scientific reports

OPEN

Towards estimation of CO_2 adsorption on highly porous MOF-based adsorbents using gaussian process regression approach

Majedeh Gheytanzadeh[1], Alireza Baghban[2✉], Sajjad Habibzadeh[1,3✉], Amin Esmaeili[4], Otman Abida[5], Ahmad Mohaddespour[5] & Muhammad Tajammal Munir[5]

In recent years, new developments in controlling greenhouse gas emissions have been implemented to address the global climate conservation concern. Indeed, the earth's average temperature is being increased mainly due to burning fossil fuels, explicitly releasing high amounts of CO_2 into the atmosphere. Therefore, effective capture techniques are needed to reduce the concentration of CO_2. In this regard, metal organic frameworks (MOFs) have been known as the promising materials for CO_2 adsorption. Hence, study on the impact of the adsorption conditions along with the MOFs structural properties on their ability in the CO_2 adsorption will open new doors for their further application in CO_2 separation technologies as well. However, the high cost of the corresponding experimental study together with the instrument's error, render the use of computational methods quite beneficial. Therefore, the present study proposes a Gaussian process regression model with four kernel functions to estimate the CO_2 adsorption in terms of pressure, temperature, pore volume, and surface area of MOFs. In doing so, 506 CO_2 uptake values in the literature have been collected and assessed. The proposed GPR models performed very well in which the exponential kernel function, was shown as the best predictive tool with R^2 value of 1. Also, the sensitivity analysis was employed to investigate the effectiveness of input variables on the CO_2 adsorption, through which it was determined that pressure is the most determining parameter. As the main result, the accurate estimate of CO_2 adsorption by different MOFs is obtained by briefly employing the artificial intelligence concept tools.

Abbreviations

ANFIS	Adaptive neuro fuzzy inference system
ANN	Artificial neural network
BDP	1,4-Benzenedipyrazolate
Be-BTB	Beryllium benzene tribenzoate
BTC	Benzene-1,3,5-tricarboxylate
BTTri	1,3,5-Benzenetristriazolate
GPR	Gaussian process regression
H	Hat matrix
H^*	Critical leverage limit
I_n	Unit array
k	Covariance (kernel) function

[1]Surface Reaction and Advanced Energy Materials Laboratory, Chemical Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. [2]Chemical Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), Mahshahr Campus, Mahshahr, Iran. [3]Department of Chemical Engineering, McGill University, 3610 University Street, Montreal, QC H3A 0C5, Canada. [4]Department of Chemical Engineering, School of Engineering Technology and Industrial Trades, College of the North Atlantic—Qatar, Doha, Qatar. [5]College of Engineering and Technology, American University of the Middle East, Egaila, Kuwait. ✉email: Alireza_baghban@alumni.ut.ac.ir; sajjad.habibzadeh@mail.mcgill.ca

1

K	Modified Bessel function
ℓ	Length scale
LS-SVM	Least square-support vector machine
m	Mean function
MSE	Mean square error
MOF	Metal organic framework
MRE	Mean relative error
P	Pressure (bar)
T	Temperature (K)
r	Relevancy factor
R^2	Difference between the experiments and the calculated values
RBF	Radial basis function
RF	Random forest
RMSE	Root mean square error
S	Surface area (m^2/g)
STD	Standard deviation
U	Matrix dimensional of i * j
V	Volume (cm^3/g)
x	Input variable
$\underline{X}$	Input parameter in r eq.
$\overline{X}$	Input average in r eq.
y	Target variable
$\underline{Y}$	Output parameter in r eq.
$\overline{Y}$	Average of outputs in r eq

Greek symbols

f	A random function
ε	Observation noise
σ^2_{noise}	Variance of the noise
μ_T	Mean value
Σ_T	Covariance
α	Scale mixture
σ	Amplitude
Γ	Gamma function

Suscripts

E	Exponential kernel function
i	Number of parameters
j	Number of training points
k	Number of input parameter in r eq.
M	Matern kernel function
p	Pore
RQ	Rational quadratic kernel function
SE	Squared exponential kernel function
T	Testing data
v	Positive parameter

The concentration of atmospheric CO_2 has increased from 270 ppm before the industrial revolution to more than 400 ppm today, mainly due to the increasing consumption of fossil fuels[1]. In addition, it is widely believed that CO_2 has a major role in global climate change[2]. Thus, carbon capture technology has been employed as a promising route to reduce the CO_2 concentration into the atmosphere and inhibit global warming[3,4]. Several approaches have been studied for CO_2 capture: membranes[5,6], chemical absorption[7,8], physical adsorption[9], and fluidized bed technologies[10]. However, these methods suffer from some drawbacks, such as high energy consumption, complex regeneration processes, and low CO_2 capture capacity. In order to build up a long-lasting chance in CO_2 elimination, an appropriate adsorption medium should provide the following conditions: (1) a periodical structure for both the capture and release of CO_2 reversibly, (2) high CO_2 selectivity, (3) optimized CO_2 adsorption capacities through modifying by chemical functionalization, and (4) thermal, chemical, and mechanical stabilities[11,12]. Metal–organic frameworks (MOFs) have been one of the most applicable porous compounds due to their regulating chemical structure, adjustable chemical functionality, and high thermal stability, allowing potential applications in gas adsorption[13–16].

MOFs are formed by a combination of two main parts of metal ions or clusters and organic ligands, creating a 3D structure with a network of channels and uniform pores. In addition to the robust 3D structure, the main characteristics of the MOFs are their permanent porosity and modular nature. These features of MOFs support them in adsorbing other molecules as a guest and sustaining their structures with negligible damage[17,18]. In comparison to the other porous materials, the most important advantage of the MOFs is their possibility to design the functionality and the pore size by choosing the metal ion, the functional group, the organic ligand, and the activation method[19]. The properties of MOFs depend on the metal of interest and the linker. For example, MOF-5

or IRMOF-1, containing zinc atoms linked to terephthalic acid molecules, possess a big void for gas capture, while M-dobdc or M-MOF-74 (M = Mg, Ni, Co, Zn), with unsaturated metal centers in their 3D structures, provide extra sites to bond with guest molecules[20,21]. Besides, the pore sizes of MOFs change from several angstroms to a few nanometers based on the diverse organic linkers[15].

Several studies reported high CO_2 adsorption capacity for MOF materials, ranging from 8.0 to 10.2 mol/kg at 298 K and 15 bar. CuBTC or HKUST-1 is one of the most explored MOFs for gas adsorption and storage[17,22–24]. To compare the adsorption capacity in zeolites and MOFs, at higher pressures, the adsorption capacity of the benchmark zeolite 13X is much lesser than that of MOFs[22]. Additionally, when the micropore diffusion is the rate control mechanism for CO_2 adsorption, the adsorption process in NaX and 5A zeolites proceeds slower than in MOF materials[25]. MOFs are promising candidates for gas adsorption applications among the various porous materials based on the mentioned features.

Despite numerous studies reported about gas–solid adsorption systems, investigating this phenomenon from a cohesive viewpoint is still challenging[26]. The experimental studies are time-consuming and costly, through which the instruments' errors affect the adsorption results. On the other hand, many adsorption isotherms are usable just for a specific range of data because they have been developed under simplified conditions assumptions[27]. Accordingly, a comprehensive and accurate model for examining the adsorption of a gas on MOFs should be developed. Intelligent methods (machine learning algorithms), namely, least-square support vector machine (LS-SVM), artificial neural network (ANN), random forest (RF) adaptive neuro-fuzzy inference system (ANFIS), and radial basis function network (RBF), can be possibly hired as an alternative to mathematical models for solving problems precisely and without the experimental works' troubles[28,29]. Compared to the conventional mathematic approaches, the smart models have gained excellent success in solving complex and non-linear optimization problems[30–39].

In the current study, an intelligent model is used to predict the non-linear system of CO_2 capture by MOFs materials. For the first time, a machine learning algorithm of GPR with four various kernel functions was developed to evaluate the CO_2 uptake on MOFs. Thirteen MOFs with different porosity and structural features including: $Cu_3(BTC)_2$, MOF-505, MOF-74, IRMOF-11, beryllium benzene tribenzoate (Be-BTB), MOF-177, IRMOF-1, IRMOFs-3, IRMOFs-6, MOF-2, Cu-BTTri (BTTri^{3-} = 1,3,5-benzenetristriazolate), Mg^{2-} (dobdc) (dobdc^{4-} = 1,4-dioxido-2,5-benzenedicarboxylate), and Co(BDP) (BDP^{2-} = 1,4-benzenedipyrazolate) based on experimental data were studied[17,40]. Pressure, temperature, pore volume, and surface area of MOFs are considered the model's inputs. Several statistical analyses were applied to investigate the established model, while analysis of sensitivity was used to determine the effective factors on the CO_2 adsorption by MOFs. Additionally, to assess the precision of the proposed GPR models, the predicted results were compared with the experimental CO_2 adsorption values in the literature.

Methodology

Gaussian process regression. This study used the machine learning technique, GPRs model, because they are able to deal with uncertainty in a probabilistic framework (Bayesian) and overcome the complex issues straightforwardly[41,42]. The non-linear GPR models need less training data and can combine new evidence when the available data increases. Typically, the low number of hyper-parameters to optimize through training makes this model less affected by the "overfitting" problem[43]. In the GPR technique, the training sample information determines the parameters of the model. Then, the GPR model is developed via adding the previous information to the modeling procedure and merging the actual (laboratory-measured) data[41]. In contrast to the traditional learning models, the GPR works through computing posterior distributions over models instead of finding the most acceptable match to the experimental data[44].

Generally, the GPR model is established in this way: if the input and the target variables are represented by x and y, assume $T = \{x_{T \cdot i} \cdot y_{T \cdot i}\}_{i=1}^n$ and $L = \{x_{L \cdot i} \cdot y_{L \cdot i}\}_{i=1}^n$ as the arbitrarily chosen test and training data sets, respectively. The starting step in the GPR modeling is the following general equation:

$$y_{L \cdot i} = f(x_{L \cdot i}) + \varepsilon_{L \cdot i} \quad i = 1, 2, 3, \ldots, n \tag{1}$$

where x_L indicates the independent variables and y_L represents the targets of the learning data points. The $\varepsilon \sim N(0 \cdot \sigma_{noise}^2 I_n)$, σ^2_{noise}, and I_n are the observation noise, the variance of the noise, and the unit array, respectively. Therefore, each measured y is connected to the function f(x) by Gaussian noise model[45]. GPR assumes f as a random function that can be entirely defined by its covariance and mean functions. Likewise, we can write:

$$y_{T \cdot i} = f(x_{T \cdot i}) + \varepsilon_{T \cdot i} \quad i = 1, 2, 3, \ldots, n \tag{2}$$

where x_T denotes the independent variables, and y_T is the targets of the testing data sets. Also, the f(x) is distributed as a Gaussian process with covariance function k(x, x′) (also called kernel function) and mean function m(x) [45]:

$$f(x_{L \cdot i}) \sim GP\big(m(x) \cdot k(x \cdot x')\big) \tag{3}$$

The mean function m(x) can be specified by using the explicit basis functions. Usually, the calculations are simplified by considering m(x) to be zero because it can be challenging to identify a fixed m(x)[41,45]. Thus, we have:

$$f(x_{L \cdot i}) \sim GP\big(0 \cdot k(x \cdot x')\big) \tag{4}$$

The distribution of y is achieved by the combination of Eqs. (1) and (4):

$$y \sim N\left(0 \cdot k\left(x \cdot x'\right) + \sigma_{noise}^2 I_n\right) \tag{5}$$

Considering all the above-described parameters and noises, we have:

$$\begin{bmatrix} \overrightarrow{f_L} \\ \overrightarrow{f_T} \end{bmatrix} \sim N\left(0 \cdot \begin{bmatrix} k(x_L \cdot x_L) & k(x_L \cdot x_T) \\ k(x_T \cdot x_L) & k(x_T \cdot x_T) \end{bmatrix}\right) \tag{6}$$

$$\begin{bmatrix} \overrightarrow{\varepsilon_L} \\ \overrightarrow{\varepsilon_T} \end{bmatrix} \sim N\left(0 \cdot \begin{bmatrix} \sigma_{noise}^2 I_n & 0 \\ 0 & \sigma_{noise}^2 I_n \end{bmatrix}\right) \tag{7}$$

The summation of Eqs. (6) and (7) gives the following Gaussian expression:

$$\begin{bmatrix} \overrightarrow{y_L} \\ \overrightarrow{y_T} \end{bmatrix} \sim N\left(0 \cdot \begin{bmatrix} k(x_L \cdot x_L) + \sigma_{noise}^2 I_n & k(x_L \cdot x_T) \\ k(x_T \cdot x_L) & k(x_T \cdot x_T) + \sigma_{noise}^2 I_n \end{bmatrix}\right) \tag{8}$$

Then, the distribution of the y_T can be derived through the conditioning rule of Gaussians, in which μ_T and Σ_T are the mean value and the covariance:

$$\left(y_T | y_L\right) \sim N(\mu_T \cdot \Sigma_T) \tag{9}$$

$$\mu_T = m\left(\overrightarrow{y_T}\right) = k(x_T \cdot x_L)\left(k(x_L \cdot x_L) + \sigma_{noise}^2 I_n\right)^{-1} \overrightarrow{y_T} \tag{10}$$

$$\begin{aligned} \Sigma_T = k(x_T \cdot x_T) &= k(x_T \cdot x_T) + \sigma_{noise}^2 I_n \\ &- k(x_T \cdot x_L)\left(k(x_L \cdot x_L) + \sigma_{noise}^2 I_n\right)^{-1} k(x_L \cdot x_T) \end{aligned} \tag{11}$$

e given independent variable and the training data set can obtain the outputs prediction of the test data. In training, choosing a powerful kernel function, which has an invertible and symmetric matrix, could significantly affect the estimation power of the established GPR model. To find the most appropriate kernel function for the current study, the learning method was conducted, through which four common and diverse kernel functions of Matern, Exponential, Squared exponential, and Rational quadratic are manipulated. These functions have the following forms:

- Matern kernel function:

$$k_M\left(x \cdot x'\right) = \sigma^2 \frac{2^{1-v}}{\Gamma(v)}\left(\sqrt{2v}\frac{x-x'}{\ell}\right)^v K_v\left(\sqrt{2v}\frac{x-x'}{\ell}\right) \tag{12}$$

- Exponential kernel function:

$$k_E\left(x \cdot x'\right) = \sigma^2 exp\left(-\frac{x-x'}{\ell}\right) \tag{13}$$

- Rational quadratic kernel function:

$$k_{RQ}\left(x \cdot x'\right) = \sigma^2\left(1 + \frac{x-x'^2}{2a\ell}\right)^{-a} \tag{14}$$

- Squared Exponential kernel function:

$$k_{SE}\left(x \cdot x'\right) = \sigma^2 exp\left(-\frac{x-x'^2}{\ell^2}\right) \tag{15}$$

where ℓ, $\alpha > 0$, σ, and σ^2 are the length scale, scale-mixture, amplitude, and variance. Also, the K_v and v represent the modified Bessel function and a positive parameter, respectively, while the symbol Γ indicates the gamma function. The exponential and squared exponential kernel functions are two special cases in the Matern function, where if $v = 0.5$ or 1 Matern function becomes exponential or squared exponential function.

Data collection. A total number of 506 experimental data of CO_2 adsorption by various structured MOFs, including pores decorated with open metal sites $Cu_3(BTC)_2$ and (MOF-505), hexagonally packed cylindrical channels (MOF-74), interpenetration (IRMOF-11), square channels (MOF-2), $Mg_2(dobdc)$, Cu-BTTri, the extra-high porosity MOF-177, Be-BTB, IRMOF-1, amino- and alkyl-functionalized pores (IRMOFs-3 and-6), and Co(BDP), were collected from reported studies (see Table S1)[17,40]. The pressure (P, bar), the temperature (T, K), the pore volume (V_p, cm³/g), and the surface area (S, m²/g) of the MOFs are the model input variables, while the CO_2 uptake (xCO_2 mmol/g) is the output of the model. In order to establish the most accurate model, arbitrarily, 20% of the total data was separated as the testing set, which was used to study the validity of the

model. The rest (80%) of the total data was utilized as the training set to investigate the MOF-CO_2 systems. Five statistical parameters (Eqs. 16–20), including R^2 (difference between the experiments and the calculated values), mean-square error (MSE), the standard deviation (STD), root-mean-square error (RMSE), and mean relative error (MRE) were used to evaluate the precision of the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left[x_i^{predicted} - x_i^{experimental} \right]^2}{\sum_{i=1}^{n} \left[x_i^{predicted} - x_m \right]^2} \qquad (16)$$

$$STD = \sqrt{\sum_{i=1}^{n} \frac{\left(x_i^{predicted} - x_m \right)^2}{n}} \qquad (17)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(x_i^{predicted} - x_i^{experimental} \right)^2 \qquad (18)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i^{predicted} - x_i^{experimental} \right)^2}{n}} \qquad (19)$$

$$MRE = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| x_i^{predicted} - x_i^{experimental} \right|}{x_i^{experimental}} \qquad (20)$$

Estimation of the precision of the collected data. Some data have inconsistent behavior in the data bank with the remainder of the data points identified as the suspected data. The suspected data mainly makes mention of the experimental errors. Recognizing the suspected data is crucial because its presence in the data bank can result in an inappropriate forecast for the established model. Thus, to seek the suspected or outlier data and advance the data bank quality, the Leverage method is used. In this method, Hat matrix (H) and critical leverage limit (H*) are used for identification of the outlier data, which are defined as follow[46].

$$H = U \left(U^T U \right)^{-1} U^T \qquad (21)$$

$$H^* = \frac{3j}{i+1} \qquad (22)$$

where U, i, and j are a matrix dimensional of i * j, the number of the model parameters, and the number of training points, respectively. To investigate the precision of the CO_2 adsorption data bank, the standardized residuals are represented against Hat values in Fig. 1, namely William's plot. The bounded zone between the critical leverage limit and standardized residuals of − 3 to 3 is known as the reliable region in William's plot. It is clear that all the extracted data points for the CO_2 uptake by different MOFs are reliable. Therefore, the dataset is excellent for testing and training models.

Results and discussion

Analysis of sensitivity. In order to propose a precise model, identification of the effects of the input on the CO_2 uptake by MOFs is vital. A sensitivity analysis is the needed technique to obtain the relevancy factor of each input parameters, which is calculated as follow[47,48]:

$$r = \frac{\sum_{i=1}^{n} \left(X_{k.i} - \overline{X}_k \right) \left(Y_i - \overline{Y} \right)}{\sqrt{\sum_{i=1}^{n} \left(X_{k.i} - \overline{X}_k \right)^2 \sum_{i=1}^{n} \left(Y_i - \overline{Y} \right)^2}} \qquad (23)$$

where $X_{k.i}$, $\overline{X}_k$, Y_i, and $\overline{Y}$ are the 'k' th input, input average, 'i'th output, and the average of outputs, respectively. The more value of r for an input parameter means that its efficiency on the CO_2 adsorption is higher and vice versa. The effect of the input variable on the CO_2 adsorption is shown in Fig. 2. The sensitivity analysis indicates that the pressure and the surface area of MOFs with r values of 0.68 and 0.52 are the most influential input variables on the CO_2 adsorption estimation. These inputs have a direct relationship with CO_2 uptake. Furthermore, increasing the pore volume of the MOFs results in higher CO_2 adsorption. It is worth mentioning that the small amount of r for the temperature can be related to its limited change in the experimental data.

Modeling results. In order to examine how exactly the proposed model is, the matching statistical parameters are used to specify a match between experimental and predicted CO_2 adsorption values. These parameters are determined and reported in Table 1. The R^2 values of 1.00, 0.998, 0.997, and 0.997 are obtained for GPR mod-
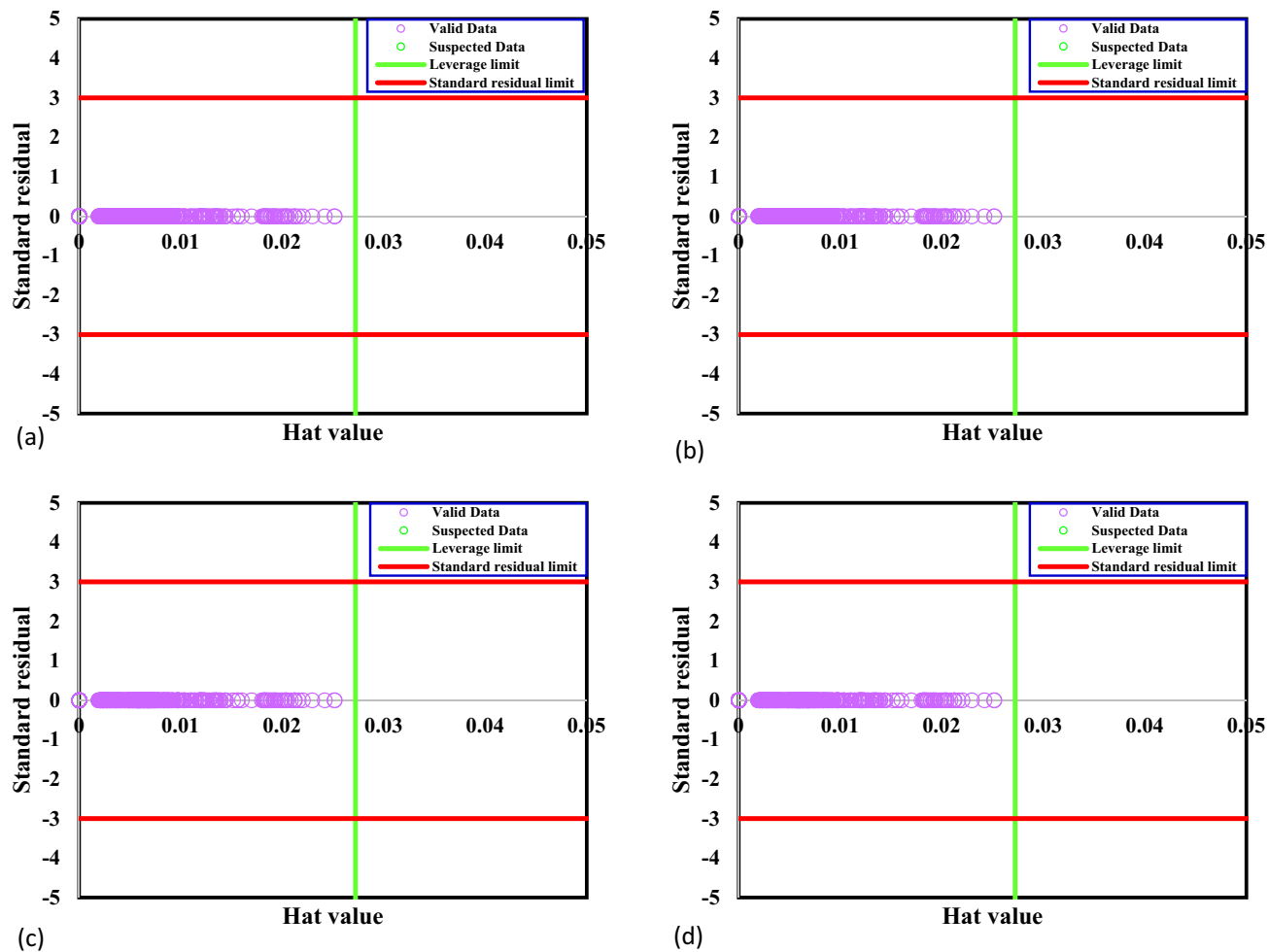
Figure 1. Detection of outliers for GPR model containing kernel function of (**a**) exponential, (**b**) matern, (**c**) squared exponential and (**d**) rational quadratic.
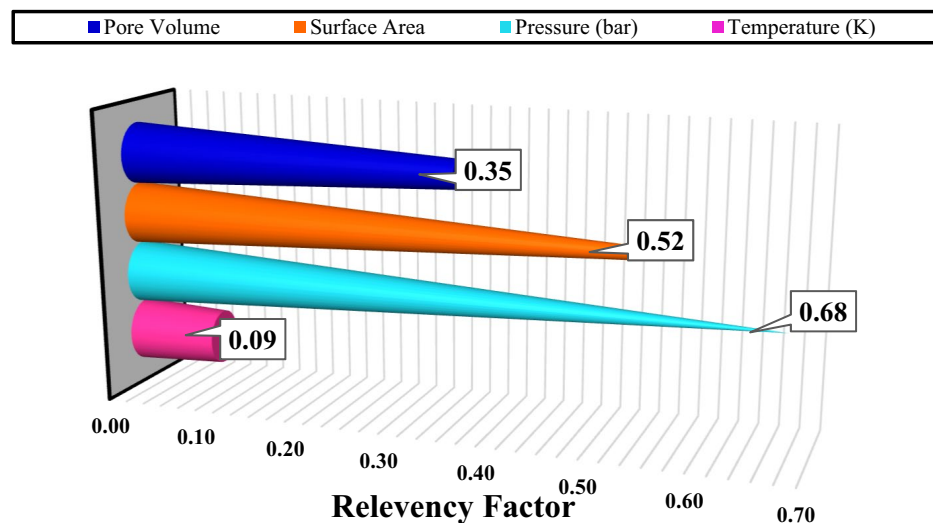
Figure 2. Analysis of sensitivity of the input parameters for CO_2 uptake by various MOFs.

	R^2	MRE (%)	MSE	RMSE	STD
GPR (Exponential)					
Train	1.000	0.51	0.00	0.02	0.02
Test	0.999	3.11	0.07	0.26	0.22
Total	1.000	1.75	0.02	0.26	0.14
GPR (Matern)					
Train	0.998	3.06	0.14	0.38	0.35
Test	0.990	62.78	0.65	0.81	0.75
Total	0.995	31.13	0.28	0.81	0.49
GPR (Squared exponential)					
Train	0.997	1.68	0.20	0.44	0.41
Test	0.992	41.81	0.56	0.75	0.69
Total	0.995	20.40	0.33	0.75	0.53
GPR (Rational quadratic)					
Train	0.997	10.27	0.20	0.45	0.40
Test	0.989	36.59	0.65	0.81	0.72
Total	0.994	22.83	0.36	0.81	0.53

Table 1. The statistical parameters of proposed GPR models.

els with Exponential, Matern, Squared exponential, and Rational quadratic kernel functions. The error parameters of MRE, MSE, RMSE, and STD in the training data indicate that the proposed GPR models have trained the data with acceptable precision. In addition to the prediction accuracy of the training data, the ability of the established models to forecast unseen CO_2 adsorption data points has critical importance. Thus, the proposed models were assessed for the testing data set. It can be seen that the GPR model containing the Exponential kernel function has the most accurate prediction of the unseen CO_2 uptake dataset, where R^2, MRE, MSE, RMSE, and STD are 0.999, 3.11%, 0.07, 0.26, and 0.22, respectively.

To further confirm the precision of the established models, the experimental and predicted CO_2 adsorption values are simultaneously shown in Fig. 3. It can be clearly observed that there is excellent agreement between the experimental CO_2 adsorptions and different GPR models. For all proposed models, the predicted CO_2 adsorption values follow the experimental CO_2 adsorption precisely. Thus, the proposed GPR models have outstanding capability in the prediction of CO_2 adsorption.

The predicted CO_2 adsorption values versus experimental data for all the models are plotted and described in Fig. 4. All the predicted CO_2 adsorption are situated to their experimental values so that the fitting lines on them have correlation coefficients higher than 0.98. The fitting lines cross considerably with 45° line representing the precision of all the GPR models for forecasting experimental CO_2 adsorption data. The bisector line (45° line) is a standard for the precision of established models. Nevertheless, the GPR model with Exponential kernel function yields the most precise results due to the correlation coefficient of 1.

Figure 5 shows the relative deviations between the experimental CO_2 adsorption and all GPR models' predicted values. As it is presented, the various kernel functions of Matern, Squared exponential, and Rational quadratic have absolute deviation points lower than 30%, while for Exponential kernel function, they are lower than 20%.

According to the results, the proposed GPR models showed excellent performance for CO_2 adsorption prediction. To ensure that the suggested models have enough precision in estimating CO_2 adsorption by different MOFs, the current study results are compared to the available correlations with the same aim reported by Dashti et al.[25]. The statistical parameters, including R^2, MSE, and STD, for the Dashti et al. study are listed in Table S2. Among the four examined algorithms, the RBF showed the best prediction with $R^2 = 0.997$, MSE = 0.204, and STD = 4.211. In comparison, all the established GPR models have better estimating of CO_2 adsorption, specifically, the GPR model with Exponential kernel function with $R^2 = 1.00$, MSE = 0.02, and STD = 0.14.

As shown in Fig. 6, MOF-177 has the highest CO_2 adsorption capacity of 33.5 mmol/g, which is much more significant than other MOFs. After that, IRMOFs-11, -1, and -3, with $Zn_4O(O_2C)_6$-type frameworks, show excellent capacities for CO_2 adsorption at room temperature. These MOFs have great effective pore sizes, which induce a sigmoidal shape(step) in their adsorption isotherms[24]. Also, the CO_2 adsorption isotherms of MOF-2, MOF-74, Norit RB2, MOF-505, and $Cu_3(BTC)_2$ are monotonic (Type I). The severe CO_2 adsorption at low pressure makes a "knee shape" in these isotherms, while the maximum capacity is gained at high pressure as the pores are saturated.

Figure 7 indicates the CO2 adsorption isotherms of Co(BDP), Cu-BTTri, BeBTB, Mg_2(dobdc), and MOF-177 at 313 K. The MOF-177 and BeBTB show much better performance than other MOFs in the CO_2 adsorption, which is due to their higher surface area (see Table S1). The isotherm of Co(BDP) has a step-like feature which might be attributed to its flexible structure, allowing gate-opening occurrence[49,50]. Cu-BTTri and Mg_2(dobdc) adsorbed high CO_2 at low pressures, which is related to their surface areas and the additional polarizing effect of metal cations on the framework surface. Due to higher polarizability and the quadrupole moment of CO_2, the surface area can affect the amount of CO_2 adsorption by MOF. Figure 8 shows the temperature effect on the CO_2 adsorption.
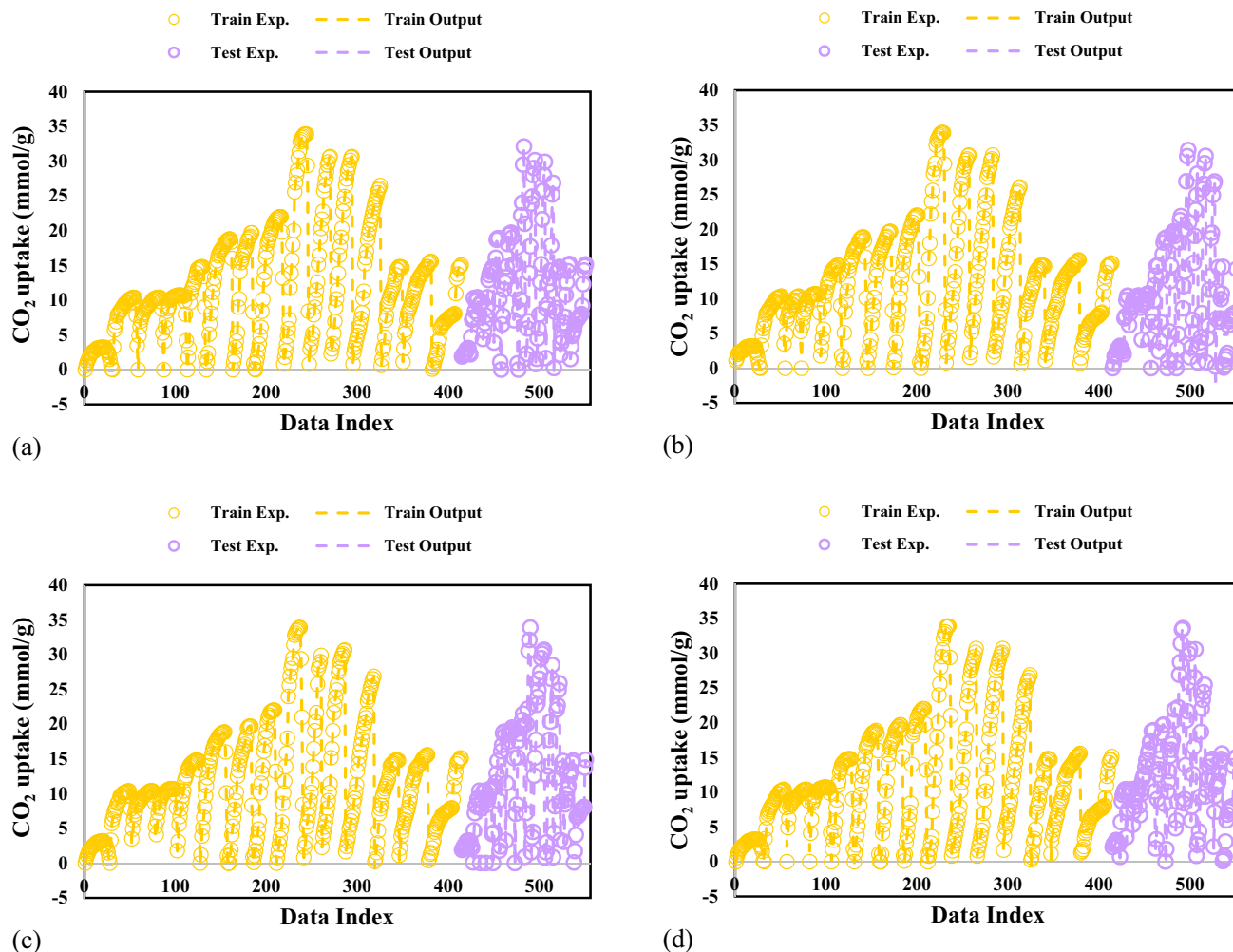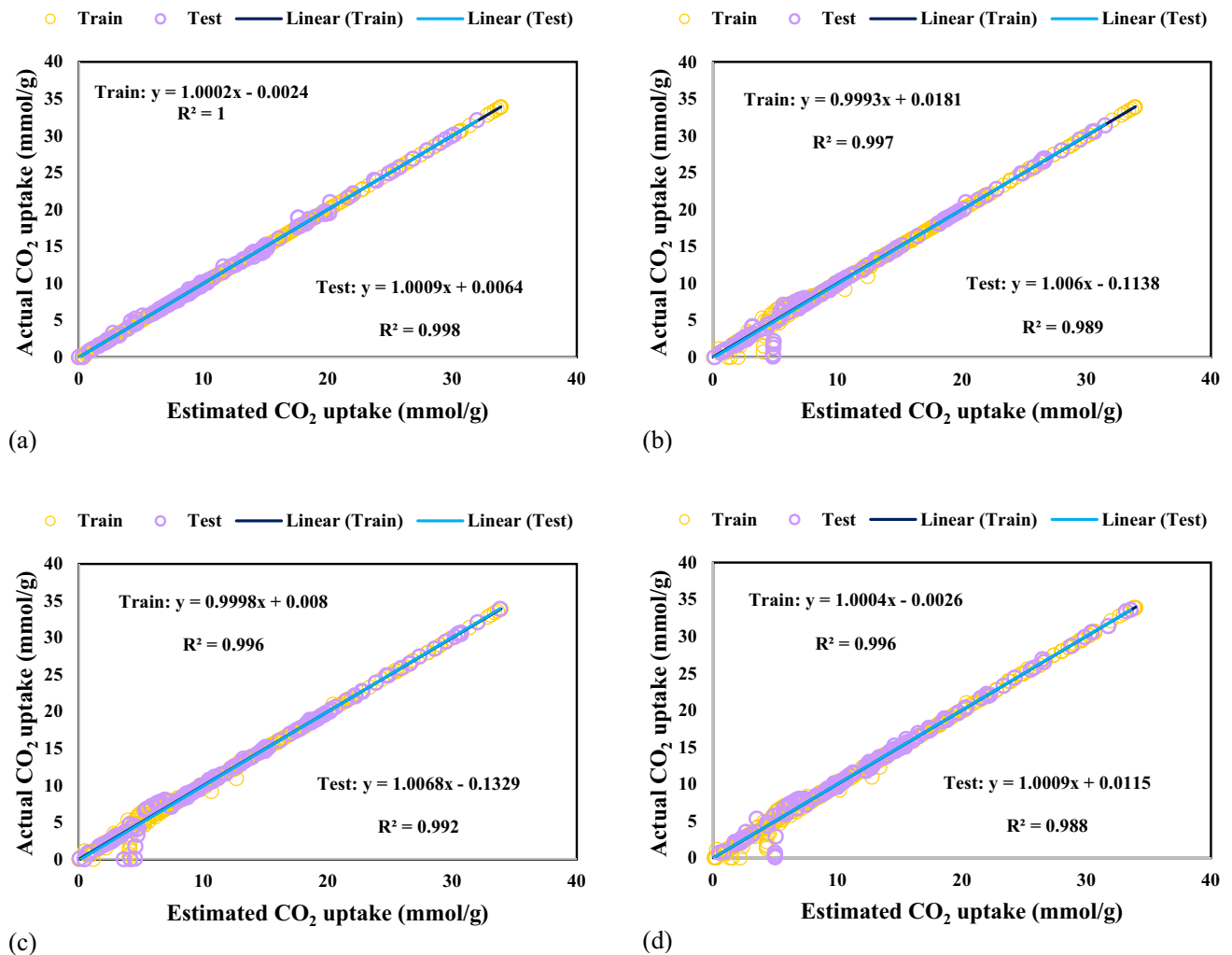
Figure 3. Comparison of experimental values and model outputs for GPR model containing kernel function of (**a**) Exponential, (**b**) Matern, (**c**) Squared exponential and (**d**) Rational quadratic.

Conclusion

In the current study, the GPR models based on different kernel functions have been established to estimate the CO_2 adsorption ability of MOFs in terms of pressure, temperature, pore volume, and surface area of MOFs. For this purpose, 506 experimental CO_2 uptake values in the literature have been collected and assessed. Four various kernel functions of Exponential, Squared exponential, Matern, and Rational quadratic have been studied. An excellent match has been detected between the experimental CO_2 adsorptions and predicted values by the developed GPR models, confirming these models' great ability in determining the CO_2 uptake. Among the proposed models, the GPR model based on exponential kernel function, was shown as the most precise predictive tool with $R^2 = 1.00$, $MSE = 0.02$, and $STD = 0.14$. Also, the suggested GPR models have better performance in comparison to the reported correlations. The sensitivity analysis indicates that the pressure is the most influential variable in CO_2 adsorption by MOFs. The surface area of the MOFs can be presented as the second determining parameter in the CO_2 capture by MOFs systems. The discussions in the current study can make it a helpful report for the engineers and researchers dealing with gas separation technologies.
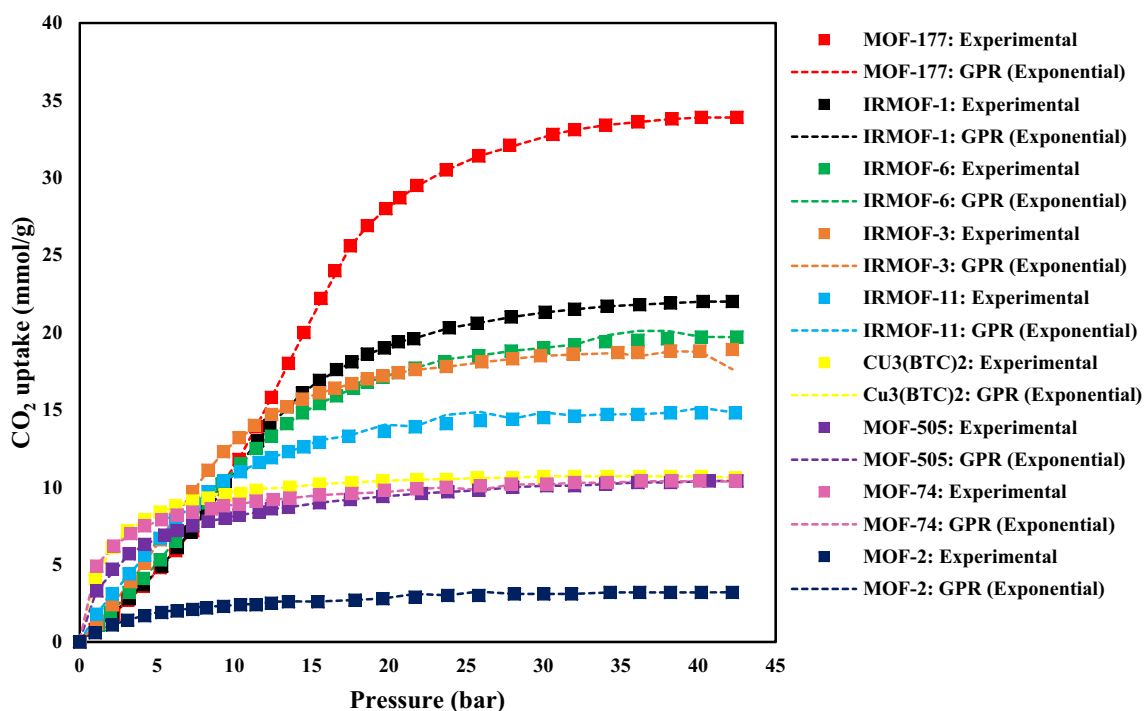
Figure 4. Cross plots for GPR model containing kernel function of (**a**) Exponential, (b) Matern, (**c**) Squared exponential and (**d**) Rational quadratic.
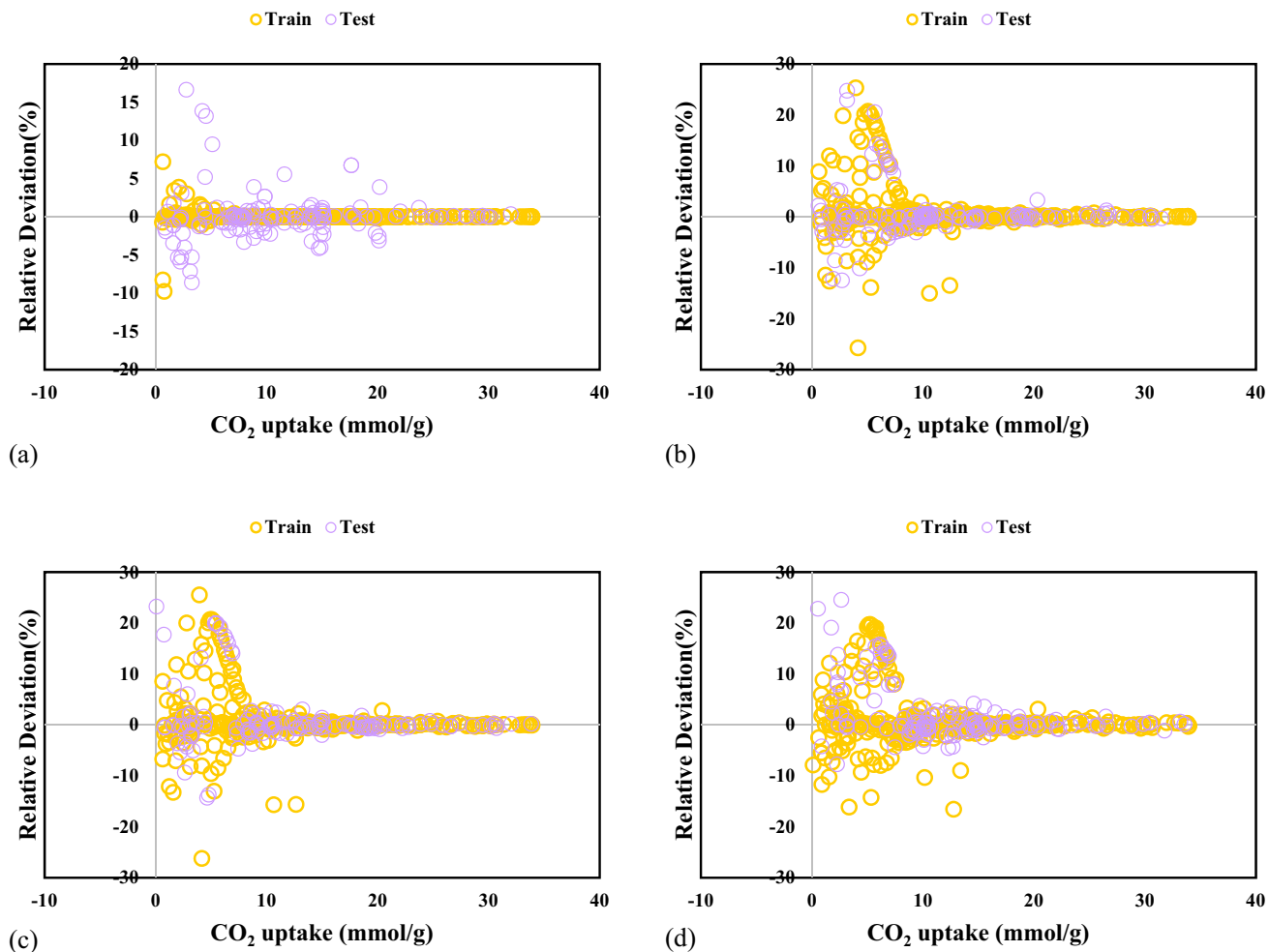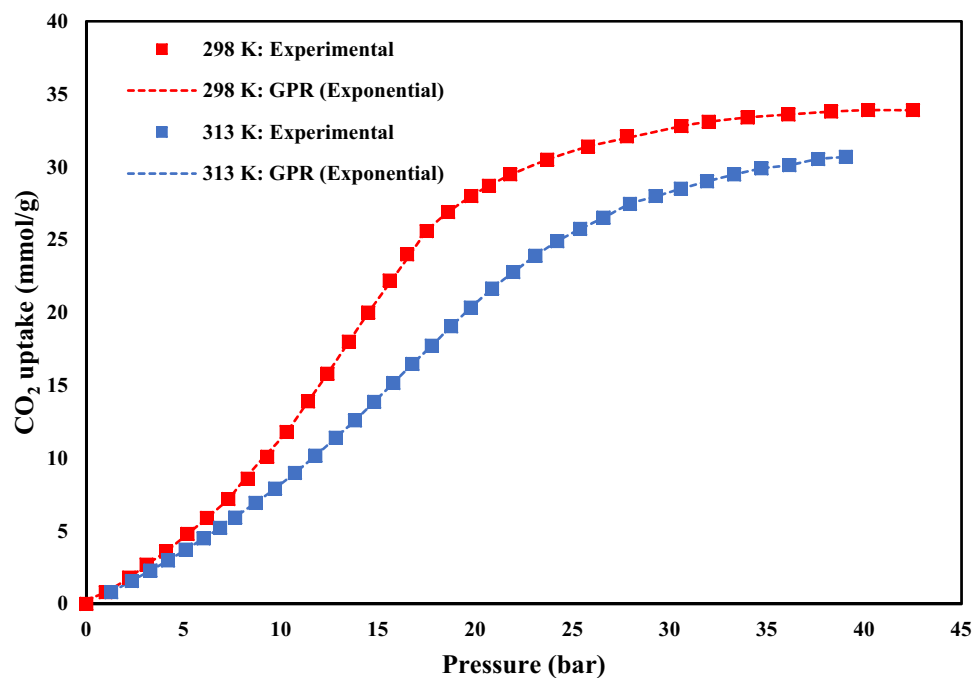
Figure 5. Comparison of experimental values and model outputs for GPR model containing kernel function of (**a**) Exponential, (**b**) Matern, (**c**) Squared exponential and (**d**) Rational quadratic.

Figure 6. The CO_2 adsorption capacities of different MOFs at 298 K.
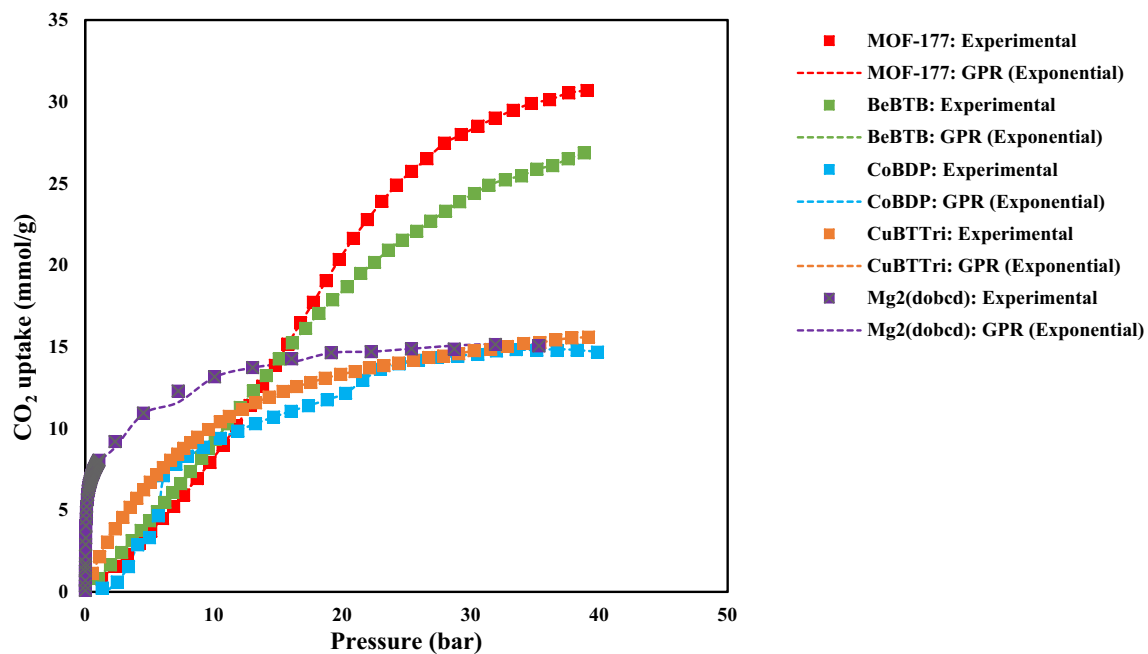
Figure 7. The CO_2 adsorption capacities of different MOFs at 313 K.

Figure 8. The CO_2 adsorption capacities of MOF-177 at 298 and 313 K.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Yang, Y., Xu, X., Guo, Y. & Wood, C. D. Enhancing the CO_2 capture efficiency of amines by microgel particles. *Int. J. Greenh. Gas Control* **103**, 103172 (2020).
2. Dashti, A. *et al.* Efficient hybrid modeling of CO_2 absorption in aqueous solution of piperazine: Applications to energy and environment. *Chem. Eng. Res. Des.* **144**, 405–417 (2019).
3. Zendehboudi, S., Bahadori, A., Lohi, A., Elkamel, A. & Chatzis, I. Practical and economic aspects of the Ex-Situ process: Implications for CO_2 sequestration. *Energy Fuels* **27**, 401–413 (2013).
4. Zendehboudi, S., Leonenko, Y., Shafiei, A., Soltani, M. & Chatzis, I. Modeling of CO_2 droplets shrinkage in ex situ dissolution approach with application to geological sequestration: Analytical solutions and feasibility study. *Chem. Eng. J.* **197**, 448–458 (2012).
5. Amirkhani, F., Harami, H. R. & Asghari, M. CO_2/CH_4 mixed gas separation using poly(ether-b-amide)-ZnO nanocomposite membranes: Experimental and molecular dynamics study. *Polym. Test.* **86**, 106464 (2020).
6. Lv, Y., *et al.* Fabrication and characterization of superhydrophobic polypropylene hollow fiber membranes for carbon dioxide absorption. *Appl. Energy* **90**, 167–174 (2012).
7. Damartzis, T., Papadopoulos, A. I. & Seferlis, P. Process flowsheet design optimization for various amine-based solvents in post-combustion CO_2 capture plants. *J. Clean. Prod.* **111**, 204–216 (2016).
8. Lv, B., Jing, G., Qian, Y. & Zhou, Z. An efficient absorbent of amine-based amino acid-functionalized ionic liquids for CO_2 capture: High capacity and regeneration ability. *Chem. Eng. J.* **289**, 212–218 (2016).
9. Wang, J. *et al.* Application of polyethylenimine-impregnated solid adsorbents for direct capture of low-concentration CO_2. *AIChE J.* **61**, 972–980 (2015).
10. Johnson, J. Putting a lid on carbon dioxide. *Chem. Eng. News* **82**, 36–42 (2004).
11. Sumida, K. *et al.* Carbon dioxide capture in metal-organic frameworks. *Chem. Rev.* **112**, 724–781 (2012).
12. Sayari, A., Belmabkhout, Y. & Serna-Guerrero, R. Flue gas treatment via CO_2 adsorption. *Chem. Eng. J.* **171**, 760–774 (2011).
13. Yaghi, O. M. *et al.* Reticular synthesis and the design of new materials. *Nature* **423**, 705–714 (2003).
14. Rowsell, J. L. C. & Yaghi, O. M. Metal-organic frameworks: A new class of porous materials. *Microp. Mesop. Mater.* **73**, 3–14 (2004).
15. Eddaoudi, M. *et al.* Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science (80-.).* **295**, 469–472 (2002).
16. Karra, J. R. & Walton, K. S. Effect of open metal sites on adsorption of polar and nonpolar molecules in metal-organic framework Cu-BTC. *Langmuir* **24**, 8620–8626 (2008).
17. Millward, A. R. & Yaghi, O. M. Metal-organic frameworks with exceptionally high capacity for storage of carbon dioxide at room temperature. *J. Am. Chem. Soc.* **127**, 17998–17999 (2005).
18. Kuppler, R. J. *et al.* Potential applications of metal-organic frameworks. *Coord. Chem. Rev.* **253**, 3042–3066 (2009).
19. Ghanbari, T., Abnisa, F. & Wan Daud, W. M. A. A review on production of metal organic frameworks (MOF) for CO_2 adsorption. *Sci. Total Environ.* **707**, 135090 (2020).
20. Dietzel, P. D. C., Morita, Y., Blom, R. & Fjellvåg, H. An in situ high-temperature single-crystal investigation of a dehydrated metal-organic framework compound and field-induced magnetization of one-dimensional metal-oxygen chains. *Angew. Chemie Int. Ed.* **44**, 6354–6358 (2005).
21. Collins, D. J. & Zhou, H. C. Hydrogen storage in metal-organic frameworks. *J. Mater. Chem.* **17**, 3154–3160 (2007).
22. Bahamon, D., Díaz-Márquez, A., Gamallo, P. & Vega, L. F. Energetic evaluation of swing adsorption processes for CO_2 capture in selected MOFs and zeolites: Effect of impurities. *Chem. Eng. J.* **342**, 458–473 (2018).
23. Xiang, S. *et al.* Microporous metal-organic framework with potential for carbon dioxide capture at ambient conditions. *Nat. Commun.* **3**, 1–9 (2012).
24. Bourrelly, S. *et al.* Different adsorption behaviors of methane and carbon dioxide in the isotypic nanoporous metal terephthalates MIL-53 and MIL-47. *J. Am. Chem. Soc.* **127**, 13519–13521 (2005).
25. Dashti, A., Bahrololoomi, A., Amirkhani, F. & Mohammadi, A. H. Estimation of CO_2 adsorption in high capacity metal-organic frameworks: Applications to greenhouse gas control. *J. CO2 Util.* **41**, 101256 (2020).
26. Langmuir, I. The adsorption of gases on plane surfaces of glass, mica and platinum. *J. Am. Chem. Soc.* **40**, 1361–1403 (1918).
27. Rostami, A., Ebadi, H., Arabloo, M., Meybodi, M. K. & Bahadori, A. Toward genetic programming (GP) approach for estimation of hydrocarbon/water interfacial tension. *J. Mol. Liq.* **230**, 175–189 (2017).
28. Britt, D., Furukawa, H., Wang, B., Glover, T. G. & Yaghi, O. M. Highly efficient separation of carbon dioxide by a metal-organic framework replete with open metal sites. *Proc. Natl. Acad. Sci. USA* **106**, 20637–20640 (2009).
29. Vafaei, A. *et al.* Removal of hydrochlorothiazide from molecular liquids using carbon nanotubes: Radial basis function neural network modeling and culture algorithm optimization. *J. Mol. Liq.* **324**, 114766 (2021).
30. Gheytanzadeh, M., Baghban, A., Habibzadeh, S., Mohaddespour, A. & Abida, O. Insights into the estimation of capacitance for carbon-based supercapacitors. *RSC Adv.* **11**, 5479–5486 (2021).
31. Bemani, A. *et al.* Modeling of cetane number of biodiesel from fatty acid methyl ester (FAME) information using GA-, PSO-, and HGAPSO–LSSVM models. *Renew. Energy* **150**, 924–934 (2020).
32. Baghban, A. & Khoshkharam, A. Application of LSSVM strategy to estimate asphaltene precipitation during different production processes. *Pet. Sci. Technol.* **34**, 1855–1860 (2016).
33. Baghban, A., Kashiwao, T., Bahadori, M., Ahmad, Z. & Bahadori, A. Estimation of natural gases water content using adaptive neuro-fuzzy inference system. *Pet. Sci. Technol.* **34**, 891–897 (2016).
34. Bahadori, A. *et al.* Computational intelligent strategies to predict energy conservation benefits in excess air controlled gas-fired systems. *Appl. Therm. Eng.* **102**, 432–446 (2016).
35. Baghban, A., Abbasi, P. & Rostami, P. Modeling of viscosity for mixtures of Athabasca bitumen and heavy n-alkane with LSSVM algorithm. *Pet. Sci. Technol.* **34**, 1698–1704 (2016).
36. Baghban, A. Application of the ANFIS strategy to estimate vaporization enthalpies of petroleum fractions and pure hydrocarbons. *Pet. Sci. Technol.* **34**, 1359–1366 (2016).
37. Rostami, A., Baghban, A., Mohammadi, A. H., Hemmati-Sarapardeh, A. & Habibzadeh, S. Rigorous prognostication of permeability of heterogeneous carbonate oil reservoirs: Smart modeling and correlation development. *Fuel* **236**, 110–123 (2019).
38. Liu, Z. & Baghban, A. Application of LSSVM for biodiesel production using supercritical ethanol solvent. *Energy Sources, Part A Recover. Util. Environ. Eff.* **39**, 1869–1874 (2017).
39. Nabipour, N. *et al.* Estimating biofuel density via a soft computing approach based on intermolecular interactions. *Renew. Energy* **152**, 1086–1098 (2020).
40. Herm, Z. R., Swisher, J. A., Smit, B., Krishna, R. & Long, J. R. Metal-organic frameworks as adsorbents for hydrogen purification and precombustion carbon dioxide capture. *J. Am. Chem. Soc.* **133**, 5664–5667 (2011).
41. Fu, Q. *et al.* Prediction of the diet nutrients digestibility of dairy cows using Gaussian process regression. *Inf. Process. Agric.* **6**, 396–406 (2019).
42. Sniekers, S. & van der Vaart, A. Adaptive bayesian credible sets in regression with a gaussian process prior. *Electron. J. Stat.* **9**, 2475–2527 (2015).
43. Leco, M. & Kadirkamanathan, V. A perturbation signal based data-driven Gaussian process regression model for in-process part quality prediction in robotic countersinking operations. *Robot. Comput. Integr. Manuf.* **71**, 102105 (2021).

44. Zhang, Y., Su, G. & Yan, L. Gaussian process machine learning model for forecasting of karstic collapse. in *Communications in Computer and Information Science* **224**, 365–372 (2011).
45. CE, W. C. K. I. R. Adaptive Computation and Machine Learning series. In *Gaussian Processes for Machine Learning* 1–272 (2005).
46. Razavi, R., Bemani, A., Baghban, A., Mohammadi, A. H. & Habibzadeh, S. An insight into the estimation of fatty acid methyl ester based biodiesel properties using a LSSVM model. *Fuel* **243**, 133–141 (2019).
47. Bemani, A., Baghban, A. & Mohammadi, A. H. An insight into the modeling of sulfur content of sour gases in supercritical region. *J. Pet. Sci. Eng.* **184**, 106459 (2020).
48. Bemani, A., Baghban, A., Mohammadi, A. H. & Andersen, P. Ø. Estimation of adsorption capacity of CO2, CH4, and their binary mixtures in Quidam shale using LSSVM: Application in CO2 enhanced shale gas recovery and CO2 storage. *J. Nat. Gas Sci. Eng.* **76**, 103204 (2020).
49. Salles, F. *et al.* Multistep N2 breathing in the metal-organic framework Co(1,4-benzenedipyrazolate). *J. Am. Chem. Soc.* **132**, 13782–13788 (2010).
50. Li, J. R., Kuppler, R. J. & Zhou, H. C. Selective gas adsorption and separation in metal-organic frameworks. *Chem. Soc. Rev.* **38**, 1477–1504 (2009).

Author contributions

All the authors collaborated in data curation, methodology, software, formal analysis, and resources, writing—review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.B. or S.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.