Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

ML-based adsorption isotherm prediction of metal-organic frameworks for carbon dioxide and methane separation adsorbent screening

Dongin Jung[a], Hyeon Yang[a], Donggeun Kang[a], Donghyeon Kim[a], Siuk Roh[a] and Jiyong Kim[a]*

[a] Sungkyunkwan University (SKKU), Department of Chemical Engineering, Suwon, Republic of Korea
* Corresponding Author: jiyongkim@skku.edu.

ABSTRACT

The efficient separation of carbon dioxide (CO_2) and methane (CH_4) is crucial for chemical processes, including biogas upgrading and natural gas purification. Metal-organic frameworks (MOFs) have gained significant attention as promising adsorbents for these processes due to their high porosity and tunable structures. Estimating the adsorption capacity of MOFs is essential for screening high performing adsorbents. While molecular simulations are commonly used to estimate the adsorption capacities, their computational intensity acts as a bottleneck in screening MOF adsorbents. In this study, we propose a machine learning (ML)-based framework for the high-throughput prediction of adsorption isotherms for CO_2 and CH_4 in MOFs. A graph neural network (GNN) model was developed to predict adsorption capacities, effectively replacing the time-consuming molecular simulations. The GNN model processes the structural graphs of MOFs, capturing their spatial configurations, such as surface structure and pore characteristics, which are closely related to adsorption performance. Based on the adsorption capacities predicted by the GNN model, an isotherm model was derived to characterize the adsorption behavior of the MOFs. This framework enables the rapid and reliable screening of MOFs with superior adsorption performance for CO_2/CH_4 separation, offering a computationally efficient approach that paves the way for broader applications of MOFs in carbon dioxide separation technologies.

Keywords: Adsorption, Machine Learning, Carbon Dioxide Capture

1. INTRODUCTION

The separation of carbon dioxide (CO_2) and methane (CH_4) from gas mixtures is a critical process in chemical processes such as biogas upgrading [1]. Efficient separation of these gases is essential for enhancing biomethane purity, reducing greenhouse gas emissions, and improving the sustainability of energy production. Pressure swing adsorption (PSA) is one of the most widely used techniques for gas separation, relying on porous solid adsorbents to selectively capture CO_2 from gas mixtures. The performance of PSA largely depends on the properties of the adsorbent, including its adsorption capacity and selectivity. Therefore, the development of high-performance adsorbents is crucial for improving the efficiency of CO_2/CH_4 separation. Metal-organic frameworks (MOFs), characterized by high porosity and structural tunability, have emerged as promising candidates for PSA adsorbents owing to their outstanding adsorption capacities and selectivity [2]. The modular nature of MOFs, which allows flexible design of their chemical composition and physical structure, enables tailored designs for specific gas separation applications. However, this tunability poses a challenge in efficiently identifying materials optimized for specific separation processes.

Traditional methods for determining MOF adsorption properties, such as molecular simulations and experimental measurements, are both time-intensive and computationally expensive, especially when screening large datasets [4]. Machine learning (ML) has recently emerged as a powerful alternative, enabling rapid and accurate predictions of MOF adsorption properties. By reducing computational costs and accelerating analysis,

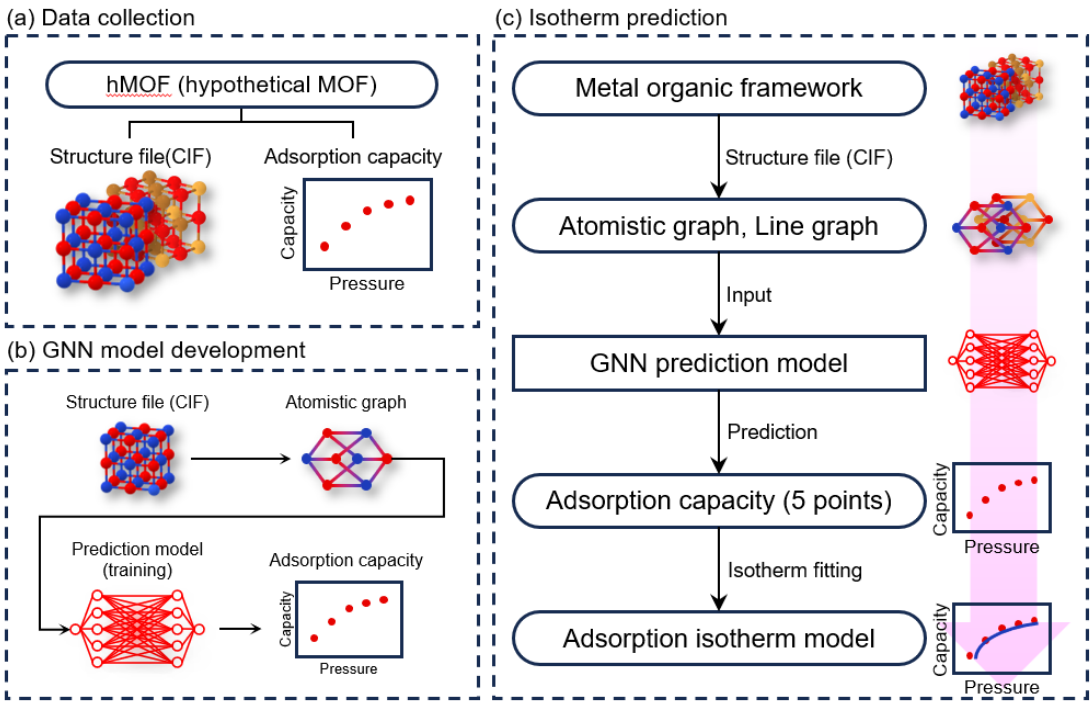

Figure 1: Research overview

ML provides a transformative approach for efficiently screening MOFs across diverse applications. Among ML approaches, graph neural network (GNN) stand out as a highly effective and specialized method for predicting MOF properties. Unlike conventional ML models, which rely on predefined descriptors or simple structural representations, GNN directly process chemical structures as graphs, where atoms are represented as nodes and bonds as edges [5]. This unique capability allows GNN to inherently capture both atomic- and bond-level features while preserving the spatial and topological complexity of MOFs. As a result, GNN excel in modeling the intricate relationships between structural characteristics and adsorption behavior, making them exceptionally suited for analyzing the complex and diverse architectures of MOFs.

This study introduces a framework for predicting the adsorption isotherms of CO_2 and CH_4 in MOFs. Using graph representations of MOFs, the GNN model predicts adsorption capacities. These predictions are subsequently processed by the pyAPEP software, which fits adsorption isotherms using established models such as the Langmuir or Dual-Site Langmuir equations [6].

By leveraging advanced GNN model and large-scale MOF datasets, this study provides a scalable framework for predicting CO_2 and CH_4 adsorption isotherms with reduced computational costs.

2. METHOD

Data collection

In this study, we used the hypothetical MOFs (hMOF) dataset from the MOFX database to train the model. The MOFX database is a comprehensive repository that consolidates diverse datasets of MOFs, including both hypothetical and experimentally synthesized structures [7]. It provides Crystallographic Information Files (CIFs) containing atomic coordinates and structural properties, along with adsorption and textural property data. Such detailed information is instrumental in facilitating the screening of MOFs for adsorption-based applications.

The hMOF dataset, comprising 137,953 hypothetical MOF structures, was selected from the MOFX database for the present study [8]. These structures were computationally generated by systematically combining a diverse set of 102 building blocks, including metal nodes and organic linkers, to explore a wide range of potential MOF configurations. The resulting hypothetical structures represent the diverse structural and chemical properties of MOFs.

In this study, preprocessing was applied to filter out structures unsuitable for adsorption studies. MOFs lacking methane adsorption data were removed, and those with a pore limiting diameter (PLD) smaller than 3.3 Å were excluded. The PLD represents the smallest pore diameter of an adsorbent, and if it is smaller than the gas molecules, adsorption cannot occur as the molecules cannot enter the pores. Since this dataset is used to

train a model for predicting CO_2 adsorption capacity, information on MOFs that are physically incapable of adsorbing CO_2 must be excluded. Therefore, a 3.3 Å constraint, corresponding to the molecular diameter of CO_2, was applied to filter out such structures. After preprocessing, 112,351 MOFs remained as training data. The dataset was then split into training, validation, and test sets with a ratio of 0.8:0.1:0.1. CIFs of the hMOF structures were utilized to construct graph-based representations, where atoms and bonds correspond to nodes and edges, respectively. These graph representations served as input data for training the GNN model.

The adsorption data of CO_2 and CH_4 from the hMOF dataset were used as target values for model training, which were computed using Grand Canonical Monte Carlo (GCMC) simulations at 298 K across multiple pressure conditions [9]. We obtained five adsorption data points for each gas, with CO_2 adsorption data including simulation results at pressures of 0.01, 0.05, 0.1, 0.5, and 2.5 bar, and CH_4 adsorption data including simulation results at pressures of 0.05, 0.5, 0.9, 2.5, and 4.5 bar. The TraPPE (Transferable Potentials for Phase Equilibria) force field was employed to model molecular interactions for the adsorption gases, and the UFF (Universal Force Field) was used for the MOF structures. Each simulation included 1,000 initialization cycles for system equilibration, followed by 1,000 production cycles to gather adsorption data.

GNN model development

We developed a GNN model to predict the adsorption capacities of CO_2 and CH_4 for MOFs, based on the Atomistic Line Graph Neural Network (ALIGNN) architecture. ALIGNN combines atomistic and line graph representations to effectively capture structural and bond-level interactions within MOFs, enabling accurate predictions of material properties [10].

In the atomistic graph representation, atoms are represented as nodes and bonds as edges. For this study, 8 atomic features were used: column number, atomic radius, ionization energy, electronegativity, electron affinity, valence electrons, atomic volume, and block (outermost shell). Edge features, representing bond distances, were encoded using a radial basis function (RBF) expansion with a cutoff of 8 Å. To ensure structural completeness, a periodic 12-nearest-neighbor configuration was applied during graph construction.

The line graph representation complements the atomistic graph by treating interatomic bonds as nodes and bond angles as edges. Bond angles were encoded using RBF features and propagated with bond distance features through Edge-gated Graph Convolutions. Alternating convolutions between atomistic and line graphs allowed ALIGNN to efficiently update atom, bond, and bond-angle features, capturing both local and higher-order structural relationships in MOFs.

The ALIGNN architecture consisted of 4 ALIGNN layers and 4 graph convolution (GCN) layers, each with 256 hidden nodes. Initial bond and bond angle features were embedded into a 64-dimensional space using 80 RBF features for bond distances and 40 RBF features for bond angles. Final atom-wise representations were averaged and passed through a linear layer to generate regression outputs. This robust architecture enabled ALIGNN to capture both local and higher-order structural relationships in MOFs, allowing accurate prediction of adsorption capacities across various pressure conditions.

Adsorption isotherm fitting

To construct adsorption isotherms for each MOF, the predicted adsorption capacities at five pressure points were fitted using the pyAPEP software package. pyAPEP is a comprehensive software designed for adsorption process simulations, automating data processing and fitting adsorption isotherms to enable accurate modeling of gas adsorption behavior [6].

pyAPEP employs a well-designed algorithm to automatically determine the most suitable isotherm model for a given pressure-uptake dataset. This is achieved by optimizing parameters, minimizing errors, and selecting the best-fitting model from five established options: Langmuir, Freundlich, Sips, Quadratic, and Dual-Site Langmuir. Each model is described by the following equations:

Langmuir model: $q(P) = \frac{K_1 K_2 P}{1+K_2 P}$ (1)

Freundlich model: $q(P) = K_1 P^{K_2}$ (2)

Sips model: $q(P) = \frac{K_1 K_2 P^{K_3}}{1+K_2 P^{K_3}}$ (3)

Quadratic model: $q(P) = K_1 \frac{(K_2+2K_3 P)P}{1+K_2 P+K_3 P^2}$ (4)

Dual-site Langmuir model: $q(P) = K_1 \frac{K_2 P}{1+K_2 P} + K_3 \frac{K_4 P}{1+K_4 P}$ (5)

This approach efficiently generates adsorption isotherms from limited data, providing a robust basis for analyzing and screening MOFs based on their adsorption performance.

Model validation

To validate the generalizability of the model, the CRAFTED dataset, which provides adsorption data for 690 experimentally synthesized MOFs extracted from the CoREMOF2014 database, was utilized [11]. This dataset includes simulated adsorption data performed for the adsorption of CO_2 with two fields (i.e., UFF and DREIDING), six partial charge schemes (i.e., no charge, Qeq, EQeq, PACMOF, MPNN, and DDEC), at three temperatures (i.e., 273, 298, 323 K).

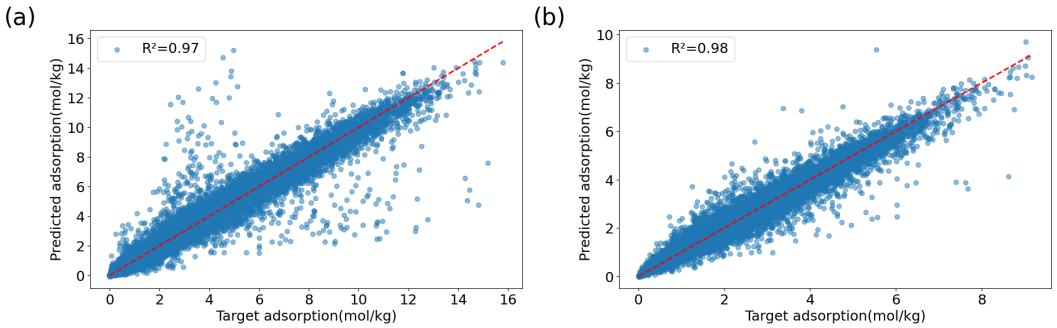

Figure 2: Comparison of target vs predicted adsorption capacities for (a) carbon dioxide adsorption with model accuracy (R^2 = 0.97), (b) methane adsorption with model accuracy (R^2 = 0.98)

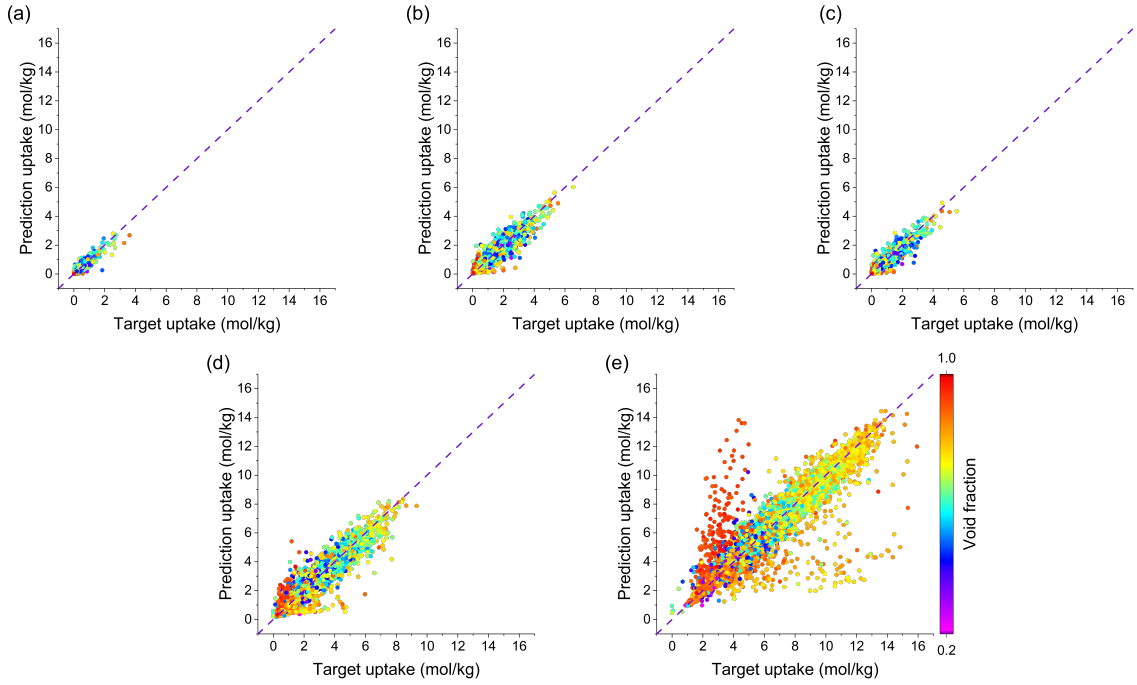

Figure 3: The plots of target vs predicted CO_2 adsorption capacity colored by void fraction at five pressure points
(a) 0.01 bar, (b) 0.05 bar, (c) 0.1 bar, (d) 0.5 bar, (e) 2.5 bar

To ensure compatibility with the GNN model's training conditions, adsorption data simulated with the UFF force field, EQeq partial charge calculation method, and a temperature of 298 K were selected. Based on this simulated data, simulated adsorption isotherms were fitted and compared with the isotherms predicted by the GNN framework.

3. RESULT ANALYSIS

Model accuracy

We evaluated the accuracy of the adsorption prediction models for CO_2 and CH_4 using the coefficient of determination (R^2), which measures the proportion of variance in the target values explained by the predictions. R^2 scores range from 0 to 1, with higher values indicating greater predictive accuracy. The evaluation results showed R^2 scores of 0.97 for CO_2 and 0.98 for CH_4, demonstrating the explanatory power of the prediction models in estimating adsorption capacities for both gases. Additionally, Figure 2 illustrates the alignment between the target and predicted adsorption capacities, highlighting the high accuracy of the models.

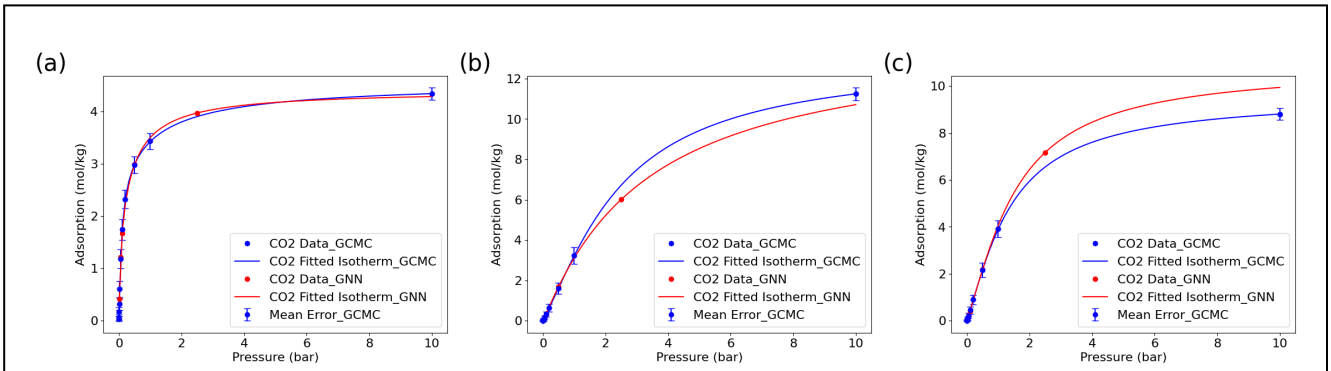

Figure 4: Three examples of CRAFTED dataset validation results: GCMC vs. GNN CO_2 isotherm

Table 1. R^2 scores for each pressure point for CO_2 adsorption model

Pressure (bar)	CO_2 model (R^2 score)
0.01	0.88
0.05	0.92
0.1	0.93
0.5	0.94
2.5	0.88

Table 2. R^2 scores for each pressure point for CH_4 adsorption model

Pressure (bar)	CH_4 model (R^2 score)
0.05	0.92
0.5	0.95
0.9	0.95
2.5	0.94
4.5	0.94

In addition to overall accuracy, Table 1 and 2 provides R^2 scores for each pressure point across the two gases, offering a detailed assessment of model performance. While the CH_4 model maintained consistently high R^2 scores across all pressure points, the CO_2 model exhibited lower accuracy at some pressure points. For example, at 0.01 bar and 2.5 bar, the model exhibited R^2 score of 0.88

To investigate the factors responsible for the lower accuracy of the CO_2 model at specific pressure points, we analyzed the textural properties of MOFs with significant prediction errors. Among the textural properties, void fraction, which is known to have a significant association with adsorption, was identified as the property most strongly influencing the model's predictive performance. The analysis revealed that MOFs with large prediction errors were predominantly those with high void fractions. The relationship between target and predicted CO_2 adsorption capacities at five pressure points is visualized in Figure 3, with void fraction represented as a color gradient. At 2.5 bar, the structures with the most significant prediction errors were mostly MOFs with void fractions greater than 0.8. This pattern suggests that the unique structural characteristics of MOFs with high void fractions are not fully captured by the current GNN model. This limitation is interpreted as resulting from the graph representation of MOFs, where a cutoff distance of 8 Å is used to define connections between atoms, assuming bonds exist between the 12 nearest neighbors within this range. For MOFs with high void fractions, the relative distances between atoms are often larger, making the graph representation less effective at capturing their structural features. These findings highlight the need to refine the model into a more robust framework that ensures high accuracy across diverse MOF characteristics.

Computational efficiency

To demonstrate the computational efficiency of our model, we compared its performance with molecular simulations by analyzing the time required for adsorption data generation. In GCMC simulations, generating adsorption data for a single pressure point requires approximately 8.6 hours [12].

While the initial training process for the GNN model requires 31.5 hours for CO_2 and 16.5 hours for CH_4, this is a one-time computational cost. Once trained, the model is capable of rapidly predicting adsorption data for any given MOF. This makes the proposed framework advantageous for handling large-scale datasets compared to GCMC simulations. This advantage is particularly significant in high-throughput screening processes, where predicting adsorption properties for a vast number of MOFs is required.

CRAFTED dataset validation

Figure 4 presents the validation results comparing the GNN-predicted and GCMC-simulated isotherms for MOFs included in the CRAFTED dataset. The results indicate a strong alignment between the two isotherms. However, minor discrepancies are observed at comparatively high-pressure ranges. For example, Figure 4 (b) and (c) show noticeable deviations at pressures above 2 bar. The reason for this discrepancy is attributed to the absence of high-pressure adsorption data in the training

set.

These results demonstrate the generalizability of the proposed framework in accurately predicting the adsorption behavior of diverse MOFs.

4. CONCLUSION

This study proposed a GNN-based framework utilizing the ALIGNN to predict adsorption isotherms for CO_2 and CH_4 in MOFs. We developed a GNN model to predict adsorption capacities using a large-scale dataset of hypothetical MOFs and applied an isotherm fitting algorithm to rapidly and accurately fit adsorption isotherms based on the predicted capacities. The developed model demonstrated high predictive accuracy. The computational efficiency of the framework was validated by comparing its calculation time with that of molecular simulations, highlighting its comparative advantage. Furthermore, the framework's generalizability to experimentally synthesized MOF datasets was confirmed through validation.

Consequently, the proposed framework is expected to be utilized for rapid high-throughput screening of MOF adsorbents. Future research should be directed toward developing a model capable of capturing complex MOF structures and establishing a framework that demonstrates high generalizability to diverse MOFs.

REFERENCES

1. Zhang, Y., Sunarso, J., Liu, S., & Wang, R. (2012). Current status and development of membranes for CO2/CH4 separation: A review. International Journal of Greenhouse Gas Control, 12, 84–107. https://doi.org/10.1016/j.ijggc.2012.10.009
2. Furukawa, H., Cordova, K. E., O'Keeffe, M., & Yaghi, O. M. (2013). The Chemistry and Applications of Metal-Organic Frameworks. Science, 341(6149). https://doi.org/10.1126/science.1230444
3. Baumann, A.E., Burns, D.A., Liu, B. et al. Metal-organic framework functionalization and design strategies for advanced electrochemical energy storage devices. Commun Chem 2, 86 (2019). https://doi.org/10.1038/s42004-019-0184-6
4. Choudhary, K., Yildirim, T., Siderius, D. W., Kusne, A. G., McDannald, A., & Ortiz-Montalvo, D. L. (2022). Graph neural network predictions of metal organic framework CO2 adsorption properties. Computational Materials Science, 210, 111388. https://doi.org/10.1016/j.commatsci.2022.111388
5. Reiser, P., Neubert, M., Eberhard, A., Torresi, L., Zhou, C., Shao, C., Metni, H., Van Hoesel, C., Schopmans, H., Sommer, T., & Friederich, P. (2022). Graph neural networks for materials science and chemistry. Communications Materials, 3(1). https://doi.org/10.1038/s43246-022-00315-6
6. Ga, S., An, N., Joo, C., & Kim, J. (2023). pyAPEP: An all-in-one software package for the automated preparation of adsorption process simulations. Computer Physics Communications, 291, 108830. https://doi.org/10.1016/j.cpc.2023.108830
7. Bobbitt, N. S., Shi, K., Bucior, B. J., Chen, H., Tracy-Amoroso, N., Li, Z., Sun, Y., Merlin, J. H., Siepmann, J. I., Siderius, D. W., & Snurr, R. Q. (2023). MOFX-DB: An Online Database of Computational Adsorption Data for Nanoporous Materials. Journal of Chemical & Engineering Data, 68(2), 483–498. https://doi.org/10.1021/acs.jced.2c00583
8. Wilmer, C., Leaf, M., Lee, C. *et al.* Large-scale screening of hypothetical metal–organic frameworks. *Nature Chem* **4**, 83–89 (2012). https://doi.org/10.1038/nchem.1192
9. Wilmer, C. E., Farha, O. K., Bae, Y., Hupp, J. T., & Snurr, R. Q. (2012). Structure–property relationships of porous materials for carbon dioxide separation and capture. Energy & Environmental Science, 5(12), 9849. https://doi.org/10.1039/c2ee23201d
10. Choudhary, K., & DeCost, B. (2021). Atomistic Line Graph Neural Network for improved materials property predictions. Npj Computational Materials, 7(1). https://doi.org/10.1038/s41524-021-00650-1
11. Oliveira, F. L., Cleeton, C., Ferreira, R. N. B., Luan, B., Farmahini, A. H., Sarkisov, L., & Steiner, M. (2023). CRAFTED: An exploratory database of simulated adsorption isotherms of metal-organic frameworks. Scientific Data, 10(1). https://doi.org/10.1038/s41597-023-02116-z
12. Ga, S., An, N., Lee, G. Y., Joo, C., & Kim, J. (2024). Multidisciplinary high-throughput screening of metal–organic framework for ammonia-based green hydrogen production. Renewable and Sustainable Energy Reviews, 192, 114275. https://doi.org/10.1016/j.rser.2023.114275