

Metal–Organic Frameworks for Water Harvesting: Machine Learning-Based Prediction and Rapid Screening

Zhiming Zhang, Hongjian Tang, Mao Wang, Bohui Lyu, Zhongyi Jiang*, and Jianwen Jiang*



Cite This: *ACS Sustainable Chem. Eng.* 2023, 11, 8148–8160



Read Online

ACCESS |

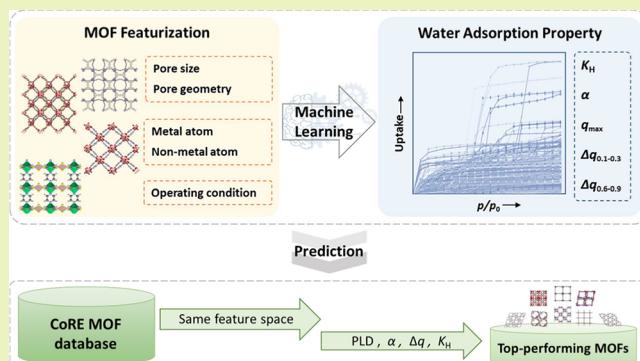
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Atmospheric water harvesting based on metal–organic frameworks (MOFs) is an emerging technology to potentially mitigate water scarcity. Because of the tremendously large number of existing MOFs, it is challenging to find suitable candidates. In this context, a data-driven approach to identify top-performing MOFs represents an important direction. Herein, we develop a machine learning (ML) method to predict water adsorption in MOFs and screen out top-performing MOFs for water harvesting. First, experimental water adsorption isotherms in MOFs are collected and water adsorption properties are extracted. Quantitative structure–property relationships are analyzed in terms of pore structure and framework chemistry, providing task-specific design principles. Then, ML models are trained and interpreted to predict water adsorption properties by using structural and chemical features, as well as operating conditions as descriptors. The transferability of the ML models is validated by out-of-sample predictions in seven newly reported MOFs. Finally, the ML models are applied to screen ~8000 “Computation-Ready, Experimental” (CoRE) MOFs. Top-performing candidates are identified including 149 MOFs with the maximum adsorption capacity ≥ 35 mmol/g, 39 MOFs with working capacity ≥ 10 mmol/g in a relative pressure window 0.1–0.3, and 139 MOFs with working capacity ≥ 8.7 mmol/g in a relative pressure window 0.6–0.9. The developed ML-based method would advance task-oriented design and rapid discovery of reticular materials for energy and environmental applications.

KEYWORDS: metal–organic frameworks, water harvesting, adsorption isotherm, machine learning, design principles



INTRODUCTION

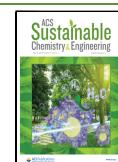
With the growing population and energy demand, water scarcity has posed a serious threat to our society and sustainable development. Water harvesting from the air in adsorption-based devices is an emerging technology to alleviate water stress for drinking, sanitarian, agricultural and industrial purposes.^{1,2} Water harvesters extract ubiquitous atmospheric moisture and release fresh water using renewable low-grade energy sources, exhibiting promising technical feasibility and energy efficiency.^{2,3} The performance of water harvesters is strongly dependent on adsorbents inside devices.^{4,5} Ideal adsorbents should possess customized water–adsorbent interactions and offer step-like reversible water adsorption isotherms with high water uptakes, tailored Henry's constants, and appropriate inflection points in operating windows.^{2,6} Hitherto, hygroscopic salts, silica gels, and zeolites have been applied as adsorbents for autonomous water harvesters.^{1,3} However, they suffer from desiccant leakage and low efficiency due to undesired water adsorption isotherms.^{1,3} Rational design of adsorbents with task-specific water adsorption properties is highly desired to develop efficient water harvesters.

In the ongoing quest for designable adsorbents for water harvesting, metal–organic frameworks (MOFs) have triggered tremendous interest.^{7,8} MOFs are reticular materials with the periodic arrangement of molecular building blocks.^{9,10} Reticular chemistry confers task-oriented water adsorption properties through diverse pore structure and versatile framework chemistry. In terms of the pore structure, water adsorption can be regulated by customizing pore size, volume, dimensionality, etc.^{11–13} With respect to the framework chemistry, water adsorption process may be manipulated by chemisorption sites like unsaturated metals and physisorption sites like hydroxyl groups in linkers.^{14–16} Therefore, MOFs can be produced with step-like water adsorption isotherms, tunable step positions, and tailorabile working capacities in task-specific operating windows. For instance, Cr-soc-MOF-1,¹¹ MIL-

Received: March 1, 2023

Revised: May 4, 2023

Published: May 19, 2023



101(Cr),^{17,18} MIL-100(Fe),^{17–19} and Y-shp-MOF-S²⁰ were proven to surpass commercial benchmark materials like Zeolite-Y and AQS0A, as a result of their lower average enthalpy, higher porosity, and enhanced water nucleation at vertices.^{6,11,13,21} However, it is challenging to identify suitable MOFs for water harvesting because of the huge number of existing MOFs and lack of theoretical design principles. It is of paramount importance to develop computational methods to accelerate the discovery of top-performing MOFs.

Over the past few years, materials science has witnessed transformation from an empirical to a data-driven paradigm.^{22–25} As a representative data-driven technique, machine learning (ML) has been explosively applied to facilitate the intuitive analysis of structure–property relationships and the rational design of reticular materials like MOFs.^{26–28} Specifically, ML has been applied to accelerate the discovery of top-performing MOFs for energy and environment-relevant applications such as carbon dioxide capture,^{29,30} methane storage,^{31,32} and paraffin/olefin separation.^{33,34} In these adsorption-based applications, the adsorption properties of simple gases (i.e., carbon dioxide, methane, and light hydrocarbons) were predicted from molecular simulations. However, water adsorption properties cannot be accurately generated in the same manner. This is because no generic, reliable, and transferable force field is currently available to describe the complex interactions between highly polar water molecules and MOFs. Consequently, there is a scarce ML study to investigate water adsorption in MOFs. Recently, the performance of adsorption chillers with water as a working fluid was examined using ML, and top-performing MOF/water working pairs were shortlisted.³⁵

The ML technique holds great potential to explore water adsorption in MOFs and derive predictive design principles. Toward this end, we develop an ML method in this study to predict water adsorption in MOFs, aiming to accelerate the discovery of top-performing MOFs for water harvesting. First, we collected an experimental data set with 344 water adsorption isotherms in 285 MOFs. The MOFs were encoded in terms of structural and chemical features. The key water adsorption properties including Henry's constant, inflection point, maximum capacity, and working capacities under two operating conditions were extracted from isotherms as prediction targets. Then, we established structure–property relationships with an emphasis on pore size, pore geometry, pore dimensionality, linker chemistry, and metal cluster. Subsequently, we developed five independent ML models for the targets by using different algorithms. The best-performing ML models were interpreted by the Gini-impurity and the Shapley Additive Explanations. The transferability of the ML models was validated by out-of-sample predictions in seven newly reported MOFs. Finally, these ML models were applied to screen ~8000 “Computation-Ready, Experimental” (CoRE) MOFs,³⁶ and top-performing MOFs were identified for water harvesting under different operating conditions.

METHODOLOGY

Data Collection. Adsorption isotherm is informative to assess underlying adsorption mechanism and provide key adsorption properties, which facilitate narrowing down adsorbent candidates for water harvesting. As illustrated in Figure S1, we collected experimentally measured water adsorption isotherms from two sources: the NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials (NIST-

ISODB) and literature articles.³⁷ The isotherms in NIST-ISODB were filtered with three criteria: (1) water as the adsorbate with its International Chemical Identifier Key of XLYOFNOQVPJJNP-UHFFFAOYSA-N; (2) “exp” as the measurement category; and (3) MOFs as the adsorbents. The types of MOFs were analyzed by matching their digital object identifiers (DOIs) and names with the Cambridge Structural Database (CSD).³⁸ Meanwhile, the crystallographic information files (cifs) of the matched MOFs were collected by utilizing the CSD Application Programming Interface (API). In addition, we supplemented water adsorption isotherms by searching literature articles on the Web of Science. The isotherm data points were extracted by using the Web Plot Digitizer.³⁹ In total, 344 adsorption isotherms (198 from the NIST-ISODB database and 146 from the literature articles) for pure water in 285 MOFs were collected. The measurement temperature was in the range of 273–373 K, as listed in Table S1, with the highest percentage (77.3%) at 298 K.

The collected isotherms were processed for feasible comparison and extraction. The units of water uptake were unified as mmol/g, and pressure was normalized as the relative pressure (p/p_0 , where p_0 refers to the saturated vapor pressure at the operating temperature). Then, water uptake values were calculated at $p/p_0 = 0\text{--}0.95$, which covers the common operating windows, with an interval of 0.05 by using the first-order Bézier function. Five key water adsorption properties were extracted including Henry's constant (K_H , mmol/(g·Pa)), maximum adsorption capacity (q_{\max} , mmol/g), inflection point (α), working capacities in operating windows $p/p_0 = 0.1\text{--}0.3$ ($\Delta q_{0.1\text{--}0.3}$, mmol/g), and $p/p_0 = 0.6\text{--}0.9$ ($\Delta q_{0.6\text{--}0.9}$, mmol/g). K_H represents the initial water uptake at low pressure and denotes the intensity of water–adsorbent interaction, it was calculated as the isotherm slope at $p/p_0 = 0\text{--}0.05$. α indicates the location of the characteristic step, and it was defined as p/p_0 approaching half of q_{\max} . q_{\max} is the maximum water adsorption capacity. $\Delta q_{0.1\text{--}0.3}$ was calculated by the water uptake difference between $p/p_0 = 0.1$ and 0.3, while $\Delta q_{0.6\text{--}0.9}$ was calculated by the water uptake difference between $p/p_0 = 0.6$ and 0.9.

Featurization of MOFs. The 285 collected MOFs were encoded as machine-readable descriptors to analyze structure–property relationships and train ML models. To eliminate disordered structures, we preprocessed the collected cifs, including computational desolvation by the CSD API and replacing them by the cifs of CoRE MOFs with the same identifiers. As listed in Table S2, the descriptors comprised structural and chemical features, as well as operating conditions. First, eight structural features including three pore sizes (LCD, PLD, and LFPD) and five pore geometry metrics (VSA, GSA, VF, PV, and ρ) were calculated by Zeo++⁴⁰ with a probe radius of 0.1 Å for a fine resolution. As illustrated in Figure S2, each of the eight structural features exhibits a single-peak distribution. The median values of LCD, PLD, and LFPD are 7.7, 4.7, and 7.7 Å, respectively. The median values of VSA and GSA are $3961.9\text{ m}^2/\text{cm}^3$ and $3602.3\text{ m}^2/\text{g}$, respectively. The median values of ρ , VF, and PV are 1.2 kg/m^3 , 0.54, and $0.46\text{ m}^3/\text{g}$. Second, 68 chemical features were quantified by different atom types with their densities calculated by the lammps interface with parameters from the force field of UFF4MOF.^{41,42} The local chemical environment like hybridization and connectivity of atoms was found important to capture framework chemistry,^{32,33} thus incorporated in this work. For instance, C_1, C_2, C_3, and C_R

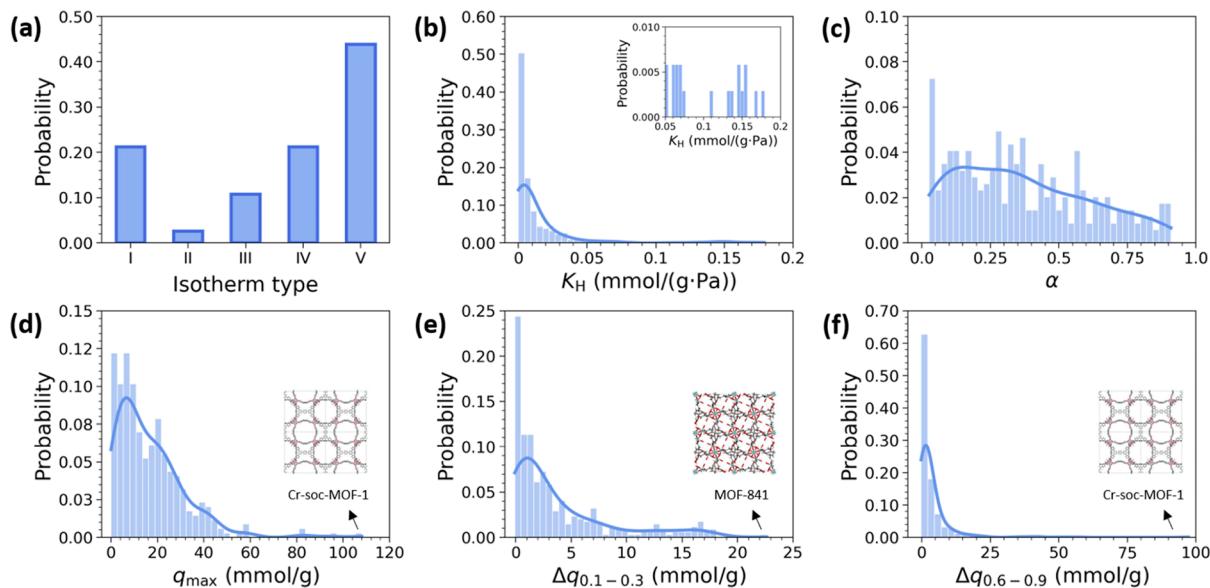


Figure 1. Distributions of (a) isotherm type, (b) K_H , (c) α , (d) q_{\max} , (e) $\Delta q_{0.1-0.3}$, and (f) $\Delta q_{0.6-0.9}$ in the collected experimental data set. The lines in (b–f) indicate kernel density estimations. Best-performing MOFs are highlighted with their structures depicted in (d–f).

represent different carbon atom types in single, double, triple, and resonant bonding states, respectively. As shown in Figure S3, the 285 MOFs cover a wide chemical feature space of 68 atom types including 46 metal atom types and 22 nonmetal atom types. In each MOF, the atom densities were normalized by the cell volume. Third, the temperature at which the experimental isotherm was measured was taken as the operating condition.

To establish the structure–property relationships, MOFs were further characterized by pore dimensionality, linker chemistry, and topology. Specifically, pore dimensionality was evaluated using Zeo++ with a probe radius of 1.35 Å, which is the radius of a water molecule. Among the 285 collected MOFs, the pore dimensionalities of 237 MOFs were successfully evaluated. The ligand chemistry of 237 MOFs and the topologies of 135 MOFs were identified through MOFid.⁴³

Machine Learning. We used the above descriptors to train five independent ML models for five targets (K_H , q_{\max} , α , $\Delta q_{0.1-0.3}$, and $\Delta q_{0.6-0.9}$). First, we randomly split the data set into training set and test set. The descriptors were centered to zero and scaled using the mean and standard deviation for each training set. Then, we applied four algorithms to train ML models: random forest regression (RFR), gradient boosting regression (GBR), kernel ridge regression with Gaussian kernel (KRR_G), and kernel ridge regression with Laplacian kernel (KRR_L) as implemented in the *scikit-learn* toolkit.^{44,45} Both RFR and GBR are ensembled algorithms. RFR is based on the average value of many decision trees, while GBR consists of multistage regression trees by the negative gradient of a loss function. KRR is a kernel-based algorithm by statistically measuring the distance between data points in a feature space. The hyperparameters of each ML model were optimized by a random grid search with five-fold cross validation. The performance of each ML model was evaluated by the coefficient of determination (R^2) and mean absolute error (MAE):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{\text{exp}}^{(i)} - y_{\text{pre}}^{(i)})^2}{\sum_{i=1}^n (y_{\text{exp}}^{(i)} - \bar{y}_{\text{exp}}^{(i)})^2} \quad (1)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |y_{\text{exp}}^{(i)} - y_{\text{pre}}^{(i)}|}{n} \quad (2)$$

After evaluating the performance, the entire data set was used to retrain the best-performing ML models. Moreover, the Pearson correlation coefficients (PCCs) were calculated to analyze the structure–property relationships. The ML models were interpreted by feature importance analysis, which was quantified by the mean decrease of Gini-impurity and the SHAP.^{46,47} The out-of-sample validation of the ML models was conducted for seven newly reported MOFs with experimentally measured water adsorption isotherms.

Predictions and Screening. The best-performing ML models were applied to predict water adsorption in CoRE MOFs.³⁶ Among 12,020 CoRE MOFs, 109 were found already existing in the collected data set and thus removed. The remaining 11,911 were featurized using the same procedure as described above. A total of 8059 CoRE MOFs possessing the same feature space as in the collected data set were used for predictions. The similarity between 285 collected MOFs and 8059 CoRE MOFs was analyzed by using principal component analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE).⁴⁸ All the features of 8059 CoRE MOFs were normalized and standard by the same parameters used to train ML models. Water adsorption properties in 8059 CoRE MOFs were predicted and top-performing ones were screened out under different operating conditions.

RESULTS AND DISCUSSION

Data Analysis. Figure S4 shows the collected experimental water adsorption–desorption isotherms. Based on the International Union of Pure and Applied Chemistry (IUPAC) convention, five types (I–V) are observed in these isotherms. The isotherm type reveals specific water–adsorbent interaction and water uptake behavior. As plotted in Figure 1a, types III

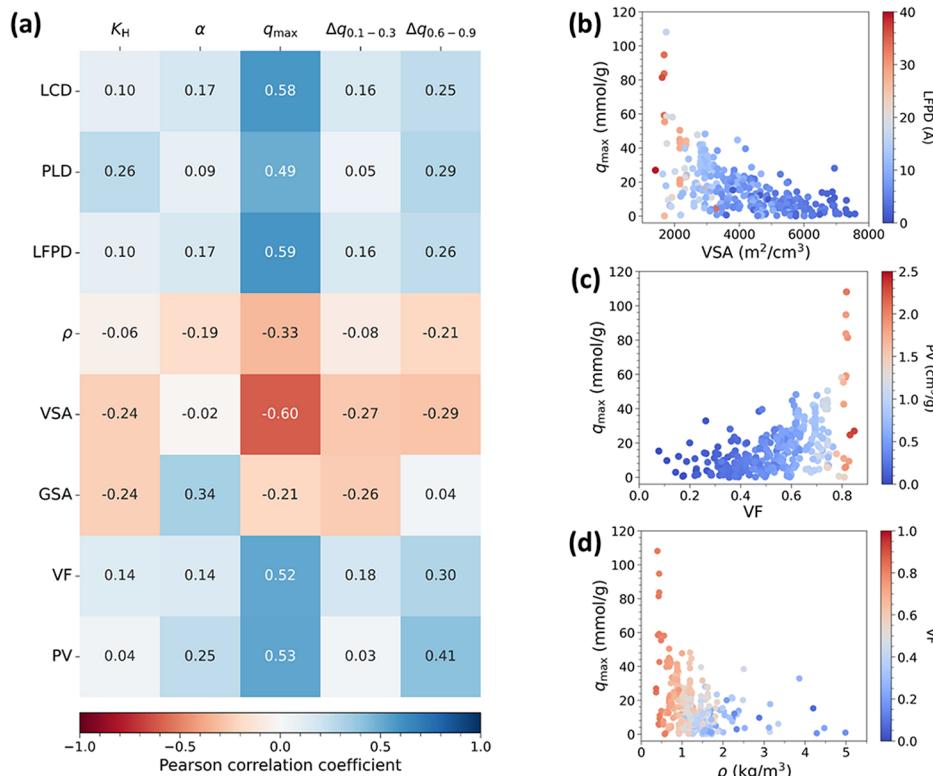


Figure 2. (a) PCC matrix between eight structural features and five adsorption properties. The color scale indicates PCC value. (b) Relationship of $q_{\max} \sim \text{VSA}$. The color scale denotes LFPD. (c) Relationship of $q_{\max} \sim \text{VF}$. The color scale denotes PV. (d) Relationship of $q_{\max} \sim \rho$. The color scale denotes VF.

and V account for ~55% of all the collected isotherms, denoting weak water–MOF affinity. Types IV and V exhibit step-like uptake and account for ~65%. Weak water affinity and step-like uptake are desirable for efficient water harvesting. However, the qualitative analysis of isotherm type cannot precisely assess adsorbent performance.

To quantitatively analyze water adsorption, five adsorption properties including K_H , α , q_{\max} , $\Delta q_{0.1-0.3}$, and $\Delta q_{0.6-0.9}$ were extracted from each isotherm. Their distributions in all the collected isotherms are plotted in Figures 1b–f and S5. (1) K_H represents the initial water uptake at a low pressure and denotes the strength of water–adsorbent interaction. K_H is found to populate at a narrow interval of 0–0.01 mmol/(g·Pa), demonstrating that most of the 285 MOFs possess weak water–adsorbent interactions, corresponding to the dominant isotherm types of III and V. This is consistent with the fact that hydrophobic MOFs with low K_H are preferred for water harvesting to prevent structure degradation, promote water-resistance stability, and reduce regeneration energy. (2) α indicates the location of the characteristic step, which is utilized to differentiate adsorbents with step-like uptake in and outside an operating window. α is observed to spread widely throughout the entire pressure range, demonstrating varied step positions in different MOFs. This is because α depends on water–adsorbent interaction, water arrangement in MOFs, as well as operating condition. (3) As for the maximum water capacity, q_{\max} exhibits a single-peak distribution centered around ~10 mmol/g. Commercial benchmark materials such as Zeolite-Y and AQSOA possess q_{\max} of 11.1–16.7 mmol/g.^{6,13} In the collected data set, 44 fall in this range, whereas 144 surpass the benchmarked $q_{\max} = 16.7$ mmol/g. Type IV isotherms (e.g., in Cr-soc-MOF-1) exhibit the highest q_{\max} as

attributed to their step-like uptake during pore filling or capillary condensation process. (4) $\Delta q_{0.1-0.3}$ and $\Delta q_{0.6-0.9}$ are the working capacities in two operating windows $p/p_0 = 0.1–0.3$ and $p/p_0 = 0.6–0.9$, respectively. The first window of $p/p_0 = 0.1–0.3$ is a commonly adopted operating condition for water harvesting, thus $\Delta q_{0.1-0.3}$ acts as a benchmark metric for screening adsorbents.^{2,6} The second window of 0.6–0.9, corresponding to the annual humidity range in Singapore,⁴⁹ is examined here as a case study for water harvesting in a tropical area. Both $\Delta q_{0.1-0.3}$ and $\Delta q_{0.6-0.9}$ exhibit a single-peak distribution. $\Delta q_{0.1-0.3}$ has a median value of ~1.9 mmol/g, suggesting a low initial water uptake and, hence low K_H in most of the 285 MOFs. The top 10% highest $\Delta q_{0.1-0.3}$ is bounded by 10 mmol/g. Notably, MOF-841 has the highest $\Delta q_{0.1-0.3}$ of 22.6 mmol/g.⁵⁰ $\Delta q_{0.6-0.9}$ has a median value of ~1.6 mmol/g. The top 10% highest $\Delta q_{0.6-0.9}$ is bounded by 8.7 mmol/g. Only 10 MOFs hold $\Delta q_{0.6-0.9}$ exceeding 20 mmol/g and they all have type V isotherms. Particularly, Cr-soc-MOF-1 has the highest $\Delta q_{0.6-0.9}$ of 108.0 mmol/g.¹¹

From the above data analysis, we infer that most of the 285 MOFs experimentally examined are hydrophobic with low K_H and type III or V isotherms. The wide distribution of α manifests that MOFs have potential to be tailored for different operating conditions. Nearly half of q_{\max} in the collected isotherms surpass that of benchmark materials. However, only a small number of MOFs among 285 have desired $\Delta q_{0.1-0.3}$ and $\Delta q_{0.6-0.9}$.

Structure–Property Relationships. To quantify the relationships between eight structural features and five adsorption properties, the PCC matrix is plotted in Figure 2a. K_H is mainly related to PLD and VSA/GSA; α exhibits major correlations with GSA and PV; q_{\max} shows high

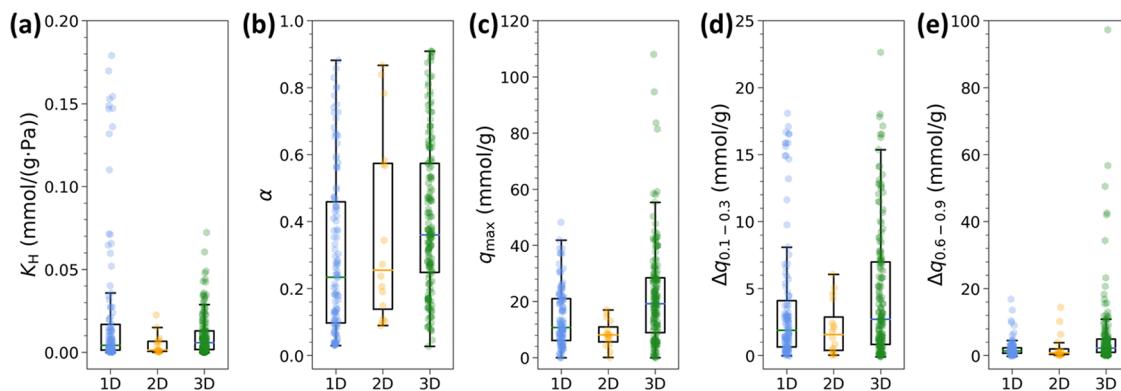


Figure 3. Box plots of pore dimensionality versus (a) K_H , (b) α , (c) q_{\max} , (d) $\Delta q_{0.1-0.3}$, and (e) $\Delta q_{0.6-0.9}$. The markers denote the minimum, first quartile, median, third quartile, and maximum values, respectively.

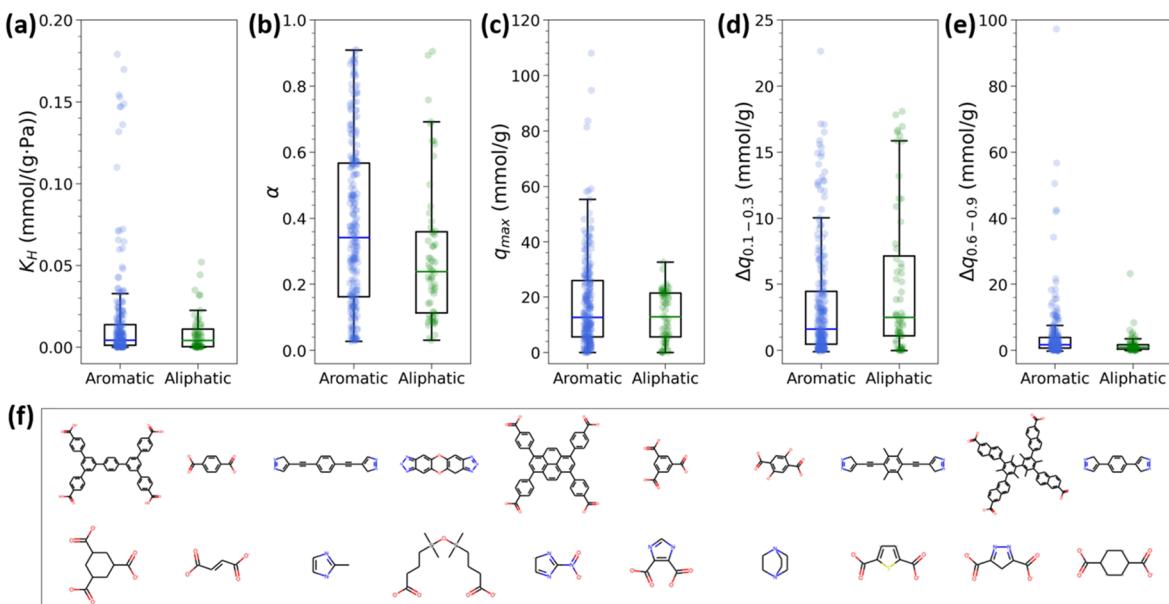


Figure 4. Box plots of linker chemistry versus (a) K_H , (b) α , (c) q_{\max} , (d) $\Delta q_{0.1-0.3}$ and (e) $\Delta q_{0.6-0.9}$. (f) Representative aromatic and aliphatic linkers in collected MOFs.

correlations with nearly all the structural features; $\Delta q_{0.1-0.3}$ is mainly related to VSA/GSA; $\Delta q_{0.6-0.9}$ has similar correlation pattern as q_{\max} except with GSA. Among all the properties, q_{\max} shows the highest PCC with nearly all the structural features, suggesting that q_{\max} is related to the overall structure, to a large extent. Subsequently, we analyze how q_{\max} is affected by different structural features.

As VSA, LFPD, PV, and VF hold the highest correlations with q_{\max} thus their relationships with q_{\max} are depicted in Figure 2b–d. Figure 2b shows that large VSA and small LFPD result in low q_{\max} . With increasing VSA in MOFs, the water-MOF interaction becomes weaker and the framework structure becomes more tortuous with smaller LFPD, thus water molecules are less likely to form a broad hydrogen bond network in the framework, resulting in low q_{\max} . It is observed that $q_{\max} > 35$ mmol/g, corresponding to the highest 10% q_{\max} occurs at VSA $< 4192 \text{ m}^2/\text{cm}^3$ and LFPD $> 36 \text{ \AA}$. From Figure 2c, MOFs with large VF and PV are more likely to provide high q_{\max} . Larger VF and PV indicate more volume to accommodate water molecules, which is a necessary prerequisite for high q_{\max} . For example, Cr-soc-MOF-1 has the highest q_{\max} in the collected data set, as it has VF and PV of up

to 0.82 and $2.1 \text{ cm}^3/\text{g}$, providing adequate volume for water adsorption. Figure 2d reveals that high ρ and small VF result in low q_{\max} . The increase of ρ is accompanied by decreased VF, resulting in fewer internal voids available for water adsorption. Moreover, high ρ indicates a small weight percentage of light organic elements, which constitute most hydrophilic functional groups like $-\text{OH}$. As a result, high ρ leads to a low density of water adsorption sites and hence low q_{\max} . Specifically, $q_{\max} > 35$ mmol/g occurs at $\rho < 2.50 \text{ kg/m}^3$ and VF > 0.47 . However, low q_{\max} is also observed in the highest range of VF, as well as the lowest range of VSA and ρ . Thus, q_{\max} is jointly determined by multiple structural features rather than a single one. In brief, q_{\max} reflects the overall topological structures of MOFs. While VSA/GSA and ρ have negative correlations with q_{\max} , LFPD, VF, and PV exhibit positive influences on q_{\max} . The difficulty lies in the comprehensive analysis of how high-dimension structural features jointly determine water adsorption properties.

Furthermore, the effect of pore dimensionality on water adsorption is explored. By decomposing 8-dimension structural features into two principal components via PCA, a blurred boundary is observed among MOFs with different pore

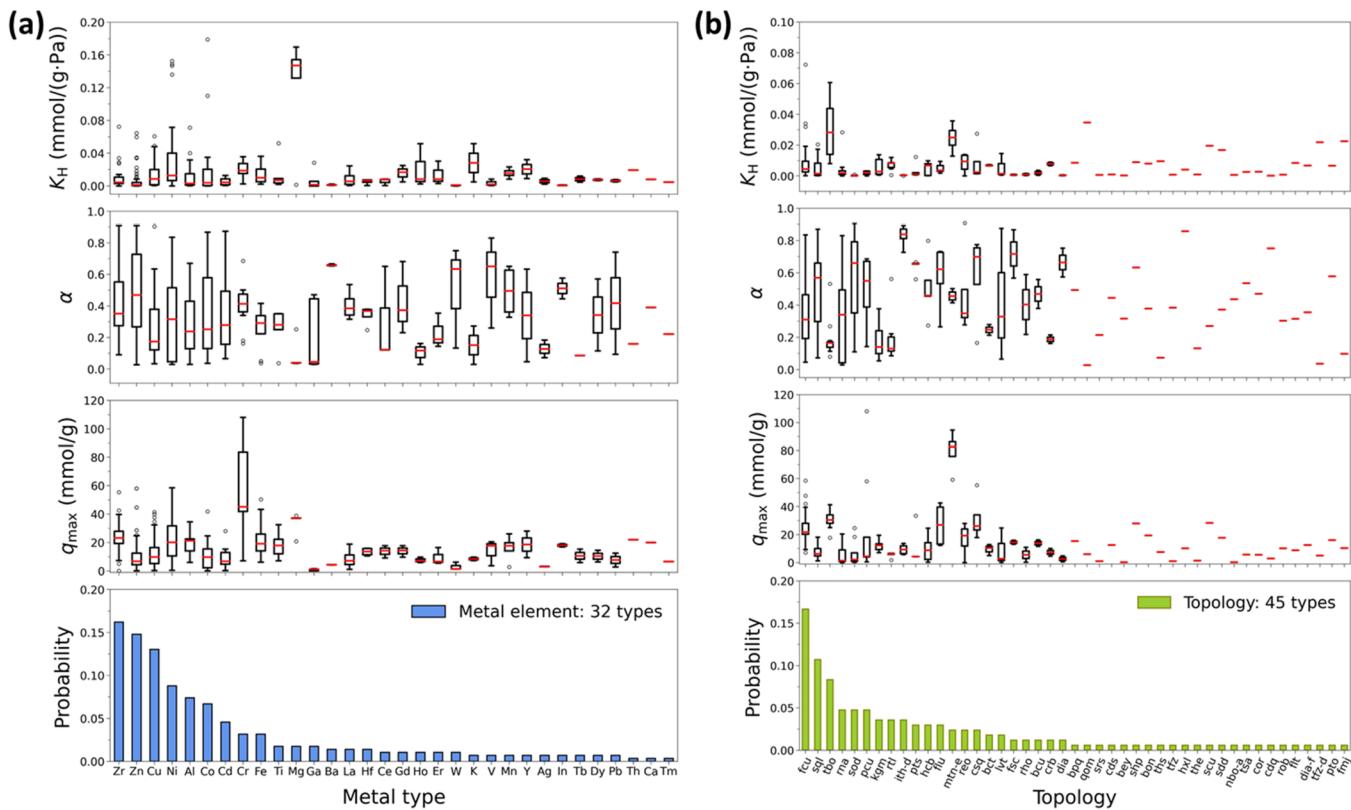


Figure 5. Water adsorption properties based on (a) metal types and (b) topologies.

dimensionalities in Figure S6. Pore dimensionality provides a probabilistic interpretation of how a high-dimension structure determines adsorption properties. As shown in Figure S7, generally 1D MOFs have high ρ , large VSA, low VF, and small PV; 3D MOFs have large LCD, large LFPD, low ρ , small VSA, high VF, and large PV; 2D MOFs have feature values in between 1D and 3D MOFs. Figure 3 shows the relationships between pore dimensionality and five adsorption properties. First, 1D MOFs exhibit high K_H , low α , moderate q_{max} , as well as moderate $\Delta q_{0.1-0.3}$ and $\Delta q_{0.6-0.9}$. On one hand, 1D MOFs with limited pore sizes provide a strong affinity for water and facilitate water adsorption at low pressure. Ten MOFs with the highest K_H are identified as 1D, and they are mostly MOF-74 coordinated with different metals including Co, Mg, Ni, and Zn. On the other hand, water uptake in the confined space of 1D MOFs is limited due to small PV. As a balance between strong water affinity and undesired PV, 1D MOFs exhibit moderate capacities. Second, only about 4% of MOFs are identified as 2D MOFs and thus statistically insignificant. We observe scarce 2D MOFs with desired adsorption properties due to median structural features. Third, 3D MOFs have the potential to achieve high water uptake (high q_{max} and high $\Delta q_{0.6-0.9}$) at high pressure, attributed to their large pore sizes, small VSA, large PV, and high VF. Ten MOFs with the highest q_{max} are 3D, such as Cr-soc-MOF-1, MIL-101(Cr), and Ni₈(L₅)₆. However, they do not show distinct advantages at low pressure ($\Delta q_{0.1-0.3}$). In summary, 1D MOFs are more likely to provide high water affinity in their confined space, whereas 3D MOFs hold the potential for favorable water adsorption at high pressure.

In addition to structural features, framework chemistry also plays an essential role in water adsorption. The basic building blocks in MOFs are organic linkers and metal clusters, which

are coordinated to form periodic networks, thus we examine their effects separately. The linkers of 285 collected MOFs can be categorized into aromatic and aliphatic. Figures 4 and S8 show the effects of linker chemistry on adsorption properties and MOF structures, respectively. Aromatic linkers have higher K_H , wider distribution of α , higher adsorption capacities at high pressure (q_{max} and $\Delta q_{0.6-0.9}$), and lower $\Delta q_{0.1-0.3}$. Despite the hydrophobic nature of phenyl rings, aromatic linkers may be functionalized with strong hydrophilic groups as shown in Figure 4f to enhance K_H and adjust the α position. High q_{max} and $\Delta q_{0.6-0.9}$ are found in aromatic linkers, attributed to desired structural features (large PV, VF, and pore size) constructed by aromatic linkers. However, aliphatic linkers possess relatively lower K_H and narrower distribution of α , indicating weaker water affinity and less versatility in functionalization. Their $\Delta q_{0.1-0.3}$ is slightly higher than aromatic counterparts, plausibly due to less steric hindrance of aliphatic linkers, which facilitates water molecules to easily approach adsorption sites at low pressure. Overall, aromatic linkers are more tunable for high water affinity, wide distribution of α , and favorable for high-pressure operation. Aliphatic linkers show potential at low pressure due to easy accessibility to water adsorption sites, but other properties are less desired because of limited functionalization versatility.

By their constituting metal types, the 285 MOFs can be classified into 32 types. As illustrated in Figure 5a, Zr-, Zn-, and Cu-based MOFs account for the largest proportion of the collected MOFs. Zr(IV) has a high oxidation state, high charge density, and bond polarization, contributing to high stability and abundant topology types;⁵¹ thus, Zr-based MOFs such as the UiO series are most extensively investigated for water adsorption. As a biological/eco-friendly transition metal, Zn exists in a wide variety of MOFs including DMOF, CID and

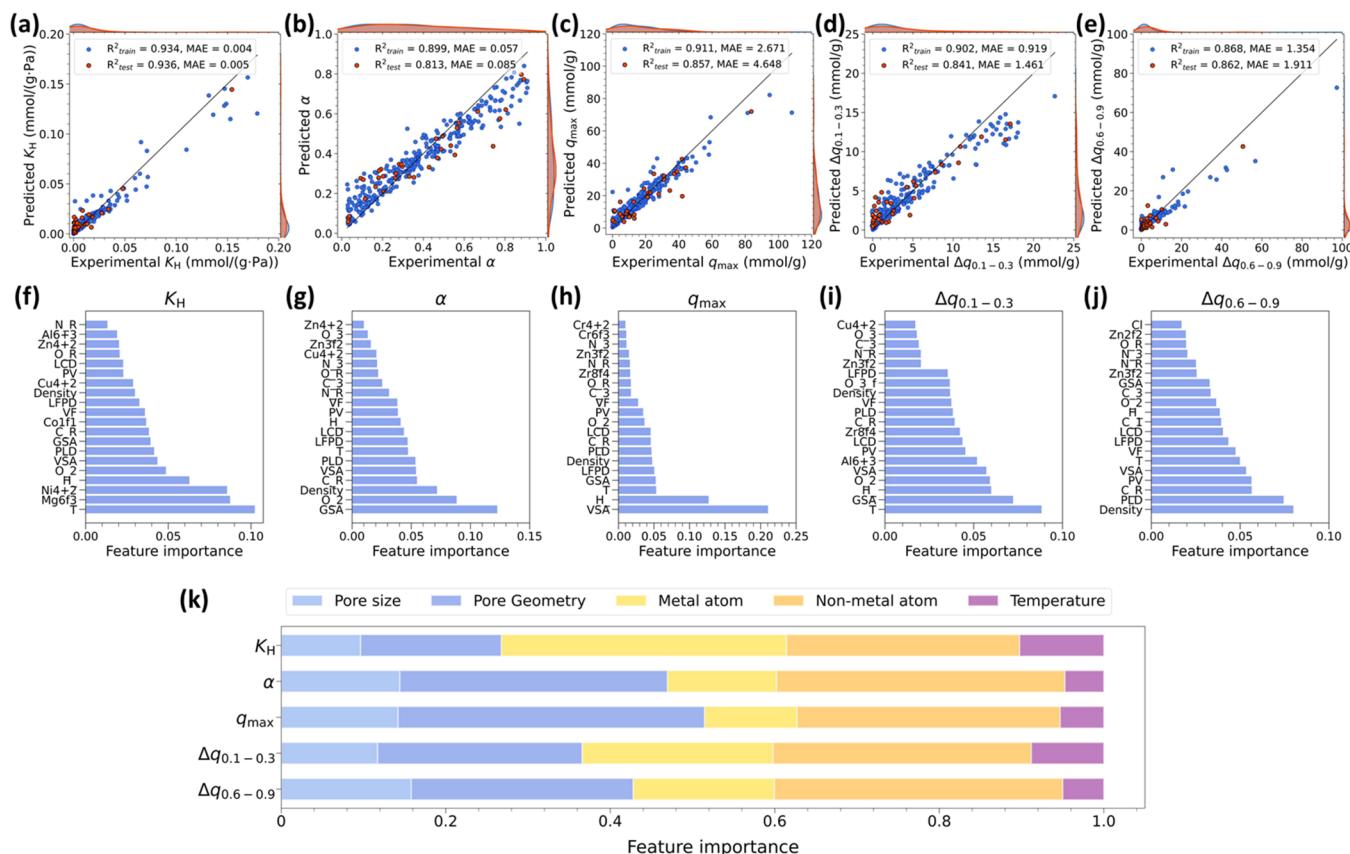


Figure 6. (a–e) Parity plots for K_H , α , q_{max} , $\Delta q_{0.1-0.3}$, and $\Delta q_{0.6-0.9}$. The distributions of experimental and predicted values are shown along the horizontal and vertical axes, respectively. (f–j) Top 20 important features and (k) cumulative feature importance for K_H , α , q_{max} , $\Delta q_{0.1-0.3}$, and $\Delta q_{0.6-0.9}$.

IRMOFs, as well as zeolithic imidazolate frameworks (ZIFs).⁹ Cu-based MOFs usually comprise coordinated unsaturated sites provided by strong electrophilic copper atoms; particularly, many of them with a paddle wheel structure like HKUST-1 are commonly investigated.⁵² In addition, Ni-, Co-, and Mg-based MOF-74 exhibit the highest K_H among all the metal types, owing to the large density of open metal sites. Cr-based MOFs (Cr-soc-MOF-1, MIL-100, and MIL-101) have the highest q_{max} , attributing to the strong Lewis acidity and low redox potential of Cr^{3+} . It is worthwhile to note that metal sites may exist as metal clusters or complexes and assemble with different secondary building units in MOFs. Their complex structures play an important role in determining the interactions (e.g., dipole–dipole interaction and hydrogen bonding) with guest molecules.¹⁰ Therefore, it is necessary to further consider their effects on adsorption in the future study. Alternatively, the 285 MOFs can be categorized into 45 types based on their framework topology, as shown in Figure 5b. Topology describes the connectivity and symmetry of metal and organic building blocks.¹⁰ There are 45 different topologies in the collected MOFs, among which the face-centered cubic (fcu), square grid-like (sql), and twisted boracite (tbo) account for the highest probability. Specifically, fcu is present in UiO-66, MIL-12S, and CAU series MOFs, sql is in CID series, and tbo is in HKUST-1. High K_H and high q_{max} are observed in MOFs with tbo and mtn-e (e.g., MIL-101(Cr)) topologies, respectively. Many topologies have a wide α range, manifesting tunable adsorption behavior. Thus,

reticulating predefined building blocks of different topologies would expand the scope of MOFs for water harvesting.

Performance, Accuracy, and Transferability of ML Models.

From the above analysis, both framework structure and chemistry have significant effects on water adsorption. Thus, MOFs were encoded as 8-dimensional structural features and 68-dimensional chemical features, along with operating conditions, to develop ML models for each of the five targets (K_H , α , q_{max} , $\Delta q_{0.1-0.3}$, and $\Delta q_{0.6-0.9}$). The ML models were trained using four algorithms (RFR, GBR, KRR_G, and KRR_L) with hyperparameter grids listed in Table S3. The performance of each ML model was quantified by R^2 and MAE based on 20 different random splits. As illustrated in Figures S9 and S10, both RFR and GBR exhibit higher performance with lower MAE and higher R^2 among the four algorithms. Furthermore, RFR outperforms GBR in circumventing the overfitting problem. On this basis, RFR is chosen for detailed discussion. From the learning curves of MAE in Figure S11, the performance is improved and the data fluctuation is reduced upon increasing the split ratio of training/test sets. As a counterbalance, the split ratio was set as 0.9 for all the ML models. Figure 6a–e compares the ML-predicted targets with experimental data. Among the five targets, the best performance is observed for K_H with MAE of 0.005 mmol/(g·Pa) and R^2 of 0.936 in the test set. For the other four targets, the MAE values of test sets are 0.085, 4.648 mmol/g, 1.461 mmol/g, and 1.876 mmol/g for α , q_{max} , $\Delta q_{0.1-0.3}$, and $\Delta q_{0.6-0.9}$, manifesting the small difference between predicted and experimental data. Compared to K_H ,

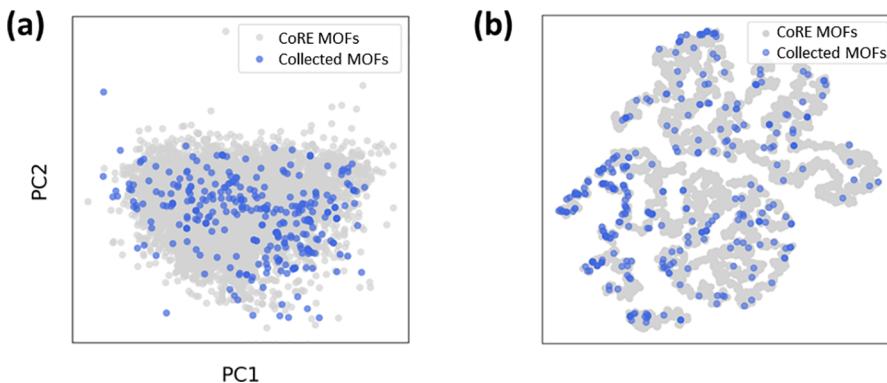


Figure 7. (a) PCA and (b) t-SNE maps. The gray points indicate the overall feature space of 8059 CoRE MOFs and the blue points indicate 285 collected MOFs.

the other four targets have relatively higher heteroscedasticities with R^2 values of 0.813, 0.857, 0.841, and 0.857 for α , q_{\max} , $\Delta q_{0.1-0.3}$, and $\Delta q_{0.6-0.9}$, respectively. The high accuracy for K_H demonstrates that the input features in this study accurately capture decisive factors for the initial water adsorption at low pressure. The accuracy for the other four targets drops slightly as the operating condition is more involved. During isotherm measurement, it is difficult to control the operating condition, thus introducing statistical errors to the experimental data set. These errors could be reduced by increasing the data amount and balancing data distribution in the future study.

To interpret feature contributions, we analyze the feature importance and the partial dependence of important features identified by the Gini-impurity, as displayed in Figures 6f–j and S12. Furthermore, the features are categorized into structural (pore size and pore geometry), chemical (metal and nonmetal atom densities), and operating (temperature). The cumulative feature importance values of different categories for five targets are summarized in Figure 6k.

(1) For K_H , chemical features account for a major contribution. The feature importance values of metal and nonmetal atom densities are 0.347 and 0.283, respectively. During initial water uptake, water molecules are adsorbed at primary adsorption sites including unsaturated metal sites and strong hydrophilic groups (e.g., μ -OH groups).³ Thus, framework chemistry, especially metal atoms, is decisive to K_H . Six metal atom types and five nonmetal atom types occur in the top 20 important features. Metal atoms like Mg, Ni, and Co have a significant positive effect on K_H via the coordination interaction. Meanwhile, O atom also shows a positive effect because it can act as a hydrogen bond acceptor to form hydrogen bonds with water thus facilitating water adsorption. Nevertheless, H atom has a negative effect. The majority of H atoms exist in hydroneutral groups, which cannot form hydrogen bonds with water; instead, they may occupy internal pore volume and hinder water adsorption. Except for chemical features, temperature holds high importance of 0.102 and a negative influence on K_H because adsorption is an exothermic process.

(2) For α , nonmetal atom density and pore geometry show similar high importance of 0.351 and 0.325. Thus, α is determined by both chemical and structural features. The former feature represents a local chemical environment and affects the interaction between framework and water, while the latter feature reflects a global framework structure and affects overall water adsorption capacity. The top 20 important

features determining α include 11 chemical features, eight structural features, as well as temperature. It is worth noting that the features show opposite effects on α and K_H because higher K_H usually corresponds to lower α . The important chemical features are mainly related to organic atom types such as O, C, H, and N. Among all the atom types, O has the highest importance of 0.088 and a negative influence on α , which is consistent with its influence on K_H .

(3) For q_{\max} , it has the highest percentage (0.515) of structural feature importance among all the five targets. As mentioned before, q_{\max} corresponds to overall water adsorption capacity; thus, it is related to the global framework structure. A handful of experimental studies for water adsorption in MOFs were focused on the influence of PV on q_{\max} .^{1,4} However, we find that the surface area (VSA and GSA) and pore size (LFPD, PLD, and LCD) have higher feature importance than PV. A lower surface area and a larger pore size imply a less tortuous framework structure, in which water molecules are more likely to form a broad hydrogen bond network and thus promote q_{\max} . A pore size of $\sim 10 \text{ \AA}$ appears to be a critical size for a sudden increase of q_{\max} which was experimentally found.⁶ Further increase of pore size does not contribute significantly to q_{\max} . In addition, a steep decrease of q_{\max} occurs at VSA of $\sim 4200 \text{ cm}^2/\text{cm}^3$. These observations can be utilized as design principles to improve water harvesting in MOFs.

(4) For $\Delta q_{0.1-0.3}$ and $\Delta q_{0.6-0.9}$, they have the same ranking of important features: nonmetal atom > pore geometry > metal atom > pore size > temperature. Compared to $\Delta q_{0.6-0.9}$, $\Delta q_{0.1-0.3}$ has higher importance on chemical features and lower importance on structural features. This unravels that framework chemistry plays a more important role at a lower pressure, whereas the framework structure is more dominant at a higher pressure. Organic atoms (O, H, and C) exhibit similar importance on $\Delta q_{0.1-0.3}$ and K_H , but metal atoms are less influential on $\Delta q_{0.1-0.3}$ than K_H . A sudden drop is observed in $\Delta q_{0.1-0.3}$ at GSA of $\sim 3400 \text{ cm}^2/\text{g}$ and VSA of $\sim 4200 \text{ cm}^2/\text{cm}^3$. For $\Delta q_{0.6-0.9}$, C atom has a positive influence because it is a major atom type in many MOFs. Most of the structural features show a significant influence on $\Delta q_{0.6-0.9}$. There is a steep rise of $\Delta q_{0.6-0.9}$ at PV of ~ 0.7 , VF of ~ 0.7 , and pore size of $\sim 10 \text{ \AA}$. In addition, temperature has about twice of importance for $\Delta q_{0.1-0.3}$ than $\Delta q_{0.6-0.9}$, showing higher importance at a low pressure.

To test the transferability of the developed ML models, out-of-sample predictions were performed in seven newly reported MOFs with 10 experimentally measured water adsorption

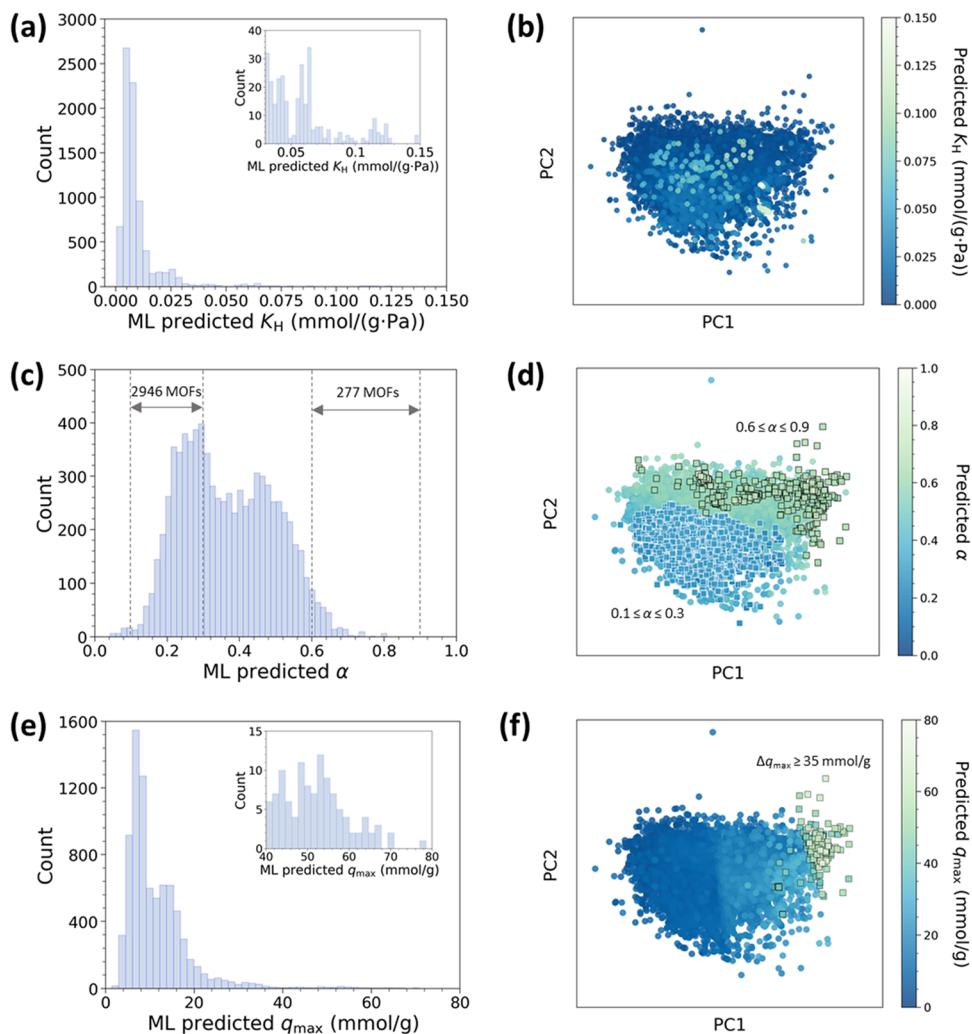


Figure 8. (a,b) Distributions of predicted K_H . The color scale varies with K_H . (c,d) Distributions of predicted α . The color scale varies with α . The squares indicate α within 0.1–0.3 and 0.6–0.9. (e,f) Distributions of predicted q_{\max} . The color scale varies with q_{\max} . The squares indicate $q_{\max} \geq 35$ mmol/g.

isotherms. As listed in Tables S4–S6. Good agreement is observed between the predicted and experimental adsorption properties, with low MAE values, thus the ML models exhibit good transferability. In conclusion, the ML models are accurate and transferable for water adsorption properties. Through the analysis of feature importance and partial dependence, the complex interplays among 77-dimensional features and five targets are elucidated. It is revealed that framework chemistry and operating condition play essential roles at a low pressure, while pore structure is more important at a high pressure.

Predictions and Screening of CoRE MOFs. We apply the best ML models to predict water adsorption properties in CoRE MOFs and screen top-performing MOFs for water harvesting. Derived from experimentally synthesized MOFs, the CoRE MOF database comprises computation-ready cifs. As plotted in Figure S13, there are 98 metal atom types and 24 nonmetal atom types in the CoRE MOFs. A total of 8059 CoRE MOFs have the same feature space as 285 collected MOFs. Figure S14 shows the distributions of structural features in 8059 CoRE MOFs. Compared to the collected MOFs (Figure S2), the CoRE MOFs have lower median pore sizes (5.9, 4.2, and 5.8 Å for LCD, PLD, and LFPD, respectively), higher median surface areas (4720.8 m²/cm³

and 3759.9 m²/g for VSA and GSA), similar median ρ (1.2 kg/m³), lower median VF (0.49), and lower median PV (0.39 m³/g). Higher maximum values of these structural features are observed in the CoRE MOFs. By reducing 77-dimensional features to two principal components, the 8059 CoRE MOFs are illustrated in Figure S15. It is observed that 285 collected MOFs are scattered among the CoRE MOFs. To further examine the similarity between 285 collected MOFs and 8059 CoRE MOFs, we plot PCA and t-SNE maps in Figure 7. In both maps, the collected MOFs span a wide and scattered regime of CoRE MOFs. This implies that the collected MOFs are diverse, cover a wide feature space of CoRE MOFs, and thus can be considered as representative of CoRE MOFs.

Figure 8 shows the ML-predicted K_H , α , and q_{\max} at 298 K in 8059 CoRE MOFs. (1) ML-predicted K_H is centered at ~0.007 mmol/(g·Pa) and there is no significant spatial distribution pattern in the PCA of K_H . This could be explained by the SHAP summary plot in Figure S16, where chemical features especially metal atoms (like Ni, Mg, Co, and Cu) play a major role in predicting K_H . They show a wide range of positive influence and a limited negative impact on K_H . Thus, the ML model is able to quantify water–adsorbent interaction by identifying water adsorption sites. (2) ML-predicted α has a

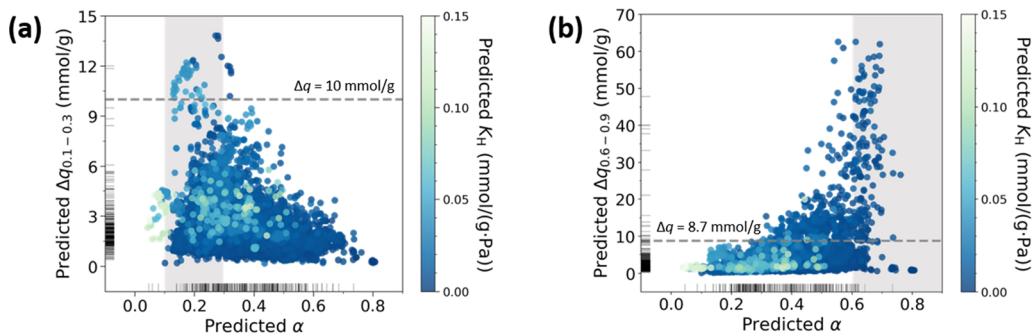


Figure 9. (a) Predicted $\Delta q_{0.1-0.3}$ versus predicted α with predicted K_H as the color scale. The dashed line indicates $\Delta q_{0.1-0.3} = 10 \text{ mmol/g}$ and the gray area indicates $0.1 \leq \alpha \leq 0.3$. (b) Predicted $\Delta q_{0.6-0.9}$ versus predicted α with predicted K_H as the color scale. The dashed line indicates $\Delta q_{0.6-0.9} = 8.7 \text{ mmol/g}$ and the gray area indicates $0.6 \leq \alpha \leq 0.9$.

Table 1. Representative Top-Performing CoRE MOFs Predicted with High $\Delta q_{0.1-0.3}$

MOF	metal	linker	predicted K_H (mmol/(g·Pa))	predicted α	predicted $\Delta q_{0.1-0.3}$ (mmol/g)
WAFKEU05	Zn		0.004	0.28	13.82
SURHEU02	Zr		0.005	0.29	12.55
NEVVAM02	Co		0.013	0.20	12.36
RIFGAQ	Cu		0.039	0.17	12.22
IKETOH	Cu		0.039	0.18	12.05

bimodal distribution centered at 0.25 and 0.5. There are 2946 and 277 CoRE MOFs with α within 0.1–0.3 and 0.6–0.9, respectively. These MOFs are potential candidates for water harvesting in the corresponding operating windows. The PCA plot shows that α increases consistently from the bottom region to the top, which is consistent with the PCA plots of GSA and ρ in Figure S15. GSA, ρ , and O_2 density have the highest and widest SHAP values, as shown in Figure S17. Therefore, structural optimization in terms of GSA and ρ as well as chemical optimization of oxygen-containing groups would have a significant effect on customizing α . (3) ML-predicted q_{\max} has a median value of ~9 mmol/g and a maximum value of ~78.56 mmol/g. There are 149 CoRE MOFs found with $q_{\max} \geq 35 \text{ mmol/g}$. The highest q_{\max} are in the top right corner of the PCA plot, which refers to the lowest surface area and the largest pore size. Figure S18 illustrates the SHAP summary plot and the force plot for q_{\max} in two MOFs (WAGQAY_clean and SICDAL_clean). WAGQAY_clean has the highest q_{\max} of 78.56 mmol/g and the SHAP force plot indicates that its low VSA of 1495.27 m²/cm³, large PV of 2.1437 m³/g, large pore size, and low H density affect q_{\max} in a positive way. On the contrary, SICDAL_clean has the lowest q_{\max} due to high VSA of 5792.34 m²/cm³, accompanying low pore size and high H density.

The ML models are further applied to screen top-performing MOFs for p/p_0 in 0.1–0.3 and 0.6–0.9 windows, respectively, as shown in Figure 9. For water adsorption, it is recognized that reversible pore filling and irreversible capillary condensation can be differentiated by a critical diameter (D_c); D_c

is approximately 20.76 Å at 298 K.² On this basis, 7953 CoRE MOFs with PLD < 20.76 Å are used for screening. Moreover, efficient water harvesting requires adsorbents to provide high working capacity, step-shaped cycling adsorption, and easy regeneration, which are closely related to Δq , α , and K_H . In the operating window of $p/p_0 = 0.1-0.3$, α of 0.1–0.3 and $\Delta q_{0.1-0.3}$ of 10 mmol/g are set as thresholds. There are 2946 MOFs with ML-predicted α within 0.1–0.3 and 39 of them have $\Delta q_{0.1-0.3} \geq 10 \text{ mmol/g}$. Table S7 lists the encoded features and ML-predicted properties in these 39 MOFs, with five representative ones in Table 1. K_H varies from 0.003 to 0.056 mmol/(g·Pa) and is mostly higher than the median value of 0.007 mmol/(g·Pa). This is because high Δq is always accompanied by strong water affinity, which is particularly significant under a low-pressure condition. We should note that high K_H is often accompanied by low water stability of MOFs. Thus, a trade-off between water affinity and stability should be weighed with the assistance of K_H values. In the collected data set, K_H for stable MOFs is in the range of 0.00004–0.05 mmol/(g·Pa). Considering 0.05 mmol/(g·Pa) as a threshold, 36 MOFs in Table S7 hold the potential to be water stable. However, water stability is jointly determined by multiple factors, rather than by K_H alone, and further investigation is desired. Figure S19 illustrates the SHAP summary plot, as well as the force plots for $\Delta q_{0.1-0.3}$ and K_H in WAFKEU05_clean. This MOF has the highest $\Delta q_{0.1-0.3}$ of 13.82 mmol/g, attributing to high densities of O and Zn atoms and favorable structures.³³ In addition, its low K_H of 0.004 mmol/(g·Pa) is desired for cyclic water harvesting.

Table 2. Representative Top-Performing CoRE MOFs Predicted with High $\Delta q_{0.6-0.9}$

MOF	metal	linker	predicted K_H (mmol/(g·Pa))	predicted α	predicted $\Delta q_{0.6-0.9}$ (mmol/g)
 WAGQAY	Ni		0.007	0.66	58.07
 EDUWAB	Zn		0.011	0.63	56.06
 ECOKAJ	Zn		0.007	0.63	54.83
 FEBXIV	Cu		0.009	0.64	53.73
 ADUROI	Zn		0.006	0.64	53.22

In the operating window of $p/p_0 = 0.6-0.9$, α of 0.6–0.9 and $\Delta q_{0.6-0.9}$ of 8.7 mmol/g are set as thresholds. There are 248 MOFs with ML-predicted α within 0.6–0.9 and 139 of them have $\Delta q_{0.6-0.9} > 8.7$ mmol/g. Table S8 lists the encoded features and ML-predicted properties of these 139 MOFs, with five representative ones in Table 2. Most of them have K_H lower than 0.007 mmol/(g·Pa). If considering K_H of 0.05 mmol/(g·Pa) as a threshold, all the MOFs in Table S8 are potentially water stable. Figure S20 shows the SHAP summary plot, as well as the force plots for $\Delta q_{0.6-0.9}$ in two MOFs (WAGQAY_clean and ADUROI_clean). WAGQAY_clean has the highest $\Delta q_{0.6-0.9}$ of 58.07 mmol/g, which is also predicted to have the highest q_{\max} . It comprises $\{\text{Ni}_5(\mu_3\text{-O})_2(\text{O}_2\text{C})_6\}$ clusters with a high density of open metal sites and benzene-1,3,5-tribenzoate linkers.⁵⁴ ADUROI_clean has a comparable $\Delta q_{0.6-0.9}$ of 53.22 mmol/g, but significant lower K_H of 0.006 mmol/(g·Pa). It is composed of two kinds of enantiotropic Zn centers and a flexible heterodonor linker, which provides suitable pore size, low ρ , and high PV.⁵⁵ These shortlisted MOFs are potential candidates with favorable adsorption properties for water harvesting.

CONCLUSIONS

In this study, we have developed an ML method to predict water adsorption in MOFs and identify top-performing CoRE MOFs for water harvesting. From 344 experimental water adsorption isotherms in 285 MOFs, five key adsorption properties are extracted. The intrinsic structure–property relationships are established with an emphasis on pore structure and framework chemistry. Then, five independent ML models are trained to predict five adsorption properties upon 77-dimension descriptors, and interpreted by feature importance and partial dependence based on the Gini-impurity and SHAP. It is unraveled that K_H is dominated by framework chemistry and operating temperature; α is determined by pore structure and framework chemistry; q_{\max} is largely governed by pore structure. Both $\Delta q_{0.1-0.3}$ and $\Delta q_{0.6-0.9}$ are jointly decided by framework chemistry, pore structure, and operating temperature. In a relative pressure window of 0.1–0.3, operating temperature and framework chemistry, especially unsaturated metal sites and nonmetal atoms related to hydrophilic functional groups, play essential roles. In a relative pressure window of 0.6–0.9, pore structure like density and pore size is more important. The ML models exhibit good accuracy and transferability. Finally, the ML models are applied

to screen CoRE MOFs for water harvesting. Among 8059 CoRE MOFs, 149 are shortlisted with $q_{\max} \geq 35$ mmol/g, 39 with $\Delta q_{0.1-0.3} \geq 10$ mmol/g, and 139 with $\Delta q_{0.6-0.9} \geq 8.7$ mmol/g. The ML-based method in this study shows great potential for adsorption property prediction and rapid screening of MOFs for water harvesting. Such data-driven informatics is promising for task-oriented materials design and discovery for energy and environmental applications.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acssuschemeng.3c01233>.

MOF data set and featurization; structure–property relationships; ML model parameters and performance; partial dependence analysis; featurization of prediction data set; feature analysis of predictions; and top-performing MOFs (PDF)

Experimental data sources (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Zhongyi Jiang – Joint School of National University of Singapore and Tianjin University, International Campus of Tianjin University, Binhai New City, Fuzhou 350207, P. R. China; Key Laboratory for Green Chemical Technology of Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China;  orcid.org/0000-0002-0048-8849; Email: zhyjiang@tju.edu.cn

Jianwen Jiang – Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117576, Singapore;  orcid.org/0000-0003-1310-9024; Email: chejj@nus.edu.sg

Authors

Zhiming Zhang – Joint School of National University of Singapore and Tianjin University, International Campus of Tianjin University, Binhai New City, Fuzhou 350207, P. R. China; Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117576, Singapore;  orcid.org/0000-0001-9444-8479

Hongjian Tang – Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117576, Singapore

Mao Wang – Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117576, Singapore;  orcid.org/0000-0001-7583-1683

Bohui Lyu – Joint School of National University of Singapore and Tianjin University, International Campus of Tianjin University, Binhai New City, Fuzhou 350207, P. R. China; Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117576, Singapore

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acssuschemeng.3c01233>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully acknowledge A*STAR LCER-FI project (LCERFI01-0015 U2102d2004), A*STAR AME IRG Grant (A20E5c0092), the Ministry of Education of Singapore (R-279-000-578-112, R-279-000-598-114, R-279-000-574-114), and the National Natural Science Foundation of China (No. U20B2023 and 22208238) for financial support.

REFERENCES

- (1) Gordeeva, L. G.; Tu, Y. D.; Pan, Q.; Palash, M. L.; Saha, B. B.; Aristov, Y. I.; Wang, R. Z. Metal–Organic Frameworks for Energy Conversion and Water Harvesting: A Bridge between Thermal Engineering and Material Science. *Nano Energy* **2021**, *84*, No. 105946.
- (2) Kalmutzki, M. J.; Diercks, C. S.; Yaghi, O. M. Metal–Organic Frameworks for Water Harvesting from Air. *Adv. Mater.* **2018**, *30*, No. 1704304.
- (3) Hanikel, N.; Prévot, M. S.; Yaghi, O. M. MOF Water Harvesters. *Nat. Nanotechnol.* **2020**, *15*, 348–355.
- (4) Byun, Y.; Je, S. H.; Talapaneni, S. N.; Coskun, A. Advances in Porous Organic Polymers for Efficient Water Capture. *Chem. – Eur. J.* **2019**, *25*, 10262–10283.
- (5) Canivet, J.; Fateeva, A.; Guo, Y.; Coasne, B.; Farrusseng, D. Water Adsorption in MOFs: Fundamentals and Applications. *Chem. Soc. Rev.* **2014**, *43*, 5594–5617.
- (6) Liu, X.; Wang, X.; Kapteijn, F. Water and Metal–Organic Frameworks: From Interaction toward Utilization. *Chem. Rev.* **2020**, *120*, 8303–8377.
- (7) Kim, H.; Rao, S. R.; Kapustin, E. A.; Zhao, L.; Yang, S.; Yaghi, O. M.; Wang, E. N. Adsorption-Based Atmospheric Water Harvesting Device for Arid Climates. *Nat. Commun.* **2018**, *9*, 1191.
- (8) Kim, H.; Yang, S.; Rao, S. R.; Narayanan, S.; Kapustin, E. A.; Furukawa, H.; Umans, A. S.; Yaghi, O. M.; Wang, E. N. Water Harvesting from Air with Metal–Organic Frameworks Powered by Natural Sunlight. *Science* **2017**, *356*, 430–434.
- (9) Furukawa, H.; Cordova, K. E.; O’Keeffe, M.; Yaghi, O. M. The Chemistry and Applications of Metal–Organic Frameworks. *Science* **2013**, *341*, No. 1230444.
- (10) Kalmutzki, M. J.; Hanikel, N.; Yaghi, O. M. Secondary Building Units as the Turning Point in the Development of the Reticular Chemistry of MOFs. *Sci. Adv.* **2018**, *4*, No. eaat9180.
- (11) Towsif Abtab, S. M.; Alezi, D.; Bhatt, P. M.; Shkurenko, A.; Belmabkhout, Y.; Aggarwal, H.; Weseliński, Ł. J.; Alsadun, N.; Samin, U.; Hedhili, M. N.; Eddaoudi, M. Reticular Chemistry in Action: A Hydrolytically Stable MOF Capturing Twice Its Weight in Adsorbed Water. *Chem.* **2018**, *4*, 94–105.
- (12) Cho, K. H.; Borges, D. D.; Lee, U.-H.; Lee, J. S.; Yoon, J. W.; Cho, S. J.; Park, J.; Lombardo, W.; Moon, D.; Sapienza, A.; Maurin, G.; Chang, J.-S. Rational Design of a Robust Aluminum Metal–Organic Framework for Multi-Purpose Water-Sorption-Driven Heat Allocations. *Nat. Commun.* **2020**, *11*, 5112.
- (13) de Lange, M. F.; Verouden, K. J. F. M.; Vlugt, T. J. H.; Gascon, J.; Kapteijn, F. Adsorption-Driven Heat Pumps: The Potential of Metal–Organic Frameworks. *Chem. Rev.* **2015**, *115*, 12205–12250.
- (14) Hanikel, N.; Pei, X.; Chheda, S.; Lyu, H.; Jeong, W.; Sauer, J.; Gagliardi, L.; Yaghi, O. M. Evolution of Water Structures in Metal–Organic Frameworks for Improved Atmospheric Water Harvesting. *Science* **2021**, *374*, 454–459.
- (15) Gong, W.; Xie, H.; Idrees, K. B.; Son, F. A.; Chen, Z.; Sha, F.; Liu, Y.; Cui, Y.; Farha, O. K. Water Sorption Evolution Enabled by Reticular Construction of Zirconium Metal–Organic Frameworks Based on a Unique [2.2]Paracyclophane Scaffold. *J. Am. Chem. Soc.* **2022**, *144*, 1826–1834.
- (16) Henninger, S. K.; Jeremias, F.; Kummer, H.; Janiak, C. MOFs for Use in Adsorption Heat Pump Processes. *Eur. J. Inorg. Chem.* **2012**, *2012*, 2625–2634.
- (17) Seo, Y.-K.; Yoon, J. W.; Lee, J. S.; Hwang, Y. K.; Jun, C.-H.; Chang, J.-S.; Wuttke, S.; Bazin, P.; Vimont, A.; Daturi, M.; Bourrelly, S.; Llewellyn, P. L.; Horcajada, P.; Serre, C.; Férey, G. Energy-Efficient Dehumidification over Hierachically Porous Metal–Organic Frameworks as Advanced Water Adsorbents. *Adv. Mater.* **2012**, *24*, 806–810.
- (18) Kim, S.-I.; Yoon, T.-U.; Kim, M.-B.; Lee, S.-J.; Hwang, Y. K.; Chang, J.-S.; Kim, H.-J.; Lee, H.-N.; Lee, U.-H.; Bae, Y.-S. Metal–Organic Frameworks with High Working Capacities and Cyclic Hydrothermal Stabilities for Fresh Water Production. *Chem. Eng. J.* **2016**, *286*, 467–475.
- (19) Jeremias, F.; Khutia, A.; Henninger, S. K.; Janiak, C. MIL-100(Al, Fe) as Water Adsorbents for Heat Transformation Purposes—a Promising Application. *J. Mater. Chem.* **2012**, *22*, 10148–10151.
- (20) AbdulHalim, R. G.; Bhatt, P. M.; Belmabkhout, Y.; Shkurenko, A.; Adil, K.; Barbour, L. J.; Eddaoudi, M. A Fine-Tuned Metal–Organic Framework for Autonomous Indoor Moisture Control. *J. Am. Chem. Soc.* **2017**, *139*, 10715–10722.
- (21) Girnik, I. S.; Aristov, Y. I. Dynamic Optimization of Adsorptive Chillers: The “AQSOA™-FAM-Z02 – Water” Working Pair. *Energy* **2016**, *106*, 13–22.
- (22) Wang, Z.; Sun, Z.; Yin, H.; Liu, X.; Wang, J.; Zhao, H.; Pang, C. H.; Wu, T.; Li, S.; Yin, Z.; Yu, X.-F. Data-Driven Materials Innovation and Applications. *Adv. Mater.* **2022**, *34*, No. 2104113.
- (23) Axelrod, S.; Schwalbe-Koda, D.; Mohapatra, S.; Damewood, J.; Greenman, K. P.; Gómez-Bombarelli, R. Learning Matter: Materials Design with Machine Learning and Atomistic Simulations. *Acc. Mater. Res.* **2022**, *3*, 343–357.
- (24) Wang, M.; Xu, Q.; Tang, H.; Jiang, J. Machine Learning-Enabled Prediction and High-Throughput Screening of Polymer Membranes for Pervaporation Separation. *ACS Appl. Mater. Interfaces* **2022**, *14*, 8427–8436.
- (25) Wang, M.; Jiang, J. Accelerating Discovery of High Fractional Free Volume Polymers from a Data-Driven Approach. *ACS Appl. Mater. Interfaces* **2022**, *14*, 31203–31215.
- (26) Moosavi, S. M.; Novotny, B. A.; Ongari, D.; Moubarak, E.; Asgari, M.; Kadioglu, Ö.; Charalambous, C.; Ortega-Guerrero, A.; Farmahini, A. H.; Sarkisov, L.; Garcia, S.; Noé, F.; Smit, B. A Data-Science Approach to Predict the Heat Capacity of Nanoporous Materials. *Nat. Mater.* **2022**, *21*, 1419–1425.
- (27) Yao, Z.; Sánchez-Lengeling, B.; Bobbitt, N. S.; Bucior, B. J.; Kumar, S. G. H.; Collins, S. P.; Burns, T.; Woo, T. K.; Farha, O. K.; Snurr, R. Q.; Aspuru-Guzik, A. Inverse Design of Nanoporous Crystalline Reticular Materials with Deep Generative Models. *Nat. Mach. Intell.* **2021**, *3*, 76–86.
- (28) Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the Diversity of the Metal–Organic Framework Ecosystem. *Nat. Commun.* **2020**, *11*, 4068.

- (29) Zhang, X.; Zhang, K.; Yoo, H.; Lee, Y. Machine Learning-Driven Discovery of Metal–Organic Frameworks for Efficient CO₂ Capture in Humid Condition. *ACS Sustainable Chem. Eng.* **2021**, *9*, 2872–2879.
- (30) Anderson, R.; Rodgers, J.; Argueta, E.; Biong, A.; Gómez-Gualdrón, D. A. Role of Pore Chemistry and Topology in the CO₂ Capture Capabilities of MOFs: From Molecular Simulation to Machine Learning. *Chem. Mater.* **2018**, *30*, 6325–6337.
- (31) Gurnani, R.; Yu, Z.; Kim, C.; Sholl, D. S.; Ramprasad, R. Interpretable Machine Learning-Based Predictions of Methane Uptake Isotherms in Metal–Organic Frameworks. *Chem. Mater.* **2021**, *33*, 3543–3552.
- (32) Tang, H.; Jiang, J. Active Learning Boosted Computational Discovery of Covalent–Organic Frameworks for Ultrahigh CH₄ Storage. *AIChE J.* **2022**, *68*, No. e17856.
- (33) Tang, H.; Xu, Q.; Wang, M.; Jiang, J. Rapid Screening of Metal–Organic Frameworks for Propane/Propylene Separation by Synergizing Molecular Simulation and Machine Learning. *ACS Appl. Mater. Interfaces* **2021**, *13*, 53454–53467.
- (34) Oviedo, F.; Ferres, J. L.; Buonassisi, T.; Butler, K. T. Interpretable and Explainable Machine Learning for Materials Science and Chemistry. *Acc. Mater. Res.* **2022**, *3*, 597–607.
- (35) Liu, Z.; Li, W.; Cai, S.; Tu, Z.; Luo, X.; Li, S. Large-Scale Cascade Cooling Performance Evaluation of Adsorbent/Water Working Pairs by Integrated Mathematical Modelling and Machine Learning. *J. Mater. Chem. A* **2022**, *10*, 9604–9611.
- (36) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64*, 5985–5998.
- (37) Siderius, D. W.; Shen, V. K.; Johnson, R. D., III; van Zee, R. D. NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials. 2020, DOI: [10.18434/T43882](https://doi.org/10.18434/T43882).
- (38) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 171–179.
- (39) Ankit, Rohatgi. WebPlotDigitizer. <https://automeris.io/WebPlotDigitizer>.
- (40) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials. *Microporous Mesoporous Mater.* **2012**, *149*, 134–141.
- (41) Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. Extension of the Universal Force Field to Metal–Organic Frameworks. *J. Chem. Theory Comput.* **2014**, *10*, 880–891.
- (42) Thompson, A. P.; Aktulg, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* **2022**, *271*, No. 108171.
- (43) Bucior, B. J.; Rosen, A. S.; Haranczyk, M.; Yao, Z.; Ziebel, M. E.; Farha, O. K.; Hupp, J. T.; Siepmann, J. I.; Aspuru-Guzik, A.; Snurr, R. Q. Identification Schemes for Metal–Organic Frameworks To Enable Rapid Search and Cheminformatics Analysis. *Cryst. Growth Des.* **2019**, *19*, 6682–6697.
- (44) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (45) Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; Layton, R.; Vanderplas, J.; Joly, A.; Holt, B.; Varoquaux, G. API Design for Machine Learning Software: Experiences from the Scikit-Learn Project. arXiv September 1, 2013, DOI: [10.48550/arXiv.1309.0238](https://doi.org/10.48550/arXiv.1309.0238).
- (46) Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2017; Vol. 30.
- (47) Archer, K. J.; Kimes, R. V. Empirical Characterization of Random Forest Variable Importance Measures. *Comput. Stat. Data Anal.* **2008**, *52*, 2249–2260.
- (48) van der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
- (49) Climate of Singapore. <http://www.weather.gov.sg/climate-climate-of-singapore/>.
- (50) Furukawa, H.; Gándara, F.; Zhang, Y.-B.; Jiang, J.; Queen, W. L.; Hudson, M. R.; Yaghi, O. M. Water Adsorption in Porous Metal–Organic Frameworks and Related Materials. *J. Am. Chem. Soc.* **2014**, *136*, 4369–4381.
- (51) Bai, Y.; Dou, Y.; Xie, L.-H.; Rutledge, W.; Li, J.-R.; Zhou, H.-C. Zr-Based Metal–Organic Frameworks: Design, Synthesis, Structure, and Applications. *Chem. Soc. Rev.* **2016**, *45*, 2327–2367.
- (52) Chen, Y.; Zhang, Y.; Huang, Q.; Lin, X.; Zeb, A.; Wu, Y.; Xu, Z.; Xu, X. Recent Advances in Cu-Based Metal–Organic Frameworks and Their Derivatives for Battery Applications. *ACS Appl. Energy Mater.* **2022**, *5*, 7842–7873.
- (53) Dybtsev, D. N.; Chun, H.; Kim, K. Rigid and Flexible: A Highly Porous Metal–Organic Framework with Unusual Guest-Dependent Dynamic Behavior. *Angew. Chem., Int. Ed.* **2004**, *43*, 5033–5036.
- (54) Gedrich, K.; Senkovska, I.; Klein, N.; Stoeck, U.; Henschel, A.; Lohe, M. R.; Baburin, I. A.; Mueller, U.; Kaskel, S. A Highly Porous Metal–Organic Framework with Open Nickel Sites. *Angew. Chem., Int. Ed.* **2010**, *49*, 8489–8492.
- (55) Yun, R.; Lu, Z.; Pan, Y.; You, X.; Bai, J. Formation of a Metal–Organic Framework with High Surface Area and Gas Uptake by Breaking Edges Off Truncated Cuboctahedral Cages. *Angew. Chem., Int. Ed.* **2013**, *52*, 11282–11285.