Supplementary Information (SI) for Nanoscale.
This journal is © The Royal Society of Chemistry 2026

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

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.

1. Methylene Blue Calibration Curve

Figure S1 showing methylene blue (MB) dye during adsorption experiments, a calibration curve was established by preparing standard MB solutions at known concentrations ranging from 1 to 5 mg/L. The absorbance of each solution was measured using UV-Vis spectrophotometry at 664 nm, the characteristic absorbance maximum for MB. A strong linear correlation between absorbance and concentration was observed, with a regression coefficient (R^2) greater than 0.99. This confirms the validity of the Beer–Lambert law for this concentration range. The resulting equation from this calibration curve was subsequently used to quantify MB concentrations in supernatants collected after adsorption experiments.

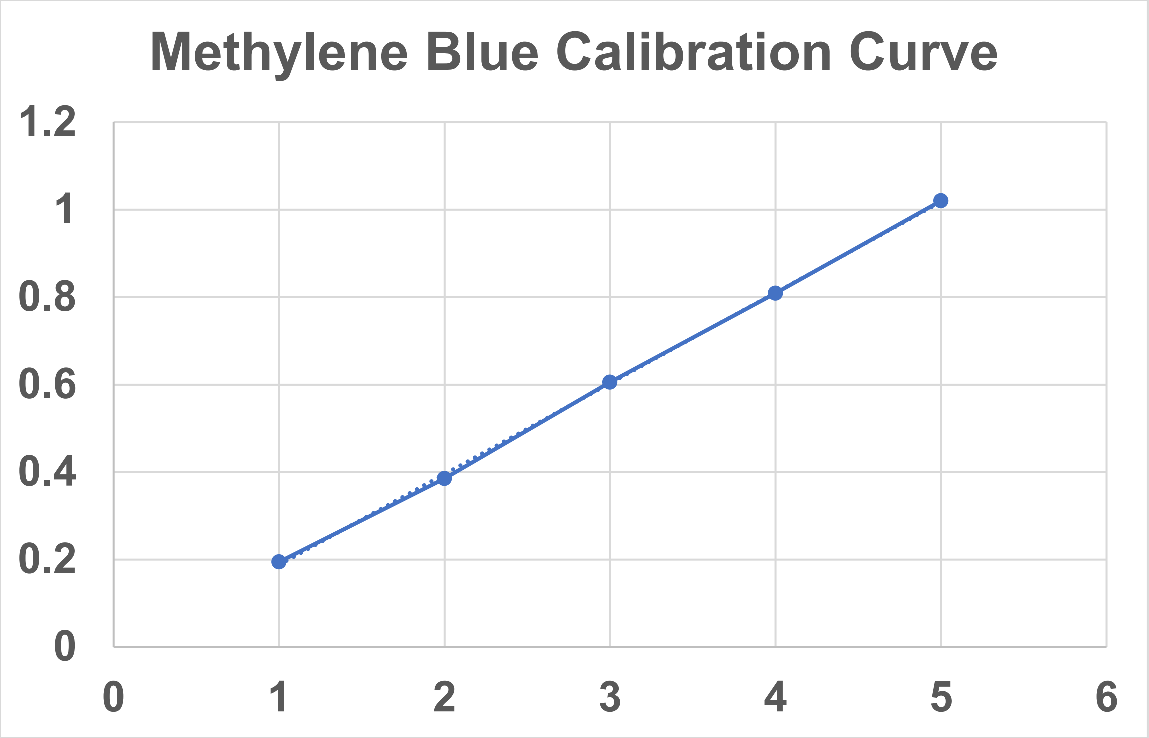

Figure S1. Calibration curve of methylene blue (MB) dye measured at 664 nm in the concentration range of 1–5 mg/L. The linearity of the curve ($R^2 > 0.99$) confirms the applicability of the Beer–Lambert law within this range and was used for all adsorption calculations.

2. Correlation Analysis: Pearson and Spearman Insights into Synthesis Parameters

To gain insights into the relationships between synthesis parameters and resultant material properties, we utilized both Pearson Correlation Coefficients (PCCs) and Spearman Rank Correlation (SRC) analyses.[1, 2] **Figure S2** shows the SRC matrix, capturing non-linear relationships. Together, these analyses provide a comprehensive understanding of how synthesis conditions impact MIL-100(Fe) properties.

To investigate the connections between raw data and different targets, we utilize Pearson Correlation Coefficients (PCCs). The Pearson correlation coefficient is used to determine the strength of a linear relationship between two data sets. It is mathematically represented as:

$$PCC = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \qquad (1)$$

In this context, x and y are the variables under study, while $\bar{x}$ and $\bar{y}$ are their respective means. The value of the Pearson Correlation Coefficient (PCC) can range from -1 to +1. A PCC value above 0 indicates a positive correlation, suggesting that as one variable increases, the other also tends to increase. Conversely, a PCC value below 0 denotes a negative correlation, where an increase in one variable typically leads to a decrease in the other. The correlation is considered stronger as the absolute value of the PCC approaches 1.[2, 3]

Figure 4 (in the main text) presents the Pearson Correlation Coefficient (PCC) matrix, which illustrates the strength and direction of the linear relationships between different synthesis parameters and the resultant properties of MIL-100(Fe). This matrix provides valuable insights into the effect of each parameter on key performance metrics such as surface area, methylene blue (MB) removal efficiency, and total pore volume. The positive or negative values indicate the direction of the relationship. For example, the matrix highlights a moderate positive correlation between synthesis time and MB removal efficiency (PCC = 0.40), indicating that longer synthesis times generally improve dye removal performance. Temperature, on the other hand, shows a strong negative correlation with yield (PCC = -0.64), suggesting that higher synthesis temperatures may result in lower material yield, likely due to structural decomposition or loss of organic ligands. Additionally, the matrix reveals that crystallinity is significantly influenced by both ion concentration (PCC = 0.63) and temperature (PCC = -0.35), emphasizing that these parameters play crucial roles in determining the crystalline nature of the material. The lack of high intercorrelation between features (e.g., temperature, molar ratio, ion concentration) is an important observation, as it implies that each feature independently contributes to material property predictions without substantial redundancy. These insights are crucial for optimizing the synthesis conditions to target specific performance metrics effectively, such as achieving higher adsorption capacities or improving crystallinity. The PCC matrix serves as a foundation for selecting important features and understanding their direct effects on performance metrics, facilitating targeted modifications during synthesis to achieve desired properties. However, it is essential to note that PCC only captures linear relationships. To understand more complex, nonlinear interactions, further analyses such as feature importance via machine learning models and Spearman rank correlations are also required.

In addition to the Pearson analysis, Spearman rank correlation was used as a complementary exploratory check for monotonic trends (**Figure S2**). As with the Pearson matrix, these results should be interpreted cautiously because of the small sample size and because the synthesis variables were defined by experimental design. Accordingly, the SRC matrix is used only to visualize broad monotonic tendencies between synthesis conditions and measured responses, rather than to support formal statistical inference or mechanistic interpretation of input–input correlations. [1-4]

Figure S2 presents the Spearman's Rank Correlation (SRC) matrix for the synthesis of MIL-100(Fe). Unlike the Pearson Correlation Coefficient (PCC) matrix, which only captures linear relationships, Spearman's rank correlation is a non-parametric analysis that measures the strength and direction of monotonic relationships between variables.[4] The matrix uses bubble size and color intensity to represent the strength of the correlation, with blue indicating a positive correlation and red indicating a negative correlation. The SRC analysis is valuable for understanding non-linear dependencies between synthesis conditions and resultant properties. For instance, the SRC analysis shows that temperature has a relatively strong negative monotonic relationship with yield (represented by larger red bubbles), which aligns with the PCC findings but is more pronounced in terms of monotonic relationships.[4] Additionally, the influence of time on MB removal and crystallite size is also observed as moderately positive, indicating that longer synthesis times tend to enhance these properties even when the relationship is not strictly linear.[3, 4]

Both the Pearson and Spearman correlation analyses offer different perspectives on the relationships between synthesis parameters and material properties. The PCC matrix primarily measures linear associations, providing a clear indication of how a change in one variable linearly affects another.[1-3] In contrast, the SRC matrix assesses monotonic relationships, which include both linear and non-linear patterns, offering a more comprehensive picture of the interconnections between synthesis variables and outcomes.[1-3] One notable difference between the PCC and SRC analyses is observed in the relationship between temperature and surface area. The PCC suggests a weak negative linear correlation, while the SRC indicates a slightly stronger non-linear negative correlation, emphasizing the complex impact of temperature on surface area that may not be purely linear.[3] Additionally, while both analyses indicate that ion concentration plays a relatively minor role in determining most material properties, the SRC matrix emphasizes the monotonic trends, which could be valuable in situations where linear trends are insufficient to capture the complexity of the underlying chemistry.[3] Overall, the SRC analysis complements the PCC by highlighting non-linear relationships, providing a more robust understanding of how synthesis conditions influence material performance. Together, these analyses are essential for optimizing synthesis conditions, as they identify both direct and complex interdependencies among features, guiding efforts towards the rational design of MIL-100(Fe) for improved functionality.[1-3]

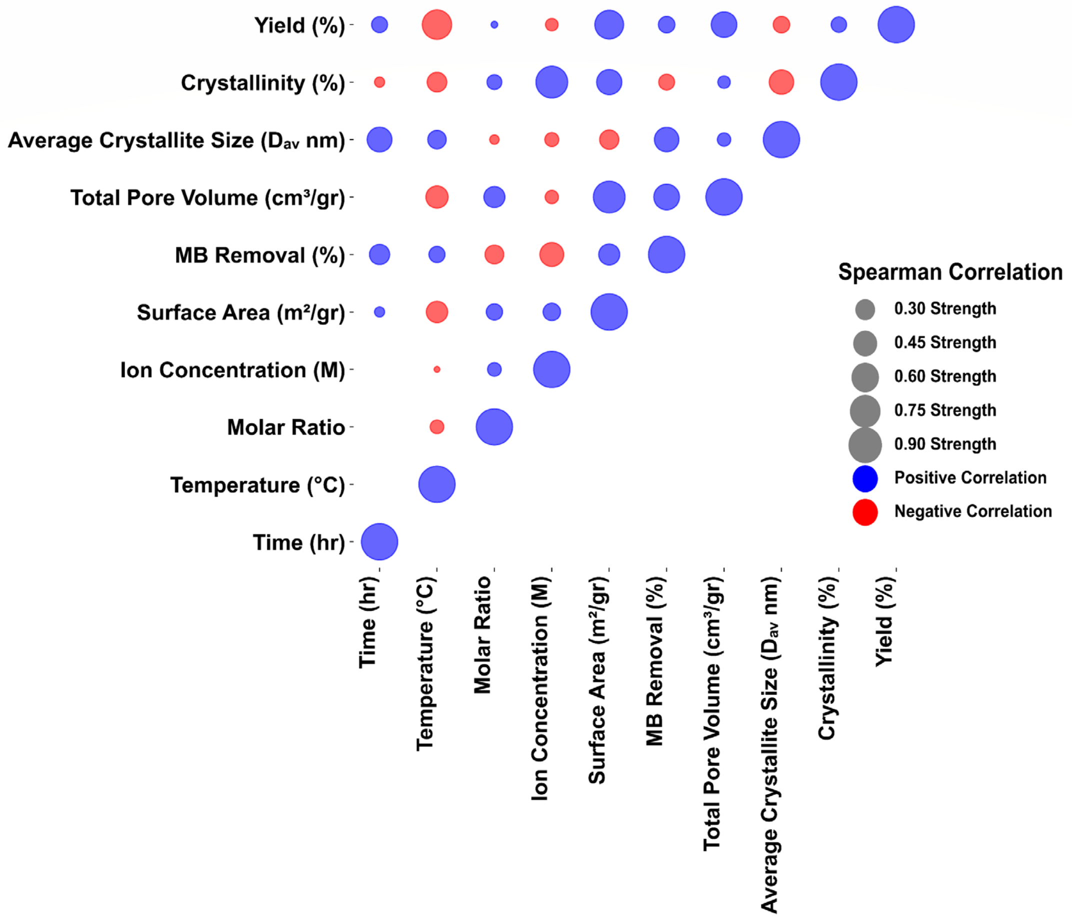

Figure S2. Spearman's Rank Correlation Matrix for All Features and Targets in the synthesis of MIL-100(Fe). This matrix represents the Spearman correlation coefficients between synthesis parameters (Time, Temperature, Molar Ratio, Ion Concentration) and material properties such as Surface Area, MB Removal Efficiency, Total Pore Volume, Average Crystallite Size, Crystallinity, and Yield.

3. Characterization Techniques

3.1 FTIR Spectra: Identification of Functional Groups in Synthesized MIL-100(Fe)

The Fourier Transform Infrared (FTIR) spectra presented in **Figure S3** provide significant insights into the chemical structure and functional groups present in the synthesized MIL-100(Fe) samples. The observed spectra indicate characteristic functional groups of MIL-100(Fe), with noticeable differences in peak intensities and positions reflecting the effects of varying synthesis conditions. The broad peak around 3400 cm^{-1} corresponds to the OH stretching vibrations, indicative of surface hydroxyl groups or adsorbed moisture. Variations in the intensity of this peak across different samples suggest differences in water adsorption or hydroxyl group availability, likely influenced by the synthesis parameters such as temperature or the metal-to-ligand molar ratio. The peak near 1650 cm^{-1} is attributed to C=O stretching vibrations from the carboxylate groups of the trimesic acid ligand, confirming successful coordination between the metal centers and the organic linker in MIL-100(Fe). Differences in the intensity of this peak may indicate variations in crystallinity or bonding efficiency, which could impact framework stability and adsorption performance. Further, peaks in the range of 1200 to 1600 cm^{-1} correspond to C-O and C=C stretching vibrations typical of aromatic carboxylates. Shifts or variations in these peaks suggest changes in bonding configurations and the presence of defects, which could have significant implications for the adsorption properties of the material. Additionally, peaks below 600 cm^{-1} are associated with Fe-O stretching vibrations, confirming the formation of the metal-oxide cluster characteristic of MIL-100(Fe). These peaks are consistently observed across all samples, indicating the preservation of the fundamental Fe-O framework, although differences in peak intensity may point to variations in the quantity of iron oxide impurities. The analysis of these FTIR spectra reveals that synthesis conditions significantly affect the chemical composition of MIL-100(Fe), particularly in terms of functional group availability and bonding environment. For example, changes in the OH and C=O peaks reflect the influence of synthesis on surface chemistry, which is crucial for adsorption efficiency in dye removal applications. The presence of functional groups such as OH and C=O suggests potential adsorption sites for dye molecules, with increased functional group availability likely enhancing the interaction with dye molecules, thereby improving the removal efficiency. Overall, the FTIR analysis provides critical insights into how synthesis parameters affect the surface chemistry and structural features of MIL-100(Fe). Understanding these relationships is essential for optimizing synthesis conditions to achieve enhanced adsorption properties, specifically for targeted applications such as pollutant removal. The variations observed in the spectra illustrate the complex interplay between synthesis conditions and material properties, highlighting the importance of a tailored approach to synthesis in order to maximize functional performance.

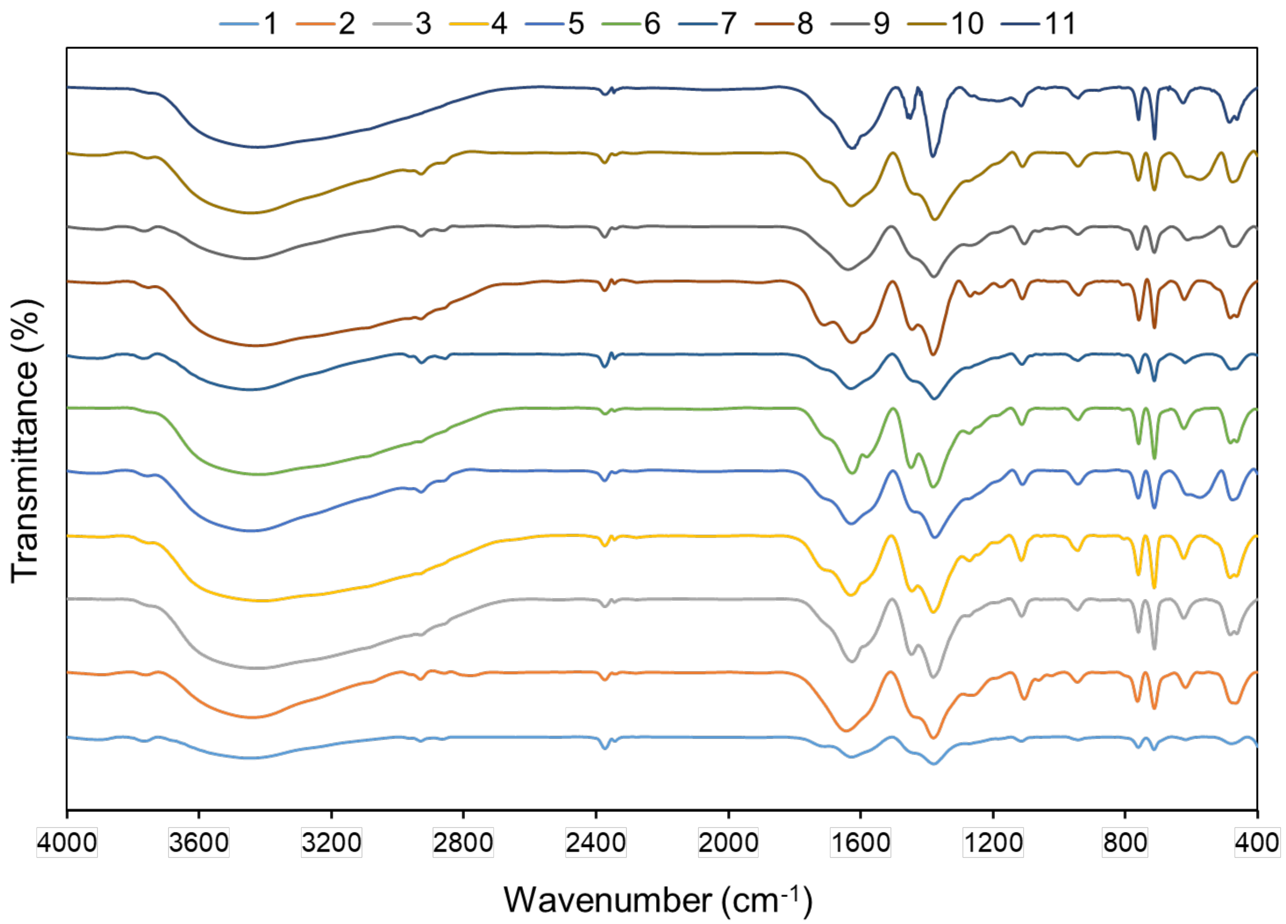

Figure S3. FTIR Spectra of Synthesized MIL-100(Fe) Samples under Different Synthesis Conditions. The spectra show characteristic peaks indicating functional groups such as OH stretching (broad peak around 3400 cm^{-1}), C=O stretching (around 1650 cm^{-1}), and Fe-O stretching (below 600 cm^{-1}), highlighting variations in the chemical composition of samples synthesized under different conditions.

3.2 Thermal Stability Analysis via Thermogravimetric Analysis (TGA)

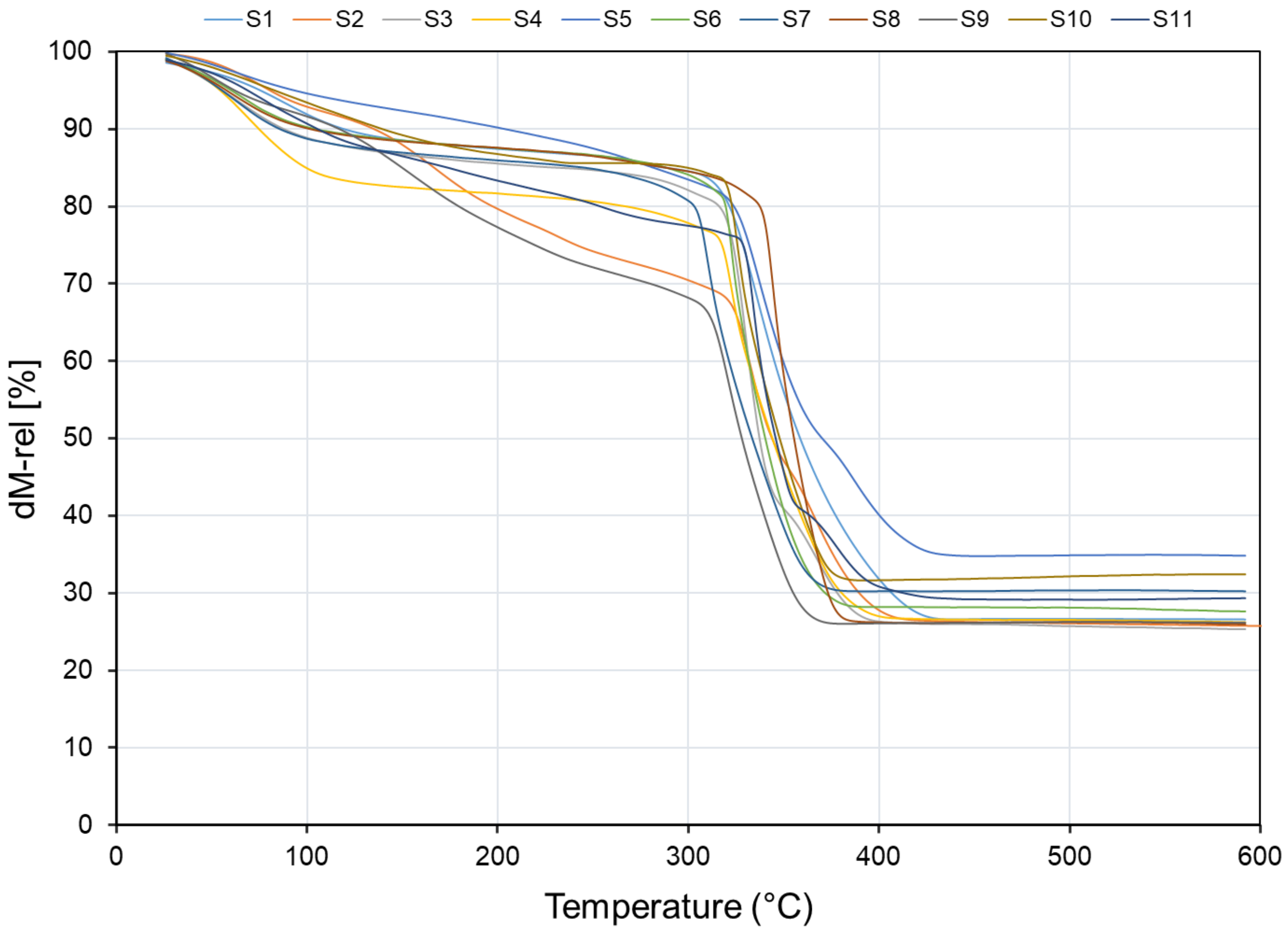

Figure S4. Thermogravimetric analysis (TGA) curves for the synthesized MIL-100(Fe) samples (S1 to S11), depicting the mass loss percentage as a function of temperature. The analysis provides insights into the thermal stability and decomposition behavior of the samples under increasing temperature conditions.

Figure S4 presents the thermogravimetric analysis (TGA) curves for the eleven synthesized MIL-100(Fe) samples (S1 to S11). The mass loss as a function of temperature provides valuable insights into the thermal stability and decomposition behavior of the materials. Initially, a gradual weight loss is observed up to approximately 110 °C, which can be attributed to the evaporation of adsorbed water molecules and the release of moisture from the porous network. This mass loss, occurring in the range of 5-15%, is consistent across all samples, indicating similar moisture absorption characteristics. As the temperature increases beyond 110 °C, a more significant mass loss is noted between 300 °C and 400 °C, which corresponds to the thermal decomposition of the organic ligand (trimesic acid) and the subsequent collapse of the MOF structure. Differences in the onset and extent of mass loss within this temperature range suggest variability in the thermal robustness and crystallinity of the synthesized MIL-100(Fe) samples. Samples such as S3 and S7 appear to exhibit greater stability, as indicated by a delayed and less pronounced decomposition event compared to other samples. Beyond 400 °C, the remaining mass becomes relatively stable, which implies that the decomposition of the organic components is largely complete, and only the stable iron-oxide residues remain. The differences in residual mass among samples reflect variations in the degree of ligand coordination and the presence of iron oxide impurities. For instance, higher residual mass may indicate incomplete decomposition or a higher content of iron-based components. Overall, TGA reveals that synthesis conditions, such as temperature and time, directly influence the thermal stability and decomposition profile of MIL-100(Fe). The thermal decomposition behavior indicates that the material is suitable for applications involving temperatures below 300 °C, beyond which significant structural degradation occurs. This information is crucial for determining optimal operating conditions for MIL-100(Fe) in potential applications, such as adsorption or catalysis, where thermal stability plays a significant role in material performance.

3.3 SEM Morphology Analysis of MIL-100(Fe) Sample 9

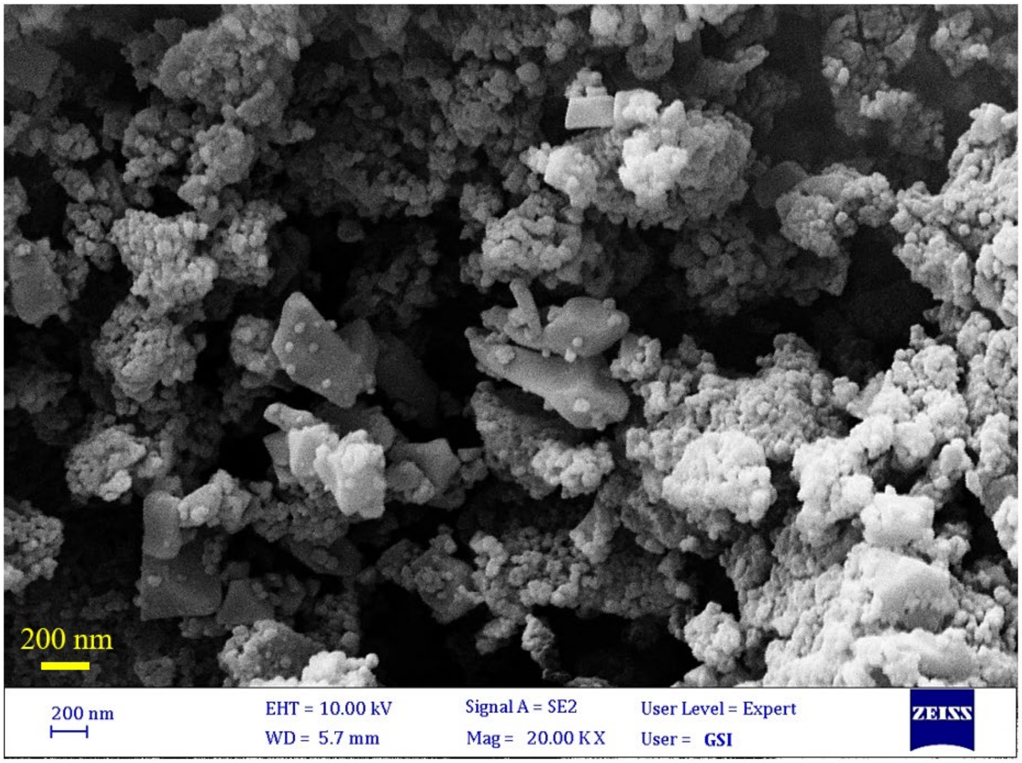

Figure S5. Scanning Electron Microscopy (SEM) image of synthesized MIL-100(Fe) sample S9, showing the microstructure at a magnification of 20,000x. The scale bar represents 200 nm, providing insights into the particle morphology and aggregation of the sample.

Figure S5 displays the SEM image of the synthesized MIL-100(Fe) sample S9, offering a detailed view of its morphology at the nanoscale. The image reveals a heterogeneous microstructure characterized by a mixture of irregular, agglomerated particles. The clusters and aggregates observed in the image suggest significant particle adhesion, likely due to Van der Waals forces or incomplete washing during synthesis, which might hinder accessibility to internal pores. The crystal structures appear predominantly as rough and irregularly shaped agglomerates, with occasional larger particles present. This non-uniformity in particle size could affect the material's adsorption properties by limiting the overall surface area accessible for dye removal or catalytic reactions. The SEM image further suggests that the synthesis parameters used for sample S9 might have led to a less desirable aggregation, which impacts the material's application efficiency. For adsorption-based applications, excessive aggregation can impede the diffusion of adsorbate molecules into the internal pores, leading to reduced performance. In contrast, smaller, well-dispersed particles would enhance surface accessibility, ultimately improving the material's adsorption efficiency.

These morphological observations from SEM are essential for understanding the physical structure and its impact on the overall performance of the material, particularly for applications requiring high surface area and easy access to active sites. Therefore, optimization of synthesis parameters, such as reaction temperature and mixing conditions, might be necessary to achieve a more favorable morphology, reducing particle aggregation and enhancing the performance of MIL-100(Fe) for specific applications.

3.4 Detailed adsorption–desorption isotherms of each 11 samples

To complement **Figure 3D** of the main text, **Figure S6** provides the full nitrogen adsorption–desorption isotherms for all 11 synthesized MIL-100(Fe) samples, complementing the BET surface area data summarized in the main text. The isotherms exhibit the expected Type I and Type IV profiles, with a steep uptake at low relative pressures ($p/p_0 < 0.1$), confirming the presence of well-developed microporosity, and a gradual increase at intermediate pressures ($0.1 < p/p_0 < 0.8$), indicative of mesoporosity or interparticle voids. A direct comparison reveals that samples with the highest BET surface areas—11 (1748 $m^2\ g^{-1}$), 10 (1423.2 $m^2\ g^{-1}$), and 4 (1436.9 $m^2\ g^{-1}$)—show markedly higher nitrogen uptake across the entire pressure range, signifying a robust microporous network and efficient pore accessibility. In contrast, samples such as 1 (499.5 $m^2\ g^{-1}$) and 8 (389.1 $m^2\ g^{-1}$) demonstrate significantly lower uptake, suggesting underdeveloped porosity, reduced surface area, or incomplete crystallization. These detailed isotherms highlight the influence of synthesis parameters—such as reaction temperature, time, and Fe:ligand ratio—on the final pore structure, providing a deeper understanding of textural variations across the MIL-100(Fe) series.

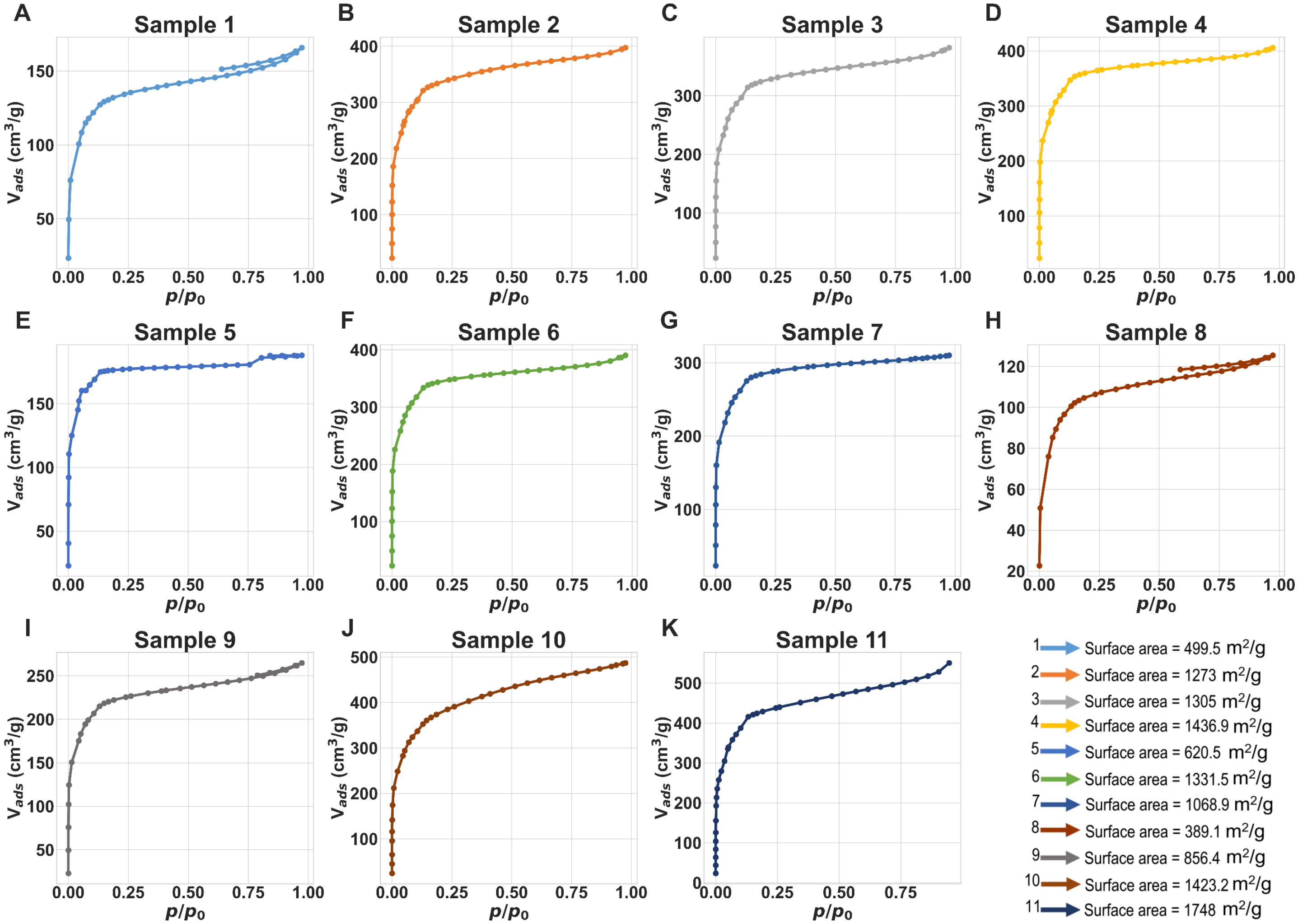

Figure S6. Nitrogen adsorption–desorption isotherms of 11 MIL-100(Fe) samples, showing characteristic Type I and Type IV features. Samples with higher BET surface areas (e.g., 11: 1748 $m^2\ g^{-1}$, 10: 1423.3 $m^2\ g^{-1}$, 4: 1436.9 $m^2\ g^{-1}$) exhibit significantly larger N_2 uptake, whereas samples with lower BET surface areas (e.g., 1: 499.5 $m^2\ g^{-1}$, 8: 389.1 $m^2\ g^{-1}$) demonstrate reduced adsorption capacity, reflecting differences in pore architecture and crystallinity.

Table S1 provides a comparative analysis of the predictive performance of several machine learning models for MIL-100(Fe) synthesis. The Root Mean Square Error (RMSE) values indicate how well each model captures the relationships between synthesis parameters and material properties. Among all models, the Gaussian Process Regressor (GPR) consistently achieves the lowest RMSE values for Surface Area (315.71 m^2/g), MB Removal Efficiency (11.18%), Total Pore Volume (0.1606 cm^3/g), and other key metrics. This highlights GPR's robustness and suitability as a surrogate model for further optimization and analysis of synthesis conditions.

Table S1. Root Mean Square Error (RMSE) values for various machine learning models used to predict the properties of MIL-100(Fe). Models are compared based on their performance in predicting surface area, methylene blue removal efficiency, total pore volume, average crystallite size, crystallinity, and yield.

Model	Surface Area (m^2/g)	MB Removal (%)	Total Pore Volume (cm^3/g)	Average Crystallite Size (nm)	Crystallinity (%)	Yield (%)
Gaussian Process Regressor	315.71	11.18	0.1606	3.20	5.74	6.76
Support Vector Regressor	322.65	12.01	0.1794	3.76	6.40	6.93
Decision Tree Regressor	334.05	12.17	0.1795	3.80	6.45	8.07
Random Forest Regressor	354.77	12.71	0.1860	3.83	6.55	8.40
K-Nearest Neighbors	392.93	13.41	0.1880	4.10	6.81	8.61
Dummy Regressor	394.67	13.92	0.1930	4.41	7.64	8.86
Linear SVR	402.73	16.51	0.2018	5.47	7.93	9.00
XGBoost	412.73	17.95	0.2018	5.57	8.09	11.24
Linear Regression	598.99	18.46	0.2996	6.35	8.10	13.79

Compared to other models, the Support Vector Regressor (SVR) and Decision Tree Regressor also show reasonable predictive accuracy but are not as consistent across all metrics as GPR. Models like the Dummy Regressor and Linear Regression show significantly higher RMSE values, indicating poor predictive capabilities and a limited ability to capture complex relationships. The high RMSE values for Linear Regression suggest that linear approaches are insufficient for accurately modeling the non-linear dependencies inherent in the synthesis of MIL-100(Fe). The Random Forest Regressor and XGBoost perform relatively better for some properties, like yield and crystallinity, but still fall short in terms of consistency when compared to GPR. These findings underscore the importance of selecting appropriate models for material property prediction—highlighting GPR's capacity to effectively capture the non-linear relationships and uncertainty often found in experimental chemical research. Ultimately, GPR emerges as the optimal model due to its balanced accuracy and ability to handle non-linearity, making it a preferred choice for tasks like feature importance analysis and synthesis condition optimization. The insights from this model will help guide the efficient synthesis of MIL-100(Fe) with tailored properties for specific applications.

4. Visualizations

4.1 Contour Plots: Effect of Synthesis Parameters on Material Performance Metrics

These contour plots**, Figures S7 to S10** depict the relationships between the molar ratio and ion concentration on different properties of MIL-100(Fe), for surface area, MB removal efficiency, total pore volume, average crystallite size, crystallinity, and yield. Each plot visually represents how synthesis parameters influence specific characteristics. The variations in colors and contour gradients indicate the effects of these parameters in maximizing or minimizing the values of key material properties. For instance, changes in the molar ratio and ion concentration distinctly affect the surface area and crystallinity, pointing out critical regions for parameter optimization. The aim is to understand how these synthesis variables interact and how they could be adjusted for optimal performance in various applications.

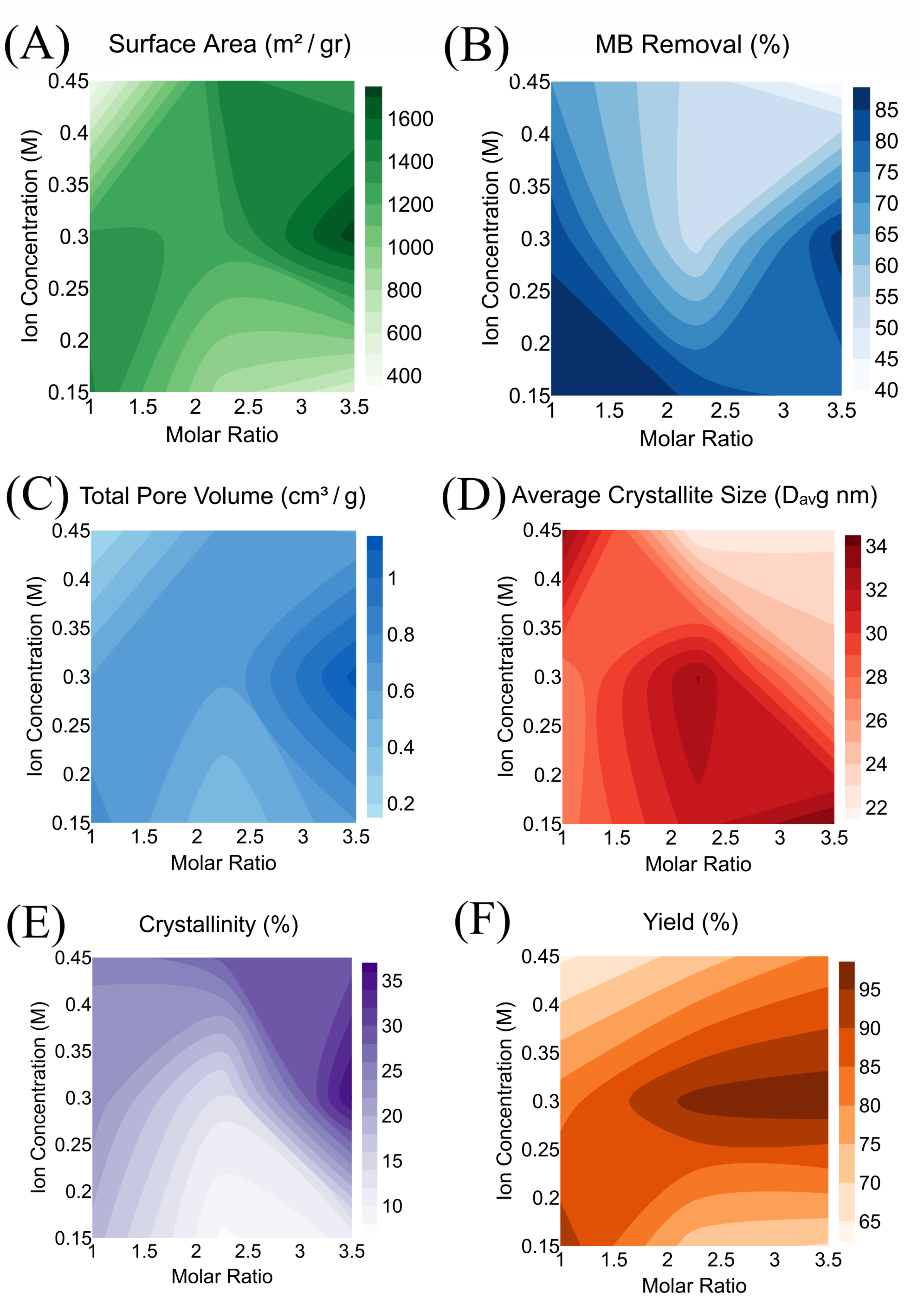

Figure S7. Contour plots depicting the relationship between synthesis parameters (Ion Concentration and Molar Ratio) and various performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots illustrate how different combinations of synthesis time and temperature influence each target property, highlighting the intricate trade-offs required to optimize material performance.

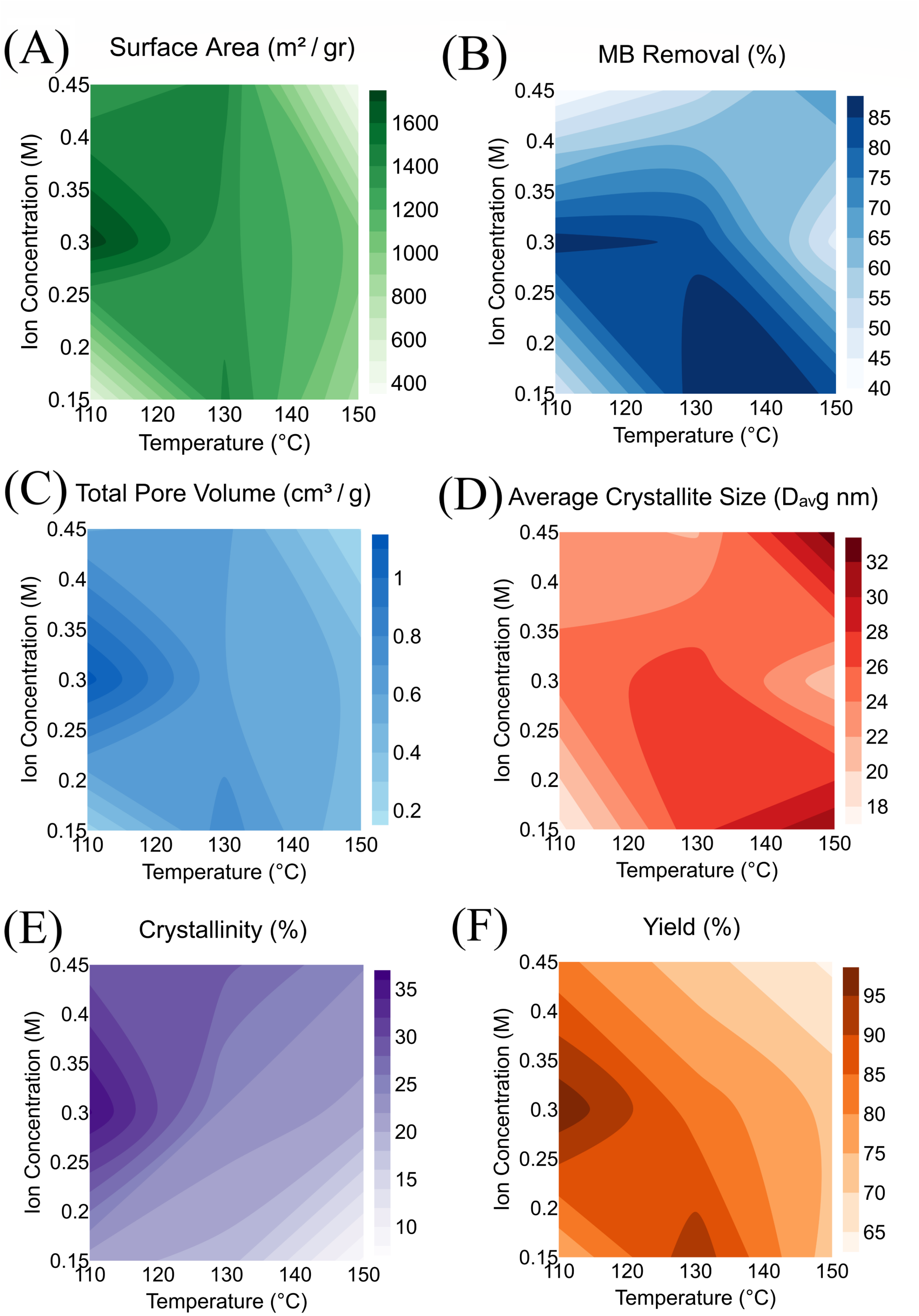

Figure S8. Contour plots depicting the relationship between synthesis parameters (Ion Concentration and Temperature) and various performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots illustrate how different combinations of synthesis time and temperature influence each target property, highlighting the intricate trade-offs required to optimize material performance.

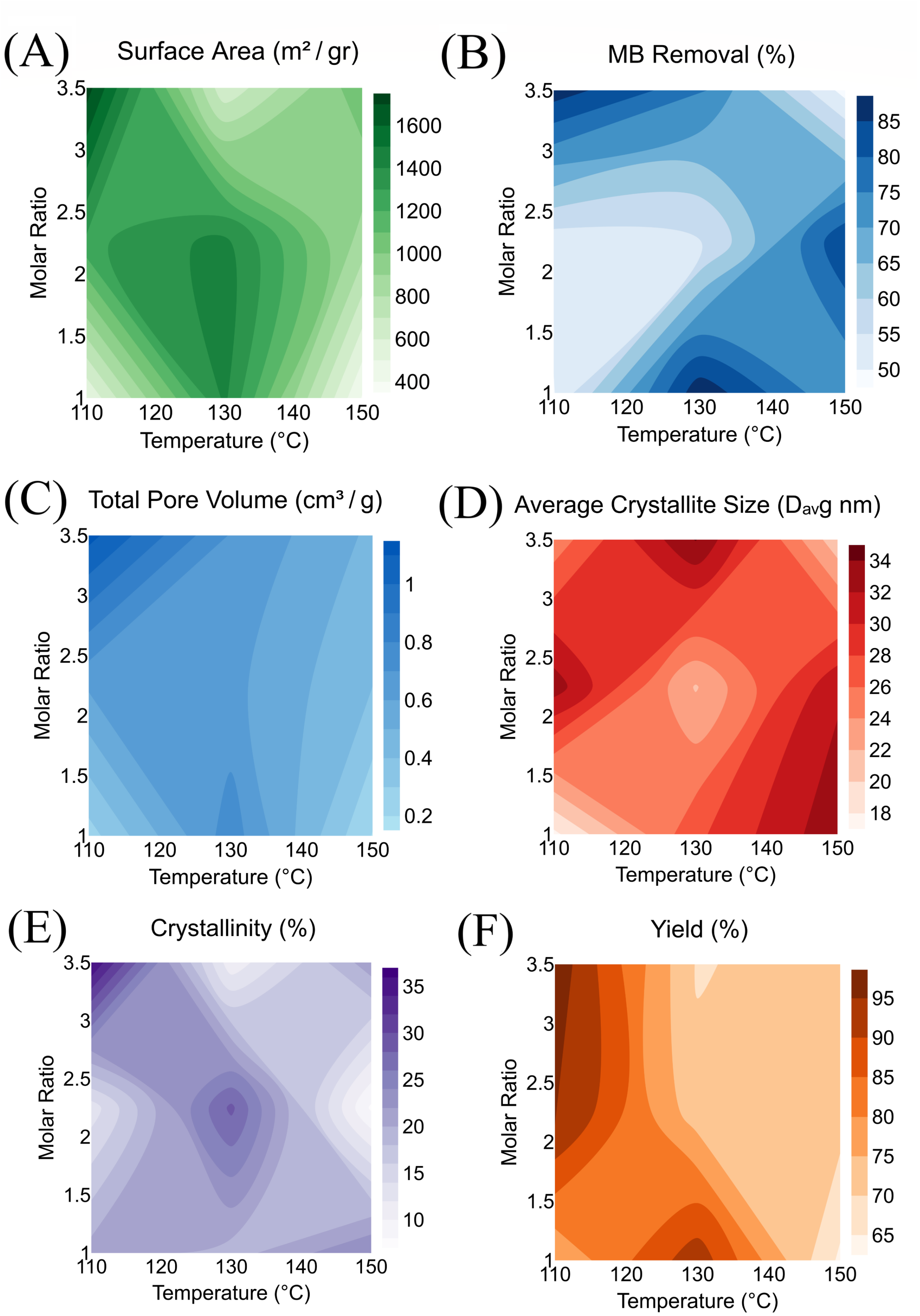

Figure S9. Contour plots depicting the relationship between synthesis parameters (Molar Ratio and Temperature) and various performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots illustrate how different combinations of synthesis time and temperature influence each target property, highlighting the intricate trade-offs required to optimize material performance.

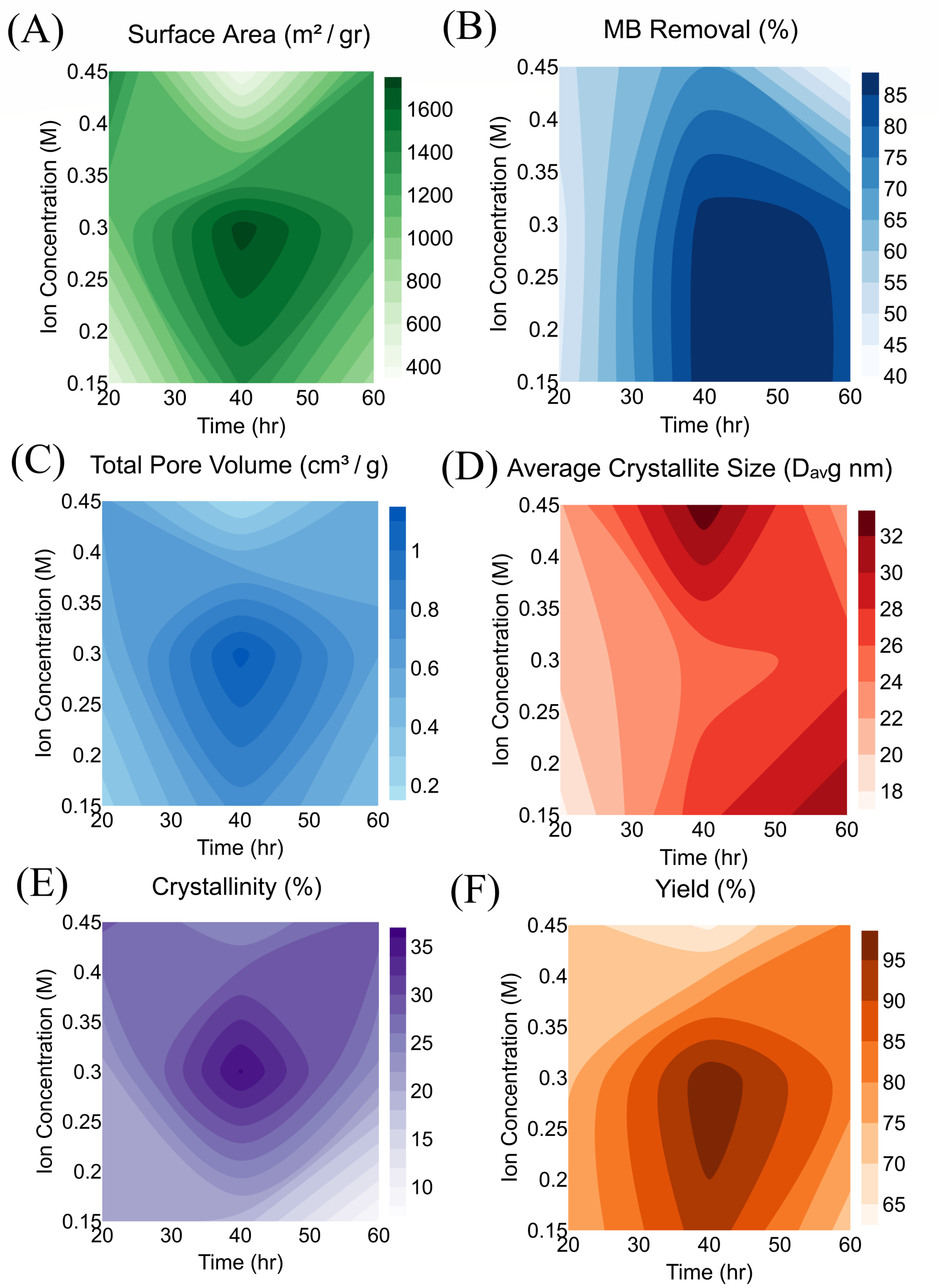

Figure S10. Contour plots depicting the relationship between synthesis parameters (Ion Concentration and Time) and various performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots illustrate how different combinations of synthesis time and temperature influence each target property, highlighting the intricate trade-offs required to optimize material performance.

4.2 3D Surface Plots: Multi-Dimensional Effects on MIL-100(Fe) Properties

The three-dimensional surface plots, **Figures S11 to S15** illustrate the relationships among temperature, molar ratio, and key properties of MIL-100(Fe). These plots provide a more comprehensive view of how multiple parameters simultaneously impact material characteristics, offering insights that are harder to observe in two-dimensional representations. The complex surfaces indicate how temperature and molar ratio interact to influence outcomes like surface area, MB removal efficiency, and yield. Notably, these 3D plots enable the visualization of peaks and valleys, which help in identifying optimal parameter combinations. They emphasize the importance of multi-dimensional tuning to maximize adsorption capacity, pore volume, and other critical properties of MIL-100(Fe) for targeted applications.

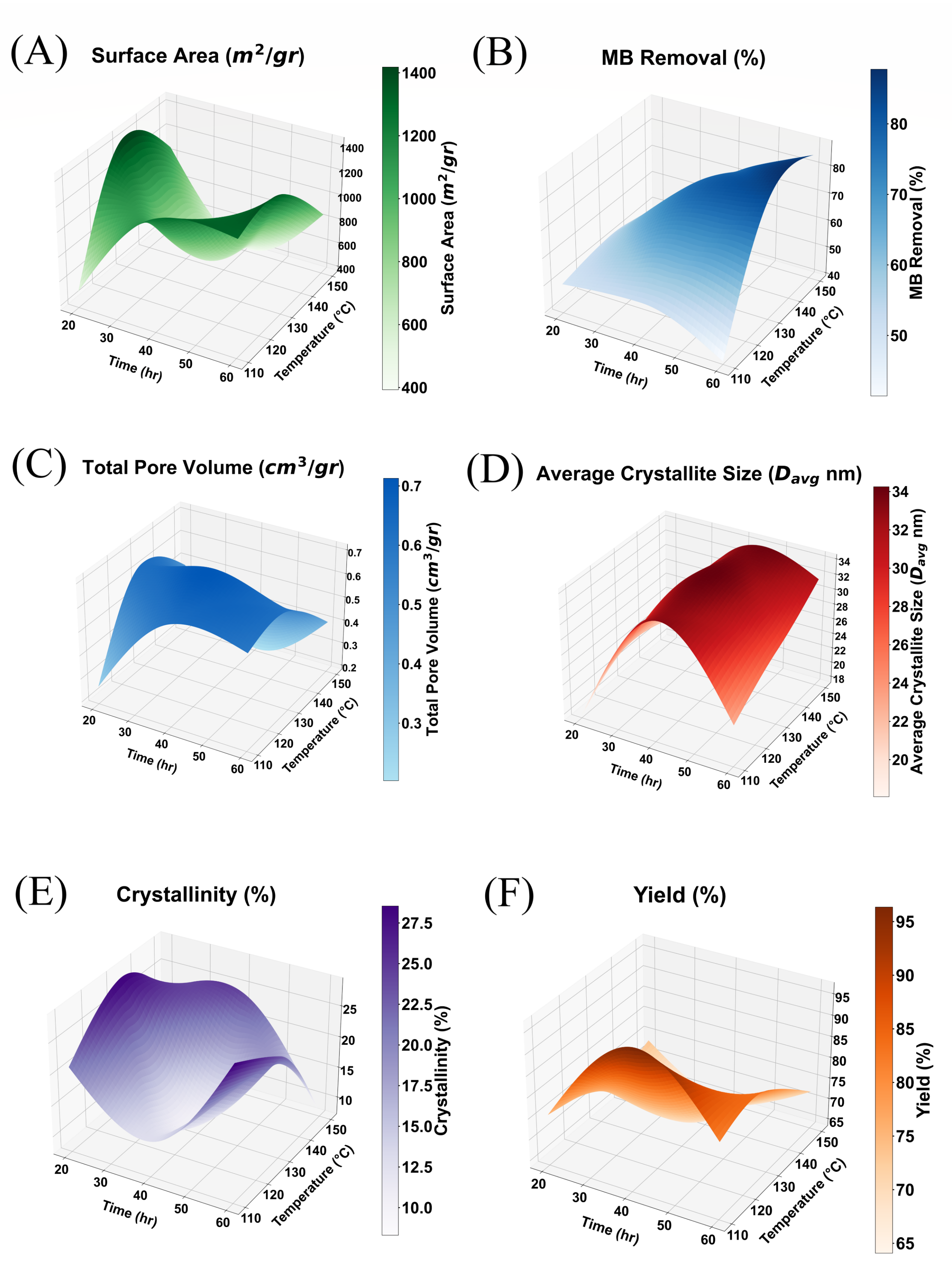

Figure S11. 3D surface plots showing the influence of synthesis parameters (Time and Temperature) on the performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots reveal the complex and non-linear interactions between synthesis parameters, providing insight into how different conditions affect each target property of the material.

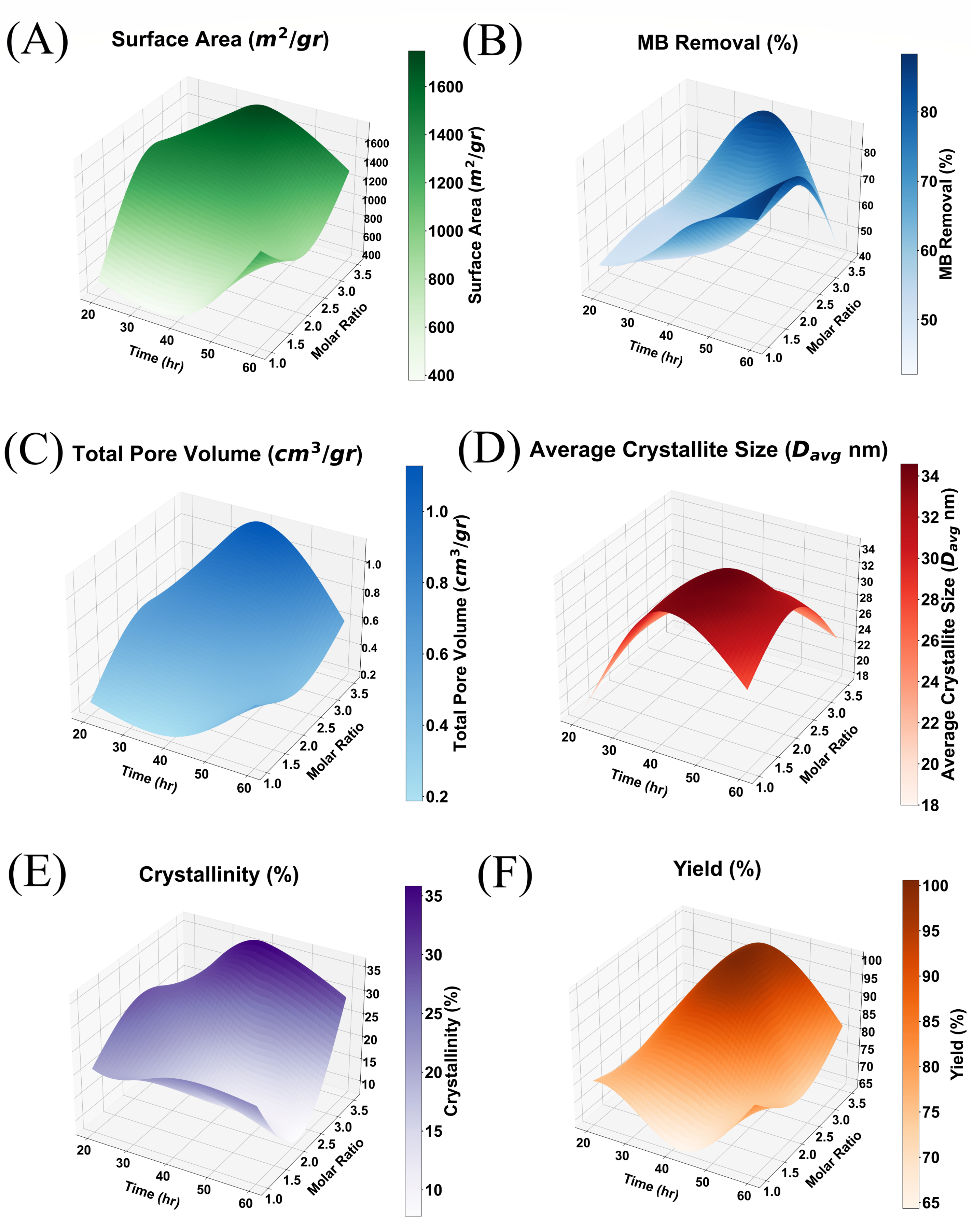

Figure S12. 3D surface plots showing the influence of synthesis parameters (Time and Molar Ratio) on the performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots reveal the complex and non-linear interactions between synthesis parameters, providing insight into how different conditions affect each target property of the material.

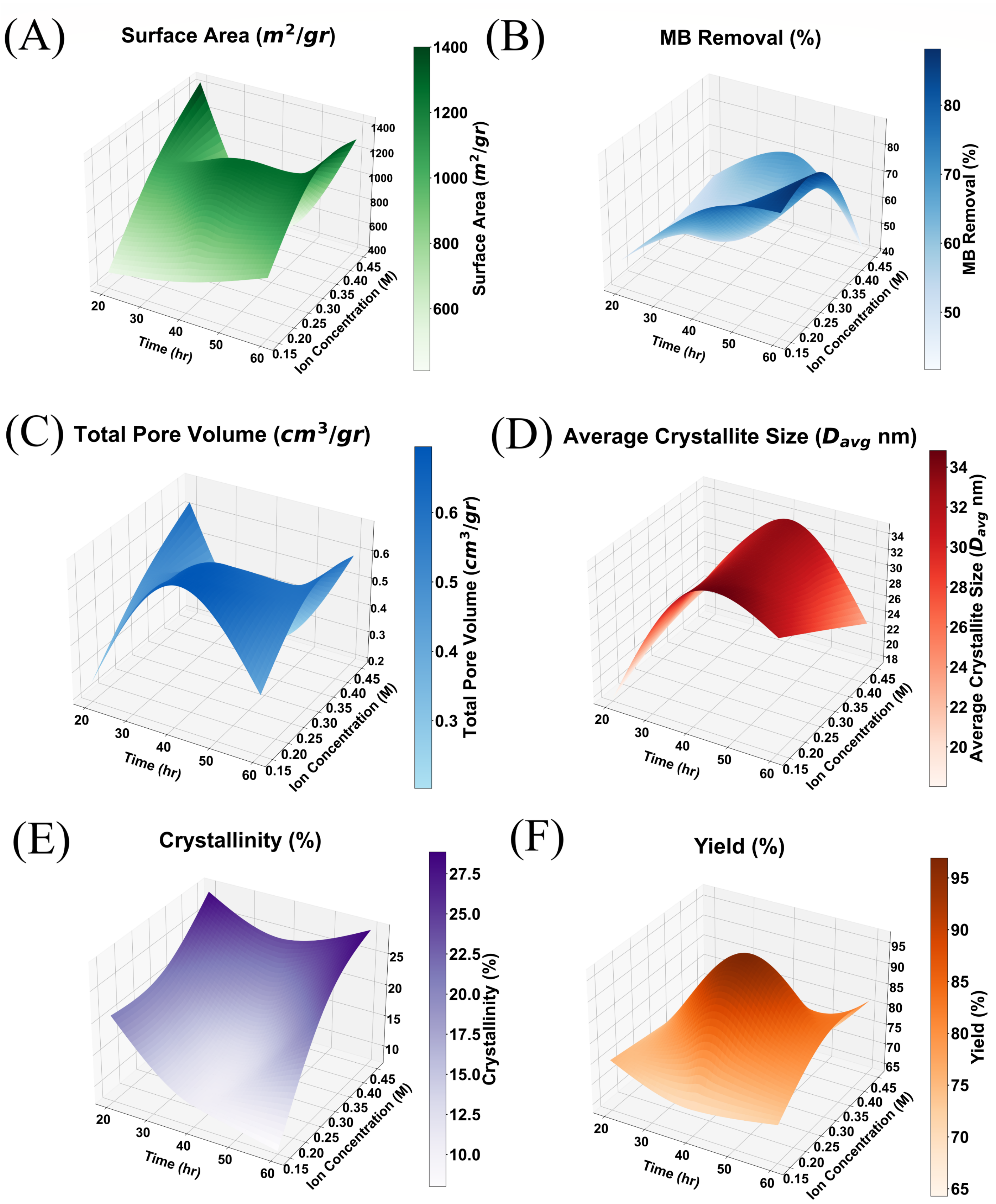

Figure S13. 3D surface plots showing the influence of synthesis parameters (Time and Ion Concentration) on the performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots reveal the complex and non-linear interactions between synthesis parameters, providing insight into how different conditions affect each target property of the material.

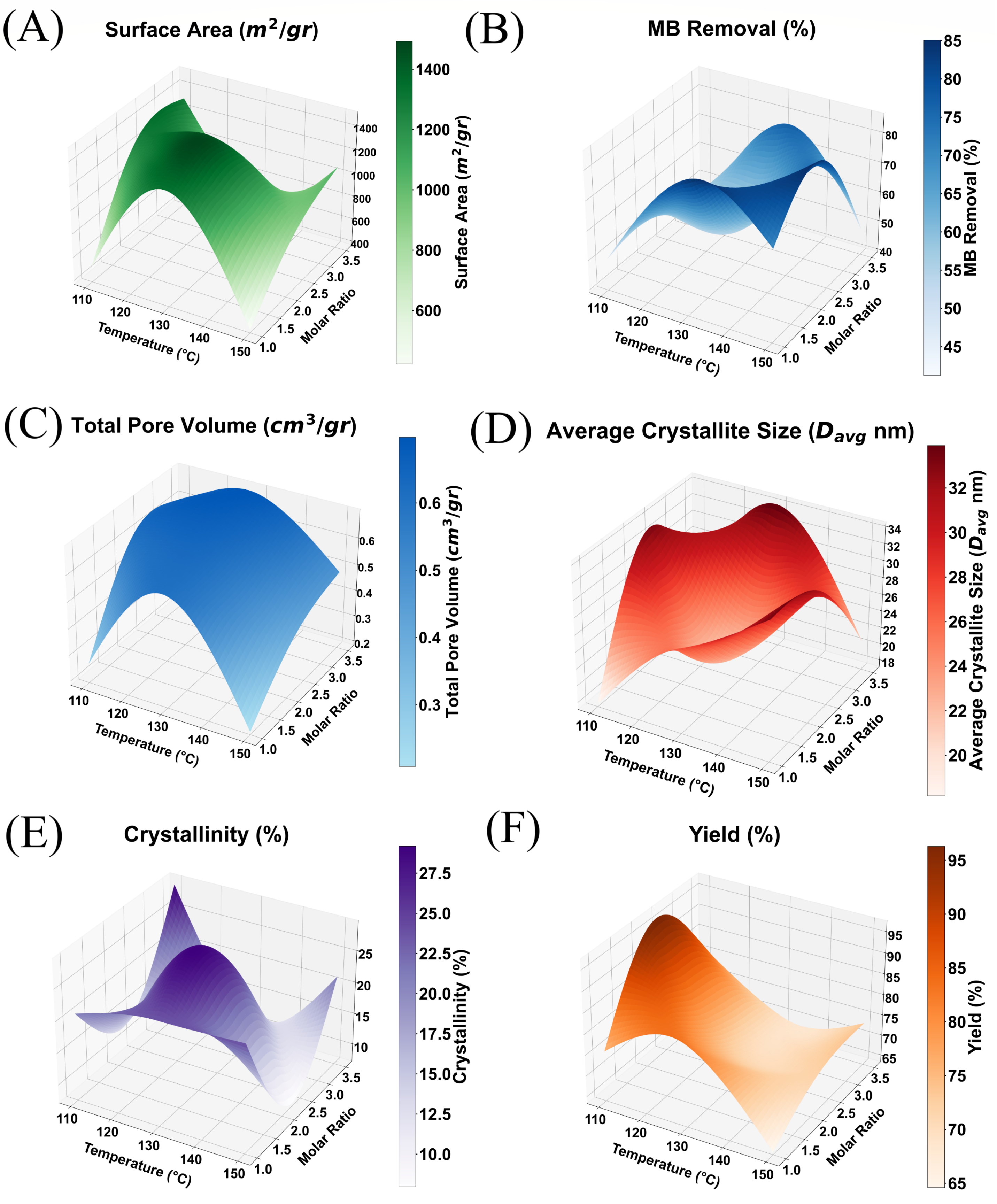

Figure S14. 3D surface plots showing the influence of synthesis parameters (Temperature and Molar Ratio) on the performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots reveal the complex and non-linear interactions between synthesis parameters, providing insight into how different conditions affect each target property of the material.

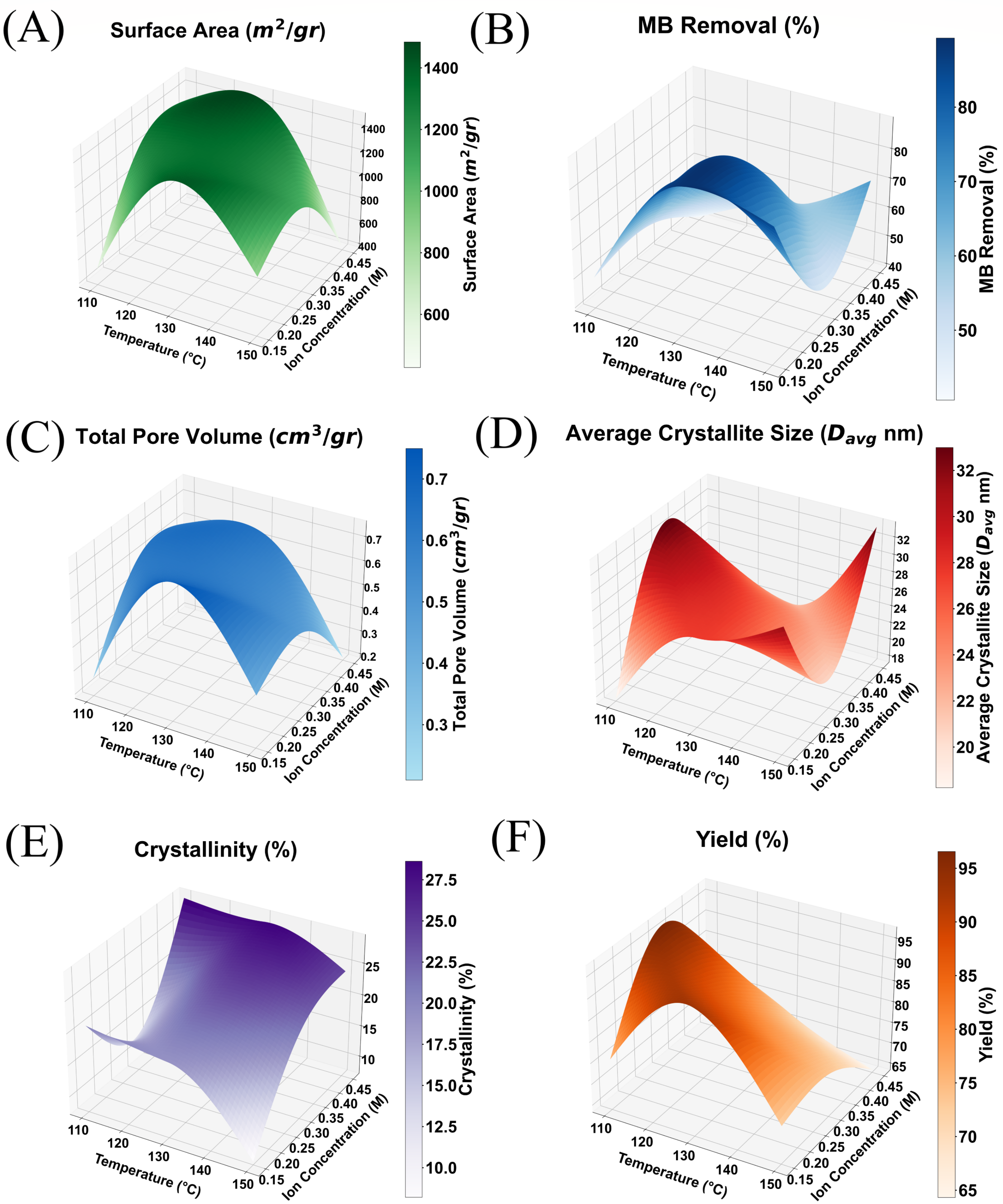

Figure S15. 3D surface plots showing the influence of synthesis parameters (Temperature and Ion Concentration) on the performance metrics of MIL-100(Fe): (A) Surface Area, (B) MB Removal Efficiency, (C) Total Pore Volume, (D) Average Crystallite Size, (E) Crystallinity, and (F) Yield. These plots reveal the complex and non-linear interactions between synthesis parameters, providing insight into how different conditions affect each target property of the material.

4.3 Partial Dependence Plots (PDPs) for Parameter Interactions

The two-variable surface plots shown in the main text (**Figures 6–8**) are experimental-data-based descriptive visualizations intended to summarize broad empirical trends within selected two-variable projections of the dataset. They are not GP models retrained using only the displayed variables. In contrast, Figures S16–S18 present the GPR-derived partial dependence plots (PDPs), which provide model-based interpretation of variable effects within the trained four-variable surrogate framework.

Figure S16 depicts the interactions between synthesis parameters as derived from the Gaussian Process Regressor (GPR) model, specifically highlighting their combined influence on the surface area of MIL-100(Fe). In panel (A), temperature and time show a synergistic effect, indicating that longer synthesis durations at elevated temperatures enhance the surface area. Similarly, panel (D) reveals that increasing the molar ratio at higher temperatures leads to optimized surface area values. Panels (B) and (C) show that both molar ratio and ion concentration have relatively stable effects when combined with synthesis time, suggesting their influence on surface area is less sensitive to time variations compared to temperature. Lastly, the combination of ion concentration and molar ratio (panel F) exhibits a linear trend in increasing surface area, indicating the potential for tuning these variables together for optimal performance. These insights are crucial in optimizing synthesis conditions for maximizing surface area, a key property for enhanced adsorption performance in dye removal and other applications. The visualizations reveal how interactions among synthesis parameters can collectively influence material characteristics, emphasizing the importance of considering multiple features during the optimization process.

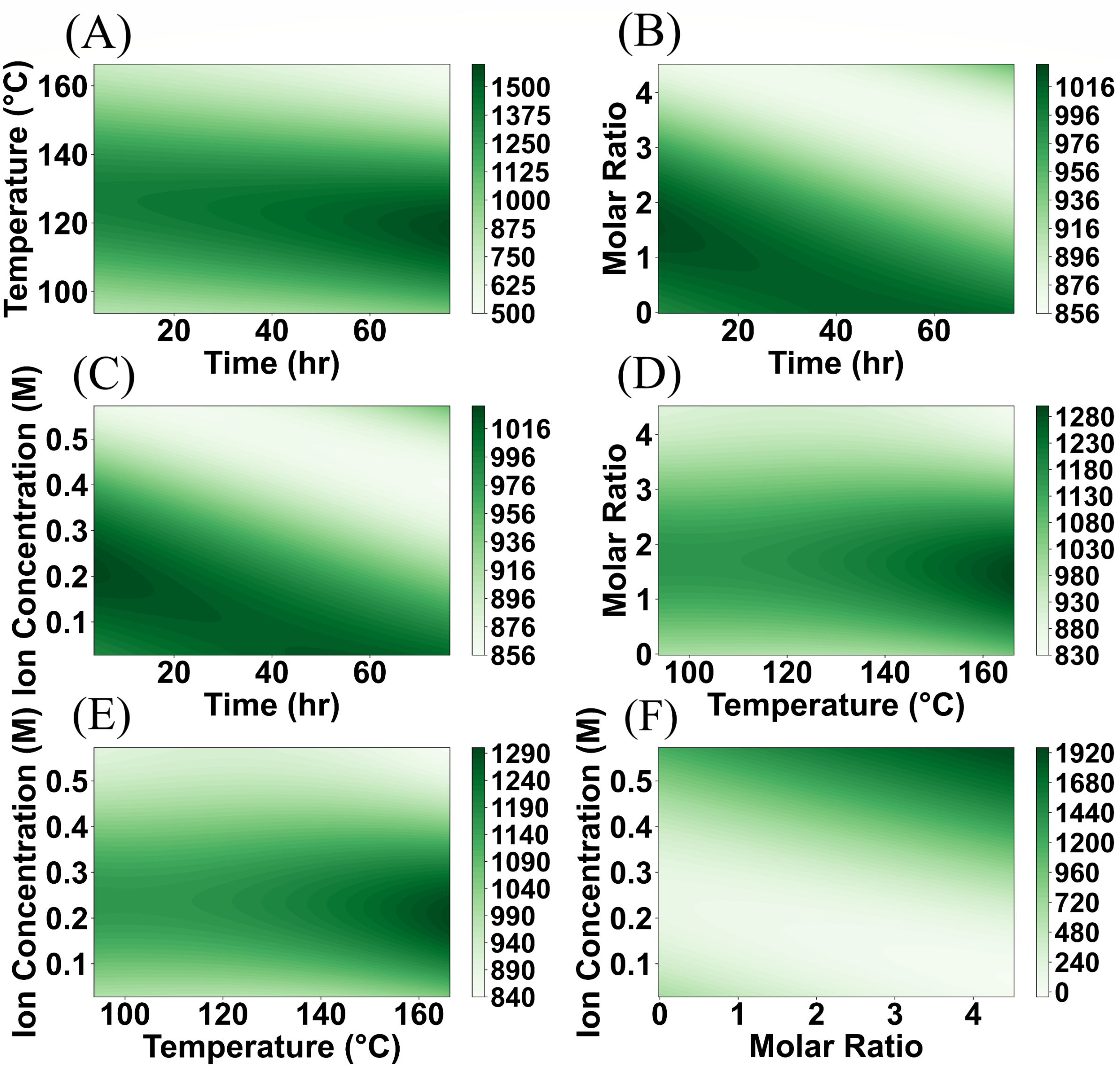

Figure S16. Partial Dependence Plots (PDPs) illustrating the interaction effects between synthesis parameters on the surface area (m²/g) of MIL-100(Fe). The parameters examined include (A) Temperature vs. Time, (B) Molar Ratio vs. Time, (C) Ion Concentration vs. Time, (D) Molar Ratio vs. Temperature, (E) Ion Concentration vs. Temperature, and (F) Ion Concentration vs. Molar Ratio.

Figure S17 presents how the interplay between different synthesis parameters impacts the MB removal efficiency of MIL-100(Fe), providing a deeper understanding of the variables governing dye adsorption. Panel (A) indicates that higher temperatures coupled with increased synthesis time result in improved MB removal efficiency, signifying that longer heating times promote the adsorption properties of the material. Panel (D) demonstrates that a higher molar ratio in combination with increased temperature enhances MB removal, suggesting temperature and molar ratio are key parameters for optimizing adsorption efficiency. Panels (C) and (E) show that increasing ion concentration in conjunction with either time or temperature also positively influences the MB removal rate. These insights underscore the non-linear relationships and combined influence of synthesis parameters on dye removal efficiency, allowing for a better understanding of how to optimize MIL-100(Fe) for applications targeting effective pollutant adsorption. This nuanced analysis enables the identification of synthesis conditions that maximize MB removal, addressing the complex interplay of parameters that affect adsorption efficiency.

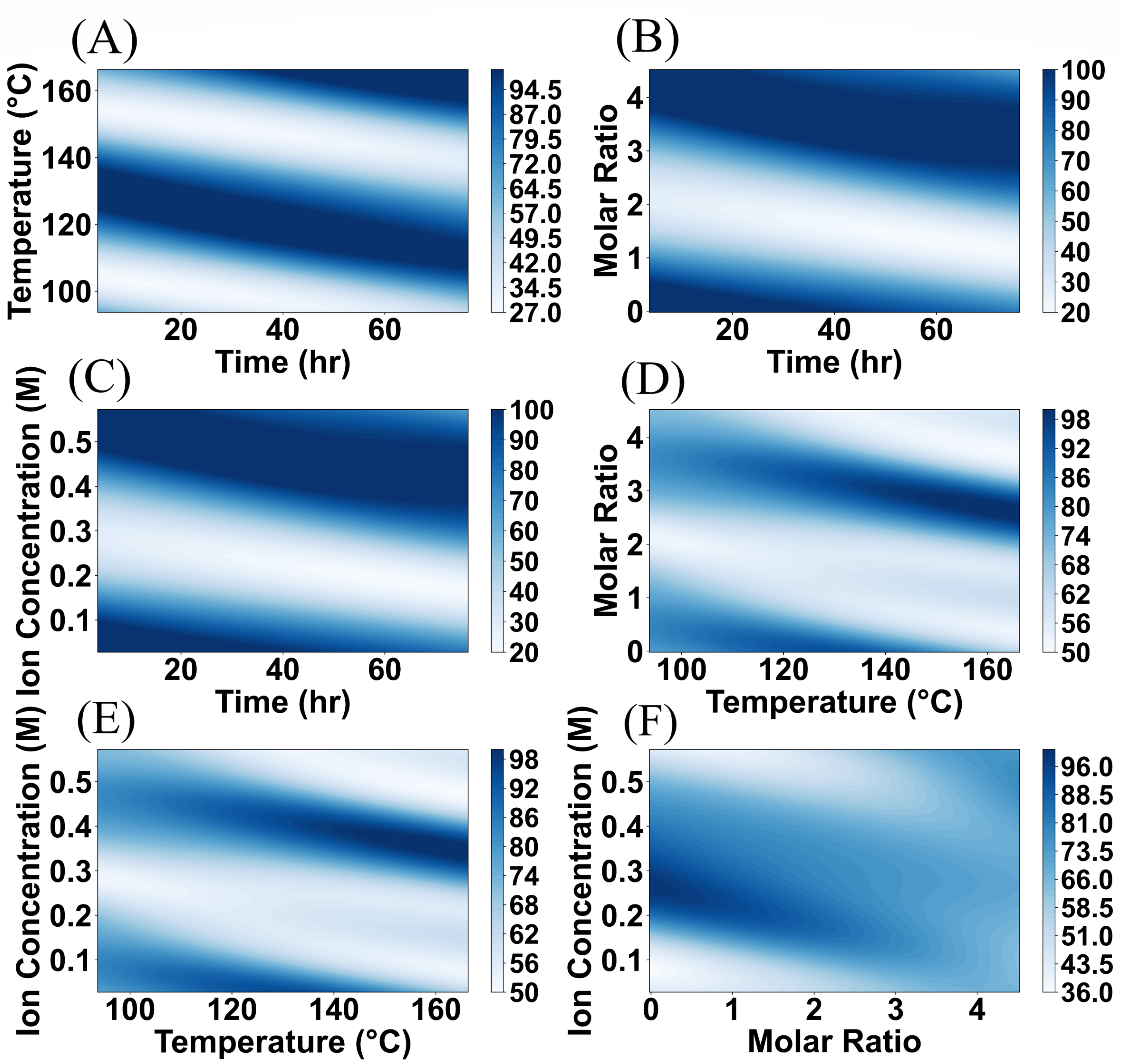

Figure S17. Partial Dependence Plots (PDPs) illustrating the interaction effects between synthesis parameters on methylene blue (MB) removal efficiency (%) of MIL-100(Fe). The parameters examined include (A) Temperature vs. Time, (B) Molar Ratio vs. Time, (C) Ion Concentration vs. Time, (D) Molar Ratio vs. Temperature, (E) Ion Concentration vs. Temperature, and (F) Ion Concentration vs. Molar Ratio.

Figure S18 explores how different combinations of synthesis parameters influence the overall yield of MIL-100(Fe). Panel (A) shows that yield increases with higher temperature and longer synthesis time, indicating the importance of controlled heating to maximize material production. Panel (D) further highlights that increasing the temperature while maintaining an optimal molar ratio results in higher yield. This suggests that synthesis temperature and molar ratio are critical parameters for optimizing product yield. Panel (E) indicates that higher ion concentrations combined with increased temperature enhance yield, reinforcing the role of ion concentration in facilitating successful material formation. These insights provide valuable guidance on tuning synthesis parameters to maximize the yield of MIL-100(Fe), highlighting the non-linear interactions between synthesis conditions and their impact on product output. Understanding these relationships enables optimization strategies that can be applied to scale up the production of MOFs with desirable characteristics, improving the practical utility and scalability of MOF synthesis.

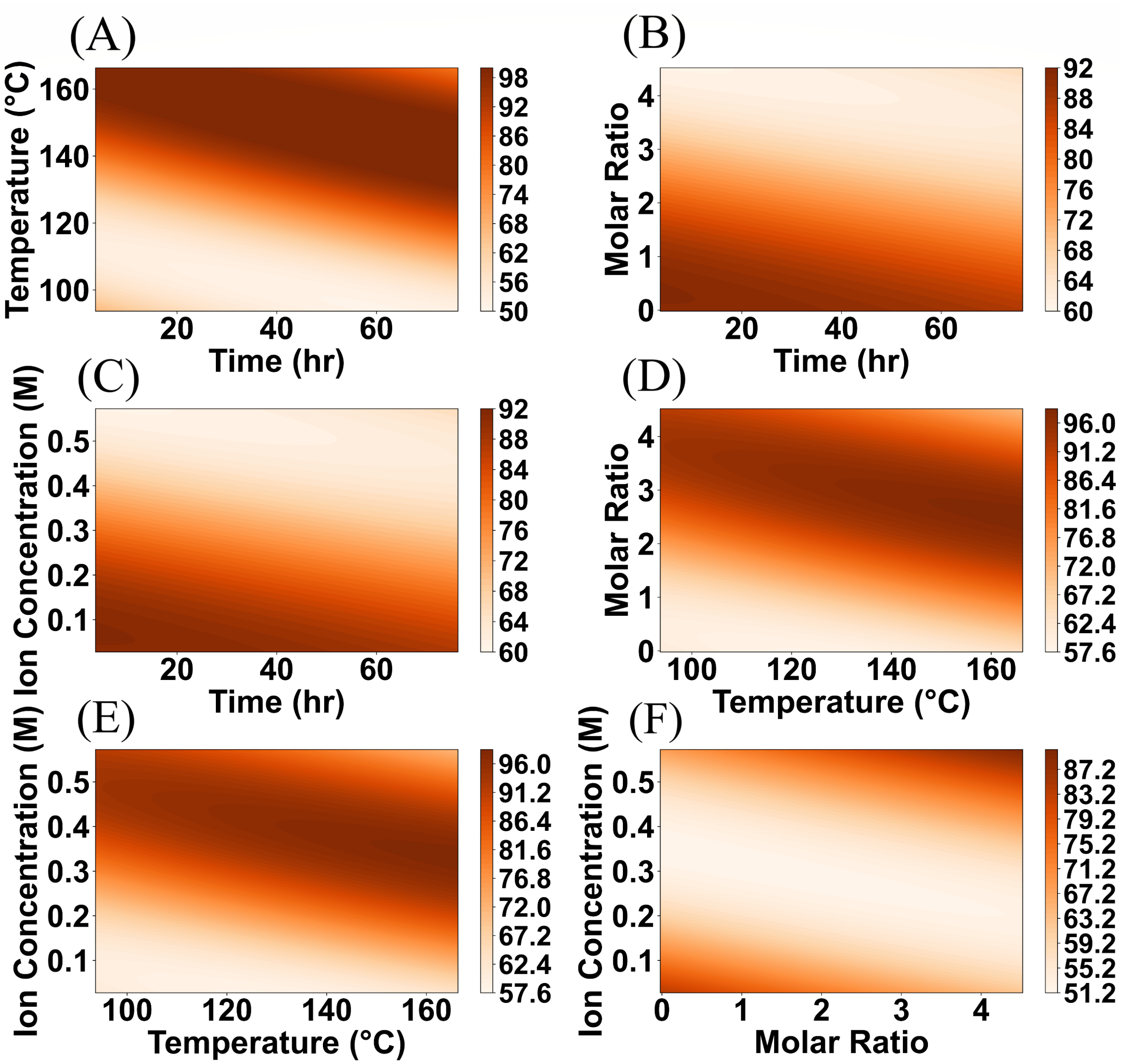

Figure S18. Partial Dependence Plots (PDPs) illustrating the interaction effects between synthesis parameters on the yield (%) of MIL-100(Fe). The parameters examined include (A) Temperature vs. Time, (B) Molar Ratio vs. Time, (C) Ion Concentration vs. Time, (D) Molar Ratio vs. Temperature, (E) Ion Concentration vs. Temperature, and (F) Ion Concentration vs. Molar Ratio.

5. Characterization techniques for analyzing optimized points

5.1 Pore Size Distribution Analysis Using BJH Method

The BJH (Barrett-Joyner-Halenda) plot presented in **Figure S19** illustrates the pore size distribution of two distinct optimization approaches for MIL-100(Fe): one that maximizes BET surface area (OPT BET) and another that maximizes methylene blue (MB) removal efficiency (OPT MB). The y-axis represents the pore volume differential with respect to the pore diameter, while the x-axis shows the pore diameter (in nanometers). This comparison aims to highlight how different optimization objectives impact pore characteristics.

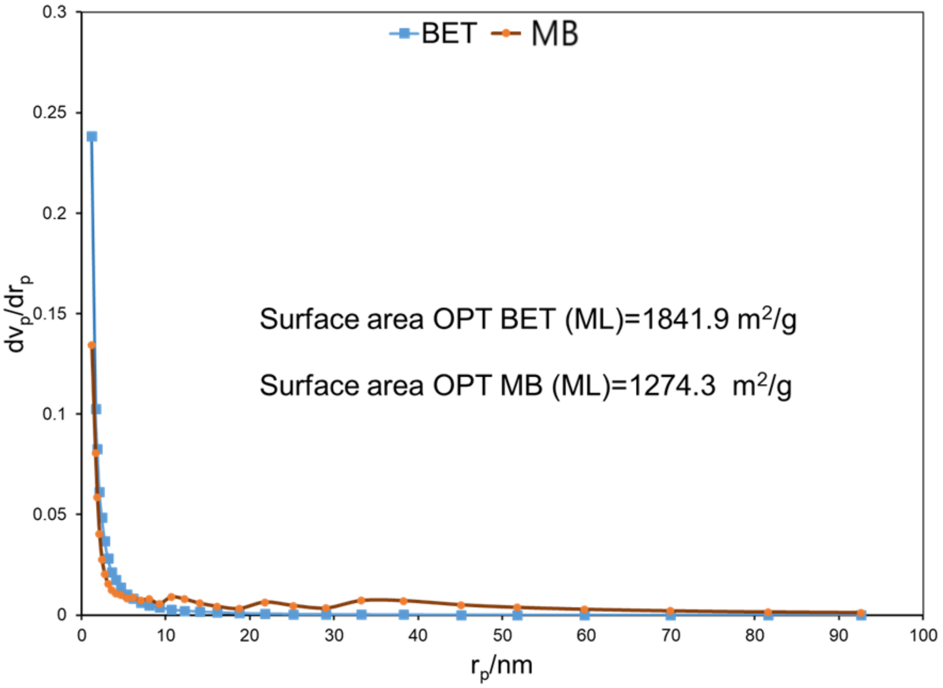

Figure S19. BJH Pore Size Distribution Plot for Optimized BET Surface Area (OPT BET) and Optimized Methylene Blue (OPT MB) Removal Efficiency. This figure compares the pore size distribution of MIL-100(Fe) samples optimized for maximum BET surface area and those optimized for MB removal efficiency.

This distribution is crucial for providing an extensive adsorption surface, which often contributes to high BET surface area values. The analysis reveals a significant difference in the pore size distributions optimized for these two targets. The OPT BET sample has a broader and higher pore volume distribution, while the OPT MB sample shows a narrower distribution, favoring smaller and more accessible pores. The narrower BJH distribution of OPT MB may be advantageous not only because of improved pore accessibility, but also because it provides a more favorable size-matching environment for methylene blue adsorption. Given the reported dimensions of MB and the known cage/window architecture of MIL-100(Fe), adsorption is expected to benefit from accessible pore features that are large enough to host the dye while maintaining steric selectivity at the pore-entry scale [5, 6].

5.2 FTIR Spectra Comparison for Optimized Samples

Figure S20 presents the FTIR spectra for MIL-100(Fe) samples optimized for BET surface area (OPT BET) and methylene blue (MB) removal efficiency (OPT MB). The spectra help identify key functional groups present in the materials, providing insights into the surface chemistry that could affect adsorption and overall performance.

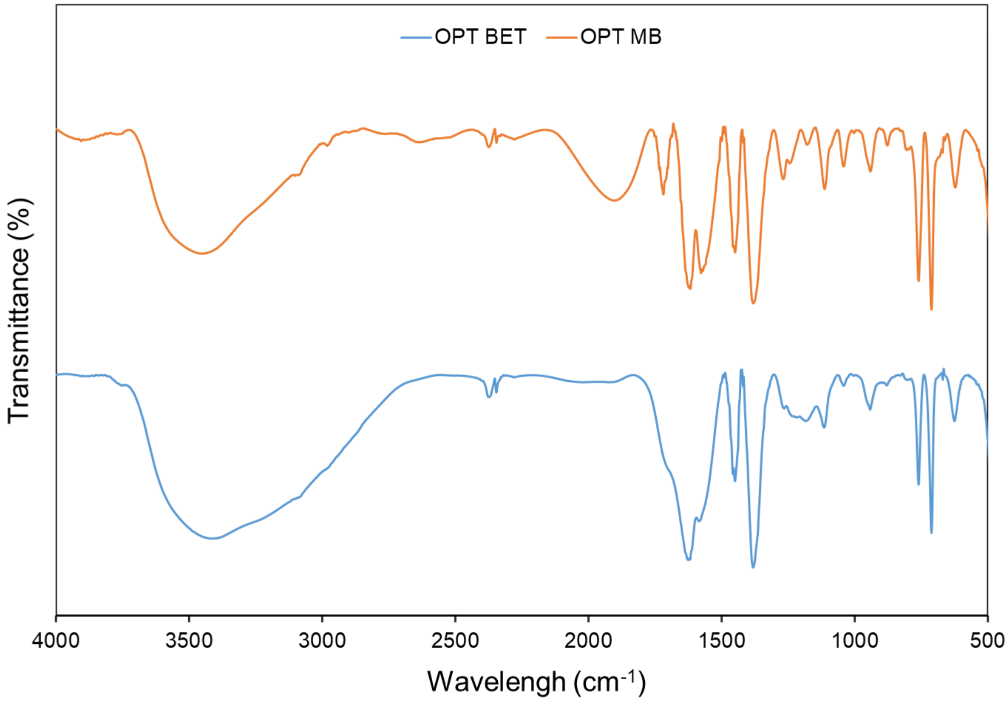

Figure S20. FTIR Spectra Comparison for Optimized BET Surface Area (OPT BET) and Optimized Methylene Blue (OPT MB) Removal Efficiency. The figure shows the Fourier-transform infrared (FTIR) spectra of MIL-100(Fe) samples optimized for maximum BET surface area and methylene blue (MB) removal efficiency, highlighting the differences in functional group presence and intensity.

The OPT BET (blue) and OPT MB (orange) spectra exhibit characteristic bands of MIL-100(Fe). Broad bands around 3400 cm^{-1} are attributed to O-H stretching, indicating the presence of hydroxyl groups, which can play a role in hydrogen bonding with adsorbates. However, the intensity is higher for the OPT BET sample, suggesting more surface hydroxyl groups, which may contribute to a higher surface area by allowing more interaction with the adsorbing gas molecules during the BET measurement. In the region of 1650 cm^{-1}, both spectra show peaks corresponding to the C=O stretching vibration from carboxylate groups, confirming the presence of trimesic acid ligands in the framework. These carboxylate groups are essential for dye adsorption, as they can form interactions with the charged dye molecules. The OPT MB spectrum shows sharper and more defined peaks in this region, indicating a higher density of accessible functional groups, which might enhance MB removal efficiency. The fingerprint region (1500–500 cm^{-1}) shows differences in the intensity and sharpness of peaks between the two samples. The OPT MB spectrum reveals more prominent bands related to C-H bending and Fe-O stretching, which suggest better-defined structural integrity and a higher number of active sites for adsorption. This indicates that optimizing for MB removal has resulted in a more chemically active surface, favoring interactions with dye molecules. In General, the comparison of the FTIR spectra suggests that while the OPT BET sample has more hydroxyl groups, the OPT MB sample exhibits a higher number of functional groups directly involved in adsorption processes. The enhanced presence of carboxylate groups and the better-defined Fe-O vibrations in the OPT MB sample are likely responsible for its superior dye removal efficiency, demonstrating the importance of optimizing surface chemistry in addition to surface area to achieve desired application performance.

5.3 TGA Analysis: Thermal Stability of Optimized MIL-100(Fe) Samples

Figure S21 shows the TGA thermograms for MIL-100(Fe) samples optimized for maximum BET surface area (OPT BET) and methylene blue removal efficiency (OPT MB). Thermogravimetric analysis provides insights into the thermal stability and decomposition characteristics of the synthesized materials, highlighting differences influenced by the optimization conditions.

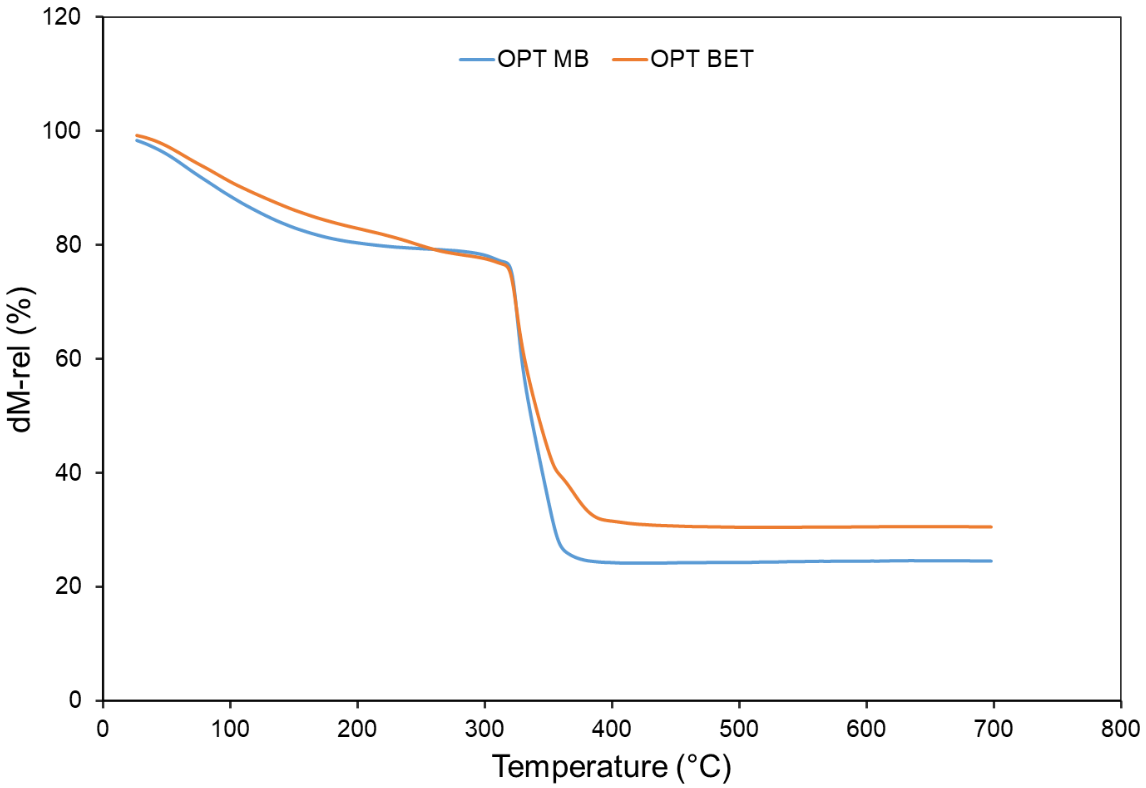

Figure S21. Thermogravimetric Analysis (TGA) of MIL-100(Fe) Samples Optimized for BET Surface Area (OPT BET) and Methylene Blue (MB) Removal Efficiency (OPT MB). The thermograms display the thermal stability and decomposition behavior of MIL-100(Fe) synthesized under different optimization conditions.

Both samples exhibit similar thermal behavior up to around 300 °C, with a gradual weight loss due to the evaporation of adsorbed water and release of trapped solvent molecules. The weight loss in this region is approximately 15–20%, which corresponds to the removal of physically adsorbed moisture and the residual solvent. The first significant weight loss between 300–400 °C is attributed to the decomposition of the organic ligand (trimesic acid), indicating the collapse of the MIL-100(Fe) structure. This weight loss is more pronounced for the OPT MB sample, suggesting a higher content of organic ligands that are more susceptible to thermal degradation. Above 400 °C, the remaining framework becomes more stable, and no significant weight loss is observed for either sample, indicating the formation of a stable residue, likely iron oxide. The OPT BET sample appears to retain slightly more mass compared to the OPT MB sample, which suggests that the material optimized for BET surface area may have a slightly more robust framework, potentially due to differences in crystallinity or pore structure. The differences in thermal stability between the two samples are indicative of the impact of optimization conditions on the structural properties of MIL-100(Fe). The OPT MB sample, while showing better performance in dye adsorption, appears to have a less thermally stable framework, likely due to a higher density of functional groups or defects that enhance adsorption but make the material more prone to thermal degradation. Conversely, the OPT BET sample demonstrates better thermal stability, which might be a result of a more compact and uniform framework, although this comes at the expense of lower adsorption performance for methylene blue. Overall, this TGA analysis highlights the trade-offs between optimizing for surface area and optimizing for adsorption performance, with implications for the intended application of MIL-100(Fe) in environmental remediation or catalysis.

5.4 XRD Patterns: Crystallinity of MIL-100(Fe) Optimized Samples

Figure S22 presents the XRD patterns for MIL-100(Fe) samples that were optimized to maximize BET surface area (OPT BET) and methylene blue removal efficiency (OPT MB). X-ray diffraction is a critical tool used to determine the crystalline structure of the synthesized MOFs and to evaluate the influence of different synthesis conditions on crystallinity.

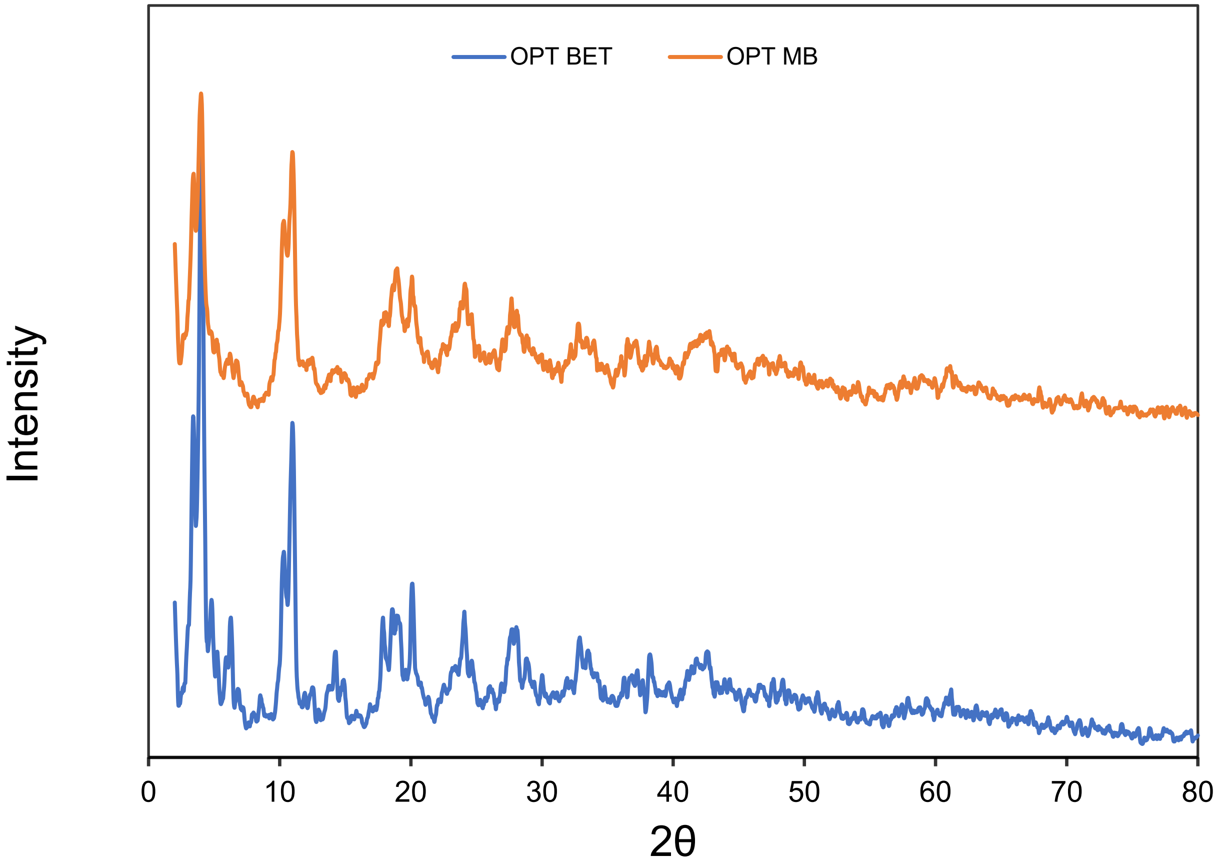

Figure S22. X-ray Diffraction (XRD) patterns for MIL-100(Fe) samples optimized for BET surface area (OPT BET) and methylene blue (MB) removal efficiency (OPT MB). The XRD patterns highlight the crystallinity and phase composition of the samples, providing insights into structural differences based on the optimization focus.

As shown in the XRD patterns, the OPT BET sample (blue) displays sharper and more intense diffraction peaks, indicative of higher crystallinity. This is consistent with the calculated crystallinity percentages shown in the corresponding table. In contrast, the OPT MB sample (orange) exhibits broader and less intense peaks, suggesting a more amorphous structure or lower crystalline order. These differences reflect the impact of synthesis conditions on structural order, where the OPT BET conditions promote well-defined crystal growth. Importantly, this structural variation aligns with the intended functional trade-offs: while higher crystallinity benefits surface area in OPT BET, the slightly disordered structure in OPT MB appears to enhance dye adsorption, potentially due to improved accessibility or defect-assisted adsorption. It should be noted that the patterns are **not normalized**, and intensity differences are presented as measured.

The broad peaks suggest a more disordered structure, which may contribute to an increased surface area due to the formation of more random pore structures. However, despite the larger surface area, the lower crystallinity of OPT BET may reduce the material's structural stability, which can have implications for applications that require robust frameworks. The differences in the XRD patterns highlight the impact of synthesis optimization on material properties. For the OPT MB sample, the higher crystallinity and presence of distinct crystal planes may facilitate efficient adsorption through better molecular interactions between the dye molecules and the MOF's surface. Conversely, the OPT BET sample, with a more disordered structure, provides a larger surface area, which can be advantageous in applications where maximizing accessible surface sites is critical. However, it is evident that higher surface area does not necessarily correspond to improved adsorption performance, as other factors such as crystallinity and pore accessibility also play significant roles. These observations underscore the need to tailor synthesis conditions based on the intended application. While optimizing for surface area might be beneficial for gas adsorption or catalysis, optimizing for higher crystallinity is essential for enhancing specific adsorption capabilities, such as dye removal. The XRD analysis further reinforces the idea that a balance between surface area, crystallinity, and other material properties is crucial for achieving optimal performance in specific applications.

5.5 Synthesis Parameters and Properties of Optimized MIL-100(Fe) Samples

Table S2 summarizes the optimized synthesis conditions for maximizing both the surface area (BET optimization) and methylene blue (MB) removal efficiency (MB optimization) of MIL-100(Fe), as determined using Gaussian Process Regressor (GPR) predictions. The table presents key synthesis parameters, including time, temperature, Fe:ligand molar ratio, and ion concentration, along with the corresponding predicted and experimentally measured values for surface area and MB removal. For **BET optimization**, the GPR model identified optimal conditions as 46.1 hours of reaction time at 110.9°C, with a Fe:ligand molar ratio of 2.7:1 and an ion concentration of 0.29 M. Under these conditions, the predicted surface area was 2010.7 m²/g, while the actual measured surface area was 1841.9 m²/g, resulting in an error of 315.7 m²/g. The corresponding real MB removal efficiency was 85.1%, with an absolute error of 13.9%. For MB optimization, the model identified synthesis conditions associated with saturation-level MB removal: 39.4 hours of reaction time at 134.1°C, a Fe:ligand molar ratio of 1.0:1, and an ion concentration of 0.19 M. For physical interpretability, the reported predicted MB removal was capped at 100%, while the experimentally measured value was 98.3%, resulting in a similar absolute error of 13.9%.

Table S2. Optimized Synthesis Conditions for MIL-100(Fe) Based on Gaussian Process Regressor (GPR) Predictions and Real Synthesis Data.

Optimization Goal	Parameter	Optimized Value	Predicted Value	Real Value	Error
Maximum Surface Area (BET Optimization)	Time (hr)	46.1	2010.7 m²/g	1841.9 m²/g	315.7 m²/g
	Temperature (°C)	110.9	-	-	-
	Fe:Ligand Molar Ratio	2.7:1	-	-	-
	Ion Concentration (M)	0.29	-	-	-
	MB Removal (%)	-	-	85.1%	13.9%
Maximum MB Removal (MB Optimization)	Time (hr)	39.4	-	-	-
	Temperature (°C)	134.1	-	-	-

	Fe:Ligand Molar Ratio	1.0:1	-	-	-
	Ion Concentration (M)	0.19	-	-	-
	Surface Area (m^2/g)	-	-	1274.3 m^2/g	315.7 m^2/g
	MB Removal (%)	-	100*	98.3%	13.9%

*Note: During optimization, the GPR surrogate response for MB removal was not explicitly clipped at 100%, although the synthesis-variable search space remained strictly bounded within the experimental design window. Consequently, the surrogate could return saturation-level predictions slightly above 100% near the optimum. Such values were not interpreted as physically realizable removal efficiencies; they were treated as indicators of approach to the upper-performance limit and capped at 100% for reporting.

Table S3 presents the synthesis parameters and resulting properties of MIL-100(Fe) samples optimized for maximum BET surface area (OPT BET) and maximum methylene blue removal efficiency (OPT MB). The synthesis conditions, including time, temperature, metal-to-ligand molar ratio, and metal concentration, were adjusted to yield different optimized properties for specific applications.

Table S3. Synthesis parameters and properties of MIL-100(Fe) samples optimized for BET surface area (OPT BET) and methylene blue (MB) removal efficiency (OPT MB). The table provides a comparison of synthesis conditions and key physical properties, such as surface area, crystallinity, total pore volume, average crystallite size, and yield.

Sample	Time (h)	Temp. (°C)	Metal-to-Ligand Molar Ratio (Fe:Ligand)	metal con. (M)	BET (m^2/g)	Crystallinity (%)	total pore volume (cm^3/gr)	Avg. crystallite size (D_{avg} nm)	Yield (%)
OPT BET	46.11	110.8	2.681:1	0.294	1841.9	34.87	0.8948	22.4	93.06
OPT MB	39.4	134.09	1.003:1	0.1951	1274.3	17.37	0.9169	30.8	70.53

For the sample optimized for BET surface area (OPT BET), the synthesis time was approximately 46.11 hours at a temperature of 110.8°C, with a metal-to-ligand molar ratio of 2.681:1 and a metal concentration of 0.294 M. This resulted in a high BET surface area of 1841.9 m^2/g, along with a crystallinity of 34.87%, a total pore volume of 0.8948 cm^3/g, an average crystallite size of 22.4 nm, and a yield of 93.06%. The combination of lower synthesis temperature and extended synthesis time contributed to the formation of a highly porous structure, which is conducive to maximizing surface area.

In contrast, the sample optimized for methylene blue removal efficiency (OPT MB) was synthesized at a higher temperature of 134.09°C for a shorter duration of 39.4 hours, with a metal-to-ligand molar ratio of 1.003:1 and a metal concentration of 0.1951 M. This resulted in a BET surface area of 1274.3 m^2/g and a crystallinity of 17.37%. The total pore volume was 0.9169 cm^3/g, while the average crystallite size was 30.8 nm, and the yield was 70.53%. Although the BET surface area of OPT MB is lower than that of OPT BET, the optimized synthesis conditions improved the exposure of functional sites and provided better pore accessibility, leading to enhanced dye removal efficiency. These results illustrate the importance of synthesis parameters in tailoring the properties of MIL-100(Fe) for specific applications.

For adsorption processes, such as dye removal, achieving a balance between surface area, crystallinity, and pore volume is essential. While a high BET surface area is often desirable, as seen in the OPT BET sample, the sample optimized for methylene blue removal (OPT MB) demonstrates that other factors, such as crystallinity and crystallite size, can significantly influence adsorption performance. This observation challenges the conventional notion that a higher surface area necessarily translates to better performance and highlights the importance of customizing synthesis conditions based on the desired functionality of the material.

Design note. The 11-sample dataset consisted of a core Taguchi-based design supplemented by two additional non-orthogonal runs (**S10** and **S11**) introduced to broaden coverage of the synthesis-variable space and strengthen low-data model development. These supplementary runs should not be interpreted as stochastic "random" points.

5.6 Comparative Analysis of Optimization Approaches

Figure S23 presents the methylene blue (MB) dye removal efficiencies of the two MIL-100(Fe) samples: OPT BET (optimized for BET surface area) and OPT MB (optimized for methylene blue removal). The bar chart shows that the OPT MB sample achieved a significantly higher MB removal efficiency of 98.29%, compared to the OPT BET sample, which had a removal efficiency of 85.05%.

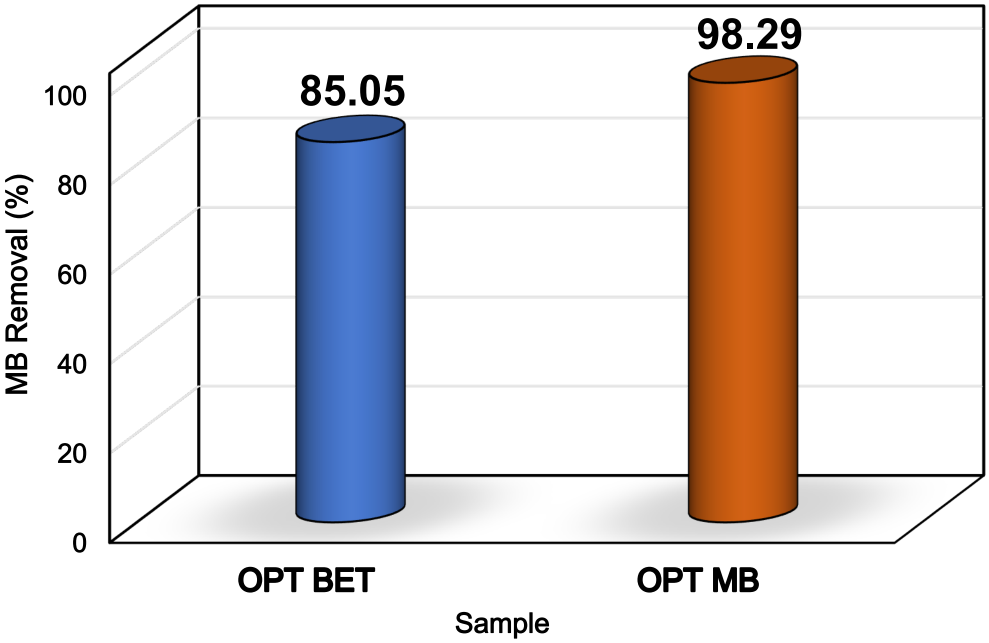

Figure S23. Methylene blue (MB) dye removal efficiency of MIL-100(Fe) samples optimized for BET surface area (OPT BET) and methylene blue removal (OPT MB). The bar chart shows the percentage of MB removal for each optimized sample, highlighting the differences in adsorption efficiency based on the synthesis optimization approach.

The differences in dye removal efficiency can be attributed to the synthesis parameters used to optimize each sample. While OPT BET was designed to maximize surface area (1841.9 m^2/g), it appears that this optimization did not necessarily lead to the best adsorption performance. In contrast, the OPT MB sample, synthesized under different conditions that led to a lower BET surface area (1274.3 m^2/g), was more effective in removing methylene blue from solution. These results suggest that a higher BET surface area is not always synonymous with enhanced adsorption capacity, emphasizing the need to consider other properties, such as pore structure, surface chemistry, and crystallinity. The OPT MB sample likely possesses a combination of pore characteristics and surface functionalities that facilitate better interaction with dye molecules, contributing to its superior performance in the dye removal test. This finding reinforces the importance of tailoring synthesis parameters based on the intended application, rather than relying solely on traditional metrics such as surface area. For adsorption applications, such as dye removal, a more nuanced approach that accounts for factors beyond surface area is essential to achieve optimal performance. The same trend is retained when performance is expressed as adsorption capacity q (mg g^{-1}), confirming the superior dye-uptake performance of OPT MB over OPT BET under the same benchmark conditions.

5.7 SEM Morphology Analysis of MIL-100(Fe) Sample 9

Figure S24 shows the SEM micrograph of the OPT MB sample recorded at 10 kV and 10,000x magnification (scale bar = 1 µm). The particles form aggregated, irregular clusters with rough surfaces and abundant interparticle mesovoids, which aligns with the broader mesopore distribution (**Figure S19**) and defect-rich XRD features (**Figure S22**). Such microstructural traits enhance external accessibility and shorten diffusion pathways, underpinning the superior MB removal efficiency (98.3%) observed for this sample.

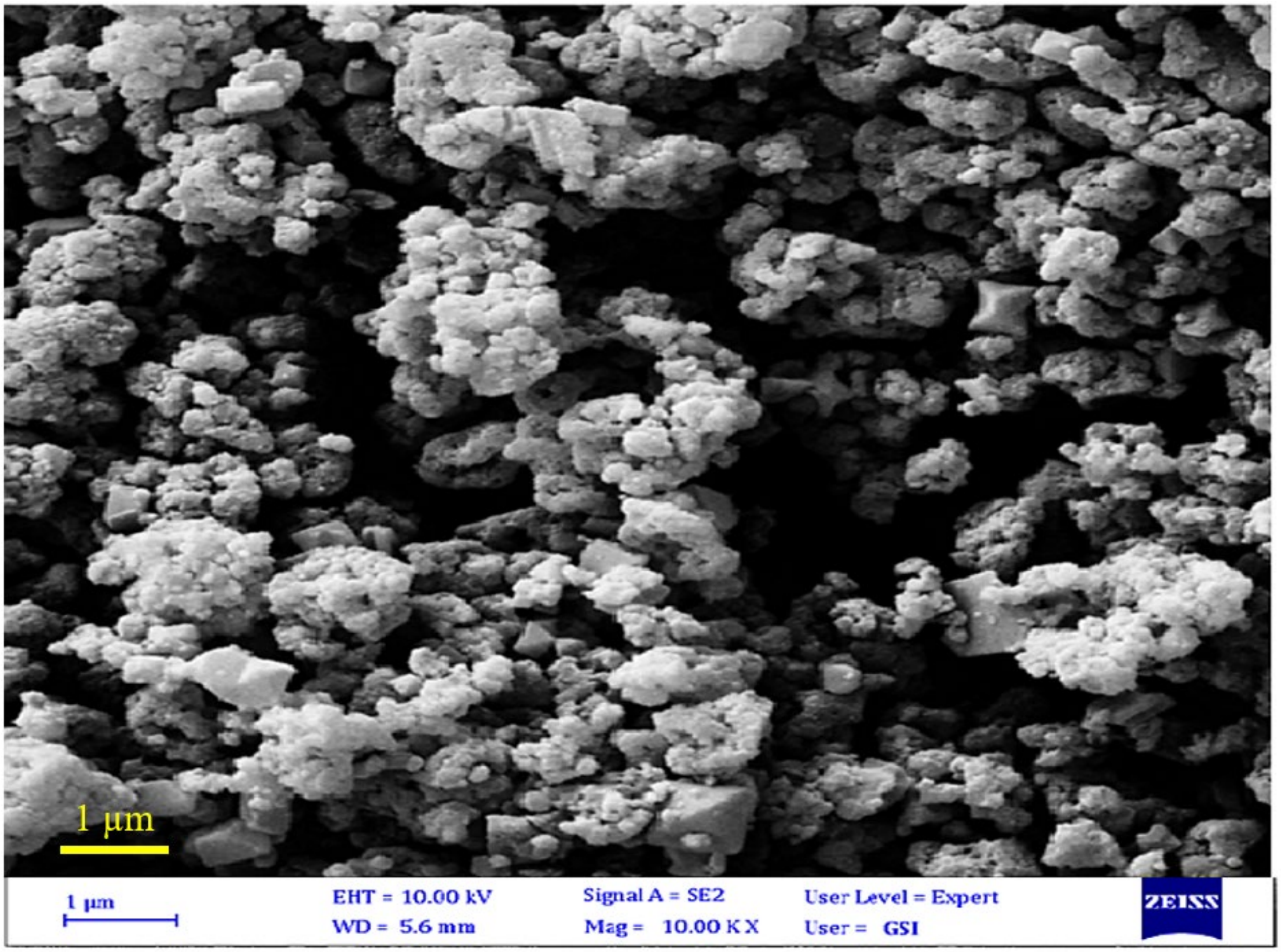

Figure S24. SEM micrograph of the OPT MB MIL-100(Fe) sample (EHT = 10 kV, WD = 5.6 mm, Mag = 10,000×, scale bar = 1 µm). The image highlights loosely packed agglomerates with rough, defect-rich surfaces and open mesovoids, consistent with the sample's enhanced mesoporous structure and high MB removal efficiency.

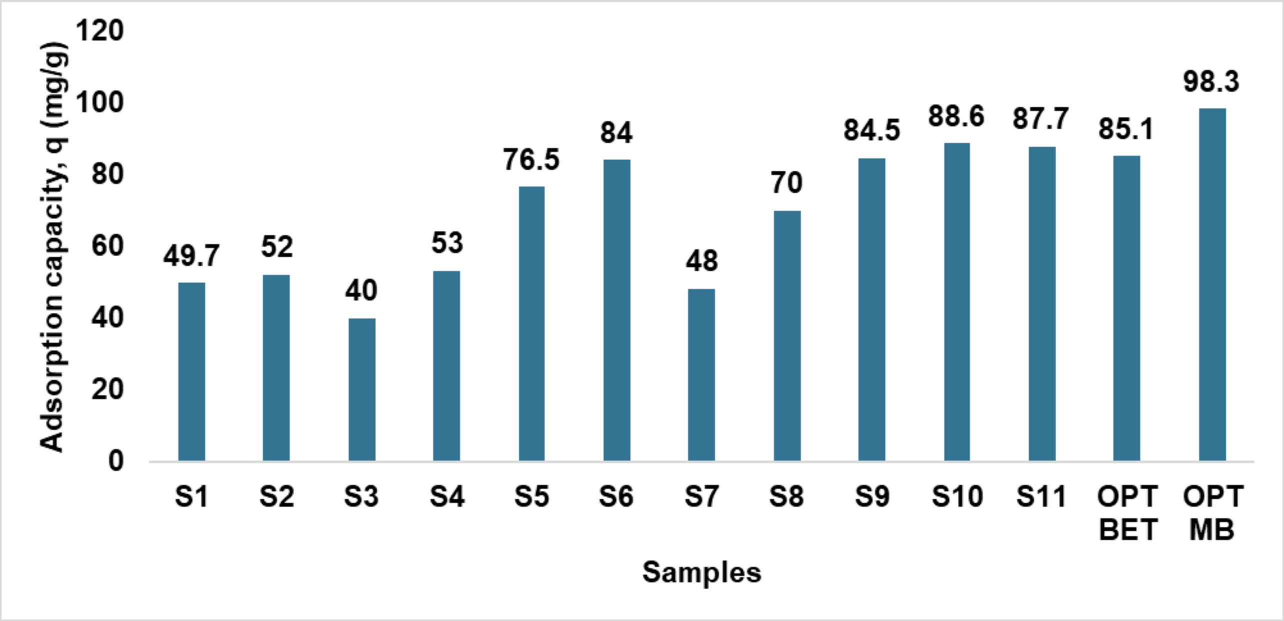

Figure S25. Adsorption capacity q (mg g^{-1}) of the 11 MIL-100(Fe) samples and the two optimized samples (OPT BET and OPT MB) under standardized MB adsorption conditions ($C_0 = 100ppm(mgL^{-1}), V = 10mL, m = 10mg, pH = 9$, 180 min, room temperature). Under these fixed conditions, the numerical value of q is equal to the corresponding MB removal value, enabling direct comparison of uptake performance across the sample set.

To complement the removal-percentage comparison, adsorption capacity q (mg g^{-1}) was calculated for all synthesized and optimized samples using the standard mass-balance expression. This provides an additional quantitative comparison of uptake performance while remaining fully consistent with the fixed adsorption-screening conditions used throughout the present study.

To provide additional insight into the adsorption mechanism of methylene blue (MB), the point of zero charge, pH_{PZC}, of the optimized MB-removal sample (OPT MB) was determined by the pH-drift method. A series of solutions with initial pH values in the range of 3–9 was prepared using 0.1 M HCl and 0.1 M NaOH. Then, 2 mg of the sample was added to 30 mL of each solution and shaken for 24 h at room temperature. The final pH values were measured, and $\Delta pH = pH_f - pH_i$ was plotted against the initial pH. The intersection of the curve with the horizontal axis gave $pH_{PZC} = 3.36$. Because the adsorption experiments were conducted at pH 9, the surface of OPT MB is expected to be negatively charged under the test conditions, favoring electrostatic attraction toward cationic MB molecules.

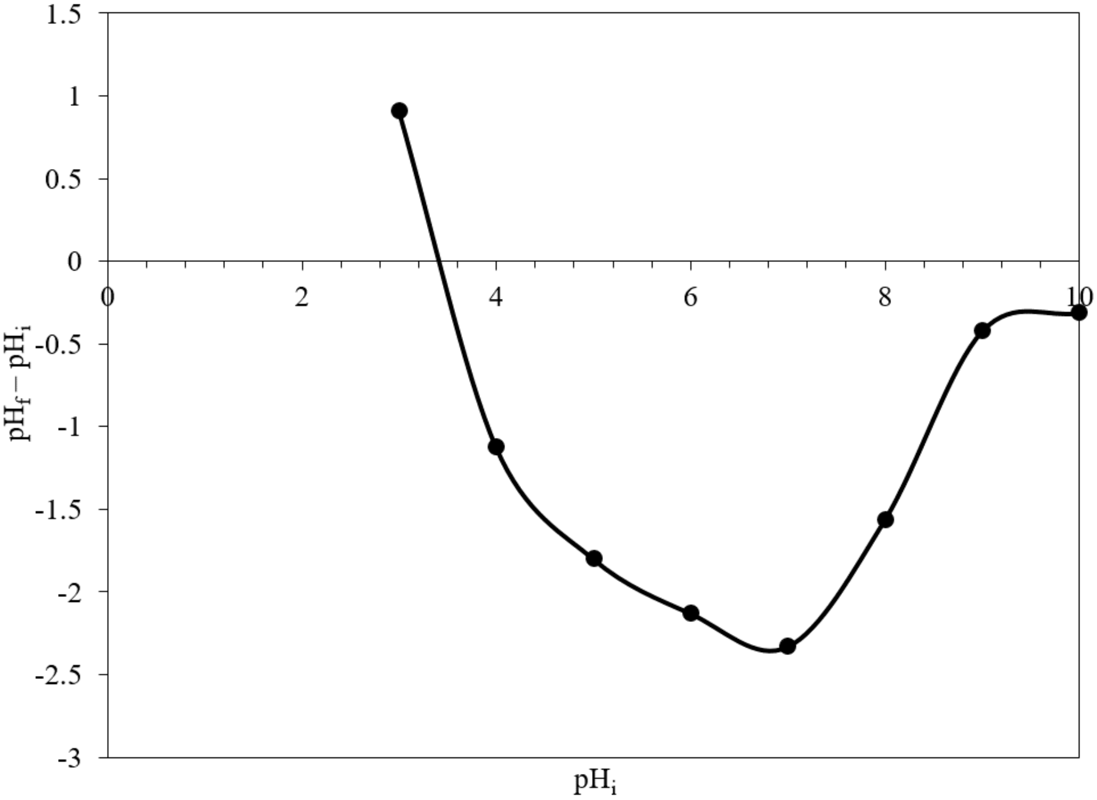

Figure S26. Determination of the point of zero charge, pH_{PZC}, of the OPT MB MIL-100(Fe) sample by the pH-drift method. The plot of $\Delta pH = pH_f - pH_i$ versus initial pH shows an intercept pH_{PZC} = 3.36. Since the adsorption experiments were performed at pH 9, the adsorbent surface is expected to be negatively charged during MB uptake.

5.8. Preliminary pH-Dependent Adsorption Screening of Methylene Blue

To justify the selection of the working pH for the adsorption experiments, preliminary pH-dependent screening tests were conducted for representative MIL-100(Fe) samples at pH 5, 7, and 9 using an initial methylene blue concentration of 100 ppm and a contact time of 180 min. As shown in **Figure S27**, all samples (S1–S9) followed the same qualitative trend, with MB removal increasing as pH increased and reaching its maximum at pH 9. This observation supported the selection of pH 9 as the benchmark adsorption condition used in the main manuscript.

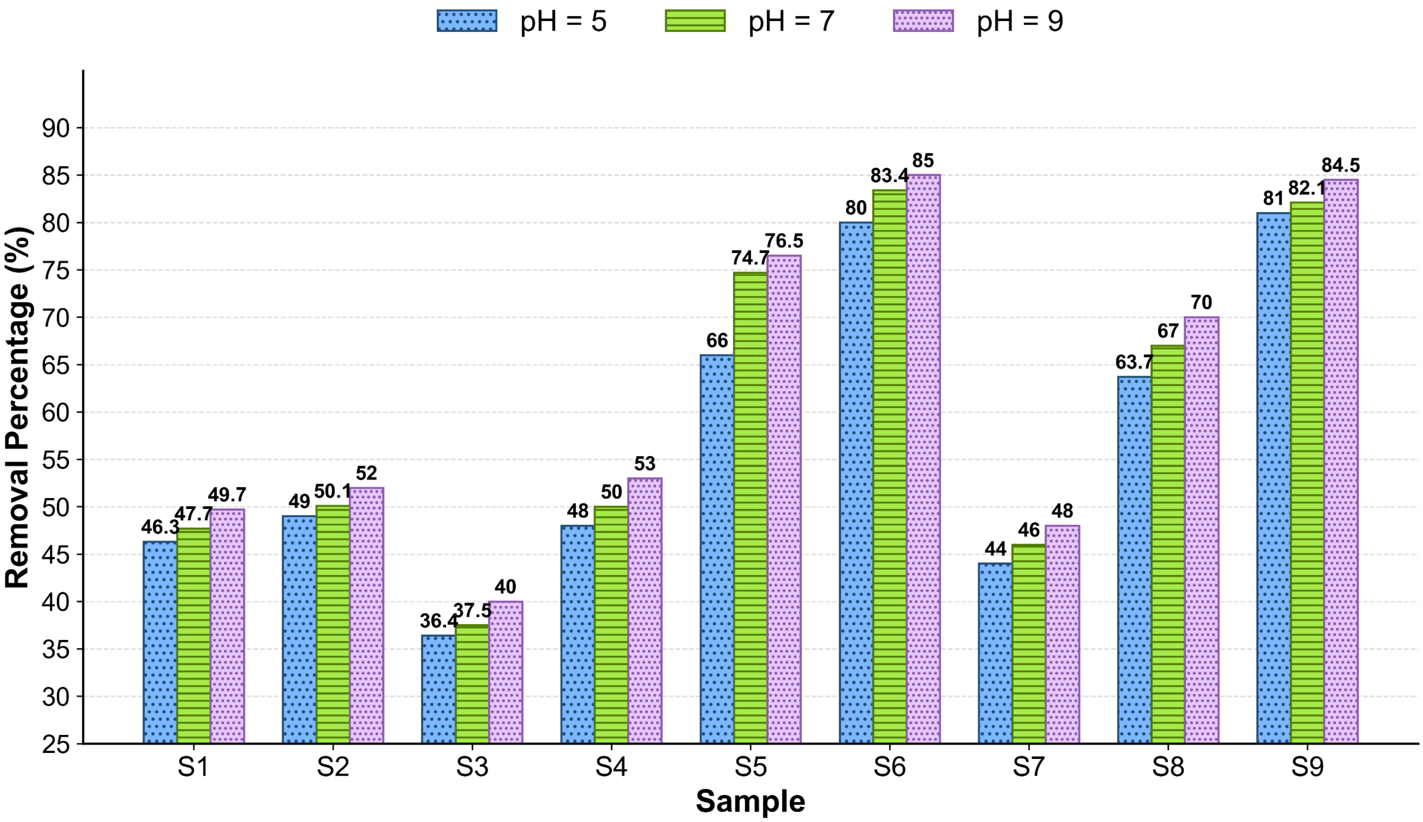

Figure S27. Preliminary pH-dependent adsorption screening of methylene blue by MIL-100(Fe) samples S1–S9 at pH 5, 7, and 9 under an initial dye concentration of 100 ppm and a contact time of 180 min. The generally higher removal observed at pH 9 motivated the selection of pH 9 as the standardized adsorption condition used in the main manuscript.

Table S4. Tentative assignment of impurity-related XRD reflections observed in samples S5, S7, and S9.

Additional weak reflections were observed in the XRD patterns of samples S5, S7, and S9 at approximately 24.3°, 33.2°, 35.8°, 41.4°, and 49.5°. These peaks were not attributed to the MIL-100(Fe) framework and are tentatively assigned here to a minor crystalline α-Fe_2O_3 impurity phase, as summarized in **Table S4**. The appearance of distinct reflections indicates that the impurity is not purely amorphous. However, the present XRD data do not resolve whether this iron oxide exists as discrete particles, intergrown domains, or partial surface deposits/coatings on MIL-100(Fe) [7].

Table S4. Tentative assignment of impurity-related XRD reflections observed in samples S5, S7, and S9. The additional reflections support the presence of a minor crystalline iron-oxide impurity phase, tentatively assigned to α-Fe_2O_3, based on peak-position agreement. Because these features appear as distinct diffraction peaks, the impurity is unlikely to be purely amorphous; however, XRD alone cannot determine whether the oxide exists as discrete particles, intergrown domains, or partial surface deposits on MIL-100(Fe).

Observed 2θ (°)	**Tentative plane**	**Proposed phase**
24.3	(012)	α-Fe_2O_3
33.2	(104)	α-Fe_2O_3
35.8	(110)	α-Fe_2O_3
41.4	(113)	α-Fe_2O_3
49.5	(024)	α-Fe_2O_3

5.9 Residual mass comparison from TGA as supportive evidence of structural modulation

To provide additional support for the discussion of structural modulation in the synthesized MIL-100(Fe) samples, the residual mass at 600 °C derived from TGA was compared across S1–S11 and the two optimized samples (**Figure S28**). In MOF studies, final oxide residue is commonly used as a supportive indicator of inorganic fraction and may, in suitable cases, be consistent with linker deficiency or defect-rich compositions. However, TGA alone cannot uniquely quantify missing-linker defects and must be interpreted cautiously, particularly when secondary inorganic phases are present. In the present system, this caution is especially important for S5 and S7, where XRD indicates iron-oxide impurity phases; therefore, their elevated residual mass cannot be attributed solely to defects within the MIL-100(Fe) framework. The comparison suggests that OPT BET and S11 retain relatively higher final inorganic fractions, whereas OPT MB and S10 show lower residual mass despite similarly strong adsorption behavior, supporting the conclusion that increasing defect-related disorder alone is not sufficient to maximize MB removal [8, 9].

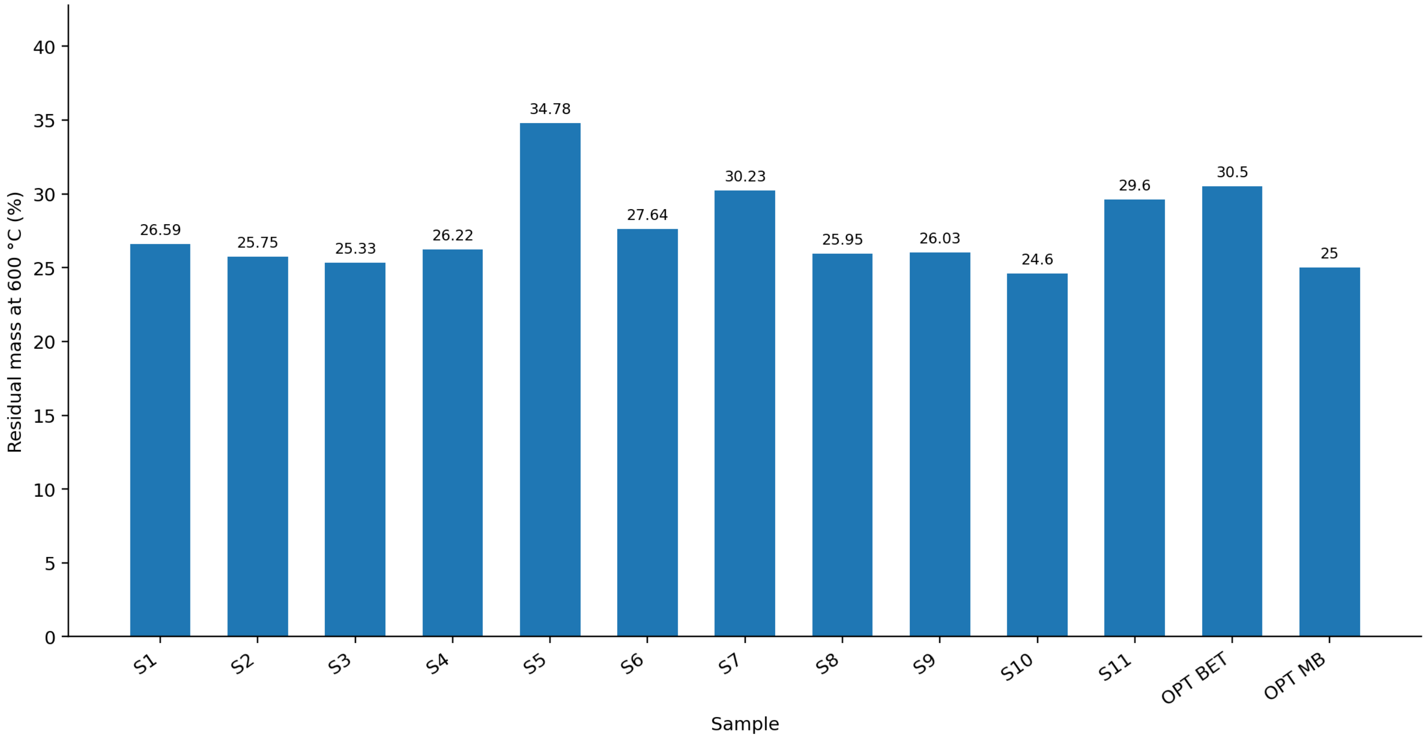

Figure S28. Residual mass at 600 °C obtained from TGA for samples S1–S11 and the optimized samples (OPT BET and OPT MB). The figure is used here as supportive evidence for differences in final inorganic fraction and defect-related structural modulation. Because TGA residue can also be influenced by secondary iron-oxide phases, especially in impurity-containing samples, these data are interpreted qualitatively rather than as a direct quantification of linker-vacancy concentration.

5.10 Recyclability of the OPT MB sample

To evaluate the practical reusability of the optimized MB-removal sample, adsorption–desorption cycling experiments were conducted over five consecutive measurements, corresponding to the initial use and four subsequent reuse cycles **(Figure S29**). The MB removal efficiency decreased gradually from 98.3% in the initial cycle to 87.4% after four adsorption–desorption cycles. This corresponds to retention of approximately 88.9% of the initial removal performance, indicating that OPT MB maintains most of its adsorption capability during repeated operation. In adsorption studies of MOF-based water-treatment materials, multi-cycle performance is widely used as a practical first-level indicator of reuse stability and application feasibility. The present result therefore provides direct evidence that OPT MB remains functionally effective over repeated dye-removal cycles.

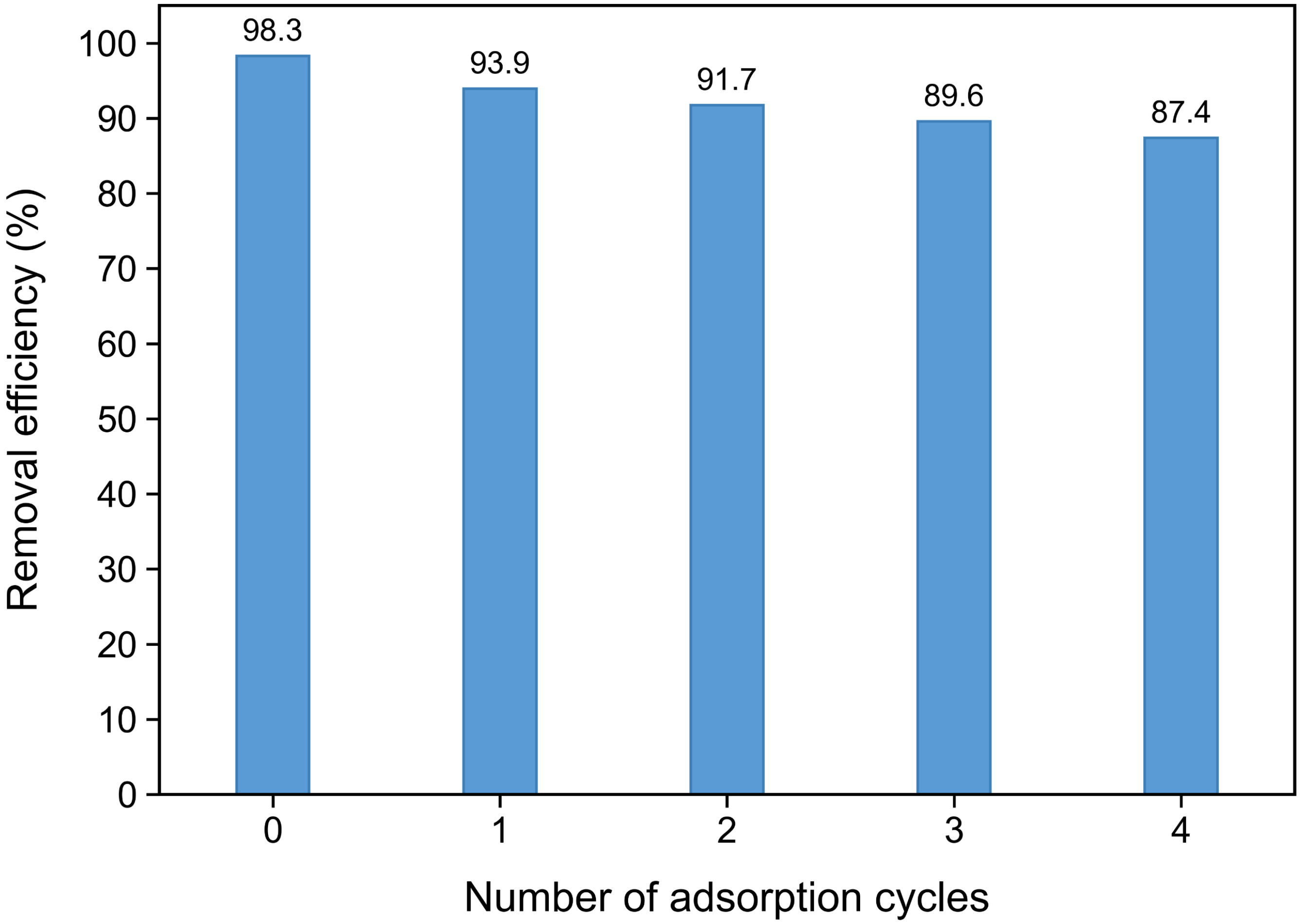

Figure S29. Reusability of the OPT MB sample over repeated adsorption–desorption cycles for methylene-blue removal. The removal efficiency decreased from 98.3% in the initial cycle to 87.4% after four reuse cycles, corresponding to retention of approximately 88.9% of the initial removal performance.

The retained removal efficiency after repeated adsorption–desorption cycles provides functional evidence of operational stability under the tested conditions. Previous MIL-100(Fe)-based aqueous adsorption studies have also reported negligible iron leaching and successful reuse, supporting the relevance of recyclability as a first-level indicator of practical water-treatment stability.

5.11 Supplementary linear embedding analysis via PCA

To complement the t-SNE visualization shown in the main text, principal component analysis (PCA) was also performed as a linear dimensionality-reduction check. Given the small number of samples in the present dataset, PCA provides a transparent linear projection of the same descriptor space. As shown in **Figures S30** and **S31**, the PCA representation gives a broadly consistent qualitative organization of the samples: the high-BET region remains associated with S11 and OPT BET, while OPT MB occupies a distinct lower-PC2 region closer to samples with stronger adsorption behavior. These PCA results therefore support the main conclusion of the t-SNE analysis, while providing a more conservative linear comparison.

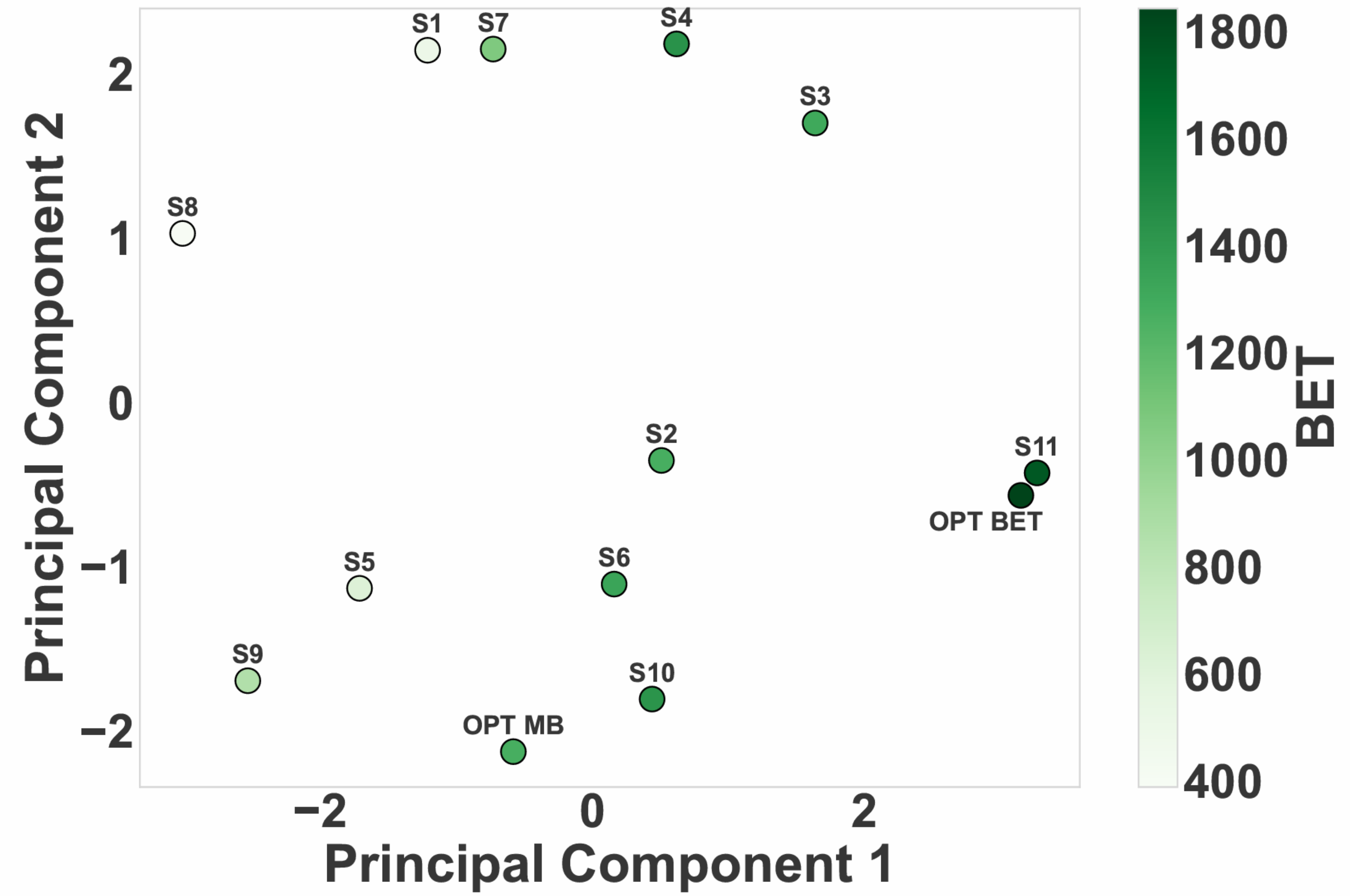

Figure S30. PCA projection of the original 11 experimental samples with the two optimized samples (OPT BET and OPT MB) overlaid, colored by BET surface area. The PCA provides a linear embedding of the descriptor space and shows that OPT BET remains positioned near the high-BET region associated with S11.

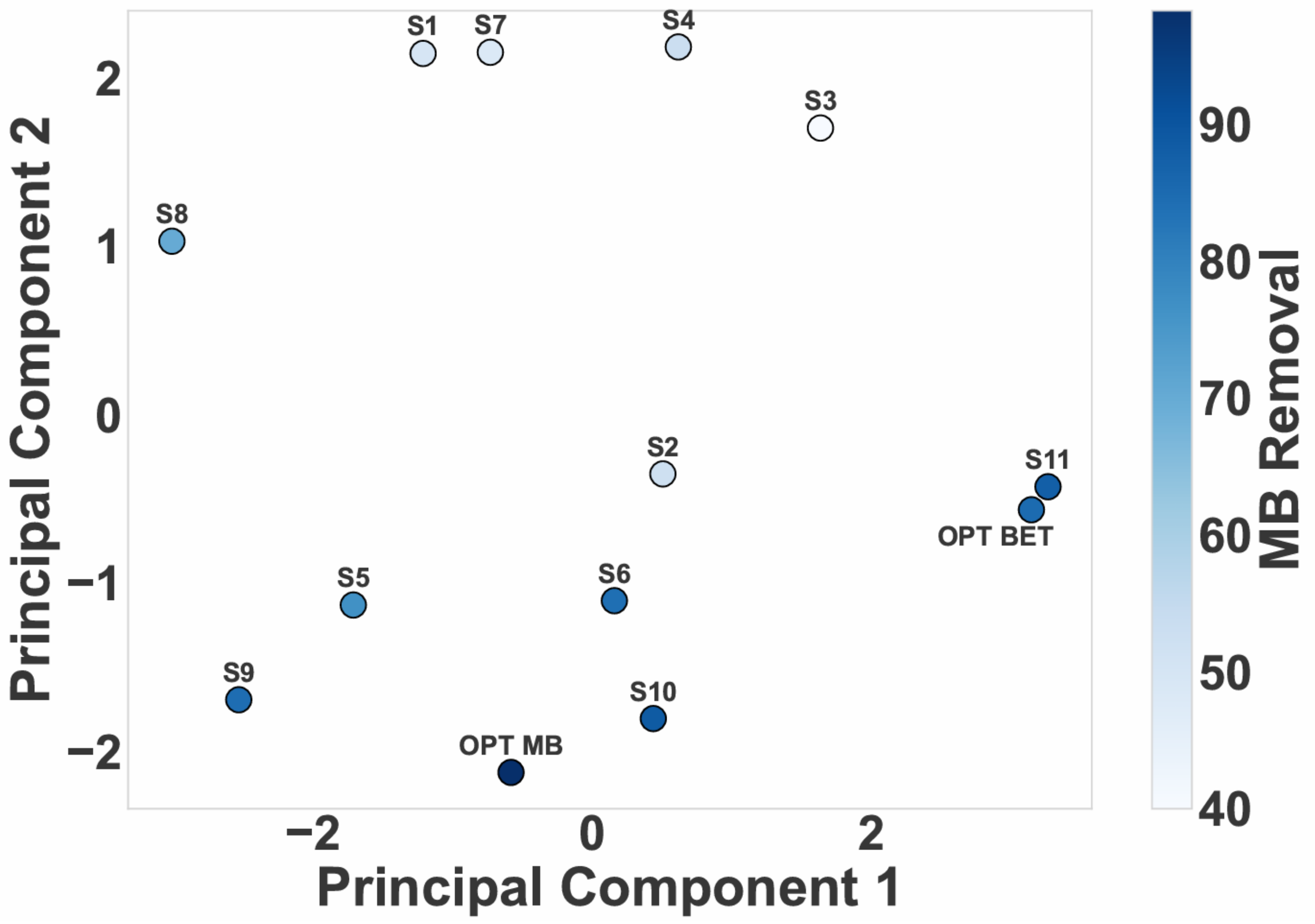

Figure S31. PCA projection of the original 11 experimental samples with the two optimized samples (OPT BET and OPT MB) overlaid, colored by MB removal efficiency. The PCA provides a linear embedding of the descriptor space and shows that OPT MB occupies a distinct lower-PC2 region closer to relatively strong adsorption samples such as S10.

5.12 Visualization of the experimental design space

To provide a direct visualization of the experimental design space, a lower-triangle pairwise scatter-matrix of the four synthesis variables was constructed (**Figure S32**). The figure shows the distribution of the 11 experimental runs across all pairwise combinations of synthesis time, temperature, metal-to-ligand molar ratio, and metal concentration. This visualization is intended to help the reader assess the spread of the experimental points and the overall coverage of the design space used for model development.

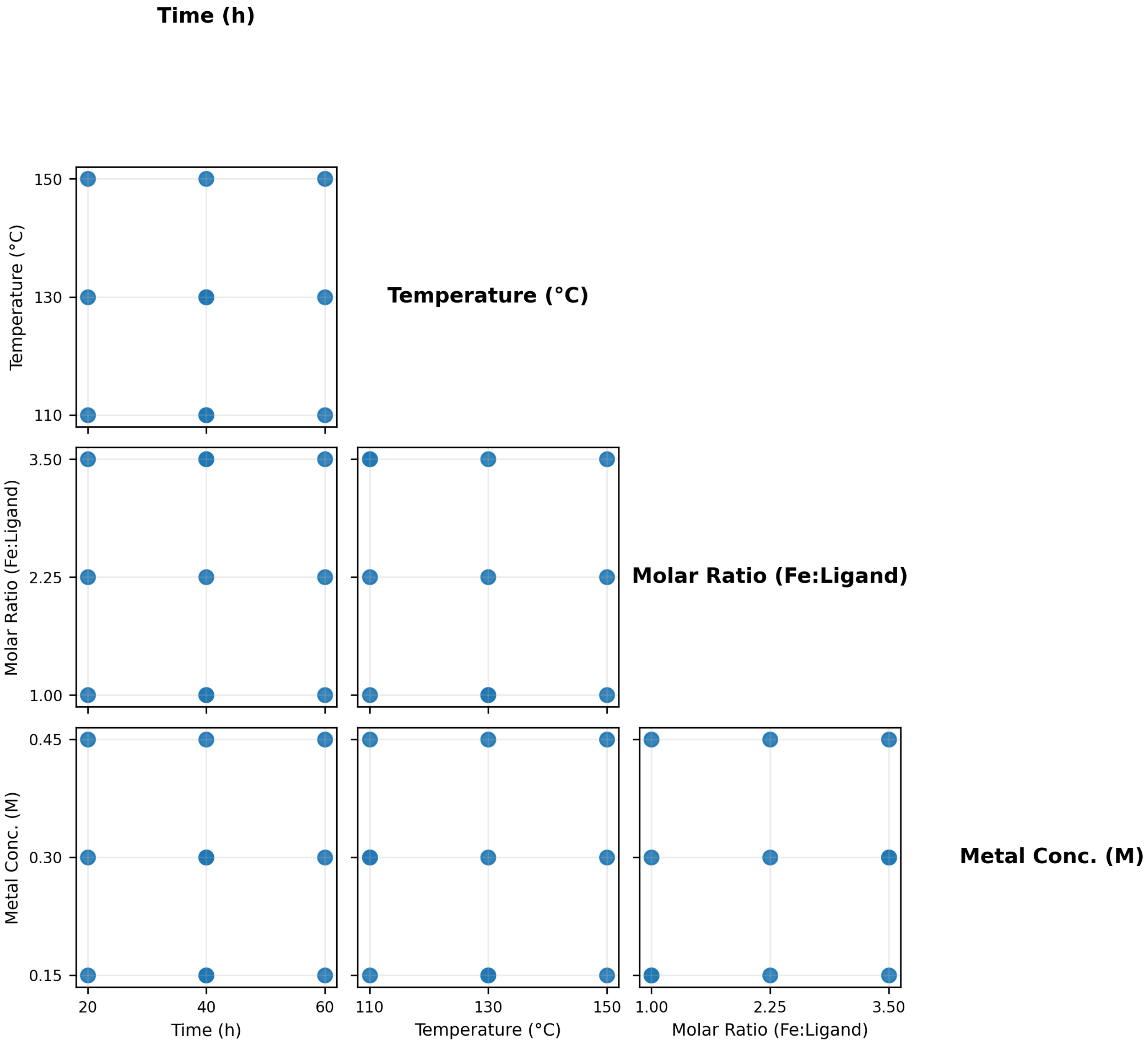

Figure S32. Lower-triangle pairwise scatter-matrix of the four synthesis variables used for the 11 experimental MIL-100(Fe) runs: synthesis time (h), temperature (°C), metal-to-ligand molar ratio (Fe:Ligand), and metal concentration (M). Each point represents one experimental sample. The figure visualizes the spread of the experimental runs across all pairwise combinations of the synthesis variables and illustrates the coverage of the explored design space.

5.13. Taguchi-based design structure and design-space visualization

The core experimental design was constructed using a Taguchi L9 orthogonal array for four synthesis variables at three levels: hydrothermal time, hydrothermal temperature, Fe:H_3BTC molar ratio, and Fe ion concentration. A full factorial design for these four three-level factors would require 81 experiments, whereas the L9 array provides a reduced but systematic set of nine experiments in which factor levels are distributed in a balanced manner. Samples S1–S9 correspond to the core Taguchi-based design. Samples S10 and S11 were added as supplementary non-orthogonal design points to broaden coverage of the synthesis-variable space and support the subsequent low-data modeling workflow. The pairwise scatter-matrix in **Figure S32** visualizes the distribution of the final 11 experimental runs across all pairwise combinations of synthesis variables. A brief summary existed in the **Table S5**.

Table S5. Factor levels used in the Taguchi-based experimental design for MIL-100(Fe) synthesis.

Factor	**Level 1**	**Level 2**	**Level 3**
Hydrothermal time (h)	20	40	60
Hydrothermal temperature (°C)	110	130	150
Fe:H_3BTC molar ratio	1:1	2.25:1	3.5:1
Fe ion concentration (M)	0.15	0.30	0.45

5.14 Leave-one-out cross-validation audit of GPR and baseline models for the two primary optimization targets

To complement the initial cross-validated model comparison in Table S1, an additional leave-one-out cross-validation (LOOCV) audit was performed for the two primary optimization targets: BET surface area and MB removal. This analysis was added to provide a more conservative evaluation for the 11-sample dataset. 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. No separate validation split was introduced inside the LOOCV audit because this would further reduce the already limited training set. A Dummy Regressor predicting the mean response of the training fold was evaluated under the same protocol as a constant-property baseline, and Linear Regression was included as a simple parametric baseline.

The LOOCV results are summarized in **Table S6** and **Figure S33**. For BET surface area, GPR improved over the Dummy Regressor across RMSE, MSE, and MAE. For MB removal, GPR reduced MAE relative to the Dummy Regressor and Linear Regression but did not reduce RMSE relative to the Dummy Regressor, indicating that the MB-removal target is more sensitive to individual held-out samples in the present low-data setting. For this reason, the MB-removal model is treated as a candidate-generation surrogate, and the final assessment of the optimized MB sample is based on experimental validation.

In addition to RMSE, MSE, and MAE, the coefficient of determination (R^2) was also calculated for the LOOCV predictions to facilitate comparison with standard regression reporting practice. However, for an 11-sample dataset, R^2 can be more sensitive to individual held-out points and to the variance of the target values than direct error-based metrics. For this reason, RMSE, MSE, and MAE are emphasized in the discussion, while R^2 is provided for completeness.

Table S6. Leave-one-out cross-validation (LOOCV) audit of Gaussian Process Regressor (GPR), Dummy Regressor, and Linear Regression for BET surface area and MB removal. In each LOOCV fold, one sample was held out as the test point and the remaining ten samples were used for model fitting. The Dummy Regressor predicts the mean response of the training fold and therefore serves as a constant-property baseline. For BET surface area, GPR improves over both baseline models across RMSE, MSE, and MAE. For MB removal, GPR reduces MAE relative to both baselines but does not reduce RMSE relative to the Dummy Regressor, indicating stronger sensitivity of MB removal to individual held-out samples in this small-data setting. R^2 is reported for completeness but is interpreted cautiously because, in this small-data setting, it is more sensitive to individual held-out points and target variance than the direct error metrics.

Target	Model	LOOCV RMSE	LOOCV MSE	LOOCV MAE	LOOCV R^2
BET surface area	Dummy Regressor	459.28	210941.12	399.67	**-0.2100**
BET surface area	Linear Regression	662.17	438474.15	565.00	**-1.5152**
BET surface area	Gaussian Process Regressor	403.93	163155.65	327.94	**0.0641**
MB removal	Dummy Regressor	19.34	373.89	18.19	**-0.2100**
MB removal	Linear Regression	23.03	530.54	17.49	**-0.7169**
MB removal	Gaussian Process Regressor	20.51	420.53	17.02	**-0.3609**

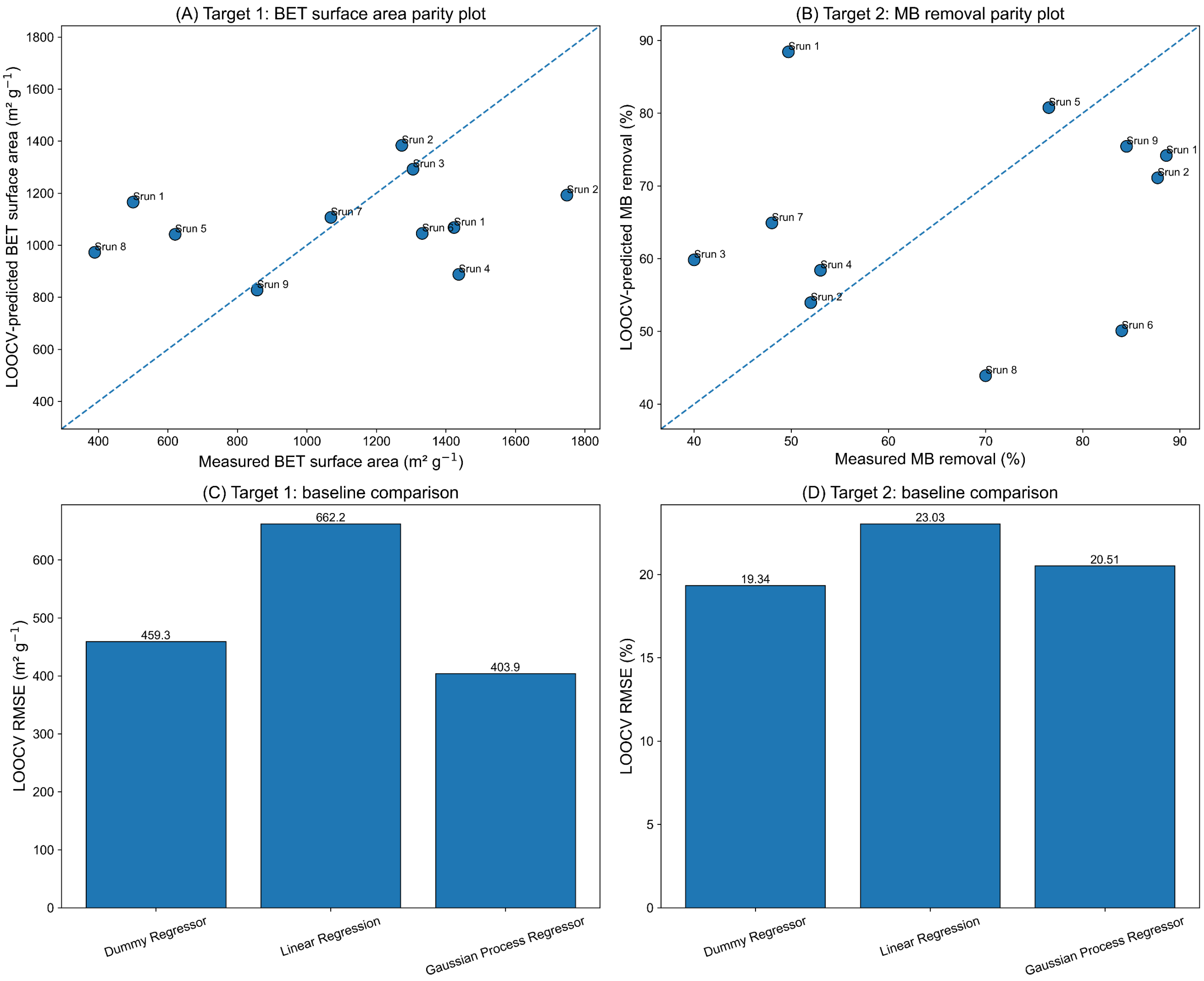

Figure S33. Leave-one-out cross-validation (LOOCV) audit of the GPR surrogate and baseline models for the two primary optimization targets. (A) Parity plot of measured versus LOOCV-predicted BET surface area using GPR. (B) Parity plot of measured versus LOOCV-predicted MB removal using GPR. (C) LOOCV RMSE comparison for BET surface area using Dummy Regressor, Linear Regression, and GPR. (D) LOOCV RMSE comparison for MB removal using the same models. In each fold, one sample was held out as the test point while the remaining ten samples were used for model fitting. The Dummy Regressor corresponds to the mean response of the training fold. Quantitative LOOCV metrics, including RMSE, MSE, MAE, and R², are reported in Table S6.

5.15 Final GPR model specifications used for the two primary targets.

Table S7. Final Gaussian Process Regressor (GPR) model specifications used for the two primary targets, BET surface area and MB removal. The kernel and hyperparameter settings are reported to make the GPR prior/model assumptions explicit and to improve reproducibility. The length-scale parameter controls the characteristic input-space variation represented by the kernel, while α\alphaα represents the numerical/noise regularization term used during fitting. These specifications were used for the final surrogate models applied in GA-guided candidate generation and for the corresponding LOOCV audit.

Target	Kernel	Length scale	Periodicity	(\alpha)	Normalize (y)	Optimizer	Restarts	Random state
Target 1: BET surface area	**ExpSineSquared**	**1.0**	**3.0**	(1.29×10^{-9})	**False**	**fmin_l_bfgs_b**	**3**	**49**
Target 2: MB removal	**ExpSineSquared**	**1.0**	**3.0**	**1.0**	**True**	**None**	**0**	**0**

To address the reviewer’s concern regarding GPR prior assumptions and hyperparameter specification, the final GPR settings used for the two primary targets are provided in **Table S6**. Reporting these settings makes the surrogate model fully reproducible and clarifies the kernel assumptions used during candidate generation. In the present work, these GPR models are interpreted as low-data empirical surrogates within the sampled MIL-100(Fe) synthesis space, rather than as universal predictors of synthesis–property relationships.

5.16. Post-reuse XRD analysis of OPT MB

To further evaluate the structural stability of OPT MB after repeated dye-removal operation, XRD patterns were compared before use and after four adsorption–desorption cycles (**Figure S34**). The post-reuse XRD pattern retains the main characteristic diffraction features observed for the original OPT MB sample, especially in the low-angle region associated with the MIL-100(Fe) framework. Although changes in relative peak intensity and baseline noise are observed after reuse, the persistence of the principal diffraction features indicates that the framework does not undergo complete crystallinity loss or catastrophic structural collapse during the cycling experiment. This post-reuse XRD result complements the recyclability data in **Figure S29** and supports the structural robustness of OPT MB under the tested adsorption–desorption conditions.

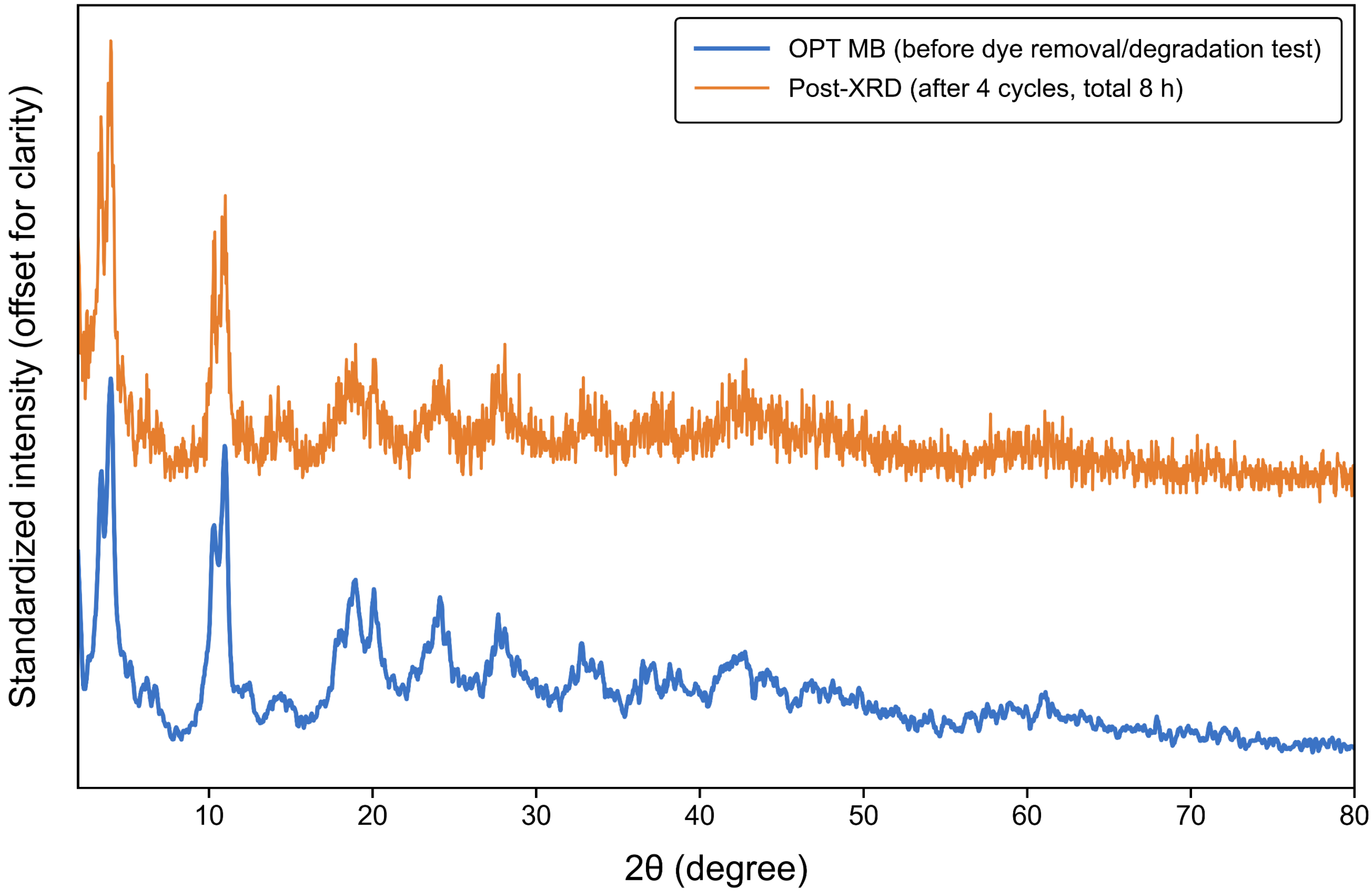

Figure S34. Standardized XRD patterns of OPT MB before dye-removal testing and after four adsorption–desorption cycles. The post-reuse pattern retains the main characteristic diffraction features of OPT MB, supporting preservation of the MIL-100(Fe) crystalline framework after repeated use. Patterns are vertically offset for clarity.

Code and data availability for ML reproducibility

To support reproducibility, the complete machine-learning dataset was converted into a CSV file containing the synthesis variables and measured target properties used in this study. The dataset includes the four synthesis inputs—reaction time, temperature, metal-to-ligand molar ratio, and metal concentration; together with the measured outputs reported in Table 3, including BET surface area, crystallinity, total pore volume, average crystallite size, yield, and MB removal. A separate primary-target CSV containing BET surface area and MB removal was used for the LOOCV audit presented in Table S6 and Table S7 and Figure S33.

The accompanying code repository includes scripts for generating the dataset, reproducing the LOOCV audit, and recreating the parity/baseline-comparison figure. The final GPR parameter files for the two primary optimization targets are also provided to make the surrogate-model specifications transparent and reusable.

References:

1. J. C. de Winter, S. D. Gosling and J. Potter, *Psychol Methods*, 2016, **21**, 273–290.
2. P. Schober, C. Boer and L. A. Schwarte, *Anesthesia & Analgesia*, 2018, **126**.
3. A. Rovetta, *Cureus*, 2020, **12**, e11794.
4. C. Xiao, J. Ye, R. M. Esteves and C. Rong, *Concurrency and Computation: Practice and Experience*, 2016, **28**, 3866–3878.
5. H. Li, L. Liu, J. Cui, J. Cui, F. Wang and F. Zhang, *RSC Advances*, 2020, **10**, 14262–14273.
6. F. Jeremias, S. K. Henninger and C. Janiak, *Dalton Transactions*, 2016, **45**, 8637–8644.
7. S. Mitra, S. Das, S. Basu, P. Sahu and K. Mandal, *Journal of Magnetism and Magnetic Materials*, 2009, **321**, 2925–2931.
8. D. K. Sannes, S. Øien-Ødegaard, E. Aunan, A. Nova and U. Olsbye, *Chemistry of Materials*, 2023, **35**, 3793–3800.
9. S. C. Moore, M. R. Smith, J. L. Trettin, R. A. Yang and M. L. Sarazen, *ACS Energy Letters*, 2023, **8**, 1397–1407.