Received: 4 April 2024 | Revised: 10 August 2024 | Accepted: 29 August 2024

DOI: 10.1002/aic.18611

AIChE JOURNAL

RESEARCH ARTICLE

Separations: Materials, Devices and Processes

Optimizing the prediction of adsorption in metal–organic frameworks leveraging Q-learning

Etinosa Osaro | Yamil J. Colón

Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana, USA

Correspondence
Yamil J. Colón, Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA.
Email: ycolon@nd.edu

Funding information
National Science Foundation, Grant/Award Number: CBET-2143346

Abstract

The application of machine learning (ML) techniques in materials science has revolutionized the pace and scope of materials research and design. In the case of metal–organic frameworks (MOFs), a promising class of materials due to their tunable properties and versatile applications in gas adsorption and separation, ML has helped survey the vast material space. This study explores the integration of reinforcement learning (RL), specifically Q-learning, within an active learning (AL) context, combined with Gaussian processes (GPs) for predictive modeling of adsorption in MOFs. We demonstrate the effectiveness of the RL-driven framework in guiding the selection of training data points and optimizing predictive model performance for methane and carbon dioxide adsorption, using two different reward metrics. Our results highlight the integration of RL as an AL method for adsorption predictions in MFs, and how it compares to a previously implemented AL scheme.

KEYWORDS
adsorption, Gaussian process, metal–organic frameworks, reinforcement learning

1 | INTRODUCTION

In recent years, the use of machine learning (ML) has been on the rise in the field of materials, from material discovery[1–4] to predictive tasks.[5–7] This surge in ML applications has significantly contributed to the accelerated pace of materials research and design, offering innovative solutions to challenges that were once considered insurmountable. Among the materials garnering attention, metal–organic frameworks (MOFs) have emerged as a particularly promising class of materials due to their tunable porous structures and versatile chemical composition.[8–10]

MOFs, characterized by their crystalline structures comprising metal nodes interconnected by organic ligands, can exhibit exceptional surface areas and tailorable functionalities. These attributes make MOFs highly desirable for applications in gas adsorption, separation, and storage.[10,11] However, the extensive database of MOFs[12–16] calls for the adoption of more advanced computational methods to efficiently screen these materials for their potential applications.

In the realm of ML, various methods have been utilized to screen MOFs and discern their adsorption behavior. These methods encompass support vector machines,[17,18] neural networks,[19–24] random forests,[17,25,26] among others. However, this paper directs its focus towards one particularly potent approach—reinforcement learning (RL). RL stands out by enabling an agent to learn from sequential experiences and adjust its strategies through trial and error, showcasing distinct capabilities for effectively exploring the extensive landscape of MOF-adsorption studies.

Optimizing the prediction of adsorption in MOFs through an RL framework offers a compelling approach to be studied within the context of active learning (AL) strategies: AL makes requests for data to be labeled as a surrogate model is being developed. Unlike conventional models, RL provides a dynamic method for selecting the training data set, which is vital in scenarios where the data set size is vast. Compared to other models, RL adaptively navigates through extensive data sets, iteratively identifying training points. Recently, RL has been used as a tool for the inverse design of MOFs for direct air capture of carbon dioxide.[27] In this study, the goal shifts from exploring

structural configurations to determining optimal training points for a Gaussian process (GP) model, which, in turn, predicts the adsorption isotherms of MOFs.

GPs provide a flexible and probabilistic framework for capturing complex relationships within data sets, taking features as inputs and have mean predictions and uncertainties as outputs.[28,29] In the context of predicting adsorption behaviors in MOFs, GPs offer a versatile approach to model adsorption isotherms, resulting in predictions of adsorption loading and uncertainties.[30–33] The integration of GPs into the proposed framework enhances the robustness and reliability of the predictive modeling, making it a key component in predicting the adsorption in MOFs.

Q-learning, also known as Quality learning, treats the selection of training points as a dynamic decision-making process.[34,35] Within this framework, the RL agent, representing the exploration strategy, learns to iteratively choose training points based on a reward metric, thereby optimizing its predictions over the agent's experience. This approach not only enhances the efficiency of predicting adsorption isotherms within MOFs but also offers a fresh perspective on leveraging RL in data-driven materials science. The integration of Q-learning into the RL-driven design framework signifies a departure from traditional methodologies, highlighting the model's adaptability to the complexities inherent in MOF adsorption studies. This research demonstrates the potential of Q-learning to redefine our approach to training predictive models for intricate phenomena like adsorption in MOFs.

In this study, we initially apply the Q-learning framework across eight distinct systems involving methane (CH_4) and carbon dioxide (CO_2) within two MOFs: Cu-BTC and IRMOF-1, at a temperature of 298K. This approach enables us to explore the applicability of the Q-learning framework across different systems, providing insights into its versatility and effectiveness. Subsequently, we extend our analysis by applying RL to a broader set of MOFs from the CoRE MOF database[16] for both gases, further expanding the scope of our investigation and evaluating the generalizability of our approach.

2 | METHODS

2.1 | Adsorption ground truth generation for Cu-BTC and IRMOF-1

The adsorption isotherms were generated through grand canonical Monte Carlo (GCMC) simulations with the software RASPA[36] at 298K. The non-bonded interactions in the MOFs were modeled using the Universal Forcefield (UFF), while the adsorbate molecules were modeled using the Transferable Potentials for Phase Equilibria (TraPPE) and using Lorentz Berthelot mixing rule.[37–39] MOF atoms were held fixed at their crystallographic positions and the charges for these two MOFs were not considered. The Monte Carlo (MC) moves were translation, rotation, reinsertion, and swap. A total of 200,000 cycles were executed, preceded by 100,000 initialization cycles.

2.2 | Adsorption ground truth generation for CoRE MOFs

The ground truth for carbon dioxide (CO_2) and methane (CH_4) was established through an AL protocol, consistent with our previous work, where we evaluated various initial training data selection schemes to predict full isotherms across 11 MOFs.[30] Building upon this methodology, we applied AL to generate complete adsorption points (64 data points) for both CO_2 and CH_4 within each MOF from the CoRE MOFs database.[16]

The AL process began with a log-spaced initial data, comprising pressures ranging from 1e−5 to 100 bar, and corresponding adsorption values generated from RASPA at 298K. This initial data set, consisting of 19 pressure points, served as the training data for the Gaussian process regression (GPR) model. Subsequently, predictions were made on an unlabeled data set, comprising the remaining pressure data points required for the full isotherms.

During the AL iterations, uncertainties from the GP predictions were collected and utilized to compute the GP relative error. This error metric guided the selection of the next data point to be added to the training data set. Specifically, the pressure data point corresponding to the highest relative error was passed to RASPA to compute the actual adsorption value, which was then incorporated into the training data set. The MOF atoms were also held fixed at their crystallographic positions and the charges were considered. The charges were taken from a study on partial charge assignment by Raza et al.[40] The MC moves and cycles are the same as in the previous section. This iterative process continued until the maximum GP relative error reached 0.01. By following this protocol, we successfully generated complete adsorption isotherms for both gases across MOFs in the CoRE MOF database.

2.3 | Q-learning

In this study, we utilize the Q-learning framework within an AL context to optimize the selection of training data points for adsorption studies. The RL agent begins with an initial limited data set and iteratively selects additional data points based on its learned policy. Only the labels of the selected data points are revealed and used to update the GP model, ensuring that the remaining data points' labels remain unknown to the GP model until selected by the agent. The reward metrics used in this study are the improvements in MRE or R^2 which directly measure the enhancement in predictive performance. This approach allows us to demonstrate the potential of RL in driving the AL process and optimizing model accuracy.

In this study, we utilize Q-learning as a strategic methodology to optimize the selection of training data points for predicting adsorption isotherms within MOFs. Q-learning, rooted in the principles of RL, operates within the framework of a Markov decision process, where initial and updated training data serve as states (s), and the selection of training data points represents actions (a). The fundamental

objective is to adapt the Q-learning algorithm iteratively to improve the efficiency of predicting adsorption loading in MOFs.

The Q-learning algorithm updates Q-values based on the Bellman equation:

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha[r + \gamma max_{a'} Q(s',a')]. \quad (1)$$

The learning rate (α), a critical parameter in Q-learning, dictates the magnitude of Q-value updates. A higher α facilitates rapid adaptation to new information but may risk overshooting optimal values, while a lower α provides stability but slower convergence. The immediate reward (r) denotes the feedback for a specific action in each state and guides the Q-value update process. In MOF adsorption, rewards are associated with the accuracy of predictions for selected training data points.

The discount factor (γ) determines the influence of future rewards on the Q-value update. A higher γ emphasizes long-term rewards, encouraging the agent to consider future consequences, while a lower γ prioritizes immediate rewards. The next state (s) reflects the updated training data after the agent's action, guiding the learning process. States (s) in the MOF adsorption context correspond to the available training data set, while actions (a) involve the selection of specific training data points for updating. The Q-learning agent systematically explores the total available data set by iteratively choosing actions that maximize the accuracy of predicted adsorption properties.

2.4 | States, actions, rewards, and states

2.4.1 | States

In the initial state (state 0) of the Q-learning process, the training data set is chosen to lay the foundation for subsequent exploration within the pressure–adsorption space study. The configuration of this initial training data set involves the selection of lower and upper bounds for pressures which are set at 1e−5 and 100 bar, respectively. This choice results in an initial training data set comprising just two data points.

2.4.2 | Actions

The actions in the Q-learning context refer to the selection of specific training data points. Each action influences the subsequent state and the agent's understanding of the environment. In the described methodology, the actions correspond to choosing data points for training the GP model, thereby refining its predictions. The action as performed by the agent is done through an exploration–exploitation tradeoff, which is a fundamental concept in RL and, specifically, Q-learning. It involves a delicate balance between exploring new actions and exploiting known high-reward actions. In this study, exploration entails selecting data points that have not been queried much, allowing the algorithm to gather more information about the data space. Exploitation, on the other hand, involves choosing actions based on the current knowledge to maximize short-term rewards. The tradeoff is managed through an epsilon-greedy strategy, where the agent occasionally explores new actions (with probability ε) and mostly exploits known high-reward actions.

Also, episodes represent discrete iterations of the Q-learning process. Each episode involves a sequence of actions, where the agent dynamically explores and exploits the possible actions. The Q-learning algorithm refines its strategies across episodes, updating the Q-values in the process. The goal is to iteratively improve the model's understanding of the environment and enhance the accuracy of adsorption predictions.

The agent is the decision-making entity within the Q-learning algorithm. It interacts with the environment by selecting actions and updating its understanding of the MOF adsorption space. The Q-table, a critical component, maintains Q-values for each state-action pair. The Q-values represent the expected cumulative rewards the agent anticipates from choosing a specific action in each state. During exploration and exploitation, the agent relies on the Q-table to inform its decision-making. The action selected by the agent is often the one with the highest Q-value in each state. The dynamics of the Q-table are shaped by the Q-learning update rule, which incorporates immediate rewards and anticipated future rewards. The agent learns from sequential experiences, adjusting its strategies through trial and error. By maximizing the Q-values, the agent hones its decision-making skills, progressively selecting actions that lead to higher rewards and, consequently, more accurate predictions of adsorption isotherms.

2.4.3 | Rewards

In the context of RL, rewards play a pivotal role in guiding the learning process of an agent. Rewards in RL encapsulate the desirability or utility associated with the outcomes of an agent's actions. They act as a form of reinforcement, providing feedback to the agent and influencing its decision-making processes. In the Q-learning approach tailored for MOFs, the purpose of rewards is to encourage the agent to learn a policy that maximizes cumulative rewards over time.

In this methodology, rewards are evaluated using two metrics: the improvement in the R^2 score or mean relative error (MRE) between predicted and actual adsorption data. The reward (r) is calculated based on the difference in MRE or R^2 between states t and $t+1$, reflecting the enhancement in predictive accuracy upon adding a selected data point to the training set. These rewards drive the update of Q-values within the training loop, influencing the agent's decision-making process.

In summary, the methodology incorporates a reward-driven approach within the Q-learning framework, aligning the agent's learning objectives with the goal of iteratively enhancing its decision-making strategies. By prioritizing actions that contribute positively to the model's accuracy, the methodology aims to achieve refined predictions of adsorption isotherms in MOFs.

2.4.4 | States

In the completion of each episode, as outlined above, the addition of a training data point triggers the formation of a new state within the Q-learning framework. The concept of episodes, crucial to the iterative learning process, encapsulates a sequence of actions taken by the agent as it dynamically explores and exploits the MOF adsorption space.

2.5 | Gaussian process regression

GPs provide a probabilistic and nonparametric approach to regression, allowing for the modeling of complex relationships between input parameters and the corresponding output values. In the GP framework, functions are considered random variables, and the joint distribution over functions is governed by a multivariate normal distribution. The choice of mean function (m) and the covariance function (K) defines the GP's prior distribution, reflecting beliefs about the functions before observing any data. This can be mathematically described in Equation (1).

$$f \approx GP(m(x), K(x,x')). \tag{2}$$

For this study, the covariance function (kernel) implemented was the rational quadratic kernel[28,41,42] as described by the mathematical form in Equation (2):

$$K(x,x') = \left(1 + \frac{(x-x')^2}{2\alpha l^2}\right)^{-\alpha}. \tag{3}$$

The kernel is defined by $x - x'$ representing the Euclidean distance between x and x', while l denotes the length scale parameter, which defines the characteristic length over which variations in the function occur. Additionally, α influences the balance between large-scale and small-scale fluctuations within the function by adjusting their respective weights. The Scikit-learn library was used for the GP implementation.[43]

In the Q-learning methodology, a surrogate model is used in approximating the adsorption behavior and guiding the agent's decision-making process. The GP, acting as this surrogate model, is trained on the available data, consisting of selected training points fed into the model during the Q-learning episodes. As the Q-learning algorithm explores the ground truth data, the GP model refines its predictions based on the newly acquired training data points. The agent's decisions, influenced by the rewards obtained during exploration, guide the augmentation of the GP model, enhancing its accuracy in predicting adsorption isotherms. The integration of GPs within the Q-learning loop ensures a continuous interplay between exploration, exploitation, and model refinement. The surrogate GP model becomes an adaptive guide, steering the agent toward selecting informative training data points.

To ensure the efficiency and accuracy of the RL-GP method, we implemented two key adjustments. First, initial data points used for training are excluded from the list of possible actions at the start of the training process. This prevents the RL agent from selecting these points again, ensuring that each selected point contributes new information to the model. Second, we implemented a mechanism to track evaluated data points and remove them sequentially from the action list. This prevents repeated selection and ensures continuous exploration of new areas in the data space.

In Figure 1 below, we show the sequence of the RL algorithm involving the states, actions, rewards, and the GP in use.

In our modeling approach, it is important to acknowledge the role of random seeds in influencing the outcomes of the exploration process. The random seed affects the selection of exploration points in the data space. This initial setting can lead to variability in which points are selected for exploration, potentially impacting the learning trajectory and the efficiency with which the model converges to an optimal solution. Recognizing this, we used a random seed of 42, enabling reproducibility. However, in Section 3, we offer a comprehensive analysis of different random seeds.

3 | RESULTS AND DISCUSSIONS

In this section of the study, our attention was directed toward the utilization of Q-learning to scrutinize the adsorption characteristics of two gases, CH_4 and CO_2, within two distinct MOFs: Cu-BTC and IRMOF-1. Our investigation encompassed the exploration of various hyperparameters, including the learning rate, the tradeoff between exploration and exploitation (ε), and discount factor, to discern their impact on the learning process. Table 1 offers an overview of the hyperparameters employed in the Q-learning exercises, facilitating a systematic evaluation of their efficacy across different adsorbate–adsorbent pairs. The primary objective was to identify the optimal combination of parameters tailored to each adsorbate–adsorbent pair in this paper.

In this investigation, the Q-learning algorithm was capped at a maximum of 10 episodes. We evaluated two reward metrics: the difference in MRE between consecutive states, and the difference in R^2 values between consecutive states. The MRE reward metric demonstrated superior performance compared to the R^2 reward metric. Consequently, we primarily focus on the MRE metric in this paper, while presenting results for the R^2 metric in the Supporting Information (SI). The determination of the optimal hyperparameter combination relied on the achieved final MRE for the MRE reward metric and the final R^2 values for the R^2 reward metric.

3.1 | MRE as reward metric

In this section, rewards are based on MRE as calculated between predicted adsorption and actual adsorption data. The reward (r) is

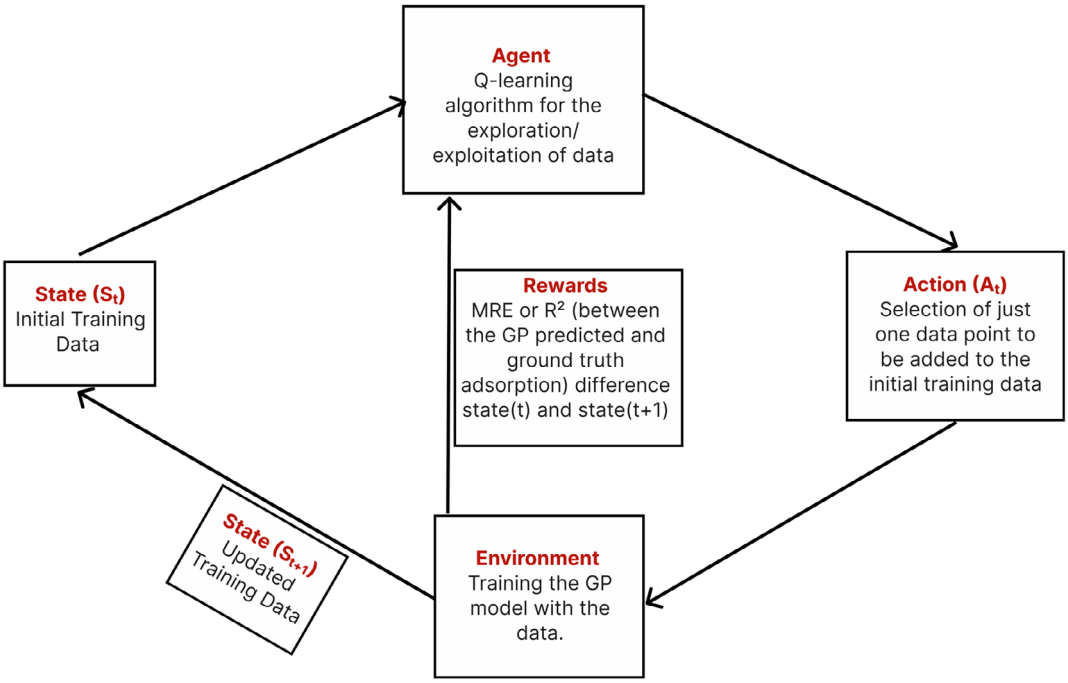

FIGURE 1 Q-learning Schematic: The diagram illustrates the Q-learning process. The initial state (t) represents the starting training data, where the agent dynamically balances exploration and exploitation influenced by the discount factor. The agent takes action by selecting a data point from the ground truth, updating the prior data set. Subsequently, these new data points contribute to the training of the GP model. The reward, quantified as the metrics (MRE or R^2) difference between the current state (t) and the next state ($t+1$), is computed. GP, Gaussian process; MRE, mean relative error.

TABLE 1 Q-learning hyperparameters.

Learning rate (α)	Discount factor (γ)	Epsilon (ϵ)
0.001	0.1	0.1
0.01	0.5	0.5
0.1	0.9	0.9

calculated as the difference in MRE between states t and $t+1$. Once the reward has been calculated, it is fed into Equation (1) to construct the Q-table.

The MRE can be calculated as described by the equation:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{|y_{\text{actual}_i}-y_{\text{predicted}_i}|}{y_{\text{actual}_i}}. \quad (4)$$

3.1.1 | Q-learning on two pure gases within two MOFs

For CH_4 adsorption in Cu-BTC, the optimal hyperparameter combination was determined to be a learning rate (α) of 0.001, a discount factor (γ) of 0.1, and an exploration/exploitation (ϵ) value of 0.9, with 10 episodes.

Figure 2 illustrates the outcomes for this specific adsorbate–adsorbent pair based on the identified best hyperparameter combinations, as stated above. In Figure 2A, the evolution of the MRE between the GP-predicted adsorption and ground truth adsorption is depicted as Q-learning episodes progress. Notably, the algorithm achieved a final MRE of 0.043. Figure 2B presents a comparative analysis between the predicted and ground truth adsorption, demonstrating the model's predictive performance, where an R^2 of 0.99 and MAE of 0.237 cm^3 STP/g was achieved. Additionally, Figure 2C visualizes the pressure space sampled by the RL agent. Finally, Figure 2D shows a good comparison between the simulated isotherm and the GP-predicted isotherm, as directed by Q-learning.

For the case of CH_4 in IRMOF-1, a final MRE of 0.044 was achieved by the Q-learning agent after the maximum episodes of 10. The best Q-learning parameters were found to be α of 0.001, γ of 0.5, and ϵ of 0.9. In Figure 3A, we show the evolution of MRE with the episodes, with a significant rise in the MRE in the second episode. Additionally, Figure 3B displays the resulting comparison between the GCMC and GP-predicted adsorptions, showing good agreements and a R^2 of 0.99 and MAE of 0.763 cm^3 STP/g was achieved. In Figure 3C, we observe the pressure-sampled regions by the agents showing a distribution across the pressure space. Finally, Figure 3D also shows a good comparison between the simulated isotherm and the RL-based GP-predicted isotherm.

CO_2 adsorption prediction in both Cu-BTC and IRMOF-1 were optimized in the same number of Q-learning episodes of 10, to achieve desirable outcomes. Specifically, in the case of Cu-BTC, the optimal hyperparameter combination was determined to be a α of

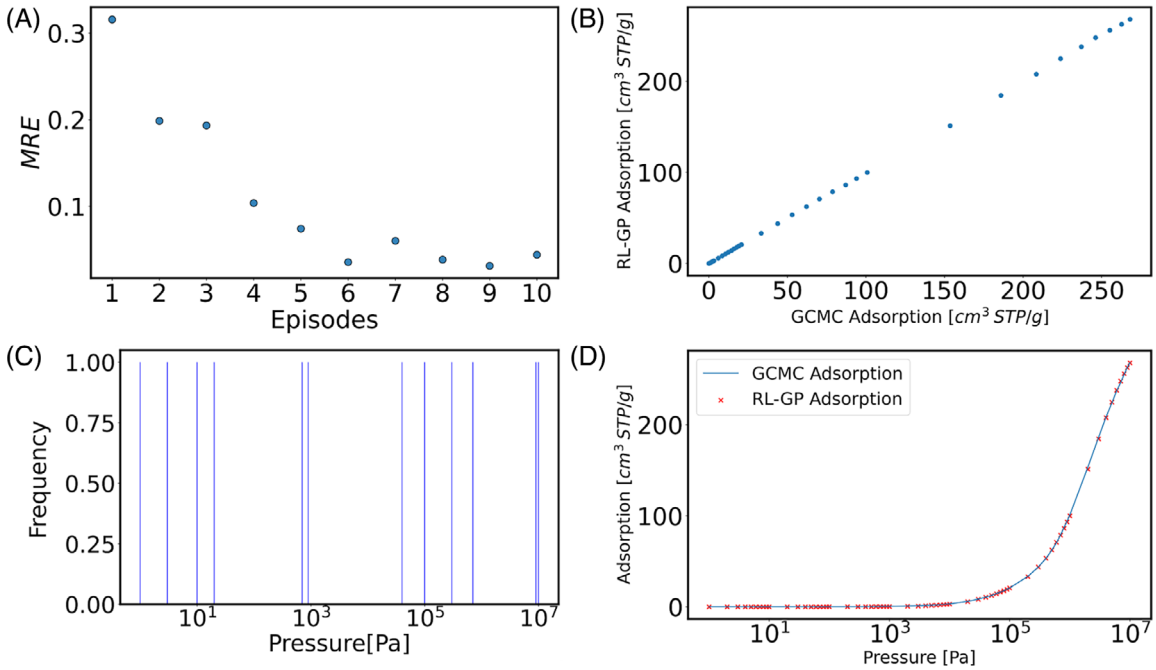

FIGURE 2 Results for CH_4 in Cu-BTC. (A) Evolution of MRE with the Q-learning episodes. (B) Comparison between GCMC adsorption and the RL-GP predicted adsorption values. (C) RL sampled region of pressures by the Q-learning agent for the final GP model. In addition to the initial training data, 10 more pressure points were sampled by the RL Scheme making a total of 12 pressure points. (D) GCMC and RL-GP predicted isotherms comparison. GCMC, grand canonical Monte Carlo; GP, Gaussian process; MRE, mean relative error; RL, reinforcement learning.

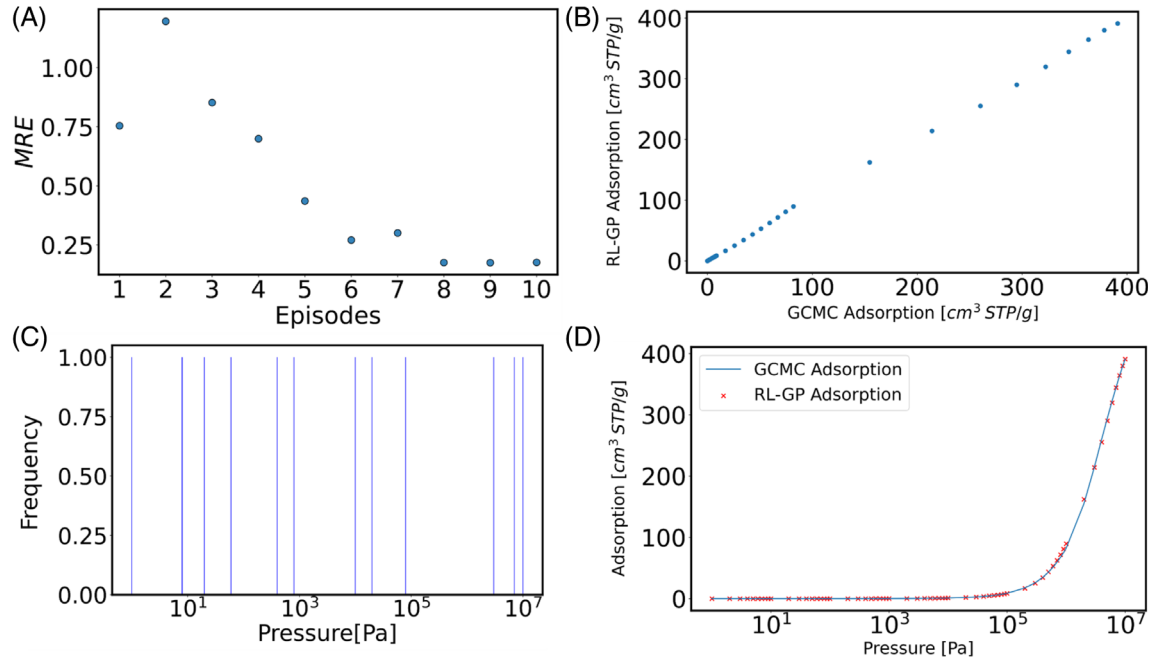

FIGURE 3 Results for CH_4 in IRMOF-1. (A) Evolution of MRE with the Q-learning episodes. (B) Comparison between GCMC adsorption and the RL-GP predicted adsorption values. (C) RL sampled region of pressures by the Q-learning agent for the final GP model. In addition to the initial training data, 10 more pressure points were sampled by the RL Scheme making a total of 12 pressure points. (D) GCMC and RL-GP predicted isotherms comparison. GCMC, grand canonical Monte Carlo; GP, Gaussian process; MRE, mean relative error; RL, reinforcement learning.

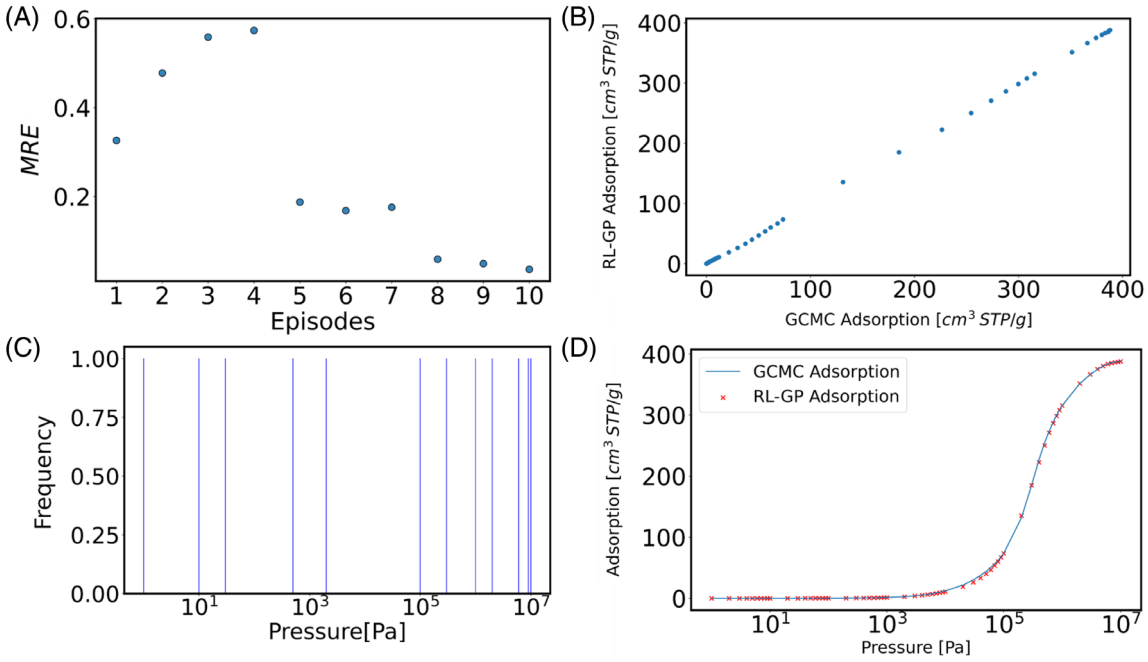

FIGURE 4 Results for CO_2 in Cu-BTC. (A) Evolution of MRE with the Q-learning episodes. (B) Comparison between GCMC adsorption and the RL-GP predicted adsorption values. (C) RL sampled region of pressures by the Q-learning agent for the final GP model. In addition to the initial training data, 10 more pressure points were sampled by the RL Scheme making a total of 12 pressure points. (D) GCMC and RL-GP predicted isotherms comparison. GCMC, grand canonical Monte Carlo; GP, Gaussian process; MRE, mean relative error; RL, reinforcement learning.

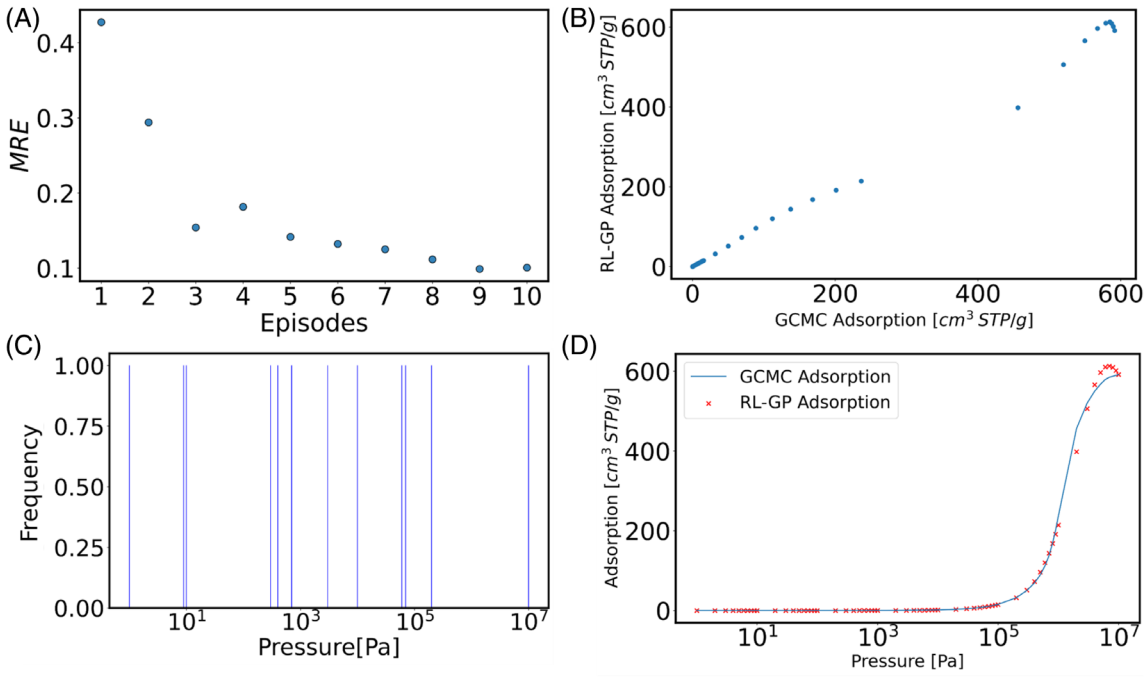

FIGURE 5 Results for CO_2 in IRMOF-1. (A) Evolution of MRE with the Q-learning episodes. (B) Comparison between GCMC adsorption and the RL-GP predicted adsorption values. (C) RL sampled region of pressures by the Q-learning agent for the final GP model. In addition to the initial training data, 10 more pressure points were sampled by the RL Scheme making a total of 12 pressure points. (D) GCMC and RL-GP predicted isotherms comparison. GCMC, grand canonical Monte Carlo; GP, Gaussian process; MRE, mean relative error; RL, reinforcement learning.

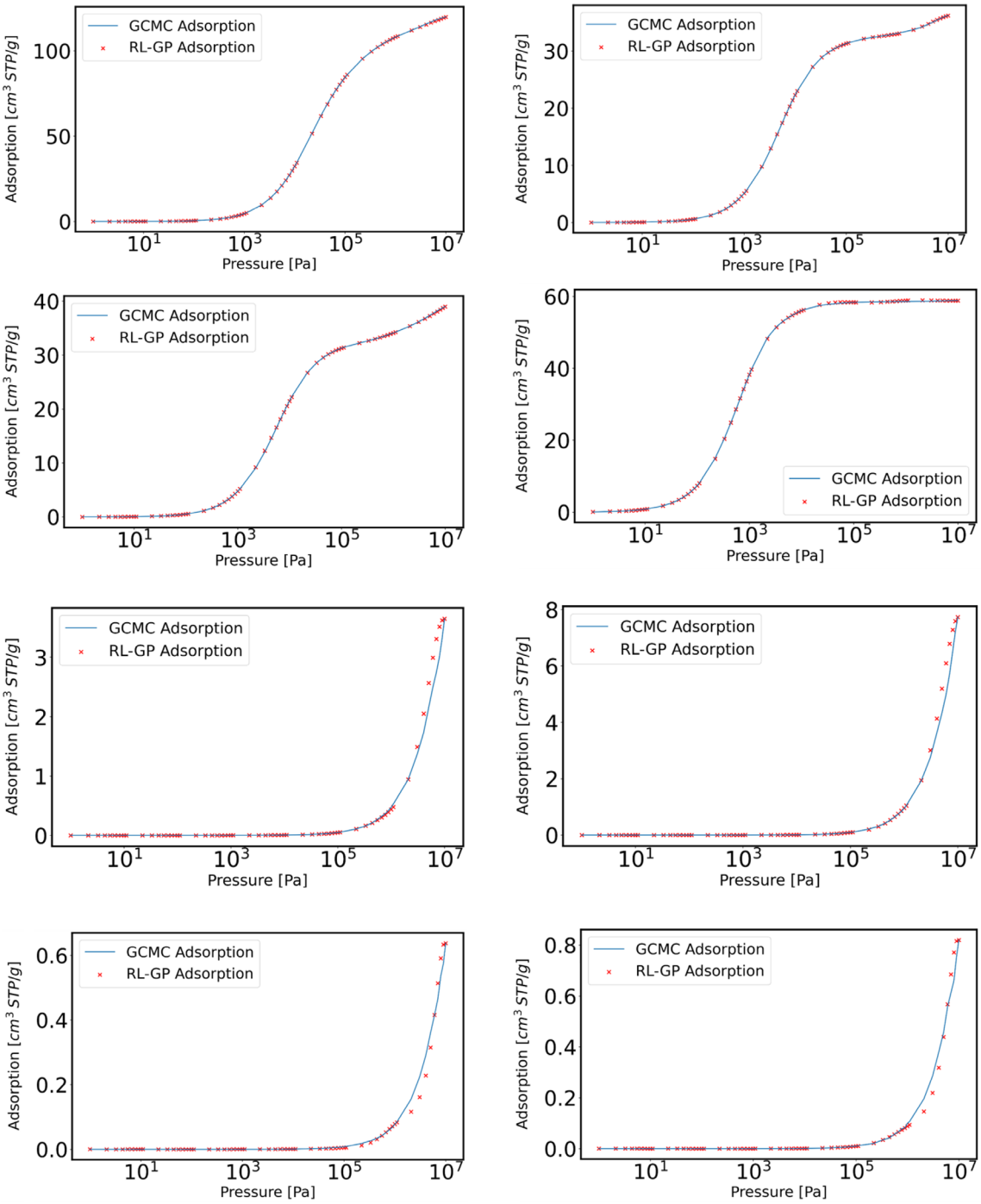

FIGURE 6 The comparison between RL-based Gaussian process (GP) predicted and actual CH_4 isotherms. The plot showcases the first four MOF models with the lowest MRE values, followed by the next four MOF models with the highest MRE values. MOF, metal–organic framework; MRE, mean relative error; RL, reinforcement learning.

0.001, γ of 0.9, and an ϵ value of 0.9, with 10 episodes. In Figure 4A, a final MRE value of 0.046 was achieved after a consistent increase in MRE from the second to fourth episodes, followed by a drop in MRE from the fifth episode, indicating the dynamic nature of the learning process. Furthermore, Figure 4B showcases a comparison between GP-predicted and GCMC adsorption in Cu-BTC, demonstrating commendable agreement between the two. An R^2 of 0.99 and MAE of 0.764 cm^3 STP/g. Figure 4C provides insights into the pressure space sampled by the agent, highlighting the agent's ability to determine the optimal points. In Figure 4D, we show a good prediction of the isotherm as compared to the GCMC-simulated isotherm, indicating that the RL-based GP can predict the isotherm.

Similarly, for IRMOF-1, the optimal hyperparameter combination was determined to be a α of 0.001, γ of 0.5, and ϵ value of 0.9, with

10 episodes. A final MRE of 0.1 was accomplished by the last episode. The evolution of MRE scores is depicted in Figure 5A with notable increases in MRE observed during the fourth episode, which shows the iterative nature of the learning process. Figure 5B illustrates the comparison between GP-predicted and GCMC adsorption in IRMOF-1, revealing robust agreement between the two predictions, with failure in the highest-pressure regions. This failure is evident in Figure 5C, where the algorithm fails to explore high-pressure regions. In Figure 5C, the pressure space sampled by the agent in IRMOF-1 indicates an exploration of the diverse pressure regions. In Figure 5D, we show a good agreement between the simulated and RL-GP predicted isotherms. An R^2 of 0.99 and MAE of 4.232 cm^3 STP/g was achieved.

3.1.2 | Q-learning on CH_4 gas within CoRE MOFs

The optimal Q-learning parameters established for CH_4 adsorption in Cu-BTC and IRMOF-1 were α of 0.001, γ of 0.5, and s of 0.9. Using these established parameters, we extended the RL framework to the CoRE MOF database,[16] encompassing 8671 MOFs using those parameters.

The initial state is the initial training data containing the minimum (1 Pa) and maximum pressure (1E7 Pa) and their corresponding adsorption. Employing this RL criteria, we conducted RL for each structure within the database, enabling a comparison between RL-GP predicted adsorption and actual adsorption values across various MOFs within the data set. The actual adsorption values are obtained from an AL-based method as described in Section 2.2, and we recognize the potential for inaccuracies in the surrogate AL-GP models trained on few GCMC data. Figure 6 visualizes the comparison for select MOFs within the database, shedding light on the efficacy of RL in predicting adsorption. The initial four frames showcase the top-performing MOFs with the lowest MRE, while the subsequent four frames present the worst-performing MOFs characterized by the highest MRE values. Upon aggregating all predictions across the entirety of MOFs, we achieved a conclusive R^2 value of 0.99, as depicted in Figure 7A. Additionally, the mean absolute error was 0.113 cm^3 STP/g, which indicates a good performance of the model. Upon delving deeper, we examined the critical points sampled by the RL agent, revealing a diverse pattern in their distribution across the pressure input space. This sampling of several pressures signifies the RL agents' adeptness in identifying pivotal regions influencing adsorption behavior. By strategically targeting certain pressures (particularly the lower and higher-pressure regions) regions, as shown in Figure 7B, RL agents facilitate the acquisition of crucial data points essential for constructing accurate GP models.

3.1.3 | Q-learning on CO_2 gas within CoRE MOFs

Given the established Q-learning parameters for CO_2 from IRMOF-1 and Cu-BTC, the RL scheme is applied to the same CoRE MOFs as in the previous sections. Figure 8 below shows the isotherms of some selected MOFs for CO_2. The first four frames indicate the best performance in terms of lowest MRE(s) and the next four frames are the least performing cases in terms of highest MRE(s).

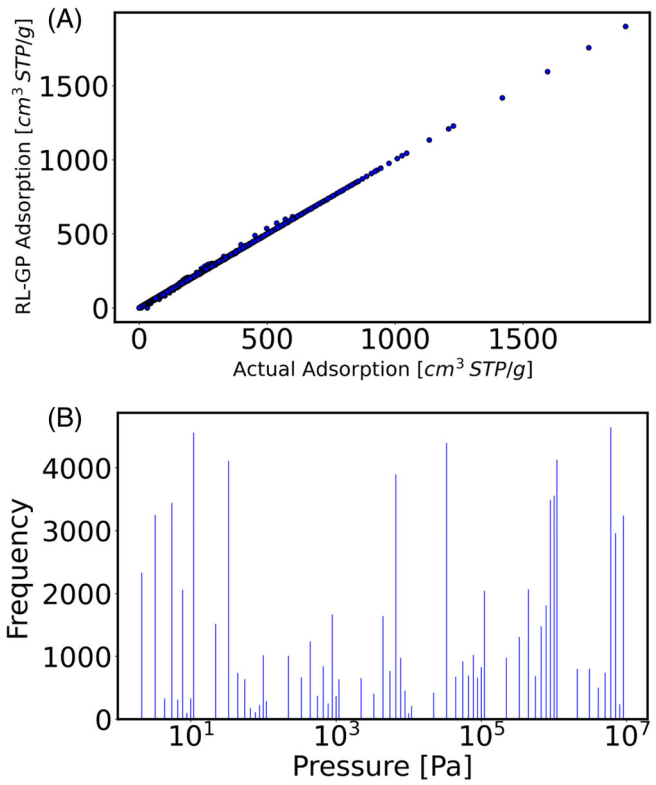

FIGURE 7 (A) Comparison between the CH_4 actual adsorption and the RL-GP predicted adsorption across the entire CoRE MOFs structures studied in this work. The final R^2 and mean absolute error attained is 0.99 and 0.113 cm^3 STP/g. (B) Pressure space exploration for CH_4. The data points collected by the RL agent across all 8671 MOFs are utilized for constructing individual GP models. Many of these points cluster towards the low-pressure and high-pressure regions, with fewer points distributed in other pressure regions. These points exclude the initial state points of 1 and 1E7 Pa. GP, Gaussian process; MOF, metal–organic framework; MRE, mean relative error; RL, reinforcement learning.

In the case of CO_2, our analysis of the combined actual and RL-based GP predictions across all MOFs yielded an R^2 value of 0.99, as depicted in Figure 9A below. The MAE stood at 0.4667 cm^3 STP/g. The differences in R^2 and MAE values between CO_2 and CH_4 can be attributed to the stronger interactions of CO_2, leading to more complex adsorption isotherms with sharper transitions. This complexity poses a challenge for accurate prediction, potentially resulting in higher MAE value, compared to CH_4. Upon closer examination, as depicted in Figure 9B, we analyzed the critical points sampled by the RL agent, revealing a distribution across the pressure input space, particularly in the lower-pressure regions. The convergence in sampling behavior underscores the robustness of the RL-guided exploration approach in identifying critical regions within the pressure space with significant implications for adsorption behavior.

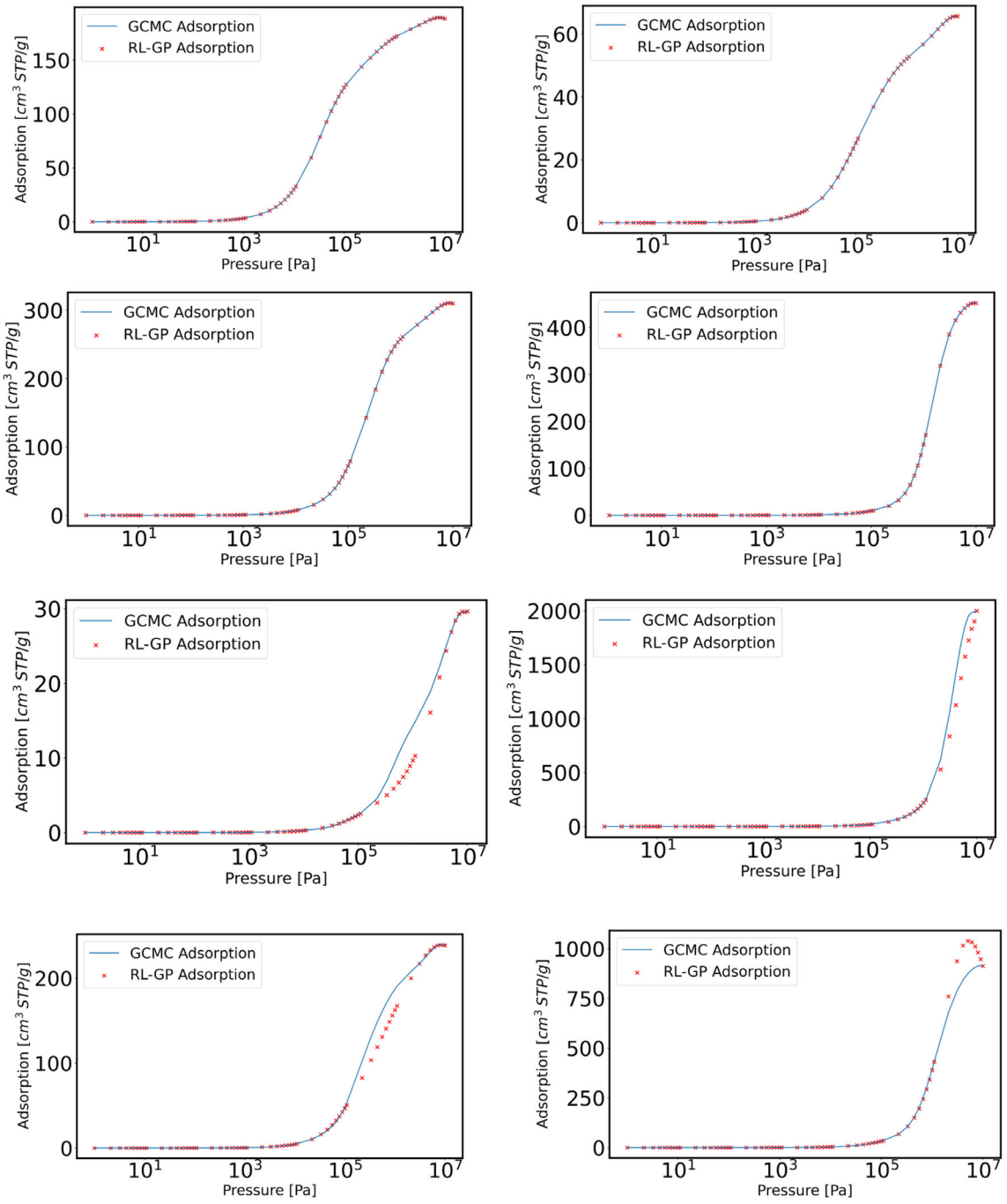

FIGURE 8 The comparison between RL-based Gaussian process (GP) predicted and actual CO_2 isotherms. The plot showcases the first four MOF models with the lowest MRE values, followed by the next four MOF models with the highest MRE values. MOF, metal–organic framework; MRE, mean relative error; RL, reinforcement learning.

In the SI, we show the results of using an R^2 reward metric in the sequence used in this paper. First, we applied the RL framework to CH_4 both in Cu-BTC and IRMOF-1 and used the optimal Q-learning parameters (based on highest R^2) in the case of the CoRE MOFs.

For CH_4 adsorption in Cu-BTC (Figure S1), we determined the optimal combination of hyperparameters to be $\alpha = 0.001$, $\gamma = 0.9$, and $\epsilon = 0.9$. For the case of CH_4 in IRMOF-1 (Figure S2), the parameters $\alpha = 0.001$, $\gamma = 0.1$, and $\epsilon = 0.9$, yielded the best results.

The results for CO_2 in Cu-BTC and IRMOF-1 are shown in Figures S5 and S6. The optimal parameters were $\alpha = 0.001$, $\gamma = 0.1$, and $\epsilon = 0.9$ for CuBTC and parameters $\alpha = 0.001$, $\gamma = 0.1$, and $\epsilon = 0.5$, for IRMOF-1.

Based on the consistency of the hyperparameters ($\alpha = 0.001$, $\gamma = 0.1$, and $\epsilon = 0.9$) across the four systems, we employed the established Q-learning parameters for CH_4 and CO_2 to the 8671 CoRE MOFs. For both CH_4 and CO_2, RL was conducted for each structure.

For CH_4, the combined actual and RL-based GP predictions yielded an R^2 of 0.99, with an MAE of 0.278 cm^3 STP/g (Figure S4A), which is higher than the MAE value realized using the MRE reward metric. For CO_2 (Figure S8A), R^2 of 0.99 and MAE of 0.914 cm^3 STP/g.

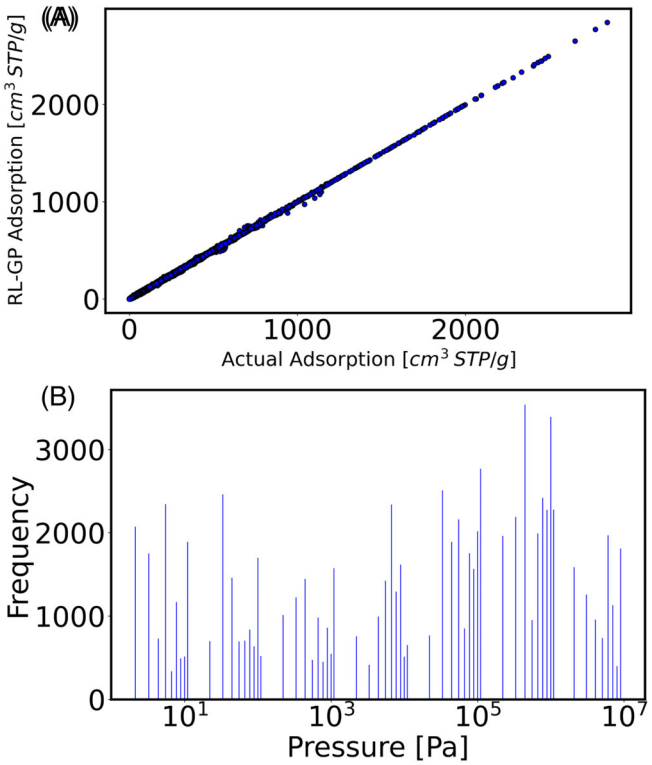

FIGURE 9 (A) Comparison between the CO_2 actual adsorption and the RL-GP predicted adsorption across the entire CoRE MOFs structures studied in this work. The final R^2 and mean absolute error attained is 0.99 and 0.4667 cm^3 STP/g framework. (B) Pressure space exploration for CO_2. The data points collected by the RL agent across all 8671 MOFs are utilized for constructing individual GP models. Many of these points cluster toward the low-pressure and high-pressure regions, with fewer points distributed in other pressure regions. These points exclude the initial state points of 1 and 1E7 Pa. GP, Gaussian process; MOF, metal–organic framework; MRE, mean relative error; RL, reinforcement learning.

3.1.4 | Influence of random seed number on RL results

As mentioned in Section 2, the choice of random seed plays a crucial role in shaping the outcomes of the RL framework, particularly in the process of selecting exploration points. While this paper maintains consistency by using random seed 42, we conducted a comprehensive analysis to evaluate the influence of 1000 distinct random seeds across four systems: CH_4 in CuBTC and IRMOF-1, as well as CO_2 in CuBTC and IRMOF-1. Table 2 below offers a comparison between the performance metrics (MAE, MRE, and R^2) obtained using seed 42 and the averaged metrics derived from the 1000 random seeds. These results shed light on the robustness of our approach, indicating that variations in the choice of random seed have minimal impact on the overall outcomes. This observation underscores the stability and reliability of the RL framework in selecting exploration points across diverse systems.

In the SI, we show a table of the pressure points selected for five different MOFs from the CoRE MOFs, for five different random seeds, showing the impact of the random seed selection.

3.1.5 | Performance comparison of AL and RL: CH_4 and CO_2 adsorption in Cu-BTC and IRMOF-1

In this section, we compare the outcomes of AL and RL (random seed of 42) on CH_4 and CO_2 adsorption in two MOFs: Cu-BTC and IRMOF-1. Initially, a training data set with two pressure points (1 and 1E7 Pa) was utilized to train a GP model. Predictions were made on an unlabeled data set, and uncertainties were computed, followed by calculating the relative error. The unlabeled data set with the highest relative error was selected, and actual adsorption was simulated at that point. Subsequently, the initial training data was updated with this pressure point and actual adsorption, and this AL protocol continued iteratively until the maximum relative error across all predictions was less than 2%. It is important to note that for the CoRE MOFs used in our study (previous sections), the initial data set consisted of 19 points, which were derived from GCMC simulations.

TABLE 2 Comparison between average metrics from 1000 different seeds to a random seed of 42.

Mean MAE (cm^3 STP/g)	MAE for random seed of 42 (cm^3 STP/g)	Mean MRE	MRE for random seed of 42	Mean R^2	R^2 for random seed for 42
CH_4 in CuBTC					
0.3301	0.237	0.048	0.043	0.99	0.99
CH_4 in IRMOF-1					
0.761	0.763	0.040	0.044	0.99	0.99
CO_2 in CuBTC					
0.791	0.764	0.049	0.046	0.99	0.99
CO_2 in IRMOF-1					
4.3	4.232	0.1	0.1	0.99	0.99

Abbreviations: MAE, mean absolute error; MRE, mean relative error.

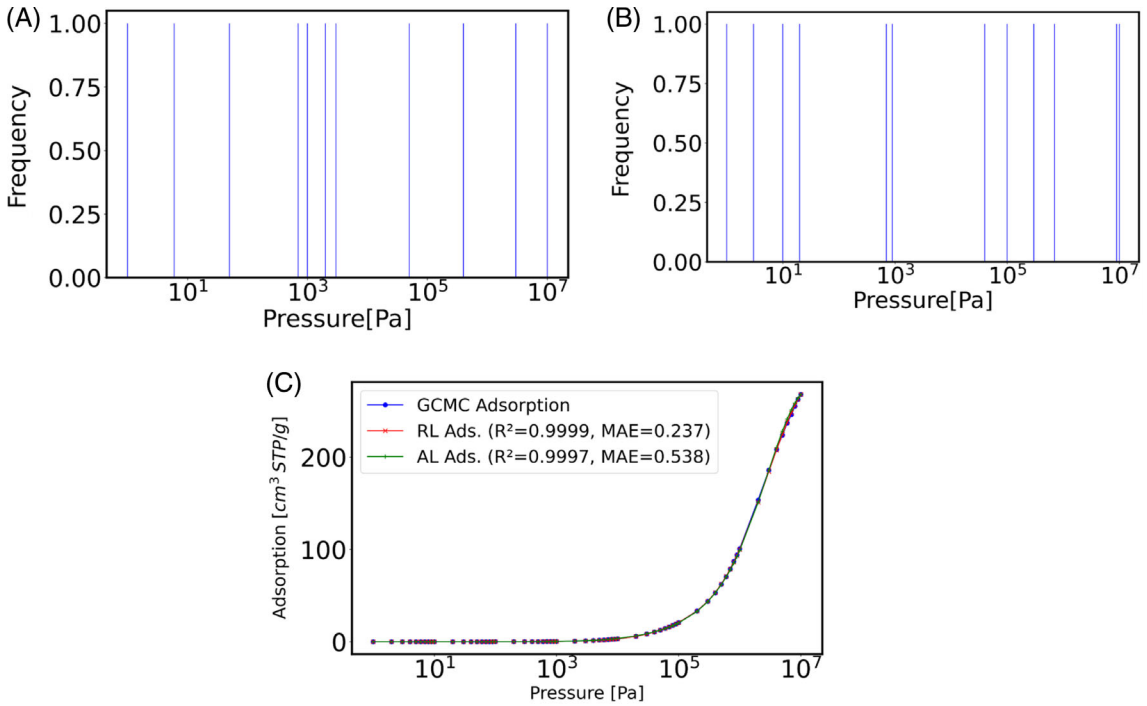

FIGURE 10 (A) AL final GP model training data. (B) RL sampled region of pressures by the Q-learning agent for the final GP model. (C) Isotherm comparison between GCMC simulation, RL-GP, and AL-GP predictions. AL, active learning; GCMC, grand canonical Monte Carlo; GP, Gaussian process; RL, reinforcement learning.

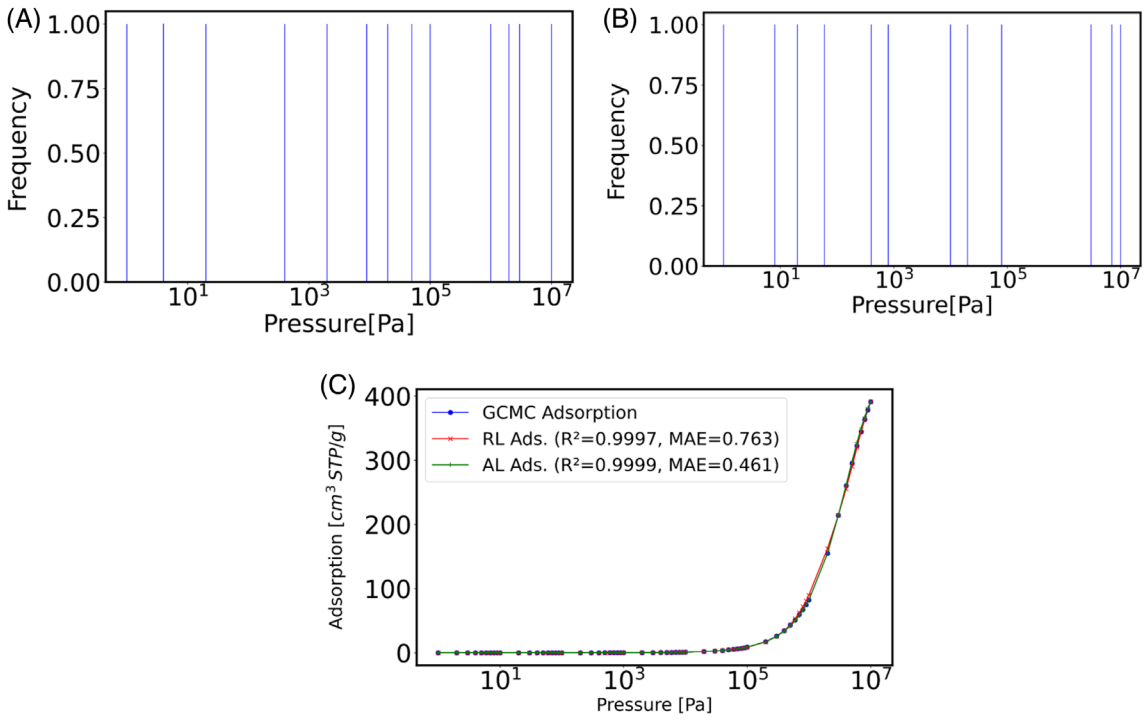

FIGURE 11 (A) AL final GP model training data. (B) RL sampled region of pressures by the Q-learning agent for the final GP model. (C) Isotherm comparison between GCMC simulation, RL-GP, and AL-GP predictions. AL, active learning; GCMC, grand canonical Monte Carlo; GP, Gaussian process; RL, reinforcement learning.

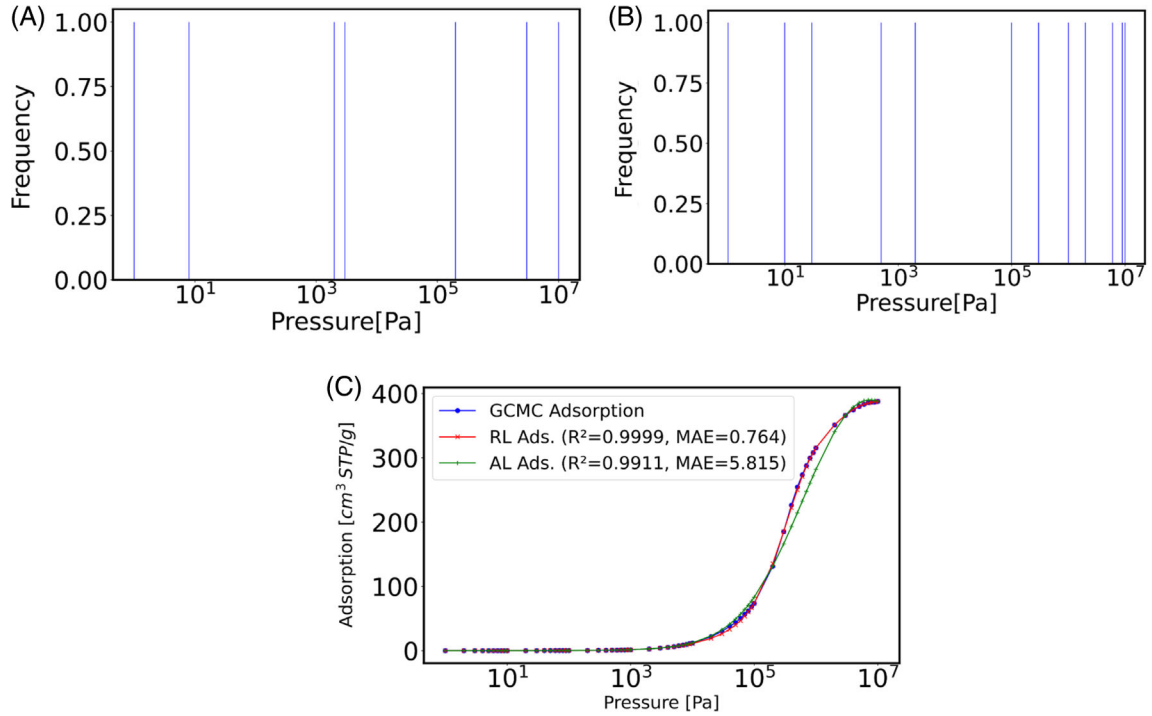

FIGURE 12 (A) AL final GP model training data. (B) RL sampled region of pressures by the Q-learning agent for the final GP model. (C) Isotherm comparison between GCMC simulation, RL-GP, and AL-GP predictions. AL, active learning; GCMC, grand canonical Monte Carlo; GP, Gaussian process; RL, reinforcement learning.

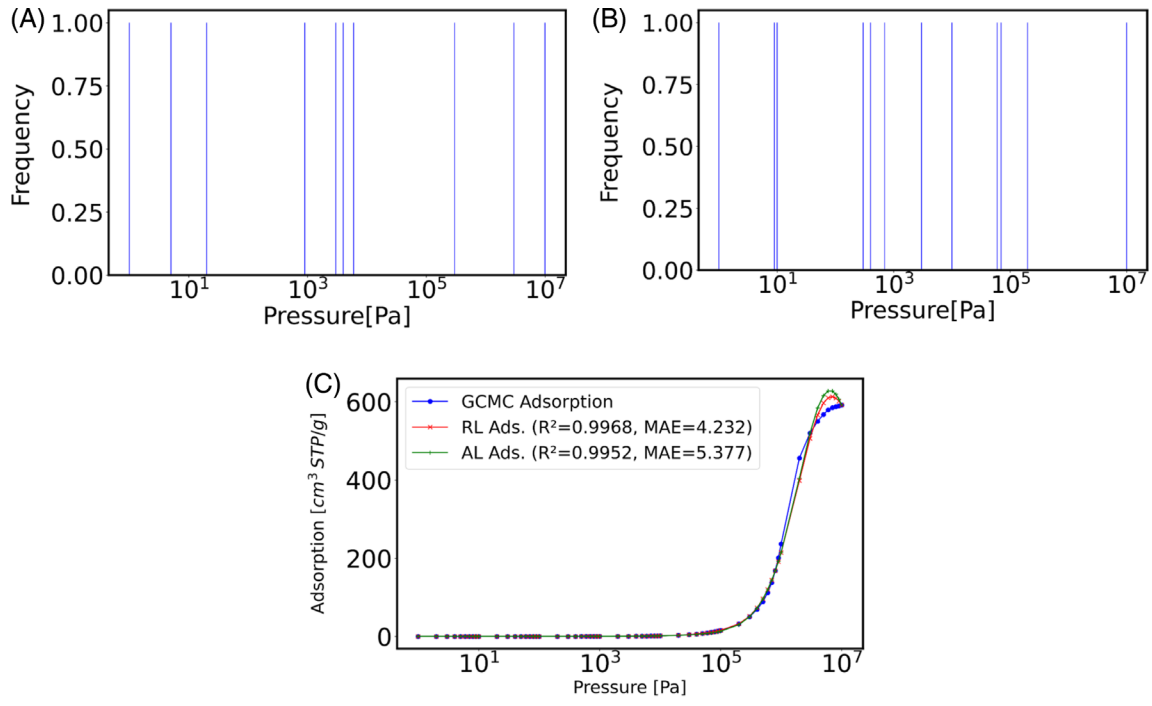

FIGURE 13 (A) AL final GP model training data. (B) RL sampled region of pressures by the Q-learning agent for the final GP model. (C) Isotherm comparison between GCMC simulation, RL-GP, and AL-GP predictions. AL, active learning; GCMC, grand canonical Monte Carlo; GP, Gaussian process; RL, reinforcement learning.

Generally, while both approaches yield favorable comparisons to the GCMC ground truth data, RL generally outperforms AL. An exception to this trend is observed in the prediction of CH_4 adsorption in IRMOF-1, where AL achieved a mean absolute error (MAE) of 0.538 cm^3 STP/g compared to RL's MAE of 0.763 cm^3 STP/g.

Although RL outperformed AL in three of the studied cases, this enhanced performance can be attributed to RL having access to all labels (data) upfront, whereas AL does not have this advantage.

CH4 in Cu-BTC

Figure 10A shows the pressure points collected by AL for the final GP model, while Figure 10B shows the RL sampled point by the Q-learning agent. For this case, we observe that the RL approach with a lesser MAE of 0.237 cm^3 STP/g outperforms that of AL with an MAE of 0.538 cm^3 STP/g, as shown in Figure 10C.

CH4 in IRMOF-1

Figure 11A displays the pressure points collected by AL for the final GP model, and Figure 11B shows the RL sampled point by the Q-learning agent. In this scenario, it was observed that the AL approach with a lesser MAE of 0.461 cm^3 STP/g outperforms that of RL with an MAE of 0.763 cm^3 STP/g, as shown in Figure 11C.

CO_2 in Cu-BTC

Figure 12A shows the pressure points collected by AL for the final GP model, and Figure 12B shows the RL sampled point by the Q-learning agent. We observed that the RL approach with a lesser MAE of 0.764 cm^3 STP/g outperforms that of AL with an MAE of 5.815 cm^3 STP/g, as shown in Figure 12C.

CO_2 in IRMOF-1

Figure 13A shows the pressure points collected by AL for the final GP model, while Figure 13B shows the RL sampled point by the Q-learning agent. For this case, we observe that the RL approach with a lesser MAE of 4.232 cm^3 STP/g outperforms that of AL with an MAE of 5.377 cm^3 STP/g, as shown in Figure 13C.

4 | CONCLUSION AND RECOMMENDATIONS

In this study, we introduced an approach that integrates Q-learning with GPs to optimize predictive modeling in MOF adsorption studies. By strategically selecting training data points based on rewards derived from predictive accuracy, the Q-learning algorithm efficiently explores the vast MOF adsorption space, offering a dynamic and adaptive framework for materials research. Our investigations across different MOF systems and gas adsorbates demonstrate the adaptability and effectiveness of this approach, showcasing its potential.

Through several experiments, we identified the best tradeoff between an RL agent exploration and exploitation. This balance proved pivotal in optimizing the training data set selection process, enhancing the adaptability and effectiveness of the learning mechanism. Moreover, our results highlight the robustness and reliability of Q-learning-guided exploration, as evidenced by the consistent performance across different systems and conditions.

A pivotal aspect of our approach lies in the strategic sampling of critical points by the RL agent within the pressure input space. By systematically exploring the pressure space, our methodology shows that there are informative pressure regions to develop an accurate adsorption model across all MOFs. Our findings underscore the significance of systematically probing these regions. The general RL code for both the MRE and R^2 reward metrics, and the isotherms generated through AL for both CH_4, and CO_2 can all be found via the GitHub repository.

In this study, we acknowledge that RL is computationally expensive compared to previous AL schemes or random sampling and requires access to all labels for its learning process. While fitting all the data or randomly selected points into a GP model can be sufficient, RL distinguishes itself by learning a policy that balances exploration and exploitation of the data points.

The adoption of RL for modeling adsorption isotherms as an AL method is justified by its capabilities in managing the exploration–exploitation tradeoff. This exploration is guided by an algorithm that not only seeks to uncover potentially more efficient adsorption points but also that each iteration contributes maximally to the accuracy of the resulting model. Hence, the integration of RL in this context enhances the accuracy of adsorption isotherm surrogate models. Looking ahead, the advancements presented in this study lay a solid foundation for future research endeavors in predictive modeling and materials science. The seamless integration of RL with GPs opens new avenues for accelerating materials discovery; RL can be used for other studies where determining a model training data is of extreme importance. Further, alternative RL techniques beyond Q-learning, such as Deep Q-Networks[44] or Actor-Critic methods,[45] could enhance the efficiency and scalability of predictive modeling. Additionally, the integration of transfer learning techniques could facilitate knowledge transfer between related tasks or domains, accelerating model convergence and improving generalization capabilities, especially in scenarios with limited training data.

AUTHOR CONTRIBUTIONS

Etinosa Osaro: Conceptualization; investigation; writing – original draft; methodology; validation; visualization; writing – review and editing; software; formal analysis; data curation. **Yamil J. Colón:** Conceptualization; investigation; funding acquisition; methodology; writing – review and editing; software; supervision; resources.

ACKNOWLEDGMENTS

E.O would like to thank the Lucy Family Institute for Data and Society, and The Patrick and Jana Eilers Graduate Student Fellowship for Energy Related Research at the University of Notre Dame. Y.J.C gratefully acknowledges NSF CAREER Award No. CBET-2143346. E. O and Y.J.C also thank the Center for Research Computing at the University of Notre Dame for the computational resources. The authors would also like to thank Reviewer 2 for their comments as which significantly improved the paper.

DATA AVAILABILITY STATEMENT

The RL codes for both the MRE and R^2 reward metrics can be found in the GitHub repository for this project. Also, the tabulated data for

the manuscript figures are found in the SI. The data for Figures 6 and 8 can be found in the repository. The Core MOFs AL isotherms for both CH_4 and CO_2 are found in the repository under the AL (GCMC + GP) subfolder. That folder also contains the files required for the AL protocol for the MOFs. The GitHub repository link for this project is: https://github.com/theOsaroJ/ReinforcementLearning/tree/main/Qlearning/MOFs/Pressure_Adsorption.

ORCID

Etinosa Osaro https://orcid.org/0000-0003-2744-339X

REFERENCES

1. Zhang X, Zhang K, Yoo H, Lee Y. Machine learning-driven discovery of metal-organic frameworks for efficient CO_2 capture in humid condition. *ACS Sustain Chem Eng*. 2021;9(7):2872-2879.
2. Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. *Nature*. 2023;624(7990):80-85.
3. Cai J, Chu X, Xu K, Li H, Wei J. Machine learning-driven new material discovery. *Nanoscale Adv*. 2020;2(8):3115-3130.
4. Liu Y, Zhao T, Ju W, Shi S. Materials discovery and design using machine learning. *J Materiomics*. 2017;3(3):159-177.
5. Pétuya R, Durdy S, Antypov D, et al. Machine-learning prediction of metal–organic framework guest accessibility from linker and metal chemistry. *Angew Chem Int Ed*. 2022;61(9):1-6.
6. Nandy A, Terrones G, Arunachalam N, Duan C, Kastner DW, Kulik HJ. MOFSimplify, machine learning models with extracted stability data of three thousand metal–organic frameworks. *Sci Data*. 2022;9(1):1-11.
7. Suyetin M. The application of machine learning for predicting the methane uptake and working capacity of MOFs. *Faraday Discuss*. 2021;231:224-234.
8. Islamov M, Babaei H, Anderson R, et al. High-throughput screening of hypothetical metal-organic frameworks for thermal conductivity. *NPJ Comput Mater*. 2023;9(1):1-12.
9. Feng L, Wang KY, Lv XL, Yan TH, Zhou HC. Hierarchically porous metal-organic frameworks: synthetic strategies and applications. *Natl Sci Rev*. 2020;7(11):1743-1758.
10. Li H, Li L, Lin RB, Zhou W, Zhang Z, Xiang S. Porous metal-organic frameworks for gas storage and separation: status and challenges. *EnergyChem*. 2019;1(1):100006.
11. Lin RB, Xiang S, Zhou W, Chen B. Microporous metal-organic framework materials for gas separation. *Chem*. 2020;6(2):337-363.
12. Sturluson A, Huynh MT, Kaija AR, et al. The role of molecular modelling and simulation in the discovery and deployment of metal-organic frameworks for gas storage and separation. *Mol Simul*. 2019;45(14–15):1082-1121.
13. Colón YJ, Gómez-Gualdrón DA, Snurr RQ. Topologically guided, automated construction of metal–organic frameworks and their evaluation for energy-related applications. *Cryst Growth des*. 2017;17(11):5801-5810.
14. Bobbitt NS, Shi K, Bucior BJ, et al. MOFX-DB: an online database of computational adsorption data for nanoporous materials. *J Chem Eng Data*. 2023;68(2):483-498.
15. Majumdar S, Moosavi SM, Jablonka KM, Ongari D, Smit B. Diversifying databases of metal organic frameworks for high-throughput computational screening. *ACS Appl Mater Interfaces*. 2021;13(51):61004-61014.
16. Chung YG, Haldoupis E, Bucior BJ, et al. Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: CoRE MOF 2019. *J Chem Eng Data*. 2019;64(12):5985-5998.
17. Chong S, Lee S, Kim B, Kim J. Applications of machine learning in metal-organic frameworks. *Coord Chem Rev*. 2020;423:213487.
18. Pardakhti M, Moharreri E, Wanik D, Suib SL, Srivastava R. Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs). *ACS Comb Sci*. 2017;19(10):640-645.
19. Hung TH, Xu ZX, Kang DY, Lin LC. Chemistry-encoded convolutional neural networks for predicting gaseous adsorption in porous materials. *J Phys Chem C*. 2022;126(5):2813-2822.
20. Yulia F, Chairina I, Zulys A, Nasruddin. Multi-objective genetic algorithm optimization with an artificial neural network for CO_2/CH_4 adsorption prediction in metal–organic framework. *Therm Sci Eng Prog*. 2021;25:100967.
21. Rosen AS, Fung V, Huck P, et al. High-throughput predictions of metal–organic framework electronic properties: theoretical challenges, graph neural networks, and data exploration. *NPJ Comput Mater*. 2022;8(1):112.
22. Anderson R, Gómez-Gualdrón DA. Deep learning combined with IAST to screen thermodynamically feasible MOFs for adsorption-based separation of multiple binary mixtures. *J Chem Phys*. 2021;154(23):234102.
23. Anderson R, Biong A, Gómez-Gualdrón DA. Adsorption isotherm predictions for multiple molecules in MOFs using the same deep learning model. *J Chem Theory Comput*. 2020;16(2):1271-1283.
24. Zhang X, Zhou T, Sundmacher K. Integrated metal–organic framework and pressure/vacuum swing adsorption process design: descriptor optimization. *AIChE J*. 2022;68(2):e17524.
25. Jose A, Devijver E, Jakse N, Poloni R. Informative training data for efficient property prediction in metal–organic frameworks by active learning. *J Am Chem Soc*. 2024;146(9):6134-6144.
26. Tang H, Jiang J. Active learning boosted computational discovery of covalent–organic frameworks for ultrahigh CH_4 storage. *AIChE J*. 2022;68(11):e17856.
27. Park H, Majumdar S, Zhang X, Kim J, Smit B. Inverse design of metal-organic frameworks for direct air capture of CO_2 via deep reinforcement learning. *Digit Discov*. 2024;3:728-741.
28. Deringer VL, Bartók AP, Bernstein N, Wilkins DM, Ceriotti M, Csányi G. Gaussian process regression for materials and molecules. *Chem Rev*. 2021;121(16):10073-10141.
29. Noack MM, Doerk GS, Li R, et al. Autonomous materials discovery driven by Gaussian process regression with inhomogeneous measurement noise and anisotropic kernels. *Sci Rep*. 2020;10(1):1-16.
30. Osaro E, Mukherjee K, Colón JY. Active learning for adsorption simulations: evaluation, criteria analysis, and recommendations for metal–organic frameworks. *Ind Eng Chem Res*. 2023;62(33):13009-13024.
31. Mukherjee K, Osaro E, Colón YJ. Active learning for efficient navigation of multi-component gas adsorption landscapes in a MOF. *Digit Discov*. 2023;2:1506-1521.
32. Mukherjee K, Dowling AW, Colón YJ. Sequential design of adsorption simulations in metal organic frameworks. *Mol Syst Des Eng*. 2021;7(3):248-259.
33. Osaro E, Fajardo-Rojas F, Cooper GM, Gómez-Gualdrón D, Colón YJ. Active Learning of Alchemical Adsorption Simulations: Towards A Universal Adsorption Model. 2024.
34. Clifton J, Laber E. Q-learning: theory and applications. *Annu Rev Stat Appl*. 2020;7(1):279-301.
35. Jang B, Kim M, Harerimana G, Kim JW. Q-learning algorithms: a comprehensive classification and applications. *IEEE Access*. 2019;7:133653-133667.
36. Dubbeldam D, Calero S, Ellis DE, Snurr RQ. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol Simul*. 2016;42(2):81-101.
37. Rappé AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc*. 1992;114(25):10024-10035.

38. Eggimann BL, Sunnarborg AJ, Stern HD, Bliss AP, Siepmann JI. An online parameter and property database for the TraPPE force field. *Mol Simul*. 2014;40(1–3):101-105.
39. Martin MG, Siepmann JI. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J Phys Chem B*. 1998; 5647(97):2569-2577.
40. Raza A, Sturluson A, Simon CM, Fern X. Message passing neural networks for partial charge assignment to metal–organic frameworks. *J Phys Chem C*. 2020;124(35):19070-19082.
41. Dudek A, Baranowski J. Gaussian processes for signal processing and representation in control engineering. *Appl Sci*. 2022;12(10):4946.
42. Wilson AG, Adams RP. Gaussian process kernels for pattern discovery and extrapolation. 30th International Conference on Machine Learning, ICML 2013. 2013;28(PART 3):2104–2112.
43. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825-2830.
44. Huang Y. Deep Q-Networks. *Deep Reinforcement Learning*. Springer Singapore; 2020:135-160.
45. Grondman I, Busoniu L, Lopes GAD, Babuska R. A survey of actor-critic reinforcement learning: standard and natural policy gradients. *IEEE Trans Syst Man Cybern B*. 2012;42(6):1291-1307.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Osaro E, Colón YJ. Optimizing the prediction of adsorption in metal–organic frameworks leveraging Q-learning. *AIChE J*. 2024;70(12):e18611. doi:10.1002/aic.18611