DOI: 10.19884/j.1672-5220.202411016

QSAR Model of Activated Carbon Adsorption Based on Langmuir Adsorption Isotherm

TAN Ting[1, 2], WEI Qunshan[1, 2*], LIU Qiong[1, 2*], SHEN Zhemin[3], SONG Xinshan[1, 2], WANG Yuhui[1, 2], CHARLES Nzila[4], CHRISTOPHER W. K. Chow[5]

1. College of Environmental Science and Engineering, Donghua University, Shanghai 201620, China
2. State Environmental Protection Engineering Center for Pollution Treatment and Control in Textile Industry, Donghua University, Shanghai 201620, China
3. School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
4. School of Engineering, Moi University, Eldoret 3900-30100, Kenya
5. Sustainable Infrastructure and Resource Management (SIRM), UniSA STEM, University of South Australia, Mawson Lakes, 5095, Australia

Abstract: From a quantum chemistry standpoint, the impact of the structural properties of the compounds on activated carbon's adsorption ability was specifically investigated. The compounds whose adsorption behavior followed the Langmuir isotherm model were selected as the research objects. An optimal quantitative structure-activity relationship (QSAR) model was built by using the multiple linear regression (MLR) method, with the saturation adsorption capacity Q_m from the Langmuir adsorption isotherm as the response variable and the structural parameters of 50 organic compounds as independent variables. The results show that the optimal model exhibits good stability, reliability and robustness, with a regression coefficient R^2 of 0.88, an adjusted regression coefficient R^2_{adj} of 0.87, an internal validation coefficient q^2 of 0.81, and an external validation coefficient Q^2_{ext} of 0.68. The variables included in the optimal model indicate that the polarity of the molecule, the molecular potential energy, and the stability and bonding strength of the organic compound are the main factors affecting the adsorption on activated carbon. The results provide key information for predicting the adsorption capacity of organic compounds on activated carbon and offer a theoretical reference for adsorption treatment in water environments.

Keywords: quantitative structure-activity relationship (QSAR); adsorption; activated carbon; multiple linear regression (MLR)

CLC number: TQ424; X52; R99 **Document code**: A

Article ID: 1672-5220(2025)06-0628-11

Open Science Identity (OSID)

0 Introduction

Currently, the problem of environmental pollution is becoming more and more serious, and organic pollutants such as dyes[1], pesticides[2], solvents and drugs[3] have attracted much attention because of their difficulty in degradation and potential threat to the ecosystem and human health. Activated carbon, as an efficient adsorbent material, plays an important role in removing organic pollutants from water[4]. This study aims to investigate the effect of the nature of compounds on the adsorption performance of activated carbon through the development of a quantitative structure-activity relationship (QSAR) model. In adsorption studies, the Langmuir adsorption isotherm is commonly used to quantify the relationship between the adsorbate concentration and adsorption capacity[5]. Selecting compounds that follow the Langmuir adsorption isotherm for QSAR modeling facilitates a deeper understanding of how molecular properties influence adsorption performance. This approach also enhances the model's predictive ability.

QSAR models are mathematical models that predict a compound's biological activity by establishing a relationship with its structural parameters[6-7]. Their advantages lie in reducing the experimental cost and time. There have been many studies on QSAR models for adsorption. Blum et al.[8] established a QSAR model based primarily on molecular connectivity indices for predicting the adsorption behavior of aromatic and aliphatic compounds on activated carbon. Zhao et al.[9] developed a QSAR model based on the adsorption affinity and capacity derived from the treatment of cationic drugs with activated carbon. The modeling results showed that the molecular mass, polar surface area (PSA), and octanol-water partition coefficient were the main factors affecting the adsorption affinity and capacity of cationic drugs. De Ridder et al.[10] developed a model based on the parameter of the molecular connectivity index

---

Received date: 2024-11-21

Foundation items: National Natural Science Foundation of China (No. 21876025); National Key R&D Program of China (No. 2023YFC3207204); Shanghai Municipal Education Commission Artificial Intelligence-Enabled Scientific Research Plan, China (No. SMEC-AI-DHUZ-07)

* Correspondence should be addressed to WEI Qunshan, email: qswei@dhu.edu.cn; LIU Qiong, email: liuqiong@dhu.edu.cn

Citation: TAN T, WEI Q S, LIU Q, et al. QSAR model of activated carbon adsorption based on Langmuir adsorption isotherm[J]. *Journal of Donghua University (English Edition)*, 2025, 42(6): 628-638.

(MCI), which was mainly used to predict the equilibrium adsorption of organic micropollutants with different properties on activated carbon. However, this model is limited to the adsorption behavior of a specific type of activated carbon (F400) under ultrapure water conditions. Gong et al.[11] developed a QSAR model for predicting the reduction rate of mono-nitroaromatic compounds catalyzed by activated carbon in an anaerobic system by using a partial least-squares regression (PLSR) method. By summarizing the adsorption models that have been studied, we find that most of them focus on a single type of compounds. In addition, the results of the current study have shown that the MCI and octanol-water partition coefficient are important factors affecting the adsorption performance on activated carbon. However, the structural parameters of the compounds are also important. Therefore, this study focuses on the structural parameters of the organic compounds and establishes QSAR models for various types of substances, such as dyes, pesticides, solvents, drugs, and odorants. This approach aims to investigate how molecular structure affects adsorption performance on activated carbon and to fill the gaps in existing QSAR studies.

The organic compounds selected for this study all conformed to the Langmuir adsorption isotherm, and the optimal QSAR model for the adsorption of compounds on activated carbon was established by applying multiple linear regression (MLR)[12], with the saturation adsorption capacity Q_{m} in the Langmuir adsorption isotherm as the response value. The model is primarily employed to explore the influence of the nature of compounds on the adsorption effect in depth and to provide valuable information for subsequent research and application.

1 Materials and Methods

1.1 Experimental dataset

The Q_{m} values of 50 organic compounds selected for this study, including dyes, pesticides, solvents, drugs, and odorants, were obtained from Refs. [13-24]. The details are shown in Table 1.

The maximum observed Q_{m} values are 8.55 (drugs), 28.57 (solvents), 4.18 (odorants), 6.85 (dyes), and 29.00 mg/g (pesticides), while the corresponding minimum values are 1.10, 0.12, 0.21, 1.61, and 3.02 mg/g, respectively. The dataset is further analyzed after logarithmic transformation (Fig. 1), and it is found that the overall pattern is close to a normal distribution.

Table 1 Adsorption results for 50 organic compounds matching Langmuir adsorption isotherm

Name	Chemical formula	Q_{m}/(mg/g)	Name	Chemical formula	Q_{m}/(mg/g)
Amitriptyline	$C_{20}H_{23}N$	1.81	Pefloxacin	$C_{17}H_{20}FN_3O_3$	2.16
Verapamil	$C_{27}H_{38}O_4N_2$	1.39	Enrofloxain	$C_{19}H_{22}FN_3O_3$	2.75
Chlorpromazine	$C_{17}H_{19}N_2SCl$	2.37	Ciprofloxacin	$C_{17}H_{18}FN_3O_3$	2.89
Clomipramine	$C_{19}H_{23}ClN_2$	2.07	Ofloxacin	$C_{18}H_{20}FN_3O_4$	2.30
Desipramine	$C_{18}H_{22}N_2$	1.32	Chloroform	$CHCl_3$	0.24
Diphenhydramine	$C_{17}H_{21}ON$	1.31	Atrazine	$C_8H_{14}ClN_5$	20.52
Imipraminum	$C_{19}H_{24}N_2$	1.62	Dicamba	$C_8H_6Cl_2O_3$	17.39
Procaine	$C_{13}H_{20}N_2O_2$	1.34	Aldicarb	$C_7H_{14}N_2O_2S$	4.22
Promethazine	$C_{17}H_{20}N_2S$	2.40	Dinoseb	$C_{10}H_{12}N_2O_5$	3.02
Propafenone	$C_{21}H_{27}NO_3$	1.80	m-Xylene	C_8H_{10}	1.85
Propranolol	$C_{16}H_{21}NO_2$	1.47	2,4-Dinitrotoluene	$C_7H_6N_2O_4$	2.86
Thioridazine	$C_{21}H_{26}N_2S_2$	3.85	Carbofuran	$C_{12}H_{15}NO_3$	4.13
Sulfathiazole	$C_9H_9N_3O_2S_2$	8.55	(*E*)-2-Octenal	$C_8H_{14}O$	4.18
Sulfapyridine	$C_{11}H_{11}N_3O_2S$	3.41	β-Cyclocitral	$C_{10}H_{16}O$	5.62
Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	1.10	Acid blue 113	$C_{32}H_{23}N_5O_6S_2$	6.85
Sulfamethazine	$C_{12}H_{14}N_4O_2S$	8.14	Acid blue 40	$C_{22}H_{17}N_3O_6SNa$	1.61
Norfloxacin	$C_{16}H_{18}FN_3O_3$	2.37	Acid blue 62	$C_{20}H_{20}O_5N_2S$	1.94
2,4-Dichlorophenol (2,4-DCP)	$C_6H_4Cl_2O$	28.57	2-Methylisoboreneol (2-MIB)	$C_{11}H_{20}O$	0.91
3-Methylindole (3-MLD)	C_9H_9N	1.08	Pentachlorophenol (PCP)	C_6Cl_5OH	29.00
Dimethyl trisulfide (DMTS)	$C_2H_6S_3$	1.10	Dimethyl disulfide (DMDS)	$C_4H_{10}S_2$	0.12
2-Isopropyl-3-methoxy pyrazine (IPMP)	$C_8H_{12}N_2O$	1.47	2,4,6-Trichloroanisole (2,4,6-TCA)	$C_7H_5Cl_3O$	0.28
trans,trans-2,4-Heptadienal (T,T-2,4-HPDA)	$C_7H_{10}O$	3.89	trans,trans-2,4-Decadienal (T,T-2,4-DDA)	$C_{10}H_{16}O$	1.97
trans,trans-2,4-Octadienal (T,T-2,4-ODA)	$C_8H_{12}O$	1.19	Bis (2-chloroisopropyl) ether (Bis)	$C_6H_{12}Cl_2O$	5.09
2,4,5-Trichlorophenoxyacetic acid (2,4,5-T)	$C_8H_5Cl_3O_3$	12.82	Geosmin	$C_{12}H_{22}O$	0.21

Note: underlined compounds indicate those in the test set.

To ensure the quality and applicability of the model, the data are selected according to uniform criteria: the experimental temperature is controlled in a range of 20–25 ℃, and the final data that reach adsorption equilibrium are selected.

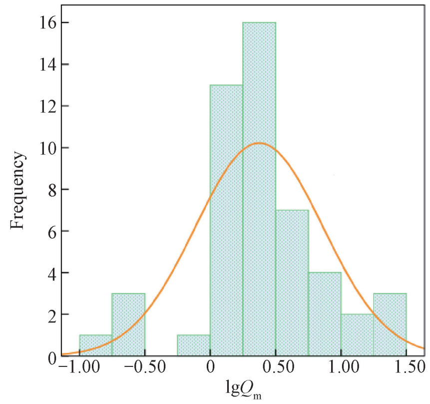

Fig. 1 Histogram of Q_m values for the organic compounds

1.2 Calculation method of structural parameters

In this study, the molecular structures of the organic compounds were constructed using ChemDraw and Chem3D. Their structural parameters were then calculated based on density functional theory (DFT)[25], primarily utilizing Gaussian 09 and Materials Studio 2017 software. Specifically, in Gaussian 09, geometry optimization and frequency calculations were performed at the B3LYP/6-311G (d, p) level to obtain the structural parameters[26]. For calculations in Materials Studio 2017, the DMol3 module with the GGA/B3LYP method was employed for geometric optimization and subsequent energy calculations[27]. The computational methods used above are applicable to the selected 50 organic compounds. The calculated structural parameters and their connotations are shown in Table 2.

Table 2 Parameters of organic compounds and their connotations

Parameter	Connotation	Unit
μ	Dipole moment in vacuum	Debye
$q(\mathrm{CH}^+)_{\max}$	Maximum value of the positive partial charge of a hydrogen atom connected to a carbon atom	e
$q(\mathrm{CH}^+)_{\min}$	Minimum value of the positive partial charge of a hydrogen atom connected to a carbon atom	e
$q(\mathrm{C}^-)_{\max}$	Maximum value of the partial charge on a carbon atom	e
$q(\mathrm{C}^-)_{\min}$	Minimum value of the partial charge on a carbon atom	e
E	Total molecular energy under B3LYP calculations	kcal/mol
$f(+)_{\max}$	Maximum value of the nucleophilic Fukui index in carbon atoms	e
$f(+)_{\min}$	Minimum value of the nucleophilic Fukui index in carbon atoms	e
$f(-)_{\max}$	Maximum value of the electrophilic Fukui index in carbon atoms	e
$f(-)_{\min}$	Minimum value of the electrophilic Fukui index in carbon atoms	e
$f(0)_{\max}$	Maximum value of the free radical affinity Fukui index in carbon atoms	e
$f(0)_{\min}$	Minimum value of the free radical affinity Fukui index in carbon atoms	e
E_{HOMO}	Highest energy of an occupied molecular orbital	eV
E_{LUMO}	Lowest energy of an unoccupied molecular orbital	eV
E_{GAP}	Difference between E_{LUMO} and E_{HOMO}	eV
$B(\mathrm{C{-}C})_{\max}$	Maximum value of C—C bond level	—
$B(\mathrm{C{-}C})_{\min}$	Minimum value of C—C bond level	—
$B(\mathrm{C{-}H})_{\max}$	Maximum value of C—H bond level	—
$B(\mathrm{C{-}H})_{\min}$	Minimum value of C—H bond level	—
$W(\mathrm{C{-}C})_{\max}$	Maximum value of the Wiberg index for the strongest bond between any two carbon atoms in a molecule	—
$W(\mathrm{C{-}C})_{\min}$	Minimum value of the Wiberg index for the weakest bond between any two carbon atoms in a molecule	—
$W(\mathrm{C{-}H})_{\max}$	Maximum value of the Wiberg index for the strongest bond between any carbon atom and a hydrogen atom in a molecule	—
$W(\mathrm{C{-}H})_{\min}$	Minimum value of the Wiberg index for the weakest bond between any carbon atom and a hydrogen atom in a molecule	—
$V(\mathrm{C})_{\max}$	Maximum value of the valence of carbon atoms in the carbon skeleton of a hydrocarbon	—

(Table 2 continued)

Parameter	Connotation	Unit
$V(\mathrm{C})_{\max}$	Minimum value of the valence of carbon atoms in the carbon skeleton of a hydrocarbon	—
$V(\mathrm{H})_{\max}$	Maximum value of the valence of hydrogen atoms in a hydrocarbon	—
$V(\mathrm{H})_{\min}$	Minimum value of the valence of hydrogen atoms in a hydrocarbon	—
$E_p(\mathrm{C})$	Potential energy produced by a single carbon atom in a molecule	Ha
$E_p(\mathrm{H})$	Potential energy produced by a single hydrogen atom in a molecule	Ha
$E_p(\mathrm{C—H})$	Potential energy produced by the interaction forces between all the carbon and hydrogen atoms in a molecule	Ha
E_p	Sum of the potential energy produced by the interaction forces between all the molecules	Ha
M	Molecular mass	g/mol

Notes: 1 kcal = 4.184 kJ; 1 Ha ≈ 4.36×10^{-18} J.

1.3 Model validation methods

The calculated parameters of organic compounds were modeled with the experimental results by using the SPSS software. The following evaluation metrics were used to assess the quality of their constructed QSAR model[28-29], including the regression coefficient R^2, standard deviation (SD), t-statistic and Fisher's test, root mean square error (RMSE), internal validation coefficient q^2, and external validation coefficient Q^2_{ext}. The constructed models were finally validated by Y-randomization.

1.4 Application domains

The Williams plot was employed in this study to evaluate the range of applicability of the model. The predictive accuracy of the optimal QSAR model was evaluated based on the criterion that an ideal range of standardized residuals (−3, 3) and the leverage value h_i was less than the warning threshold h^*.

2 Results and Analyses

2.1 Experimental results of adsorption isotherms of organic compounds

The adsorption process of the above 50 organic compounds on activated carbon was described by using the Langmuir adsorption isotherm,

$$Q_e = (Q_m b C_e)/(1 + bC_e), \tag{1}$$

where Q_e is the equilibrium adsorption amount, mg/g; C_e is the equilibrium concentration, mg/L; b is the equilibrium constant of adsorption.

Observing Q_m values in Table 1, it can be found that the range of Q_m values is relatively wide. The adsorption properties of different kinds of organic compounds vary, which indicates that the types and structures of compounds determine their different adsorption effects. The organic compound studied with the largest Q_m value is PCP (29.00 mg/g), and that with the smallest Q_m value is DMDS (0.12 mg/g). The difference in their Q_m values is 28.88 mg/g, meaning that the Q_m value of PCP is about 240 times that of DMDS. It indicates that the adsorption properties are significantly different between the compounds studied. These differences can better elucidate the relationship between the properties of the compounds and the amount of adsorption, which is more conducive to the application of the constructed QSAR model to adsorption in different fields.

2.2 Calculation results of compound structures

Quantum chemical parameters of compounds are key to reflecting the structural properties of compounds in QSAR modeling. We analyzed the structural parameters of these 50 organic compounds using computational software and obtained the following specific results.

In this study, the compound with the largest dipole moment is pefloxacin, with a value that differs from the minimum by 14.47 Debye. This suggests that there is a great difference in polarity between the organic compounds[30].

The parameter E represents the total energy of the optimized molecular structure. In general, a larger absolute value of E indicates a larger molecule size and a more complex structure of an organic compound. The absolute value of E for the compound with the maximum value (acid blue 113, 2 754.78 kcal/mol) is about 8 times larger than that of the compound with the minimum value (m-xylene, 310.10 kcal/mol), highlighting the significant structural diversity among the organic compounds studied. The parameters $q(\mathrm{CH}^+)_{\max}$, $q(\mathrm{CH}^+)_{\min}$, $q(\mathrm{C}^-)_{\max}$, and $q(\mathrm{C}^-)_{\min}$ are mainly used to characterize the atomic charges in the compounds, which in turn reflect the polar magnitude of the molecules. In the study, the maximum values of $q(\mathrm{CH}^+)_{\max}$ and $q(\mathrm{CH}^+)_{\min}$ are 0.21 e (acid blue 113) and 0.18 e (PCP), respectively, and the minimum values are 0.04 e (m-xylene) and 0.01 e (clomipramine), respectively.

The parameter E_p is a physical quantity that describes the internal stability and interactions of a molecule. The maximum and minimum of E_p and $E_p(\mathrm{C—H})$ correspond to the same compound, which indicates that the sum of the energies of the molecules of the compound is basically determined by the forces between the carbon and hydrogen atoms in the molecule. $E_p(\mathrm{C})$ ranges from −38.88 Ha (m-xylene) to −637.27 Ha (DMTS), with a difference of 598.39 Ha.

The Wiberg index, which reflects the bond strength, is a key indicator of a molecule's stability and reactivity[31-32]. These properties, in turn, determine how the compound interacts with surfaces and thereby influence its adsorption behavior. In this study, the maximum values of $W(\text{C—C})_{\max}$ and $W(\text{C—C})_{\min}$ are 4.04 (geosmin) and 3.92 (acid blue 113), respectively, and the minimum values of $W(\text{C—C})_{\max}$ and $W(\text{C—C})_{\min}$ are 3.85 (DMDS) and 1.07 (Bis), respectively. Based on these results, it is not difficult to find out the differences in the adsorption effect between the compounds.

The Fukui index f consists of electrophilic $f(-)$, nucleophilic $f(+)$, and free radical $f(0)$ parameters[33-34]. Among them, the average values of $f(-)_{\max}$, $f(-)_{\min}$, $f(+)_{\max}$, $f(+)_{\min}$, $f(0)_{\max}$, and $f(0)_{\min}$ are 0.17, 0.01, 0.25, 0, 0.18, and 0.01 e, respectively. Bond order is mainly used to reflect the stability of chemical bonds in compound molecules[35]. The average values of $B(\text{C—C})_{\max}$, $B(\text{C—C})_{\min}$, $B(\text{C—H})_{\max}$, and $B(\text{C—H})_{\min}$ are −0.50, −1.17, −0.51, and −0.63, respectively. The valence connectivity topological parameter is derived from the topological information of atoms and bonds in a molecule to predict its physicochemical properties[36]. The average values of $V(\text{C})_{\max}$, $V(\text{C})_{\min}$, $V(\text{H})_{\max}$, and $V(\text{H})_{\min}$ are 6.62, 0.62, 0.79, and 0.62, respectively.

E_{HOMO} and E_{LUMO} are used to indicate the charge transfer that occurs within the molecule of an organic compound. The larger the value of E_{HOMO}, the more readily the electrons in the highest occupied molecular orbital can escape. The larger the value of E_{LUMO}, the easier it is for electrons to enter the lowest unoccupied molecular orbital. E_{GAP} is mainly used to describe the chemical reaction properties and the kinetic stability of molecules for compounds[37]. The larger the E_{GAP} of a molecule, the more stable it is during physicochemical reactions.

The molecular mass M values range from a minimum value of 94.19 g/mol for DMDS to a maximum value of 637.69 g/mol for acid blue 113, with a mean value of 244.54 g/mol. This distribution suggests that the majority of the organic compounds have a molecular mass of around 200−300 g/mol.

2.3 QSAR model analysis

The 50 organic compounds in this study were used in a 4∶1 ratio for the training set and the test set, and the compounds in the test set are labeled in Table 1. The results of this study, according to the SPSS software, are shown in Table 3.

Table 3 Results of models for the prediction of Q_{m} by SPSS

Number	Model	R^2	R^2_{adj}	SD	RMSE	F	Sig.	q^2	Q^2_{ext}	95% CI
1	$Q_{\text{m}}=-0.65+127.03q(\text{CH}^+)_{\min}$	0.43	0.42	4.09	1 274.32	28.75	0.00	0.01	0.07	(2.57, 4.75)
2	$Q_{\text{m}}=6.69+95.02q(\text{CH}^+)_{\min}-1.75W(\text{C—C})_{\min}$	0.47	0.44	4.00	1 498.39	16.45	0.00	0.05	0.59	(2.57, 4.75)
3	$Q_{\text{m}}=19.87-180.83q(\text{CH}^+)_{\min}-4.97W(\text{C—C})_{\min}+0.09E_{\text{p}}(\text{C})$	0.77	0.75	2.68	2 767.65	39.71	0.00	0.42	0.59	(2.43, 4.90)
4	$Q_{\text{m}}=25.35+199.30q(\text{CH}^+)_{\min}-6.55W(\text{C—C})_{\min}+0.14E_{\text{p}}(\text{C—H})-0.02E_{\text{p}}(\text{C})$	0.83	0.81	2.33	2 318.52	42.56	0.00	0.63	0.62	(2.15, 5.17)
5	$Q_{\text{m}}=24.50+181.15q(\text{CH}^+)_{\min}-6.85W(\text{C—C})_{\min}+0.150E_{\text{p}}(\text{C—H})-0.02E_{\text{p}}(\text{C})-0.003E$	0.88	0.87	1.96	507.99	51.15	0.00	0.81	0.68	(2.10, 5.22)

Notes: F represents the value of Fisher's exact test, which is a specific numerical statistic; sig. is short for significance; 95% CI represents the 95% confidence interval.

With the gradual introduction of the number of independent variables, the R^2 and R^2_{adj} of models 1−5 gradually increase, and the SD values gradually become smaller, which indicates that the models show better regression effects and are statistically significant. Generally speaking, the regression coefficient of the optimal model requires $R^2 \geqslant 0.6$, from which models 1 and 2 can be excluded; the coefficients of internal and external validation require that q^2 and Q^2_{ext} are both higher than 0.5, so models 3, 4, and 5 satisfy the above requirements. The regression coefficients of model 5 ($R^2=0.88$, $R^2_{\text{adj}}=0.87$), and the coefficients of internal and external validation ($q^2=0.81$, $Q^2_{\text{ext}}=0.68$) are higher, and SD (1.96) and RMSE (507.99) are the smallest. Model 5 has a more accurate and reliable predictive ability compared to the other models. Therefore, model 5 is identified as the optimal model for the prediction of Q_{m} (denoted as Q_{m}-model).

The comparison of the predicted Q_{m} of model 5 with the experimental Q_{m} is shown in Fig. 2. The overall distribution of the predicted results under this model is more concentrated, which indicates that model 5 has a good predictive ability. There are five outliers: 2,4-DCP, Bis, 2,4,6-TCA, chlorpromazine, and clomipramine.

Most of these organic compounds are insoluble or slightly soluble in water, making them less prone to adsorption and leading to underestimated adsorption capacities. Furthermore, their structures are more complex than those in the training set, often featuring chlorine substituents or ether bonds. This structural complexity prevents the model from accurately capturing their adsorption behavior, consequently leading to prediction deviations. However, these deviations remain within an acceptable range. In conclusion, model 5 demonstrates a superior fit, as evidenced by the dense distribution of most predicted data points along the 1∶1 line, with only a few outliers falling outside the 95% confidence interval of the residuals.

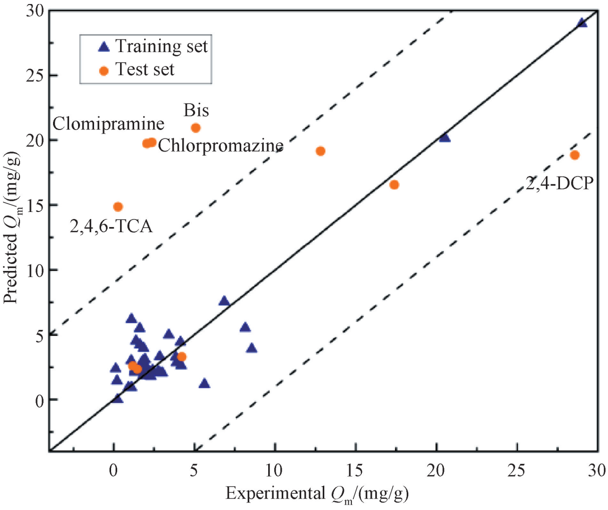

Fig. 2 Scatter plots of predicted Q_m versus experimental Q_m for Q_m-model

3 Discussion and Comparison

3.1 Mechanism explanation

According to the results of the Q_m-model, it can be seen that an increase in $q(CH^+)_{min}$ and $E_p(C—H)$ has a facilitating effect on Q_m, while an increase in $W(C—C)_{min}$, $E_p(C)$, and E has an inhibitory effect. A correlation analysis was conducted between Q_m and structural parameters for these 50 organic compounds, with the results shown in Fig. 3.

The parameter $q(CH^+)_{min}$ reflects the polarity of the molecule. As the $q(CH^+)_{min}$ value increases, the polarity of the molecule decreases, which results in the easier adsorption of the compound on activated carbon. Figure 3 shows that Q_m is significantly positively correlated with $q(CH^+)_{min}$. For example, PCP exhibits the maximum $q(CH^+)_{min}$ value (0.18 e) and the highest Q_m (29.00 mg/g), whereas clomipramine has the minimum $q(CH^+)_{min}$ (0.01 e) and the lowest Q_m (2.07 mg/g).

When $W(C—C)_{min}$ is large, it means that the bonding strength between carbon atoms is greater and the electron cloud inside the molecule is more stable, thus reducing the interaction between the molecule and activated carbon. On the contrary, a smaller $W(C—C)_{min}$ value, indicative of a weaker bond and greater molecular flexibility, appears to facilitate adsorption on activated carbon. For example, acid blue 62, amitriptyline, verapamil, procaine, propafenone, and propranolol all have large $W(C—C)_{min}$ values, and their adsorption capacities are less than the average level (4.44 mg/g).

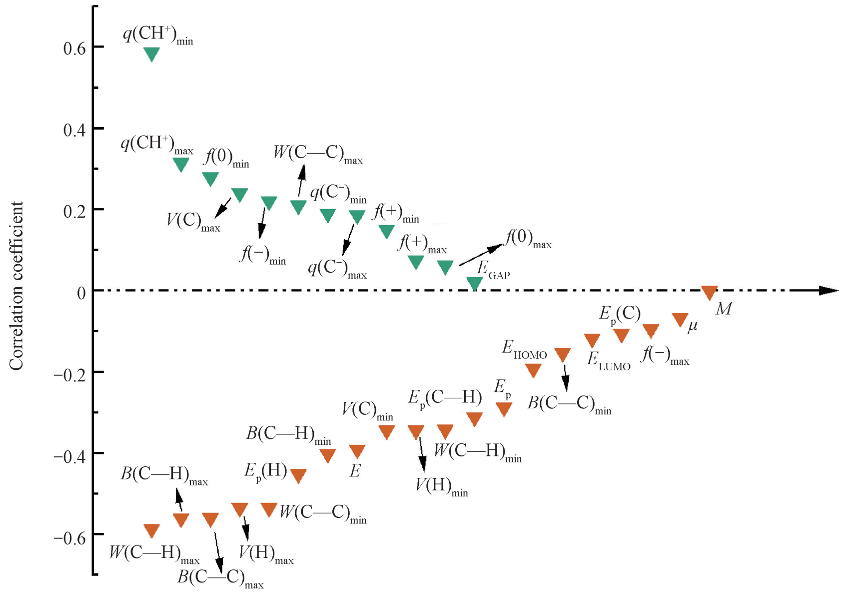

Fig. 3 Correlation coefficients of Q_m with 32 structural parameters

The larger the value of $E_p(C)$, the stronger the interaction force between the carbon atoms in the compound, which may make the molecule more compact, resulting in a lower adsorption capacity. For example, 2,4,5-T has a smaller $E_p(C)$ (−174.02 Ha) and the corresponding Q_m (12.82 mg/g) is much larger than the average value (4.44 mg/g); the same is true for atrazine and 2,4-DCP.

E_p(C—H) is related to the strength of the carbon-hydrogen bonding in the molecules, and larger values imply that there is a stronger interaction between the carbon and hydrogen atoms in the compounds, which may contribute to more stable adsorption between the molecules and activated carbon. However, we found an interesting phenomenon: T,T-2,4-ODA, T,T-2,4-DDA, and *m*-xylene with larger values of E_p(C—H) (−19.37, −17.93, and −17.28 Ha) have smaller values of Q_m (1.19, 1.97, and 1.85 mg/g). This also shows that it is not comprehensive to judge the adsorption capacity of a compound only by a single parameter in the model.

The lower the value of the total energy E, the more stable the molecule is and the more readily it is adsorbed on activated carbon. From the correlation analysis, it can be seen that E is negatively correlated with Q_m. Although the correlation (0.39) is not strong, the model suggests that the compound's stability still influences the adsorption capacity to a certain extent.

Take the example of atrazine, as shown in Fig. 4. The triazine (C_3N_3) in this compound ensures adsorption stability, and $q(CH^+)_{min}$ corresponds to a possible adsorption site, which is susceptible to electrostatic attraction with oxygen-containing functional groups on the surface of the activated carbon. The smaller W(C—C)$_{min}$ attests to the structural flexibility provided by the diethylamino group, which makes the molecule more susceptible to adsorption on activated carbon. The alkyl chain is bound to the surface of the activated carbon by the van der Waals force, further expanding the adsorption contact area. In conclusion, it is the combined effect of electrostatic adsorption and van der Waals force that causes atrazine to have a high adsorption capacity.

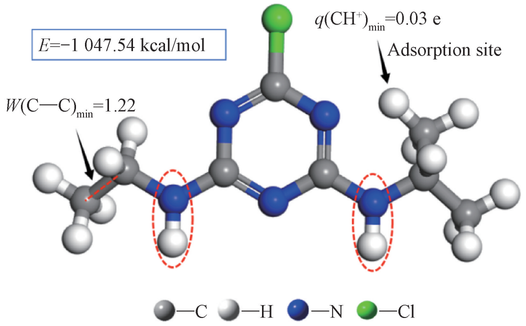

Fig. 4 Adsorption mechanism of atrazine on activated carbon

In summary, the intrinsic factors affecting the adsorption capacity of compounds on activated carbon include the molecular polarity ($q(CH^+)_{min}$), the molecular potential energy (E_p(C—H) and E_p(C)), the molecular structure (W(C—C)$_{min}$), and the molecular stability (E).

3.2 Model validation

To test the stability of the optimal model, the optimal model is tested as shown in Table 4. The sig. values of all independent variables in the model are less than 0.05, which meets the requirements. The absolute value of t (the t-statistic for testing the significance of each parameter) is higher than 2.18, which indicates that these five variables are acceptable, and the significant order is $q(CH^+)_{min}$, W(C—C)$_{min}$, E_p(C—H), E_p(C), and E. The variance inflation factor (VIF) is used to test the multicollinearity between the independent variables in the model, and its standard requirement of less than 10 means that the model equation is acceptable[38].

Table 4 Statistical values of parameters for Q_m-model

Parameter	Regression coefficient	VIF	t	Sig.
$q(CH^+)_{min}$	181.15	2.66	9.77	0
W(C—C)$_{min}$	−6.85	3.37	−9.40	0
E_p(C—H)	0.15	9.65	9.49	0
E_p(C)	−0.02	3.62	−4.65	0
E	−0.003	1.40	−3.93	0
Criterion	—	<10	>2.18	<0.05

All the parameters included in the model meet the requirement, which indicates that the Q_m-model is statistically significant. In addition, the external validation (Q^2_{ext} = 0.68) satisfies the criterion of being greater than 0.5, which suggests that the model has a good predictive potential.

3.3 *Y*-randomization validation

Y-randomization validation is mainly used to determine the stability of a model by testing the chance relationship between the dependent and independent variables. For a stable and reliable optimal model, the randomized regression coefficient R^2_{yrand} and external validation coefficient Q^2_{yrand} of the new model obtained after random shuffling should be much lower than those of the original model (R^2 and Q^2). As shown in Fig. 5, it is found that the R^2_{yrand} and Q^2_{yrand} values from multiple randomization tests are consistently lower than those of the Q_m-model. This result confirms that the Q_m-model developed in this study is robust and stable.

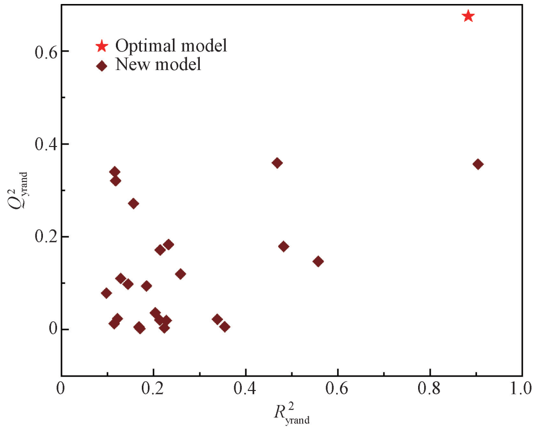

Fig. 5 *Y*-randomization validation of Q_m-model

3.4 Application domain

To define the applicability of the Q_m-model for predicting the adsorption of other organic compounds, Williams plots were employed to visualize the applicable performance domain (APD) of the Q_m-model. The detailed results are presented in Fig. 6. A reliable prediction for a compound is generally achieved only when its standardized residual falls within a range of −3 to 3 and its leverage value is below the critical threshold h^*.

As shown in Fig. 6, the standardized residuals of most of the organic compounds satisfy the above requirements, but PCP and 2,4-DCP are not in the range of the application domain. The structural properties of PCP are significantly different from those of the organic compounds in the training set. Therefore, the optimal model may not provide reliable predictions for this compound. For 2,4-DCP, it is only slightly out of the critical value range and is still within reasonable limits. In conclusion, the optimal model still has good stability and prediction ability for the compound with a similar structure to those in the training set.

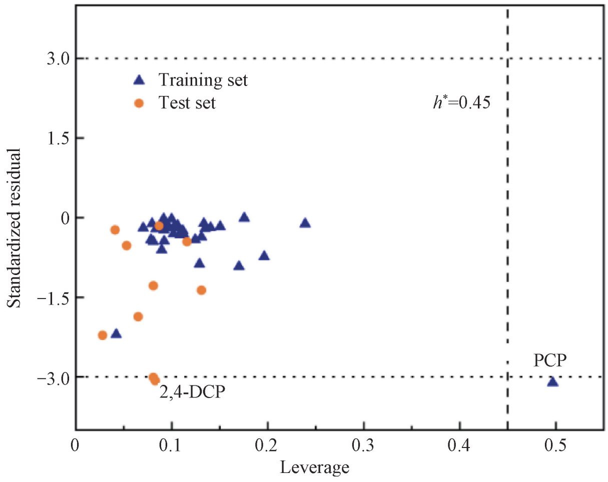

Fig. 6 Williams plot for Q_m-model

3.5 Comparison with models in literature

This study compares the Q_m-model with related studies on activated carbon adsorption, as summarized in Table 5. The comparison shows that the compound types in the datasets of the existing studies are relatively homogeneous. For instance, Blum et al.[8] only focused on aromatic and aliphatic compounds, Gong et al.[11] focused on mono-nitroaromatic compounds, and Zhao et al.[9] only studied cationic pharmaceuticals. Consequently, such models may have limited applicability and are difficult to generalize to other types of pollutants. There is a lack of model performance validation in the above studies, and the predictive ability of the models is only evaluated by indicators such as R^2 and external validation coefficients, which may not be able to adequately describe the reliability and applicability of the models. In addition, the structural parameters of the compounds are rarely considered when investigating the factors affecting activated carbon adsorption, whereas the present study focuses on the influence of the quantum chemical properties of the compounds on the adsorption mechanism.

Unlike existing modeling studies, this study breaks through the limitation of a single category and examines a wide range of pollutants, including dyes, pesticides, solvents, drugs, and odorants, to increase the applicability of the model. The Wiberg index and molecular potential energy are introduced in the selection of structural parameters of compounds. For the established model, we conduct not only internal and external validation but also Y-randomization test and Williams plot analysis to ensure its robustness. Compared with the existing studies, the model not only significantly improves the prediction accuracy but also provides a more comprehensive physical explanation of the adsorption mechanism on activated carbon, especially in the quantitative analysis of the molecular polarity and bond strength, which fills the gap in the literature.

Table 5 Modeling studies in the literature similar to this study

Ref. (author, year)	Methods	Type of dataset	R^2	Key impact factor
[8] (Blum et al., 1994)	MLR	Aromatic and aliphatic compounds	0.92	Molecular connectivity indices
[10] (De Ridder et al., 2010)	MLR	Organic micropollutants	0.61−0.84	Hansen solubility parameter for hydrogen bonding, and polarizability
[11] (Gong et al., 2015)	PLSR	Mono-nitroaromatic compounds	0.898	Molecular orbital, and atomic net charge
[9] (Zhao et al., 2018)	MLR	Cationic pharmaceuticals	0.80	Octanol-water partition coefficient, polar surface area, and molecular mass

4 Conclusions

In this study, an optimal QSAR model was developed by using the MLR method: $Q_m = 24.50 + 181.15q(\mathrm{CH}^+)_{\min} - 6.85W(\mathrm{C—C})_{\min} + 0.15E_p(\mathrm{C—H}) - 0.02E_p(\mathrm{C}) - 0.003E$. The results indicate that the polarity of the molecule, the molecular potential energy, and the stability and bonding strength of the compound have important effects on the adsorption by activated carbon. This model serves as a valuable tool for predicting and optimizing the adsorption treatment of organic pollutants in water.

References

[1] AL-TOHAMY R, ALI S S, LI F H, et al. A critical review on the treatment of dye-containing wastewater: ecotoxicological and health concerns of textile dyes and possible remediation approaches for environmental safety [J]. *Ecotoxicology and Environmental Safety*, 2022, 231: 113160.

[2] ABUQAMAR S F, EL-SAADONY M T, ALKAFAAS S S, et al. Ecological impacts and management strategies of pesticide pollution on aquatic life and human beings [J]. *Marine Pollution Bulletin*, 2024, 206: 116613.

[3] GANGAR T, PATRA S. Antibiotic persistence and its impact on the environment [J]. *3 Biotech*, 2023, 13(12): 401.

[4] WONG S, NGADI N, INUWA I M, et al. Recent advances in applications of activated carbon from biowaste for wastewater treatment: a short review[J]. *Journal of Cleaner Production*, 2018, 175: 361-375.

[5] CZEPIRSKI L, BALYS M R, KOMOROWSKA-CZEPIRSKA E. Some generalization of Langmuir adsorption isotherm [J]. *Internet Journal of Chemistry*, 2000, 3(14): 14.

[6] MONCHO S, SERRANO-CANDELAS E, DE JULIÁN-ORTIZ J V, et al. A review on the structural characterization of nanomaterials for nano-QSAR models [J]. *Beilstein Journal of Nanotechnology*, 2024, 15: 854-866.

[7] HASAN M R, ALSAIARI A A, FAKHURJI B Z, et al. Application of mathematical modeling and computational tools in the modern drug design and development process[J]. *Molecules*, 2022, 27(13): 4169.

[8] BLUM D J W, SUFFET I H, DUGUET J P. Quantitative structure-activity relationship using molecular connectivity for the activated carbon adsorption of organic chemicals in water [J]. *Water Research*, 1994, 28(3): 687-699.

[9] ZHAO Y F, CHOI J W, BEDIAKO J K, et al. Adsorptive interaction of cationic pharmaceuticals on activated charcoal: experimental determination and QSAR modelling [J]. *Journal of Hazardous Materials*, 2018, 360: 529-535.

[10] DE RIDDER D J, VILLACORTE L, VERLIEFDE A R D, et al. Modeling equilibrium adsorption of organic micropollutants onto activated carbon[J]. *Water Research*, 2010, 44(10): 3077-3086.

[11] GONG W W, LIU X H, GAO D, et al. The kinetics and QSAR of abiotic reduction of mononitro aromatic compounds catalyzed by activated carbon[J]. *Chemosphere*, 2015, 119: 835-840.

[12] ROUSTAEI N. Application and interpretation of linear-regression analysis [J]. *Medical Hypothesis, Discovery & Innovation Ophthalmology Journal*, 2024, 13(3): 151-159.

[13] ÖRÜCÜ E, TUGCU G, SAÇAN M T. Molecular structure-adsorption study on current textile dyes [J]. *SAR and QSAR in Environmental Research*, 2014, 25(12): 983-998.

[14] ZHANG P, YANG Y P, ZHAO D. Study on adsorption competition on activated carbon between chloroform and other chlorine disinfectants in drinking water[J]. *Journal of Zhejiang University (Science Edition)*, 2020, 47(1): 107-114. (in Chinese)

[15] STOYCHEVA I G, TSYNTSARSKI B G, PETROVA B N, et al. Novel polymer-based nanoporous carbon adsorbent for removal of pentachlorophenol from water [J]. *Bulgarian Chemical Communications*, 2017, 49: 44-50.

[16] NABOULSI A, EL MERSLY L, YAZID H, et al. Adsorption behaviors and mechanisms by theoretical study of herbicide 2,4,5-trichlorophenoxyacetic on activated carbon as a new biosorbent material [J]. *Journal of the Taiwan Institute of Chemical Engineers*, 2023, 142.

[17] ZHANG C R. The preparation of composite nano-structured materials and the research of their photocatalytic and adsorption characteristics[D]. Jinan: Shandong Normal University, 2012. (in Chinese)

[18] SHEN N. Emergent treatment for pesticide endocrine-disrupting chemicals in source water by powdered activated carbon[D]. Harbin: Harbin Institute of Technology, 2011. (in Chinese)

[19] AYRANCI E, HODA N. Adsorption kinetics and isotherms of pesticides onto activated carbon-cloth [J]. *Chemosphere*, 2005, 60(11): 1600-1607.

[20] LAW M D. Sorption of 2,4-dinitrotoluene and its effect on biodegradation [D]. Houston: University of Houston-Clear Lake, 2004.

[21] AN N, XIE H H, GAO N Y, et al. Adsorption characteristics of taste and odor compounds IPMP and IBMP by powder activated carbon in aqueous system[J]. *Journal of Central South University (Science and Technology)*, 2012, 43(9): 3718-3726. (in Chinese)

[22] XIAO H. Study on the influence of algae organic matters during adsorption of three typical odorants by powdered activated carbon and the mecha [D]. Beijing: Beijing University of Civil Engineering and Architecture, 2022. (in Chinese)

[23] LIAO Y. Removal effects and mechanisms of sulfide odorous by oxidation and powdered

activated carbon adsorption [D]. Beijing: Beijing University of Civil Engineering and Architecture, 2021. (in Chinese)

[24] SHAARANI F W, HAMEED B H. Ammonia-modified activated carbon for the adsorption of 2,4-dichlorophenol [J]. *Chemical Engineering Journal*, 2011, 169(1/2/3): 180-185.

[25] EL-DESOUKY M G, KHALIL M A G, EL-AFIFY M A M, et al. Effective methods for removing different types of dyes—modelling analysis, statistical physics treatment and DFT calculations: a review [J]. *Desalination and Water Treatment*, 2022, 280: 89-127.

[26] KURANOV D Y, BEDRINA M E, EGOROV N V. The structure and electronic properties of organic semiconductors [C]//2015 International Conference "Stability and Control Processes" in Memory of VIZubov (SCP). New York: IEEE, 2025: 651-652.

[27] REN Y Y, LIU S Q, TAN Y J, et al. Application of QSAR for investigation on coagulation mechanisms of textile wastewater [J]. *Ecotoxicology and Environmental Safety*, 2022, 244: 114035.

[28] ZHANG J, HOFMANN R. Modeling the adsorption of emerging contaminants on activated carbon: classical and quantum QSAR approaches [J]. *Water Supply*, 2013, 13(6): 1543-1552.

[29] GRAMATICA P, SANGION A. A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology [J]. *Journal of Chemical Information and Modeling*, 2016, 56(6): 1127-1131.

[30] JOHNSON E R, OTERO-DE-LA-ROZA A. Adsorption of organic molecules on kaolinite from the exchange-hole dipole moment dispersion model [J]. *Journal of Chemical Theory and Computation*, 2012, 8(12): 5124-5131.

[31] OSMAN ABDELKARIM O I, ASIRI A M. NBO technique as a descriptor of aromaticity [J]. *Computational and Theoretical Chemistry*, 2022, 1210: 113637.

[32] MENEKŞE B Ç. Wiberg bağ indeksi hesaplamalarını kullanarak bazı nitroaromatik bileşiklerin tetik bağ analizleri [M]. Turkey: Balikesir University, 2024.

[33] ZHU H C, SHEN Z M, TANG Q L, et al. Degradation mechanism study of organic pollutants in ozonation process by QSAR analysis [J]. *Chemical Engineering Journal*, 2014, 255: 431-436.

[34] LIU Y W, CHENG Z W, LIU S Q, et al. A quantitative structure activity relationship (QSAR) model for predicting the rate constant of the reaction between VOCs and NO_3 radicals [J]. *Chemical Engineering Journal*, 2022, 448: 136413.

[35] LANG P F. Bond order and bond energies [J]. *Foundations of Chemistry*, 2024, 26(1): 167-177.

[36] LI X H, JALBOUT A F, SOLIMANNEJAD M. Definition and application of a novel valence molecular connectivity index [J]. *Journal of Molecular Structure: Theochem*, 2003, 663(1/2/3): 81-85.

[37] CHENG Z W, CHEN Q C, PONTIUS F W, et al. Two new predictors combined with quantum chemical parameters for the selection of oxidants and degradation of organic contaminants: a QSAR modeling study [J]. *Chemosphere*, 2020, 240: 124928.

[38] MARCOULIDES K M, RAYKOV T. Evaluation of variance inflation factors in regression models using latent variable modeling methods [J]. *Educational and Psychological Measurement*, 2019, 79(5): 874-882.

基于 Langmuir 吸附等温线的活性炭吸附 QSAR 模型

谭　婷[1,2]，魏群山[1,2*]，刘　琼[1,2*]，申哲民[3]，宋新山[1,2]，王宇晖[1,2]，CHARLES Nzila[4]，CHRISTOPHER W. K. Chow[5]

1. 东华大学 环境科学与工程学院，上海 201620
2. 东华大学 纺织工业污染治理与控制国家环境保护工程中心，上海 201620
3. 上海交通大学 环境科学与工程学院，上海 200240
4. 莫伊大学 工程学院，肯尼亚 埃尔多雷特 3900-30100
5. 南澳大学 可持续基础设施与资源管理（SIRM），澳大利亚 莫森湖 5095

摘　要：从量子化学的角度探究化合物的结构特性对活性炭吸附性能的影响。选取符合 Langmuir 吸附等温模型的化合物作为研究对象，以 Langmuir 吸附等温线中的饱和吸附量 Q_m 为响应值，以 50 种不同类型化合物的结构参数为自变量，采用多元线性回归法构建了定量构效关系（quantitative structure-activity relationship，QSAR）模型。结果表明，最优模型具有良好的稳定性、可靠性和稳健性，回归系数 $R^2=0.88$，调整后回归系数 $R_{adj}^2=0.87$，内部验证系数 $q^2=0.81$，外部验证系数 $Q_{ext}^2=0.68$。最优模型所包含的变量表明分子的极性、分子势能、化合物的稳定性及键合强度是影响活性炭吸附的主要因素。研究结论为活性炭对有机化合物吸附能力的预测提供了重要依据，也为其在水环境中的吸附处理应用提供了理论支持。

关键词：定量构效关系（QSAR）；吸附；活性炭；多元线性回归