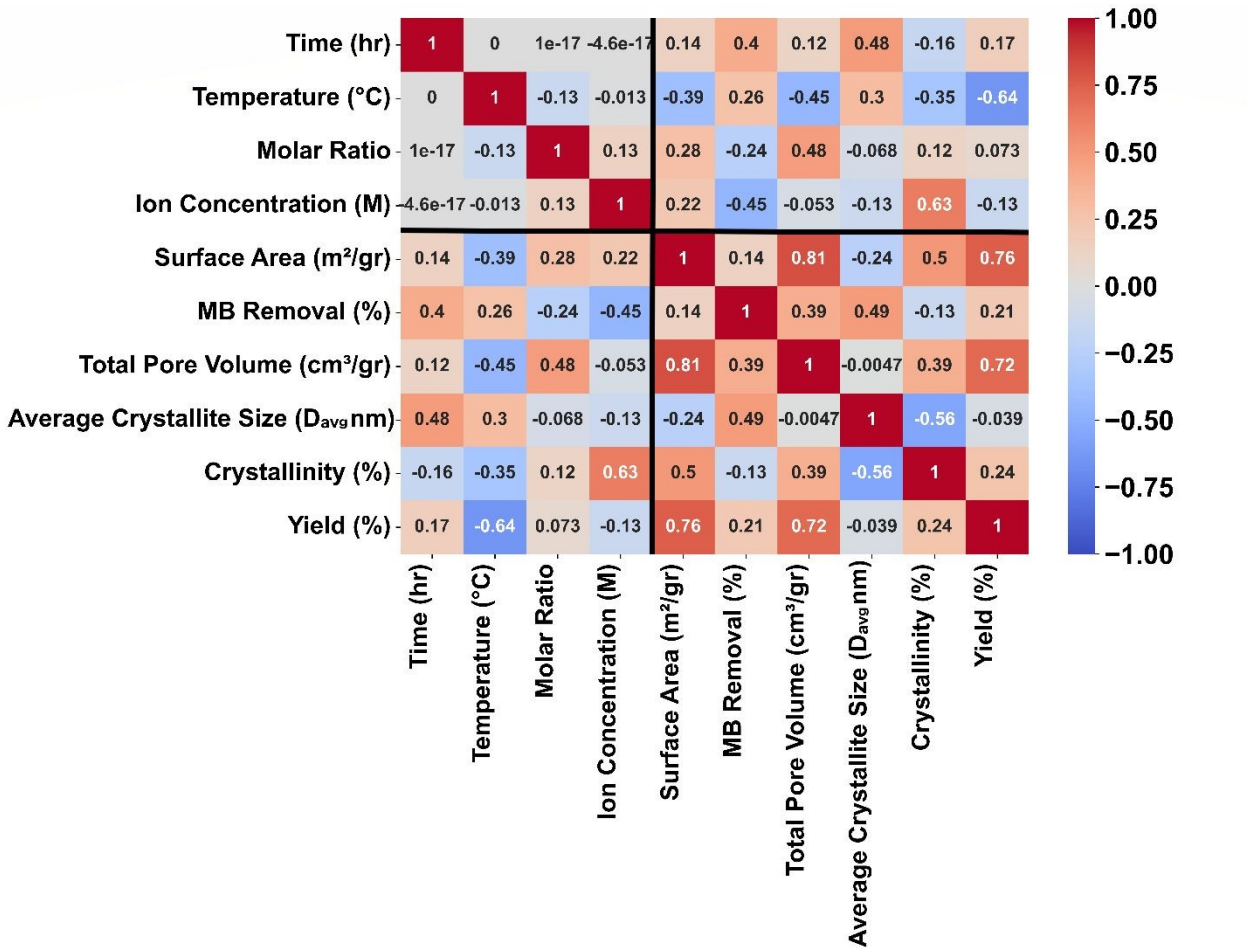


Figure 4. Pearson correlation coefficient (PCC) matrix for the 11 synthesized MIL-100(Fe) samples. The divider lines separate the synthesis variables (time, temperature, molar ratio, ion concentration) from the measured material responses (surface area, MB removal, total pore volume, average crystallite size, crystallinity, and yield). Because the synthesis variables were assigned by design and the dataset is small, the matrix is used here only as an exploratory trend visualization, with interpretation focused on the input–output relationships rather than the correlations among input variables.

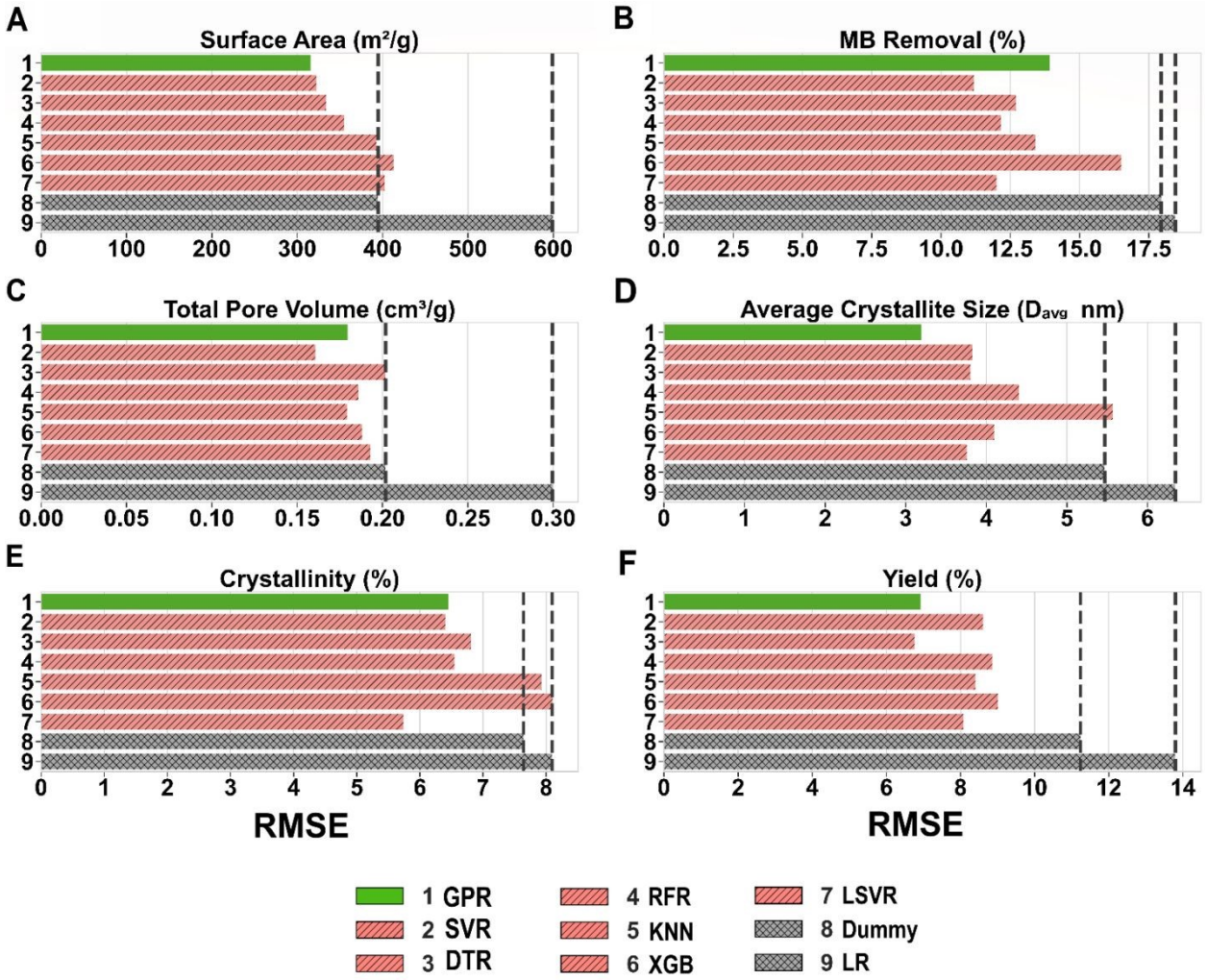
3.2. Machine Learning and Optimization Results

3.2.1. Performance Evaluation of Machine Learning Models

In order to capture chemical intuition into the interplay of parameters affecting the structural and adsorption properties of MIL-100(Fe), we applied a suite of machine learning (ML) models

View Article Online
DOI: 10.1039/D5NR04817F

including SVR, XGBoost, random forest, KNN, decision tree, lightGBM, linear SVR, and GPR, alongside dummy and linear regression baselines. These models predicted six key properties: BET surface area, MB removal efficiency, total pore volume, average crystallite size, crystallinity, and yield. Given the limited dataset (11 samples), we selected models capable of capturing non-linear dependencies and applied K-fold cross-validation to reduce overfitting. RMSE was used as the primary metric to quantify absolute prediction error, as R^2 can be misleading with small datasets. The dummy model (randomized predictions) served as a baseline, while linear regression was used for comparative ANOVA-based evaluations.



View Article Online
DOI: 10.1039/D5NR04817F

Figure 5. Initial cross-validated RMSE comparison of candidate machine-learning models for predicting key MIL-100(Fe) properties: (A) BET surface area, (B) MB removal, (C) total pore volume, (D) average crystallite size, (E) crystallinity, and (F) yield. Dummy Regression and Linear Regression were included as baseline models. A more conservative LOOCV audit for the two primary targets, BET surface area and MB removal, is provided in the Supporting Information, **Table S6** and **Figure S33**.

As it can be seen **Figure 5** illustrates the performance of nine ML models in predicting six key properties of MIL-100(Fe), evaluated using RMSE values (detailed in SI, **Table S1**). GPR consistently outperformed all other models across every target, yielding the lowest RMSE values of 315%m²/g for surface area, 11.2% for MB removal, 0.16%cm³/g for pore volume, 3.2%nm for crystallite size, 5.7% for crystallinity, and 6.76% for yield (**Table S1**). These results affirm GPR's strong capacity to capture complex, nonlinear trends within small datasets, and justify its selection as the primary model for further interpretability and optimization. SVR ranked second overall, showing competitive RMSEs for pore volume and yield (e.g., 322%m²/g surface area, 12.0% MB removal), but lacked GPR's consistency across all targets and offered limited utility for interpretability. In contrast, models such as XGBoost (RMSE: 412%m²/g surface area, 17.9% MB removal), linear regression (RMSE: 598%m²/g surface area, 18.5% MB removal), and even the dummy regressor often exhibited higher errors, indicating poor generalizability. Notably, the RMSE of linear regression surpassed that of even the dummy regressor in multiple cases. These trends emphasize GPR's robustness as a modeling choice and its vital role in guiding subsequent analysis. The results of initial cross-validated model screening justified its use as the principal surrogate model for subsequent SHAP interpretation and GA-guided candidate generation. To provide a stricter assessment for the two primary optimization targets, we further performed a

View Article Online
DOI: 10.1039/D5NR04817F

LOOCV audit against the training-mean Dummy Regressor and Linear Regression baselines. In this audit, each held-out sample was treated as a test point, while the remaining ten samples were used for model fitting. For BET surface area, GPR improved over the training-mean Dummy Regressor, reducing RMSE from 459.28 to 403.93 $m^2 g^{-1}$ and MAE from 399.67 to 327.94 $m^2 g^{-1}$. For MB removal, the LOOCV audit showed stronger target sensitivity: GPR reduced MAE from 18.19 to 17.02%, but did not reduce RMSE relative to the Dummy Regressor. Accordingly, the MB-removal model is interpreted cautiously as a candidate-generation surrogate, while the final performance claim is based on experimental validation of the OPT MB sample. The full LOOCV results and parity/baseline plots are provided in **Table S6** and **Figure S33**. The final GPR parameterization used for these targets (two primary optimization targets) is provided in **Table S6** to make the kernel assumptions and model settings explicit. In the revised manuscript, R^2 is reported together with RMSE, MSE, and MAE; however, the discussion emphasizes the error-based metrics because they more directly quantify prediction deviations in the present small-data setting.

Unlike tree-based models, which may overfit small datasets and yield discontinuous outputs, GPR was chosen as the core model due to its superior predictive accuracy, robustness to experimental noise, and ability to generate smooth, interpretable functions that align with experimental trends%%[38, 41]. Its probabilistic nature makes it ideal for MOF synthesis, where parameter uncertainty is common. Leveraging these strengths, GPR was integrated with SHAP for transparent feature-level insights and coupled with GA for multi-objective optimization. This GPR-based workflow effectively captured non-linear, high-dimensional interactions and

View Article Online
DOI: 10.1039/D5NR04817F

uncovered subtle dependencies often missed by traditional linear optimization methods [41], laying the foundation for rational, performance-driven design of MIL-100(Fe) and similar MOFs. Despite the encouraging predictive performance obtained under ten-fold cross-validation, the present model should be interpreted as a low-data, system-specific surrogate for MIL-100(Fe) within the experimentally sampled synthesis window, rather than as a universally generalizable predictor. In small-data materials machine learning, limited sample size can increase the risk of overfitting and can restrict extrapolation beyond the domain represented in the training data. Accordingly, while cross-validation supports the robustness of the model within the explored MIL-100(Fe) design space, broader transferability to other MOF families, linker chemistries, metal nodes, or synthesis environments would require additional data and dedicated retraining. The present framework should therefore be viewed as a proof-of-concept for low-data, application-specific synthesis optimization, with future expansion expected to improve generalizability [41, 58-60].

3.2.2. 2D Contour Mapping: Unveiling Synthesis-Property Interactions

To further unravel the complex, multidimensional relationships between synthesis conditions and material properties of MIL-100(Fe), we employed two-dimensional contour plots derived from GPR models. These partial dependence plots (PDPs) offer visual insights into how pairs of synthesis parameters cooperatively influence critical performance metrics, while all other variables are held constant. Among the six possible parameter pairs (combinations of time, temperature, metal-to-ligand molar ratio, and ion concentration), the two most influential sets: (i) time and temperature and (ii) molar ratio and time, were selected for in-depth discussion here. Additional combinations are provided in the Supplementary Information (**Figures S7-S10**), offering a

View Article Online
DOI: 10.1039/D5NR04817F

comprehensive view of the broader synthesis-performance landscape. Figures 6–8 present experimental-data-based two-variable trend maps constructed from the measured dataset to provide an intuitive visualization of how selected synthesis-variable pairs relate to the observed material responses. Because the remaining synthesis variables are not fixed in these empirical plots, they should be interpreted as descriptive visual summaries of the experimental design space, rather than as isolated two-factor effects. For controlled model-based interpretation, the corresponding GPR-derived partial dependence plots are provided in **Figures S16–S18 (Supporting Information)**.

Time and Temperature:

Figure 6 illustrates the effects of synthesis time and temperature on six key properties of MIL-100(Fe): (A) BET surface area, (B) MB removal efficiency, (C) total pore volume, (D) average crystallite size, (E) crystallinity, and (F) yield. In **Figure 6A**, the surface area reaches a maximum in the moderate temperature range (120-135%˚C) and extended reaction times (40-50 hours). This behavior reflects a balance between sufficient thermal energy to promote nucleation and prolonged durations that support framework growth, while avoiding pore collapse or over-densification at higher temperatures. Similar densification trends have been observed in MOF-5 under aggressive thermal treatment [61].

In contrast, MB removal efficiency (**Figure 6B**) peaks at slightly higher temperatures (130-150%˚C) and intermediate times (30-40 hours), suggesting the optimal formation of defect-engineered structures and exposure of functional sites. This decoupling from surface area points to the importance of controlled defect introduction and local framework chemistry, as supported

View Article Online
DOI: 10.1039/D5NR04817F

by studies on Ni-MOF-74 and polymer-functionalized MOFs where adsorption performance was enhanced through improved accessibility and framework stability rather than surface area alone%[62, 63].

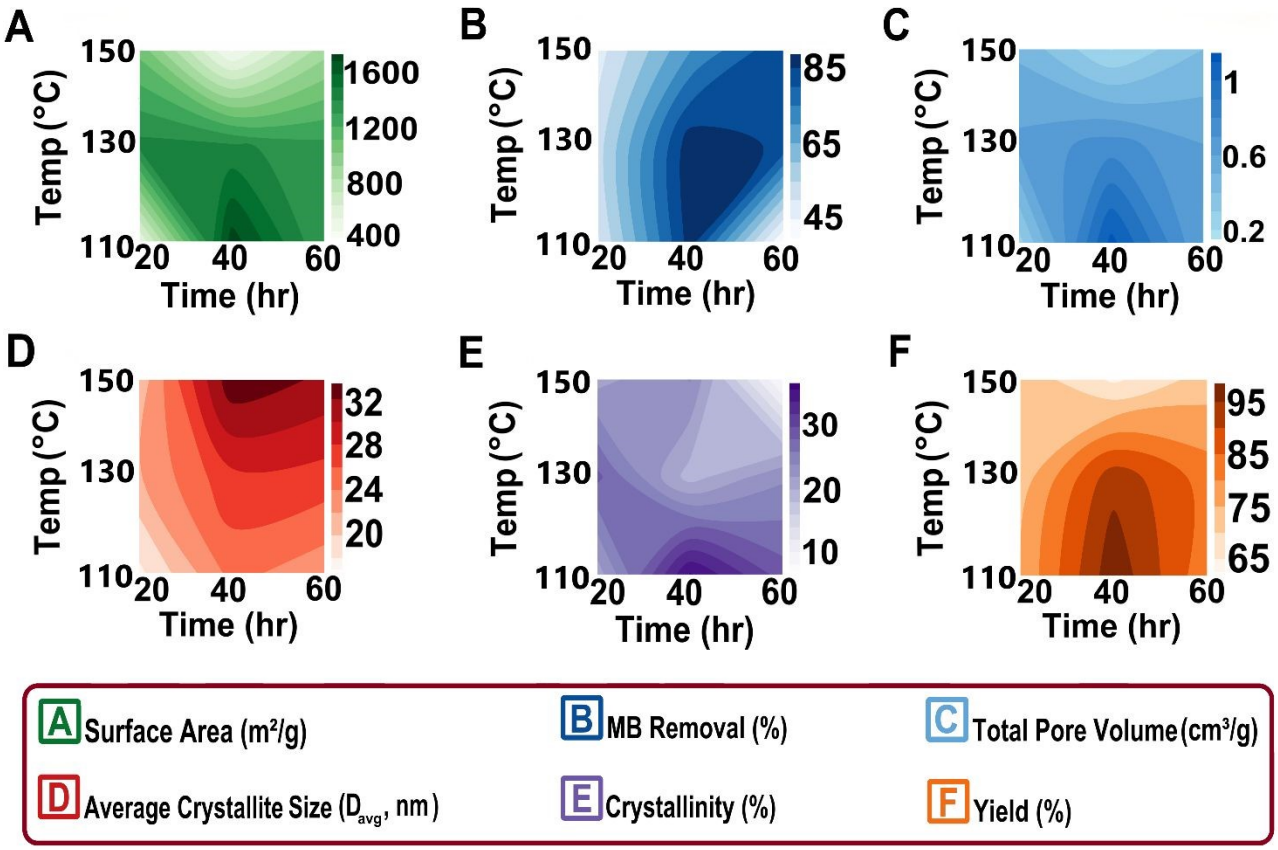


Figure 6. Contour plots depicting the relationship between synthesis parameters (Temperature and Time) and various performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots illustrate how different combinations of synthesis time and temperature influence each target property, highlighting the intricate trade-offs required to optimize material performance

Total pore volume trends (**Figure 6C**) closely follow those of surface area, as higher temperatures and longer durations improve metal-ligand coordination and promote expanded frameworks with enhanced porosity%[64]. Crystallite size and crystallinity (**Figures 6D** and **6E**) both increase with

View Article Online
DOI: 10.1039/D5NR04817F

temperature and time, reflecting accelerated coordination kinetics and thermodynamically driven crystal growth. These trends are consistent with previous findings in MOF-808 and UiO-66 systems, where synthesis under elevated thermal regimes facilitated larger, more crystalline particles and phase-pure materials%[65, 66].

Interestingly, the highest yields (**Figure 6F**) occur at moderate-to-high temperatures with shorter synthesis times (20-30 hours). This suggests that overly extended durations can lead to partial degradation, secondary phase formation, or inefficient resource utilization [65].

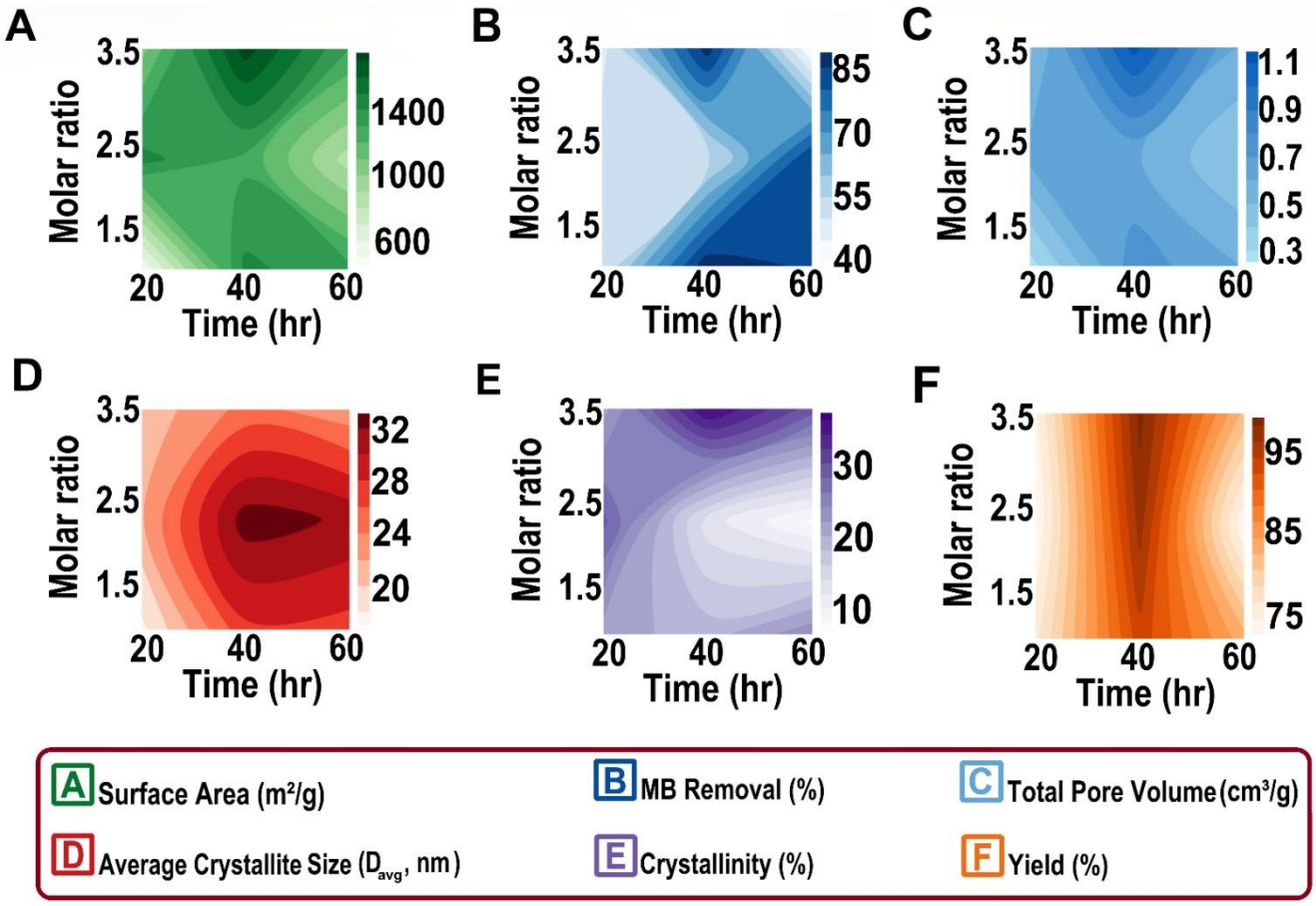


Figure 7. Contour plots depicting the relationship between synthesis parameters (Molar Ratio and Time) and various performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots illustrate how different combinations of

View Article Online
DOI: 10.1039/D5NR04817F

synthesis time and molar ratio influence each target property, highlighting the intricate trade-offs required to optimize material performance

Molar Ratio and Time:

Figure 7 presents contour maps showing the influence of metal-to-ligand molar ratio and synthesis time on the same six properties. As seen in **Figure 7A**, surface area is maximized at moderate molar ratios (about 2.0-2.5) and longer durations. These conditions likely promote balanced nucleation and gradual crystal growth. Previous work by Liu et al. and Aguiar et al. supports this finding, demonstrating that controlled precursor ratios and reaction times enhance early-stage nucleation and improve surface area in MIL-100-based systems [67, 68]. MB removal efficiency (**Figure 7B**) reaches its peak at higher molar ratios (about 3.0-3.5) and extended durations (about 50-60 hours), again suggesting the formation of accessible, defect-rich structures that promote dye diffusion and interaction. This trend aligns with the findings of Rabiee et al., who showed that quasi-2D MIL-100(Fe) structures with controlled defect density and morphology achieved significantly improved dye adsorption kinetics compared to more crystalline analogues, emphasizing the role of framework chemistry and electrostatic interactions over absolute surface area [69].

Figures 7C and **7D** show increases in total pore volume and crystallite size with higher molar ratios and prolonged synthesis, consistent with improved framework expansion and larger crystal domains. Crystallinity (**Figure 7E**) peaks at intermediate molar ratios (about 2.5-3.0) and long reaction times, reinforcing the idea that gradual, well-controlled growth favors highly ordered structures%[70]. Yield behavior (**Figure 7F**) mirrors the surface area trend, with highest values at

View Article Online
DOI: 10.1039/D5NR04817F

intermediate molar ratios (about 2.0-2.5) and moderate times (about 30-40 hours). Beyond these conditions, over-coordination and potential phase separation may reduce product yield, as previously observed in Zr-based MOFs synthesized with non-ideal stoichiometries or excessive reaction times%[71, 72].

3.2.3. 3D Surface Mapping: Capturing Multivariable Interactions

To further elucidate the complex interplay between synthesis parameters and material properties of MIL-100(Fe), three-dimensional (3D) surface plots were generated using the GPR model. These plots provide a holistic visualization of how two variables simultaneously affect key outcomes, offering richer insights than conventional 2D mappings. Among the four possible 3D combinations (choosing three parameters from four inputs), we selected molar ratio-ion concentration-property plots for detailed discussion here, based on their most substantial influence, while additional surfaces are provided in the Supplementary Information (**Figures S11-S15**).

3.2.3.1. Molar Ratio and Ion Concentration Effects

Figure 8 presents experimental-data-based two-variable trend maps showing how the measured responses vary across the molar-ratio / ion-concentration plane for six key properties of MIL-100(Fe): (A) BET surface area, (B) MB removal efficiency, (C) total pore volume, (D) average crystallite size, (E) crystallinity, and (F) yield. These plots are based on the experimental dataset rather than on retrained reduced-variable GP models. Accordingly, they are intended to provide a descriptive visualization of the observed response landscape in this two-variable projection, while the remaining synthesis variables continue to vary across the underlying experiments.

View Article Online
DOI: 10.1039/D5NR04817F

For surface area (**Figure 8A**), the highest values were achieved at elevated metal-to-ligand molar ratios (>2.5) combined with intermediate ion concentrations (0.25-0.35%M). This trend suggests that an excess of metal ions relative to ligand promotes enhanced metal-ligand coordination during framework assembly, facilitating the formation of highly porous structures. Similar behavior was reported by Schejn et al. for ZIF-8 systems, where optimized molar ratios yielded nanocrystals with improved BET surface areas%[73].

MB removal efficiency (**Figure 8B**) reached its maximum around a molar ratio of 2.5 and ion concentration of about 0.3%M, implicating the formation of defect-rich, functionalized frameworks facilitating MB uptake. This trend aligns with findings from Khan et al., who reviewed how dye adsorption in MOFs is dominated by electrostatic interactions, hydrogen bonding, π-π stacking, and pore architecture, not just textural surface area [12, 53]. Their analysis demonstrates that framework chemistry and defect engineering crucially influence adsorption performance

Total pore volume trends (**Figure 8C**) mirrored those of surface area, increasing with molar ratio and ion concentration up to an optimum, beyond which the benefits plateaued. This reflects the delicate balance required to avoid structural collapse or densification, a phenomenon observed in other porous frameworks when optimal precursor ratios are exceeded%[74, 75].

Crystallite size (**Figure 8D**) peaked at high ion concentrations paired with moderate molar ratios (about 2.0), indicating that crystal growth is highly sensitive to the balance between available metal ions and ligands. Excessive growth, while enhancing crystallinity, can restrict accessible surface area, highlighting the importance of maintaining framework openness. Similar dependencies have

View Article Online
DOI: 10.1039/D5NR04817F

been documented in ZIF-8 and Fe_3O_4 systems, where precursor ratios critically governed crystallite dimensions and stability%[76, 77].

Crystallinity (**Figure 8E**) was strongly favoured by high molar ratios, emphasizing the importance of abundant metal availability in producing ordered, defect-minimized structures. Ion concentration contributed more subtly but still enhanced framework stability. These findings align with reports on ZIF-8 and CeO_2 systems, where molar ratio tuning was key to achieving high phase purity and structural uniformity%[78, 79]. Yields were found to be maximized at intermediate molar ratios and ion concentrations, reflecting a balance between full coordination and minimizing secondary phase formation [80].

To further elucidate how synthesis variables collectively influence material properties, PDPs derived from the GPR model (**Figures%S16-S18, SI**) were analyzed. These plots visualize the non-linear interactions among temperature, time, molar ratio, and ion concentration, highlighting their combined effects on surface area, MB removal efficiency, and yield. The interpretation applies to **Figures 6** and **7** and **8**: these are empirical dataset-based visualizations in selected two-variable planes, whereas the corresponding controlled model-based trends are shown separately in the Supporting Information.

View Article Online
DOI: 10.1039/D5NR04817F

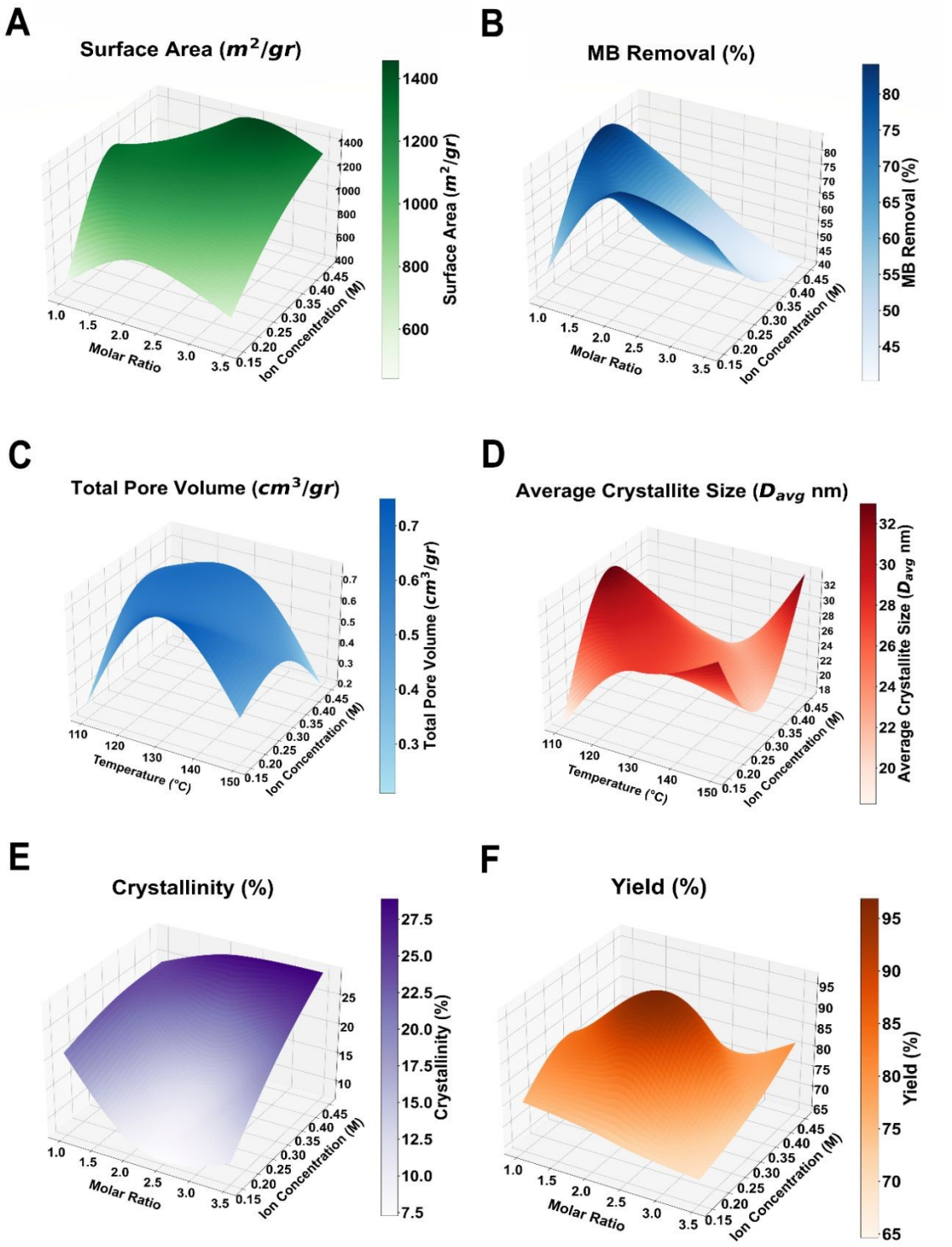


Figure 8. Experimental-data-based two-variable trend maps showing the observed variation of MIL-100(Fe) properties across the molar ratio and ion concentration plane: (A) surface area, (B) MB removal efficiency, (C) total

View Article Online
DOI: 10.1039/D5NR04817F

pore volume, (D) average crystallite size, (E) crystallinity, and (F) yield. These plots are intended as descriptive visual summaries of the measured dataset in this two-variable projection. Because the remaining synthesis variables are not fixed in these empirical surfaces, the plots should not be interpreted as isolated two-factor effects. The corresponding model-based GPR/PDP visualizations are provided in Figures S16–S18 (Supporting Information).

3.2.4. Feature Importance Analysis Using SHAP: Decoding the Drivers of MIL-100(Fe) Performance

To systematically unveil the influence of synthesis parameters on the functional performance of MIL-100(Fe), we employed SHAP analysis [81]. This powerful tool enables interpretable, model-agnostic insights into how individual variables including synthesis time, temperature, molar ratio, and ion concentration, govern key material properties. The resulting SHAP value-based feature importance plots are presented in **Figure 9**, illustrating the impact of each parameter on (A) BET surface area, (B) MB removal efficiency, (C) total pore volume, (D) average crystallite size, (E) crystallinity, and (F) yield. For compact visualization across the six panels, the synthesis variables are indexed as 1–4 in Figure 9, with the index mapping provided in the figure caption.

Across multiple properties, temperature and time consistently emerged as the dominant variables, while ion concentration showed less significance. For surface area (**Figure 9A**), temperature and metal-to-ligand molar ratio were identified as the most impactful parameters. Elevated temperatures and optimized stoichiometries promote efficient nucleation and crystal growth, leading to well-formed frameworks with enhanced porosity, as previously reported for HKUST-1 synthesized under supercritical and aerosol conditions%[82].

View Article Online
DOI: 10.1039/D5NR04817F

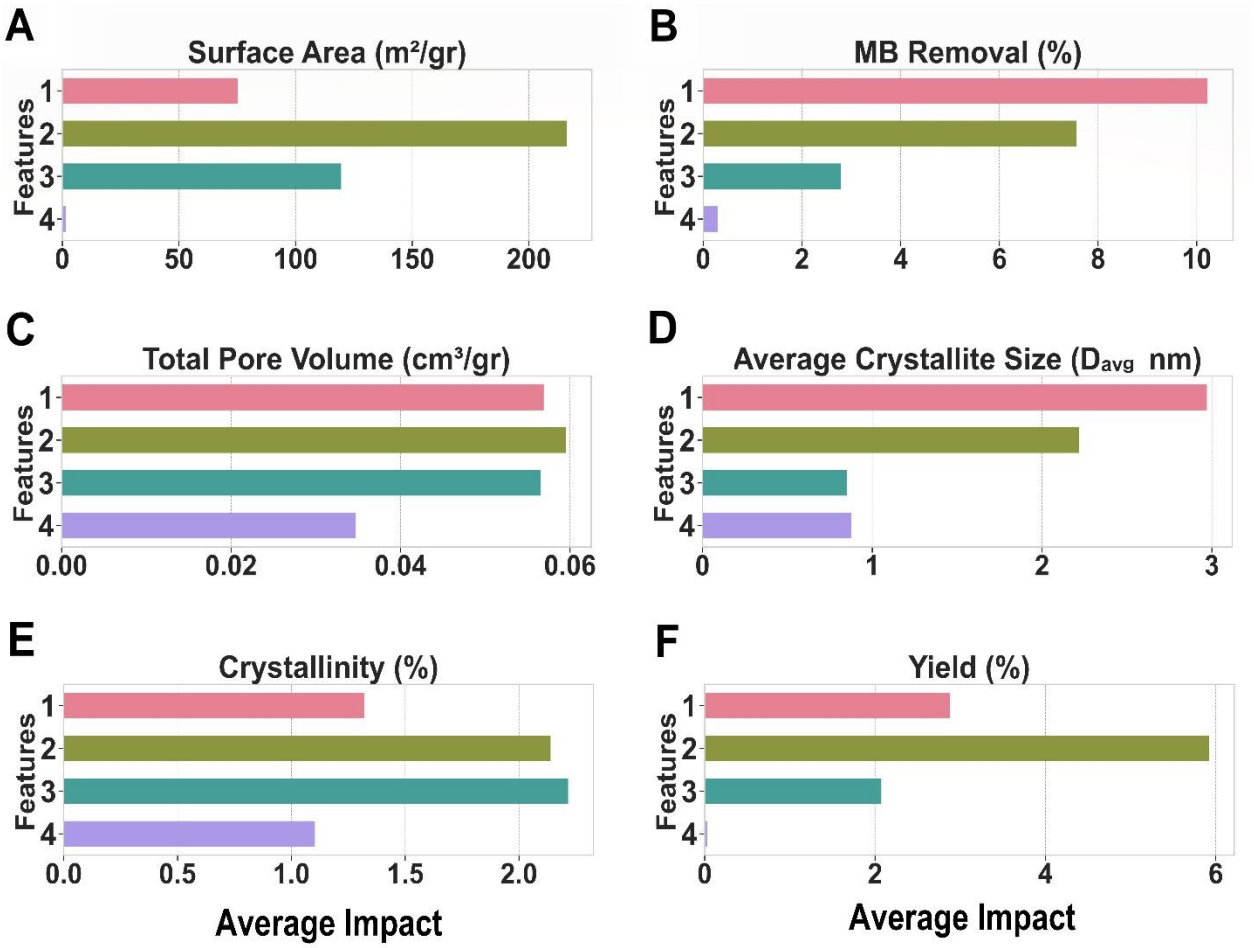


Figure 9. SHAP value-based feature importance analysis for MIL-100(Fe) synthesis parameters and their impact on various performance metrics. The average impact of synthesis features, Time (hr), Temperature (˚C), Molar Ratio, and Ion Concentration (M), on (A) Surface Area (m²/g), (B) MB Removal (%), (C) Total Pore Volume (cm³/g), (D) Average Crystallite Size (D_{avg}, nm), (E) Crystallinity (%), and (F) Yield (%). The y-axis indices correspond to the synthesis variables as follows: 1 = time (hr), 2 = temperature (˚C), 3 = molar ratio, and 4 = ion concentration (M). Results indicate that Time and Temperature were the most influential parameters across multiple targets, particularly for MB removal and surface area, whereas Ion Concentration had the least impact on the material properties. These insights assist in understanding the dynamics of synthesis conditions and facilitate the optimization of key parameters to improve material performance.

View Article Online
DOI: 10.1039/D5NR04817F

In the case of MB removal efficiency (**Figure 9B**), synthesis time was the most influential factor. Extended reaction durations likely favor defect generation and greater exposure of functional sites, critical for dye molecule interaction. This aligns with studies on UiO-66-NH_2, where modulated synthesis created framework disorder / inorganic-residue variation and significantly boosted dye adsorption capacities through improved accessibility and framework reactivity%[83]. Total pore volume (**Figure 9C**) was governed mainly by time, temperature and molar ratio, emphasizing their role in regulating framework openness and hierarchical pore structure. These trends corroborate findings in MIL-88B and similar MOFs, where fine-tuning synthesis conditions, especially temperature and metal-to-ligand molar ratios, enabled precise control of pore size and volume, enhancing adsorption potential [84]. Crystallite size (**Figure 9D**) was found to be most sensitive to synthesis time, with longer durations promoting sustained crystal growth and reducing structural defects. These observations are consistent with prior study on $BiFeO_3$ and related materials%[85].

Similarly, crystallinity (**Figure 9E**) was strongly influenced by both temperature and molar ratio. Elevated synthesis temperatures accelerate nucleation kinetics, promoting rapid and uniform crystal growth, while optimized molar ratios ensure balanced coordination and defect minimization. This combination yields highly ordered frameworks with greater structural integrity and enhanced crystallinity [86]. Furtherly yield (**Figure 9F**), was primarily dictated by temperature. Higher thermal energy promotes faster precursor conversion and stable framework formation, enhancing synthetic efficiency. Similar observations were made in water-based, synthesis of

View Article Online
DOI: 10.1039/D5NR04817F

CPO-27 (MOF-74) family members, where precise temperature control significantly improved space-time yields and material purity, while minimizing by-products [87].

Overall, the SHAP-based feature importance analysis provides a clear understanding of MIL-100(Fe) synthesis, revealing that adsorption efficiency and other performance metrics are driven by distinct parameters rather than conventional indicators like surface area. Maximizing surface area alone does not ensure superior functional outcomes, especially for pollutant adsorption, where crystallinity, pore accessibility, and defect density play crucial roles. These insights highlight the need for property-specific optimization strategies, with synthesis parameters tuned to achieve targeted functionalities. Building on this, we employ a data-driven, multi-objective optimization framework using GA, informed by SHAP analysis, to enable the rational design of MOFs for high-performance applications such as environmental remediation and catalysis.

3.2.5. Targeted Optimization of MIL-100(Fe): Guiding Synthesis Through Property-Specific Design

As previously discussed, the optimal synthesis strategies for maximizing surface area and MB removal efficiency in MIL-100(Fe) diverge sharply, underscoring crucial structure-function nuances. To test whether a machine-learning-assisted route can deliberately target distinct properties, we coupled a GPR with GA to propose synthesis conditions for two objectives: (i) maximal BET surface area and (ii) maximal MB removal efficiency. The optimized inputs and their experimental verification are given in **Table%4** (full physicochemical descriptors in **Tables%S2** and **S3**) and summarized visually in **Figure%10.** For "OPT BET", the model selected 46.1%h, 110.9%˚C, metal-to-ligand molar ratio%=%2.7:1, ion%concentration%=%0.29%M; i.e.,

View Article Online
DOI: 10.1039/D5NR04817F

long reaction time near the upper/mean end of our design window and temperature close to the lower bound. This recipe yielded a BET surface area of 1841.9%m²/g, 85.1% MB removal, 34.9% crystallinity, 0.90%cm³%g⁻¹ pore volume, and 93.1% yield. In contrast, "OPT MB" was obtained at 39.4%h, 134.1%˚C, metal-to-ligand molar ratio%=%1.0:1, ion%concentration%=%0.19%M, i.e., shorter synthesis time near the lower bound and a temperature toward the upper bound, resulting 98.29% MB removal, a surface area of 1274.3%m²/g, 17.4% crystallinity, 0.92%cm³%g⁻¹ pore volume, and a larger average crystallite size (30.8%nm), indicative of greater functional accessibility despite lower overall crystallinity. A size-matching effect likely also contributes to the superior adsorption performance of OPT MB. Methylene blue is a relatively large cationic aromatic dye, with reported molecular dimensions of approximately 14.3 × 6.1 × 4.0 Å, meaning that adsorption depends not only on the total available surface area but also on whether the accessible pore features can accommodate the dye with a favorable orientation. MIL-100(Fe) is known to contain large cages of roughly 25 and 29 Å, connected through smaller windows of about 5.5 and 8.6 Å. Within this structural context, adsorption is expected to benefit when the accessible pore environment is large enough to host the dye while still imposing sufficient steric selectivity at the window or pore-entry level. In our data, the main pore features of the synthesized samples fall in the 1.2–1.8 nm range, and the BJH comparison for the optimized materials shows that OPT MB has a narrower pore-size distribution than OPT BET. Together, these observations suggest that OPT MB achieves a more favorable steric match between accessible pore dimensions and the cross-sectional dimensions of methylene blue, which can enhance uptake efficiency even without maximizing BET surface area alone[88, 89]. This interpretation is consistent with adsorption studies

View Article Online
DOI: 10.1039/D5NR04817F

in which appropriate pore/channel size, together with electrostatic and π–π interactions, governs high MB uptake more effectively than unrestricted enlargement of pore volume [12, 14, 55, 90].

Table 4. Optimized Synthesis Parameters for Maximum Surface area and MB removal efficiency. The optimized synthesis conditions for achieving maximum BET surface area and MB removal efficiency for MIL-100(Fe) are presented. The results are based on GPR predictions, followed by real-world synthesis verification. Errors are provided in parentheses.

Feature	Surface Area Optimization (OPT BET)	MB Removal Optimization (OPT MB)
Time (hr)	46.1	39.4
Temperature (˚C)	110.9	134.1
Metal-to-Ligand Molar Ratio	2.7:1	1:1
Ion Concentration (M)	0.29	0.19

Figure 10 makes the trade-off explicit: although OPT BET possesses 44% higher surface area than OPT MB (1841.9 vs. 1274.3 $m^2 g^{-1}$), it performs 13% worse in adsorption (85.1% vs. 98.3%). Together with Figure S23, this directly shows that surface area alone is an insufficient predictor of dye uptake. Relative to S11, the strongest baseline sample within the initial 11-sample experimental set (1748 $m^2 g^{-1}$, 87.7% MB removal), OPT BET increased surface area by 93.9 $m^2 g^{-1}$ (5.37% relative), while OPT MB improved MB removal by 10.6 percentage points (12.1% relative). Conditions that favor surface area (lower temperature, longer time, higher metal-to-ligand molar ratio) do not align with those that favor MB removal (higher temperature, shorter time, lower metal-to-ligand molar ratio). Hence, a single globally optimal recipe is unrealistic;

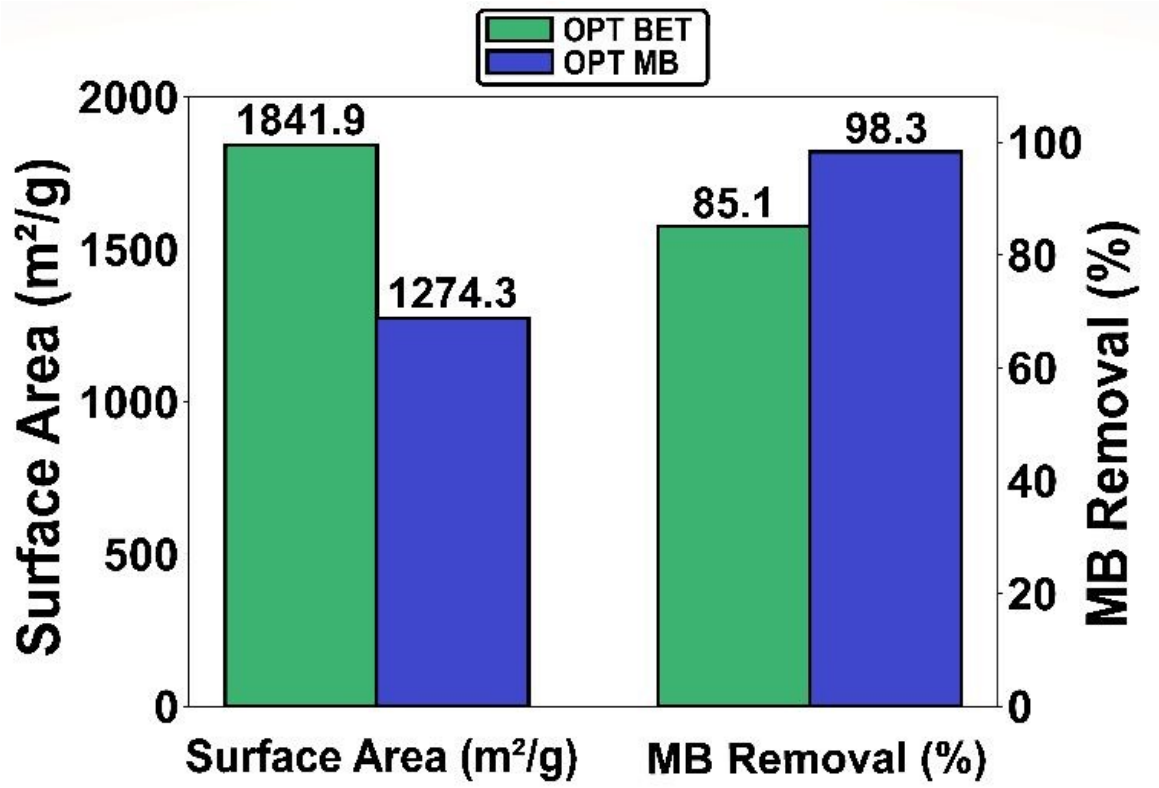
View Article Online
DOI: 10.1039/D5NR04817F

instead, ML-guided, property-specific synthesis design is essential for function-driven MOF optimization. Together with **Figure%S23**, this directly shows that surface area alone is an insufficient predictor of dye uptake. Because the purpose of **Figure 10** is to compare the experimental validation outcomes of the two optimized formulations, the plotted values are shown as direct measured results rather than as replicate-based statistical summaries. The predictive reliability of the surrogate models is evaluated separately using the GPR cross-validation metrics reported elsewhere in the manuscript. Therefore, the present comparison should be interpreted as an intra-framework optimization result for MIL-100(Fe) synthesized under different conditions, rather than as a universal claim that lower crystallinity or higher defect concentration is generally preferable across all MOFs. In a different MOF family with an intrinsically better pore/chemistry match for MB, a highly crystalline structure could indeed represent the optimal adsorbent; in the present MIL-100(Fe) synthesis space, however, the highest BET surface area did not coincide with the highest MB-removal performance. Therefore, the inferior MB removal of OPT BET relative to OPT MB should not be interpreted simply as a surface-area effect. Rather, under the fixed adsorption conditions used here, the result likely reflects the combined influence of equilibrium

View Article Online
DOI: 10.1039/D5NR04817F

site availability, pore accessibility, dye–framework interactions, and transport through the accessible pore network.

Figure 10. Experimental validation comparison of BET surface area and MB removal efficiency of MIL-100(Fe) synthesized under GPR-optimized conditions. The OPT BET sample (green) exhibits the highest surface area (1841.9%m²/g) but lower MB removal (85.1%), while the OPT MB sample (blue) achieves superior MB removal (98.3%) despite a smaller surface area (1274.3%m²/g). The plotted values represent the directly measured validation



outcomes of the two optimized formulations under controlled experimental conditions. Model predictive performance is evaluated separately using GPR cross-validation metrics and is therefore not shown as experimental error bars in this figure.

Although direct defect quantification was not performed in the present study, N_2 adsorption BET results, and TGA trends suggest that the optimized samples differ not only in textural properties but also in the degree of structural modulation associated with framework disorder and inorganic

View Article Online
DOI: 10.1039/D5NR04817F

residue. In MOF literature, TGA residue is often used as a supportive indicator of inorganic fraction and, in suitable systems, can be consistent with linker-deficient or defect-rich compositions; however, it is not a standalone proof of missing-linker defects and should be interpreted together with other structural evidence. The corresponding residual-mass comparison is provided in **Figure S28 (Supporting Information)**. In our case, the higher inorganic residue observed for OPT BET and S11 is consistent with larger linker deficiency which can be proven by higher N_2 adsorption of BET results, whereas OPT MB and S10 with lower residue and lower BET surface area show similarly strong dye-removal performance. This suggests that simply increasing defect-related disorder or inorganic content does not necessarily maximize MB adsorption, even if it may contribute to higher N_2 uptake or BET-derived surface area. Instead, the present data support the idea that controlled structural modulation, rather than defect increase alone, is more relevant to application-specific optimization. Samples S5 and S7 require additional caution because their elevated residue may partly arise from iron-oxide secondary phases, as indicated by XRD, rather than from defect formation within the MIL-100(Fe) framework itself [91-93].

The practical applicability of the OPT MB sample for water-treatment applications was further evaluated through adsorption–desorption cycling and post-reuse XRD analysis. As shown in **Figure S29** (Supporting Information), the MB removal efficiency decreased gradually from 98.3% in the initial cycle to 87.4% after four adsorption–desorption cycles, corresponding to retention of approximately 88.9% of the initial removal performance. This result demonstrates encouraging operational durability of OPT MB under repeated dye-removal conditions. In addition, the post-reuse XRD pattern of OPT MB after four cycles retains the main characteristic diffraction features

View Article Online
DOI: 10.1039/D5NR04817F

of the original OPT MB sample (**Figure S34**, Supporting Information), indicating that the framework crystallinity is largely preserved after repeated adsorption–desorption operation. These findings are consistent with the broader MIL-100(Fe)-based water-treatment literature, where recyclability and post-use structural characterization are commonly used to evaluate practical adsorbent stability, and where MIL-100(Fe)-based materials have been reported as promising platforms for aqueous contaminant removal due to their permanent porosity and comparatively good water stability. Together, the cycling and post-reuse XRD results support the functional and structural robustness of OPT MB for repeated MB-removal use. Accordingly, the present cycling and post-reuse XRD results demonstrate encouraging reuse behavior of OPT MB and supports its potential for repeated dye-removal operation which is consistent with the reported descent structural stabilities of MIL-100(Fe) under dye adsorption experiments in the literature [48, 94-98]. The retained adsorption performance over repeated cycles and post-reuse XRD (**Figure S29 and Figure S34 Supporting Information**), together with prior reports of negligible Fe leaching from MIL-100(Fe)-based adsorbents in water, supports the operational promise of OPT MB for repeated dye-removal use.

Figure 11 presents a t-SNE projection of the full descriptor space for the 11 experimental samples, with the two experimentally validated optimized samples overlaid for comparison. Here, t-SNE is used only as a qualitative visualization of local neighborhood structure in property space, rather than as a formal inferential method. Because the dataset is small, we do not interpret the embedding quantitatively or claim that t-SNE provides a unique representation of the sample relationships. Instead, the figure is used to illustrate how samples with related structural and

View Article Online
DOI: 10.1039/D5NR04817F

adsorption profiles occupy similar regions of the descriptor space. This feature-reduced view supports design and inverse design by revealing how samples cluster according to overall property profiles rather than any single metric. Three neighborhoods emerge. Samples S1, S3, S4, and S7 (surface area 1068.9–1436.9 $m^2 g^{-1}$) show lower MB removal (40–53%), demonstrating that high surface area alone does not guarantee uptake. Samples S5, S8, and S9 achieve relatively strong MB removal (70–84.5%) despite lower surface areas. A third region containing S2, S6, and S10 exhibits a more balanced response, with moderate-to-high surface area (1273–1423 $m^2 g^{-1}$) and MB removal up to 88.6%. The optimized samples fall in separate regions aligned with their targets. OPT BET (1841.9 $m^2 g^{-1}$, 85.1% removal) lies near S11 (1748 $m^2 g^{-1}$, 87.7%), whereas OPT MB (1274.3 $m^2 g^{-1}$, 98.3%) is positioned closest to S10 (1423.3 $m^2 g^{-1}$, 88.6%), within a region characterized by relatively high MB removal efficiencies and moderate surface areas. This separation provides actionable design criteria: choose the latent-space region associated with the desired objective (maximal surface area, maximal MB removal, or a trade-off), then use our GPR → GA → experiment workflow to propose and validate synthesis conditions within that region. In this way, the t-SNE map functions as a practical clustering tool for chemists, linking machine-learning insight directly to property-driven synthesis decisions without invoking detailed mechanistic claims. For clarity of interpretation, the t-SNE structure was first established from the experimental sample set, and the two optimized samples were then overlaid as prospective validation points to show how the experimentally realized optima relate to the original property landscape. To address the dependence of this visualization on the embedding method, we additionally performed a principal component analysis (PCA), which is provided in the Supporting

View Article Online
DOI: 10.1039/D5NR04817F

Information (**Figures S30** and **S31**). The PCA gives a broadly consistent qualitative picture: OPT BET remains positioned near the high-BET region associated with S11, while OPT MB lies in the lower-PC2 region closer to S10 and other relatively strong adsorption samples. Thus, the main conclusion drawn from **Figure 11** that the optimized samples occupy distinct property-space regions aligned with different objectives, does not rely solely on the use of t-SNE.

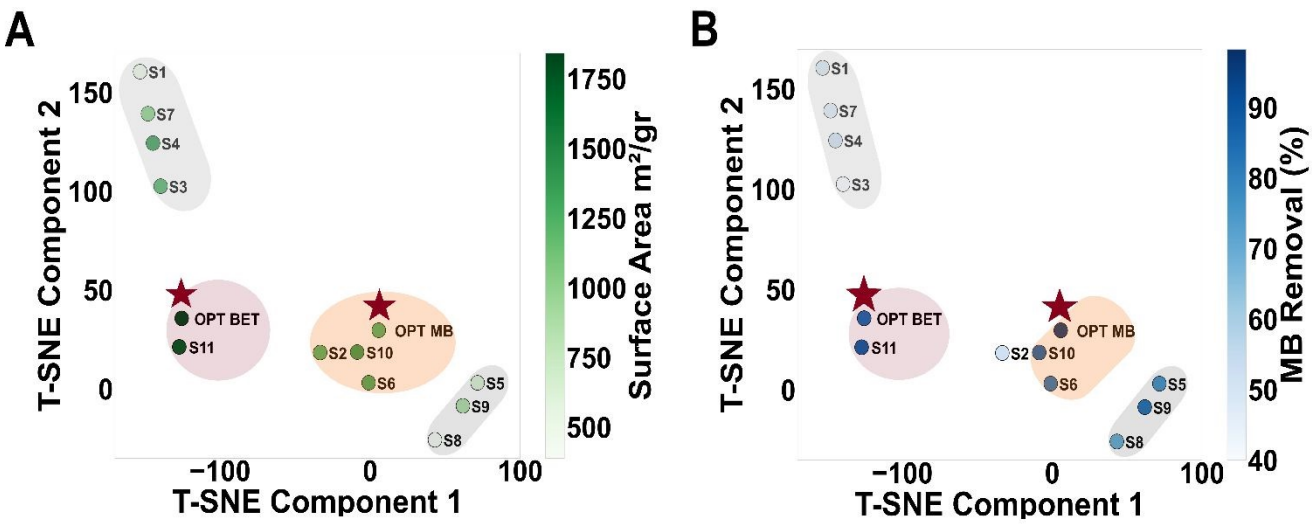


Figure 11. t-SNE projection of the 11 experimental samples and two experimentally validated optimized samples (OPT BET, OPT MB) based on their structural and performance descriptors. (A) Points colored by BET surface area (m²/g). (B) Points colored by MB removal efficiency (%). Distinct regions emerge, with OPT BET (1841.9 m²/g, 85.1% MB removal) aligning with high-surface-area samples such as S11, while OPT MB (1274.3 m²/g, 98.3% MB removal) lies near samples with relatively high MB removal efficiencies and moderate surface areas, particularly S10, emphasizing the property-specific optimization trade-off. Two optimized samples has been clarified with red stars. The optimized points are shown to indicate where the prospective validation samples fall relative to the original property-space distribution.

4. Conclusion

View Article Online
DOI: 10.1039/D5NR04817F

This study represents low-data proof-of-concept for ML-guided optimization of MIL-100(Fe) synthesis under diverse, green hydrothermal conditions. By systematically tuning temperature, reaction time, metal-to-ligand molar ratio, and ion concentration, we uncovered how these parameters govern not only structural properties including BET surface area, crystallinity, and total pore volume, but more critically, functional performance in MB dye removal. Our results decisively challenge the conventional reliance on BET surface area as a performance proxy: for example, a sample with a surface area of 1748 m^2/g did not outperform samples with lower surface areas (about 1423 m^2/g) in MB removal.

Through the integration of GPR-based surrogate modeling, GA optimization, and prospective experimental validation, we identified property-specific synthesis conditions that improved either BET surface area or MB removal. The two optimized synthesis pathways, OPT BET (46.1 h, 110.9 ˚C, 2.7:1 metal-to-ligand molar ratio, 0.29 M) and OPT MB (39.4 h, 134.1 ˚C, 1:1 metal-to-ligand molar ratio, 0.19 M), produced record performances, achieving a maximum surface area of 1841.9 m^2/g and a remarkable 98.3% MB removal, respectively. These results underscore the importance of defining the correct performance metric: optimizing for surface area alone would not have yielded the highest adsorption efficiency. The OPT MB sample also retained high MB-removal efficiency over repeated adsorption–desorption cycles and preserved its main XRD features after reuse, supporting its practical potential for repeated dye-removal operation. SHAP-based feature importance and partial dependence plots confirmed that synthesis time and temperature dominate the property landscape, while ion concentration shows comparatively less significance. Although the present study used a fixed Taguchi-based experimental design followed by GPR–GA

View Article Online
DOI: 10.1039/D5NR04817F

optimization, a natural next step would be to implement an active-learning or Bayesian-optimization workflow. In such a closed-loop strategy, the Gaussian process surrogate would not only predict target properties but also provide uncertainty estimates that guide the selection of the next most informative synthesis experiment. This would allow the experimental campaign to balance exploitation of promising synthesis regions with exploration of uncertain regions, potentially reducing the number of required experiments while improving optimization efficiency. Such adaptive experimental-design strategies are increasingly used in materials discovery and autonomous experimentation, particularly when each experiment is costly or time-consuming.

The addition of t-SNE analysis provided a holistic visualization of the property space, revealing distinct clusters of samples with contrasting structural and functional signatures. OPT BET aligned with high-surface area, samples such as S11 (1748 m²/g, 87.7% MB removal), whereas OPT MB clustered near samples like S10 (1423.3 m²/g, 88.6% MB removal). These clusters indicate that adsorption efficiency is not dictated by BET surface area alone. Samples with similar or even lower surface areas can outperform in MB removal when micropore accessibility and defect density are favorable; notably, several lower surface area samples (e.g., S5, S8, S9 with about 389 to 856 m²/g and about 70 to 85 %) delivered strong uptake despite modest BET. Thus, high surface area and high adsorption map to different structural regimes, motivating property-specific optimization rather than a single recipe.

Ultimately, this work advances a performance-driven materials design criterion, where synthesis conditions are intentionally tuned to maximize the functional property of interest, whether for environmental remediation, catalysis, or gas storage. By moving beyond surface-area-centric

View Article Online
DOI: 10.1039/D5NR04817F

paradigms, this study provides a transferable and data-efficient blueprint for intelligent MOF development. Our combination of interpretable ML, optimization algorithms, and targeted experiments offers a sustainable, resource-efficient path to designing high-performance materials, setting a foundation for future innovations at the interface of machine learning and materials chemistry.

Author contributions

Conceptualization was led by Mehrdad Asgari and Fahimeh Hooriabad Saboor. Data curation was performed by Mehrdad Asgari, Fahimeh Hooriabad Saboor, and Mohammadreza Beydaghdari, while formal analysis was carried out by Saeid Zahedi Asl and Shayan Abaei. The investigation was conducted by Hosein Alimardani, Shayan Abaei, and Saeid Zahedi Asl, and the methodology was developed by Shayan Abaei, Saeid Zahedi Asl, and Mohammadreza Beydaghdari. Project administration was managed by Mehrdad Asgari and Fahimeh Hooriabad Saboor, with resources provided by Fahimeh Hooriabad Saboor and Shayan Abaei. Software implementation was undertaken by Hosein Alimardani and Shayan Abaei. Supervision and validation were provided by Mehrdad Asgari and Fahimeh Hooriabad Saboor, and visualization was prepared by Shayan Abaei. The original draft was written by Shayan Abaei and Saeid Zahedi

View Article Online
DOI: 10.1039/D5NR04817F

Asl, with review and editing carried out by Mehrdad Asgari, Fahimeh Hooriabad Saboor, and Shayan Abaei. Saeid Zahedi Asl and Shayan Abaei contributed equally to this work (†). Correspondence should be addressed to Mehrdad Asgari (*), with Fahimeh Hooriabad Saboor (*) serving as co-corresponding author.

Conflicts of interest

There are no conflicts to declare.

Data availability

The machine-learning dataset used in this study has been provided as a machine-readable CSV file, including the synthesis variables reported in Table 2 and the measured response variables reported in Table 3. The repository also contains a data dictionary, the scripts used for the leave-one-out cross-validation audit, the corresponding LOOCV prediction/metric tables, and the GPR parameter files for the two primary optimization targets, BET surface area and MB removal. These materials are available at: https://github.com/ShayanAbaei/mil100fe-ml-ga-dye-removal

Acknowledgements

View Article Online
DOI: 10.1039/D5NR04817F

M.A. acknowledges the funding provided by UK Research and Innovation (UKRI) under the UK Government's Horizon Europe funding guarantee (grant reference: EP/Y023447/1; organization reference: 101106377).

References:

1. O. M. Yaghi, G. Li and H. Li, *Nature*, 1995, **378**, 703–706.
2. H. Li, M. Eddaoudi, M. O'Keeffe and O. M. Yaghi, *Nature*, 1999, **402**, 276–279.
3. O. M. Yaghi and H. Li, *Journal of the American Chemical Society*, 1995, **117**, 10401–10402.
4. H. Demir, H. Daglar, H. C. Gulbalkan, G. O. Aksu and S. Keskin, *Coordination Chemistry Reviews*, 2023, **484**, 215112.
5. J. B. DeCoste, G. W. Peterson, H. Jasuja, T. G. Glover, Y.-g. Huang and K. S. Walton, *Journal of Materials Chemistry A*, 2013, **1**, 5642–5650.
6. K. Leus, T. Bogaerts, J. De Decker, H. Depauw, K. Hendrickx, H. Vrielinck, V. Van Speybroeck and P. Van Der Voort, *Microporous and Mesoporous Materials*, 2016, **226**, 110–116.

View Article Online
DOI: 10.1039/D5NR04817F

7. H.-L. Wang, H. Yeh, Y.-C. Chen, Y.-C. Lai, C.-Y. Lin, K.-Y. Lu, R.-M. Ho, B.-H. Li, C.-H. Lin and D.-H. Tsai, *ACS Applied Materials & Interfaces*, 2018, **10**, 9332–9341.
8. L.-J. Han, F.-Y. Ge, G.-H. Sun, X.-J. Gao and H.-G. Zheng, *Dalton Transactions*, 2019, **48**, 4650–4656.
9. G. A. Cobian-Solorio, I. A. Aguayo-Villarreal, C. K. Rojas-Mayorga, R. Muñiz-Valencia, M. J. Emparan-Legaspi and N. E. Davila Guzman, *Journal of Molecular Structure*, 2025, **1321**, 139698.
10. J. Chen, N. Yao, Y. Tang, L. Xie, X. Zhuo and Z. Jiang, *Dalton Transactions*, 2024, **53**, 5900–5910.
11. C.-F. Zhang, L.-G. Qiu, F. Ke, Y.-J. Zhu, Y.-P. Yuan, G.-S. Xu and X. Jiang, *Journal of Materials Chemistry A*, 2013, **1**, 14329–14334.
12. Y. Zhao and Y. Yamauchi, *Nature Chemistry*, 2025, **17**, 161–162.
13. B. Ahmadi, A. Fallah, R. Ghamarpoor and M. Jamshidi, *Results in Chemistry*, 2025, **18**, 102910.
14. T. T. M. Bui, L. T. Nguyen, N. P. H. Pham, C. C. Tran, L. T. Nguyen, T. A. Nguyen, H. N. Nguyen and M. V. Nguyen, *RSC Advances*, 2021, **11**, 36626–36635.
15. J. Fito, M. Abewaa, A. Mengistu, K. Angassa, A. D. Ambaye, W. Moyo and T. Nkambule, *Scientific Reports*, 2023, **13**, 5427.
16. W. Li, J. Cao, W. Xiong, Z. Yang, S. Sun, M. Jia and Z. Xu, *Chemical Engineering Journal*, 2020, **392**, 124844.
17. K. L. Tan and K. Y. Foo, *Journal of Environmental Chemical Engineering*, 2021, **9**, 104923.
18. Z. Armstrong, A. MacRae, M. Lenertz, Q. Li, K. Johnson, A. Scheiwiller, P. Shen, L. Feng, M. Quadir and Z. Yang, *ACS Applied Materials & Interfaces*, 2023, **15**, 38124–38131.
19. C. Altintas, I. Erucar and S. Keskin, *ACS Applied Materials & Interfaces*, 2018, **10**, 3668–3679.
20. N. S. Portillo-Vélez, J. L. Obeso, J. A. de los Reyes, R. A. Peralta, I. A. Ibarra and M. T. Huxley, *Communications Materials*, 2024, **5**, 247.
21. Y. Song and S. Ma, *Chemical Science*, 2025, **16**, 11740–11767.
22. M. Liu, L. Zu and Z. M. Hudson, *ACS Nano*, 2022, **16**, 13573–13594.
23. C. R. Marshall, S. A. Staudhammer and C. K. Brozek, *Chemical Science*, 2019, **10**, 9396–9408.
24. N. Sikdar, M. Bhogra, Umesh V. Waghmare and T. K. Maji, *Journal of Materials Chemistry A*, 2017, **5**, 20959–20968.
25. S. M. Moosavi, A. Chidambaram, L. Talirz, M. Haranczyk, K. C. Stylianou and B. Smit, *Nature Communications*, 2019, **10**, 539.
26. I. Spanopoulos, C. Tsangarakis, E. Klontzas, E. Tylianakis, G. Froudakis, K. Adil, Y. Belmabkhout, M. Eddaoudi and P. N. Trikalitis, *Journal of the American Chemical Society*, 2016, **138**, 1568–1574.
27. S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit and H. J. Kulik, *Nature Communications*, 2020, **11**, 4068.

View Article Online
DOI: 10.1039/D5NR04817F

28. J.-M. Vanson, F.-X. Coudert, M. Klotz and A. Boutin, *Langmuir*, 2017, **33**, 1405–1411.
29. C. Wang, X. Liu, T. Yang, D. Sridhar, H. Algadi, B. Bin Xu, Z. M. El-Bahy, H. Li, Y. Ma, T. Li and Z. Guo, *Separation and Purification Technology*, 2023, **320**, 124144.
30. H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.
31. B. Luan Tran, H.-Y. Chin, B. K. Chang and A. S. T. Chiang, *Microporous and Mesoporous Materials*, 2019, **277**, 149–153.
32. C. G. Piscopo, F. Trapani, A. Polyzoidis, M. Schwarzer, A. Pace and S. Loebbecke, *New Journal of Chemistry*, 2016, **40**, 8220–8224.
33. J. Liu, T. A. Goetjen, Q. Wang, J. G. Knapp, M. C. Wasson, Y. Yang, Z. H. Syed, M. Delferro, J. M. Notestein, O. K. Farha and J. T. Hupp, *Chemical Society Reviews*, 2022, **51**, 1045–1097.
34. H. Wang, L. Yang, D. Leng, Y. Du and H. Ning, *Advances in Colloid and Interface Science*, 2025, **346**, 103671.
35. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.
36. K. M. Jablonka, D. Ongari, S. M. Moosavi and B. Smit, *Chemical Reviews*, 2020, **120**, 8066–8129.
37. G. Zhijiang, Z. Hongtao and J. Jinping, 2002.
38. L. P. Swiler, M. Gulian, A. L. Frankel, C. Safta and J. D. Jakeman, *Journal of Machine Learning for Modeling and Computing*, 2020, **1**, 119–156.
39. Y. Zhang and C. Ling, *npj Computational Materials*, 2018, **4**, 25.
40. Y. Xu and R. Goodacre, *Journal of Analysis and Testing*, 2018, **2**, 249–262.
41. V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, *Chemical Reviews*, 2021, **121**, 10073–10141.
42. T. Hastie, R. Tibshirani and J. Friedman, *Journal*, 2009.
43. A. Tesar and M. Drzik, *Computers & Structures*, 1995, **57**, 287–295.
44. G. E. Liepins and M. R. Hilliard, *Annals of Operations Research*, 1989, **21**, 31–57.
45. Z. Michalewicz and C. Z. Janikow, *Statistics and Computing*, 1991, **1**, 75–91.
46. N. Chakraborti, *International Materials Reviews*, 2004, **49**, 246–260.
47. A. Dhakshinamoorthy, M. Alvaro, H. Chevreau, P. Horcajada, T. Devic, C. Serre and H. Garcia, *Catalysis Science & Technology*, 2012, **2**, 324–330.
48. L. Chen, X. Zuo, S. Yang, T. Cai and D. Ding, *Chemical Engineering Journal*, 2019, **359**, 373–384.
49. S. Mitra, S. Das, S. Basu, P. Sahu and K. Mandal, *Journal of Magnetism and Magnetic Materials*, 2009, **321**, 2925–2931.
50. S. Thampy, N. Ashburn, T. J. Martin, C. Li, Y. Zheng, J. Y. Chan, K. Cho and J. W. P. Hsu, *RSC Advances*, 2018, **8**, 28–37.
51. M. Chalermnon, S. R. Thomas, J. M. Chin and M. R. Reithofer, *Inorganic Chemistry Frontiers*, 2025, DOI: 10.1039/D5QI01201E.
52. A. F. Sapnik, C. W. Ashling, L. K. Macreadie, S. J. Lee, T. Johnson, S. G. Telfer and T. D. Bennett, *Journal of Materials Chemistry A*, 2021, **9**, 27019–27027.

View Article Online
DOI: 10.1039/D5NR04817F

53. M. Shahnawaz Khan, M. Khalid and M. Shahid, *Materials Advances*, 2020, **1**, 1575–1601.
54. M. Beydaghdari, F. Hooriabad Saboor, A. Babapoor, V. V. Karve and M. Asgari, *Energies*, 2022, **15**, 2023.
55. F. Tan, M. Liu, K. Li, Y. Wang, J. Wang, X. Guo, G. Zhang and C. Song, *Chemical Engineering Journal*, 2015, **281**, 360–367.
56. V. K.-M. Au, *Frontiers in Chemistry*, 2020, **Volume 8 - 2020**.
57. J. C. F. de Winter, S. D. Gosling and J. Potter, *Psychological Methods*, 2016, **21**, 273–290.
58. P. Xu, X. Ji, M. Li and W. Lu, *npj Computational Materials*, 2023, **9**, 42.
59. J. Schmidt, M. R. G. Marques, S. Botti and M. A. L. Marques, *npj Computational Materials*, 2019, **5**, 83.
60. H. Tang, L. Duan and J. Jiang, *Langmuir*, 2023, **39**, 15849–15863.
61. B. Zheng, K. Gao, D. Tian, W. Yao, L. Zhang, L. Wang and J. Wang, *Inorganic Chemistry*, 2021, **60**, 13419–13424.
62. L. Peng, S. Yang, S. Jawahery, S. M. Moosavi, A. J. Huckaba, M. Asgari, E. Oveisi, M. K. Nazeeruddin, B. Smit and W. L. Queen, *Journal of the American Chemical Society*, 2019, **141**, 12397–12405.
63. S. Zhang, W. Zhang, A. Yadav, J. Baker and S. Saha, *Inorganic Chemistry*, 2023, **62**, 18999–19005.
64. J. Lyu, X. Zhang, Z. Chen, R. Anderson, X. Wang, M. C. Wasson, P. Bai, X. Guo, T. Islamoglu, D. A. Gómez-Gualdrón and O. K. Farha, *ACS Applied Materials & Interfaces*, 2019, **11**, 42179–42185.
65. S. D. Bagi, A. S. Myerson and Y. Román-Leshkov, *Crystal Growth & Design*, 2021, **21**, 6529–6536.
66. S. L. Griffin, M. L. Briuglia, J. H. ter Horst and R. S. Forgan, *Chemistry – A European Journal*, 2020, **26**, 6910–6918.
67. L. W. Aguiar, G. P. Otto, V. L. Kupfer, S. L. Fávaro, C. T. P. Silva, M. P. Moisés, L. de Almeida, M. R. Guilherme, E. Radovanovic, E. M. Girotto and A. W. Rinaldi, *Materials Letters*, 2020, **276**, 128127.
68. S. Liu, Y. Zhang, Y. Meng, F. Gao, S. Jiao and Y. Ke, *Crystal Growth & Design*, 2013, **13**, 2697–2702.
69. G. Rabiee, A. Abbasi and M. Behbahani, *Nanoscale*, 2024, **16**, 20738–20751.
70. Z. Hu, I. Castano, S. Wang, Y. Wang, Y. Peng, Y. Qian, C. Chi, X. Wang and D. Zhao, *Crystal Growth & Design*, 2016, **16**, 2295–2301.
71. C. Ardila-Suárez, V. D. Molina, H. Alem, V. G. Baldovino-Medrano and G. E. Ramírez-Caballero, *Phys Chem Chem Phys*, 2020, **22**, 12591–12604.
72. S. Dai, F. Nouar, S. Zhang, A. Tissot and C. Serre, *Angewandte Chemie International Edition*, 2021, **60**, 4282–4288.
73. A. Schejn, L. Balan, V. Falk, L. Aranda, G. Medjahdi and R. Schneider, *CrystEngComm*, 2014, **16**, 4493–4500.
74. L. Wang, P. Tang, J. Liu, A. Geng, C. Song, Q. Zhong, L. Xu and L. Gan, *Journal of Colloid and Interface Science*, 2019, **554**, 260–268.

View Article Online
DOI: 10.1039/D5NR04817F

75. J. Li, N. Zhang and D. H. L. Ng, *Journal of Materials Chemistry A*, 2015, **3**, 21106–21115.
76. N. Mizutani, T. Iwasaki, S. Watano, T. Yanagida, H. Tanaka and T. Kawai, *Bulletin of Materials Science*, 2008, **31**, 713–717.
77. M. Kinoshita, S. Yanagida, T. Gessei and A. Monkawa, *Journal of Crystal Growth*, 2022, **600**, 126877.
78. F. Zhou, X. Zhao, H. Xu and C. Yuan, *The Journal of Physical Chemistry C*, 2007, **111**, 1651–1657.
79. Y. Zhang, Y. Jia, M. Li and L. a. Hou, *Scientific Reports*, 2018, **8**, 9597.
80. M. Taddei, N. Casati, D. A. Steitz, K. C. Dümbgen, J. A. van Bokhoven and M. Ranocchiari, *CrystEngComm*, 2017, **19**, 3206–3214.
81. S. M. Lundberg and S.-I. Lee, presented in part at the Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, 2017.
82. J. Feng, A. Lau and I. V. Novosselov, *Nanoscale*, 2024, **16**, 22142–22151.
83. K. Wang, C. Li, Y. Liang, T. Han, H. Huang, Q. Yang, D. Liu and C. Zhong, *Chemical Engineering Journal*, 2016, **289**, 486–493.
84. A. Yurduşen and Y. Yürüm, *Industrial & Engineering Chemistry Research*, 2019, **58**, 14058–14072.
85. H. Zhang and K. Kajiyoshi, *Journal of the American Ceramic Society*, 2010, **93**, 3842–3849.
86. H.-Y. Wu, C.-L. Wu, W. Liao, B. M. Matsagar, K.-Y. Chang, J.-H. Huang and K. C. W. Wu, *Journal of Materials Chemistry A*, 2023, **11**, 9427–9435.
87. L. Garzón-Tovar, A. Carné-Sánchez, C. Carbonell, I. Imaz and D. Maspoch, *Journal of Materials Chemistry A*, 2015, **3**, 20819–20826.
88. F. Jeremias, S. K. Henninger and C. Janiak, *Dalton Transactions*, 2016, **45**, 8637–8644.
89. H. Li, L. Liu, J. Cui, J. Cui, F. Wang and F. Zhang, *RSC Advances*, 2020, **10**, 14262–14273.
90. Y. Zhao, Z. Gao, N. C. R. Chen, Y. Asakura, H. N. Nam, Q. M. Phung, Y. Kang, M. H. M. Leung, D. Jiang, L. Fu, L. Huang, T. Asahi and Y. Yamauchi, *Advanced Materials*, 2025, **37**, 2508105.
91. D. K. Sannes, S. Øien-Ødegaard, E. Aunan, A. Nova and U. Olsbye, *Chemistry of Materials*, 2023, **35**, 3793–3800.
92. C. Duan, Y. Yu, P. Yang, X. Zhang, F. Li, L. Li and H. Xi, *Industrial & Engineering Chemistry Research*, 2020, **59**, 774–782.
93. S. C. Moore, M. R. Smith, J. L. Trettin, R. A. Yang and M. L. Sarazen, *ACS Energy Letters*, 2023, **8**, 1397–1407.
94. W. Li, T. Zhang, L. Lv, Y. Chen, W. Tang and S. Tang, *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 2021, **624**, 126791.
95. T. V. F. da Silva, F. H. Silva, F. d. A. Luna, A. C. Marques, P. H. M. Andrade, A. A. Leitão, L. Tasic, M. A. L. de Oliveira and C. C. Corrêa, *Microporous and Mesoporous Materials*, 2026, **399**, 113828.

View Article Online
DOI: 10.1039/D5NR04817F

96. S. Abbasi, Z. Nezafat, S. Javanshir and B. Aghabarari, *Scientific Reports*, 2024, **14**, 14497.
97. S. H. Paiman, M. A. Rahman, T. Uchikoshi, N. Abdullah, M. H. D. Othman, J. Jaafar, K. H. Abas and A. F. Ismail, *Journal of Saudi Chemical Society*, 2020, **24**, 896–905.
98. H. Chu and C.-C. Wang, *Progress in Natural Science: Materials International*, 2023, **33**, 386–406.

View Article Online
DOI: 10.1039/D5NR04817F

Data availability

The machine-learning dataset used in this study has been provided as a machine-readable CSV file, including the synthesis variables reported in Table 2 and the measured response variables reported in Table 3. The repository also contains a data dictionary, the scripts used for the leave-one-out cross-validation audit, the corresponding LOOCV prediction/metric tables, and the GPR parameter files for the two primary optimization targets, BET surface area and MB removal. These materials are available at: https://github.com/ShayanAbaei/mil100fe-ml-ga-dye-removal