Desalination and Water Treatment
www.deswater.com

doi:10.5004/dwt.2017.21401

88 (2017) 235–256
September

Quantitative structure-activity relationship (QSAR) for the adsorption of organic contaminants by nascent state manganese dioxide

Pingru Su, Wenchao Ji, Zhemin Shen*, Jianda Zhang, Weimin Guo

School of Environmental Science and Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China, Tel. +86 15201927189, email: supingru@126.com (P. Su), Tel. +86 18818215305, email: wenchaoji@sjtu.edu.cn (W. Ji), Tel. +86 21 54741065, email: zmshen@sjtu.edu.cn (Z. Shen), Tel. +86 15831105286, email: zjdrf@163.com (J. Zhang), Tel. +86 21 54740584, email: wmguo@sjtu.edu.cn (W. Guo)
College of Resources and Environment Science, Hebei Normal University, 20# East Road of Southern 2nd Ring, Shijiazhuang, 050024, China

Received 5 December 2016; Accepted 29 August 2017

ABSTRACT

Nascent state manganese dioxide exhibits promising adsorptive capacity as it has large surface area with many water molecules and hydroxyls on. Therefore it has been applied as a treatment of organic pollution. The experimental removal rates (R_{exp}) under 3 different pH conditions and 18 quantum descriptors of 29 organic contaminants were used to construct quantitative structure-activity relationship (QSAR) models to predict the removal rate at a certain pH. The optimum models at pH 3, 7, 10 are listed as follows respectively, R_{pre} = 1 .437 – 5.627f(+)$_x$ + 1.018q(C$^-$)$_x$ + 0.099E$_{HOMO}$ (R^2 = 0.9273, pH = 3); R_{pre} = –4.153 + 2.632BO$_x$ + 0.688BO$_n$ – 0.696q(C$^-$)$_n$ (R^2 = 0.7459, pH = 7); R_{pre} = –0.057 + 0.031μ + 0.243BO$_n$ + 4.976f(0)$_n$ – 3.938f(+)$_x$ (R^2 = 0.9213, pH = 10). All of the optimum models show satisfactory stability, evident reliability and powerful predicability. Furthermore, they have no possibility of chance correlation. The results presented that affinity with solvent, charge distribution and stability of a molecule were main molecular characters influencing removal effect.

Keywords: QSAR; Organic contaminants; Adsorption; Nascent state manganese dioxide; Quantum descriptors

1. Introduction

Potassium permanganate ($KMnO_4$) process is an advanced oxidation process (AOPs) which is commonly used due to its modest cost, simple operation, safe storage and delivery, applicability over a wide range of conditions and high efficiency for pollution control [1,2]. Potassium permanganate method is mainly applied into such fields as rehabilitation of underground water [3], oxidation of dissolved manganese/ferrous ion and arsenite [4], degradation of various organic pollutants [5–7], control of taste and odor [8] and removal of algae [9]. Compared with other AOPs, the studies on potassium permanganate process are relatively insufficient, demanding further efforts on the study of its kinetics, reaction products and treatment mechanism of certain pollutants [10].

Nascent state manganese dioxide (MnO_2) is the intermediate product of potassium permanganate during the oxidation process, and exhibits favorable adsorptivity because of its high surface charge, strong ion adsorption capacity, large specific surface area [11], and the active surface hydroxyl groups [12].

Under slightly alkaline or neutral condition:

$$MnO_4^- + 2H_2O + 3e^- \rightarrow MnO_2(s) + 4OH^- \quad (1)$$

Under acid condition:

$$MnO_4^- + 4H^+ + 3e^- \rightarrow MnO_2(s) + 2H_2O \quad (2)$$

*Corresponding author.

Many studies showed that the removal efficiency of $KMnO_4$ process contains two parts: one is the oxidation by potassium permanganate, and the other one is the adsorption by nascent state MnO_2 [12,13]. Sun, Guan, Fang and Tratnyek [2] found out that MnO_4^- could be activated by HSO_3^-, resulting in a process that oxidizes organic contaminants were at extraordinary high rates as the process may involve manganese (III) species with minimal stabilization by complexation. However, we suppose the phenomenon of high rates of degradation of organic contaminants was probably due to the formation of MnO_2, as MnO_4^- was reduced by HSO_3^-. The adsorption of organic contaminants by MnO_2 was much faster than that of oxidation by MnO_4^-. Hence the observed removal effect exhibits high rates. Therefore, this new discovery is in need of further exploration and confirmation.

More and more attention has been paid on the removal mechanism of MnO_2in recent years. Xu et al. [14] indicated that MnO_2 was a favorable adsorbent for the treatment of Pb(III). Mustafa et al. [11] found out that the adsorption of phosphate anions by MnO_2 decreased when pH value increased. Qu et al. [15] pointed out that the nascent state MnO_2 presented a stronger ability to remove algal extracellular organic matter (EOM) than that of the commercial and lab-prepared MnO_2.

The mechanism of $KMnO_4$ to treat wastewater should not be merely considered as an oxidation process. While most studies, so far, just concentrate on the total removal rate and oxidation ability of $KMnO_4$. It is inevitable to separately explore the adsorption process by nascent state MnO_2 because it can provide not only an incisive comprehension of the removal mechanism of organic contaminants by $KMnO_4$ but also a valuable theoretical guidance for the application of MnO_2 to the treatment of high concentration organic wastewater.

There are huge differences between the removal efficiencies of organic compounds by manganese dioxide because of the variation of their structures [16,17] and the change of experimental conditions. In our adsorption experiment, the removal rate of Nitrobenzene (pH = 3) was merely 17.66%, while that of Rhodamine B (pH = 3) was 100.00%. The removal rates of Rhodamine B decreased to 51.35% and 7.06% with the pH value changed to 7 and 10. Consequently, not all organic contaminants are appropriate to be treated by MnO_2 and pH condition is an important factor that influences the removal efficiency a lot. Therefore, it requires the prediction of the adsorption rate and the optimal pH condition to help us choose the suitable treatment method.

Quantitative structure-activity relationship (QSAR) is a popular theoretical prediction method at present. Initially, it was applied in drug design. It has been utilized in varied areas up to now, as it is an economical and fast substitute to experimental measurement [18,19]. In the field of environmental science, many researchers measured the potential damage of chemicals to the environment by QSAR analysis [20,21]. Using QSAR models, many researchers have made predictions of the adsorption results of organic contaminants by different adsorbent [22–24], as well as estimated degradation behaviors of different oxidants [25–28]. However, as far as we know, few researches on QSAR models for adsorption behaviors of nascent state MnO_2 have been reported yet.

This study aimed to (1) investigate the properties of nascent state manganese dioxide by X-ray diffraction (XRD) and Fourier transform infrared (FT-IR) spectroscopy; (2) construct QSAR models to make predictions for the removal rates of various organic contaminants by nascent state MnO_2 at 3 different pH conditions by involvement of DFT-based descriptors; (3) ascertain the optimum pH condition for the adsorption; (4) explain the possible adsorption mechanisms under different pH conditions by the analysis of the quantum descriptors in corresponding QSAR models.

2. Materials and methods

2.1. Experimental materials and methods

29 common organic compounds (AR, Sinopharm Chemical Reagent Co., Ltd. or Aladdin Industrial Corp.) were chosen as experimental material, including dyes (Indigo, Crystal violet, Orange G et al.), indicator (Methyl orange, Methyl red, Acid chrome blue K, etc.), aromatic compounds (nitrobenzene, 2,4-dinitrophenylhydrazine, 3,4-dichloroaniline,etc.). Information about experimental materials are presented in Table 1. The adsorbent, nascent state MnO_2, was generated by the reaction of stoichiometric amount of $KMnO_4$ (AR, Sinopharm Chemical Reagent Co., Ltd.) and anhydrous Na_2SO_3 (Aladdin Industrial Corp.). Sulphuric acid (H_2SO_4) and sodium hydrate (NaOH) (AR, Sinopharm Chemical Reagent Co., Ltd.) were prepared to solutions (1.0 mol/L) to adjust pH value.

The original solutions in our study were prepared by 29 organic materials and deionized-distilled water (DDW) in triplicate, with the concentration of 100 mg/L. The pH of the triple solutions of each compound was adjusted by H_2SO_4 or NaOH to 3.0 ± 0.1, 7.0 ± 0.1 and 10 ± 0.1, respectively. MnO_2 was formed in situ in accord with the reaction below:

$$2MnO_4^- + 3SO_3^{2-} + H_2O \rightarrow 2MnO_{2(S)} + 3SO_4^{2-} + 2OH^- \quad (3)$$

The pH value of nascent state MnO_2 was adjusted and then let stand for about 10 min. The supernatant was poured away and the residue was mixed into the original solutions. The molar ratio of the adsorbent (MnO_2) to every experimental material was controlled as 20:1. After 1 h, the mixed solutions were filtered through a microporous membrane (0.45 μm) and the filtrates were analyzed. The concentrations of organic compounds in both the original solutions and the final filtrates were full wavelength (200–800 nm) scanned by a UV spectrophotometer (UV-1600, Shanghai Mapada Instruments Co., Ltd.) to get absorbancy curves.

The microstructure of the nascent state manganese dioxide was also explored by XRD and FT-IR. The samples of manganese dioxide were pretreated before the measurement, following the steps below: (1) vacuum filtration; (2) abstersion; (3) desiccation; (4) calcination: (5) grinding. X-ray diffraction powder analysis was performed with a XRD diffractometer (XRD-6100, Shimadzu Corporation) using CuKα radiation. The scanning range (2θ) is from 5° to 90° at the speed of 10°/min. The FT-IR spectrum of a tested compound was recorded on a spectrometer (Nicolet 6700, Thermo Fisher Scientific Inc.) by using KBr pellet technique.

Table 1
Information of 29 compounds used in the study

No.	Materials	SMILES	CAS no.	Relative molecular mass
1[c]	1,10-Phenanthroline monohydrate	c1cnc2c(c1)ccc3cccnc23	5144-89-8	198.22
2	1-Nitroso-2-naphthol	Oc1ccc2ccccc2c1N=O	131-91-9	173.17
3	2,4-Dichlorophenol	Oc1ccc(Cl)cc1Cl	120-83-2	163
4[bc]	2,4-Dinitrophenylhydrazine	NNc1ccc(cc1[N](=O)=O)[N](=O)=O	119-26-6	198.15
5[c]	3,4-Dichloroaniline	Nc1ccc(Cl)c(Cl)c1	95-76-1	162.02
6[c]	Acid chrome blue K	[Na+].[Na+].[Na+].Oc1ccc(cc1N=Nc2c(O)c3c(O)cc(cc3cc2[S]([O-])(=O)=O)[S]([O-])(=O)=O)[S]([O-])(=O)=O	3270-25-5	586.4
7	Acid orange	[Na+].Oc1ccc2ccccc2c1N=Nc3ccc(cc3)[S]([O-])(=O)=O	633-96-5	350.32
8[b]	Azure I	C.Nc1ccc2N=C3C=CC(=[N+])C=C3Sc2c1	531-55-5	305.83
9[a]	Basic fuchsin	[Cl-].Cc1cc(ccc1N)C(c2ccc(N)cc2)=C3C=CC(=[N+])C=C3	632-99-5	337.85
10[ab]	Bromocresol green	Cc1c(Br)c(O)c(Br)cc1C2(O[S](=O)(=O)c3ccccc23)c4cc(Br)c(O)c(Br)c4C	76-60-8	698.05
11[a]	Bromophenol blue	Oc1c(Br)cc(cc1Br)C2(O[S](=O)(=O)c3ccccc23)c4cc(Br)c(O)c(Br)c4	115-39-9	670.02
12	Chromotropic acid	Oc1cc(cc2cc(cc(O)c12)[S](O)(=O)=O)[S](O)(=O)=O	148-25-4	320.3
13	Cresol red	Cc1cc(ccc1O)C(=C2CCC(=O)C(=C2)C)c3ccccc3[S](O)(=O)=O	1733-12-6	382.42
14	Crystal violet	[Cl-].CN(C)c1ccc(cc1)C(c2ccc(cc2)N(C)C)=C3C=CC(C=C3)=[N+](C)C	548-62-9	408.03
15[ab]	Eriochrome black T	[Na+].Oc1cc(c2cc(ccc2c1N=Nc3ccc4ccccc4c3O)[N+]([O-])=O)[S]([O-])(=O)=O	1787-61-7	461.38
16	Eriochrome blue black R	[Na+].Oc1ccc2ccccc2c1N=Nc3c(O)cc(c4ccccc34)[S]([O-])(=O)=O	2538-85-4	416.39
17	Indigo	O=C1C(Nc2ccccc12)=C3Nc4ccccc4C3=O	482-89-3	262.26
18[c]	Isatin	O=C1Nc2ccccc2C1=O	91-56-5	147.13
19	Malachite green	[Cl-].CN(C)c1ccc(cc1)C(c2ccccc2)=C3C=CC(C=C3)=[N+](C)C	569-64-2	364.92
20	m-Cresol purple	[Na+].Cc1cc(O)ccc1C(=C2C=CC(=O)C=C2C)c3ccccc3[S]([O-])(=O)=O	2303-01-7	382.43
21	Metanil yellow	[Na+].[O-][S](=O)(=O)c1ccc(cc1)N=Nc2ccc(Nc3ccccc3)cc2	587-98-4	375.38
22[b]	Methyl red	[Na+].CN(C)c1ccc(cc1)N=Nc2ccc(cc2)[S]([O-])(=O)=O	547-58-0	327.33
23	Methyl orange	[Na+].CN(C)c1ccc(cc1)N=Nc2ccc(cc2)[S]([O-])(=O)=O	547-58-0	327.33
24[c]	Methylene blue	[Cl-].CCNC1CCC2N=C3C=CC(C=C3SC2C1)=[N+](C)C	7220-79-3	319.85
25[a]	Naphthol green B	[Na+].[Na+].[Na+].[O-][S](=O)(=O)c1ccc2c(C=CC(O[Fe+3](OC3C=Cc4cc(ccc4[N]3=O)[S]([O-])(=O)=O)OC5C=Cc6cc(ccc6[N]5=O)[S]([O-])(=O)=O)[N]2=O)c1	19381-50-1	878.46
26[b]	Nitrobenzene	O=[N](=O)c1ccccc1	98-95-3	123.11
27	Orange G	[Na+].Oc1ccc2cc(cc(c2c1N=Nc3ccccc3)[S](=O)(=O)[O-][Na+])[S]([O-])(=O)=O	1936-15-8	452.37
28[a]	p-dimethylaminobenzaldehyde	CN(C)c1ccc(C=O)cc1	100-10-7	149.2
29	Rhodamine B	[Cl-].CCN(CC)C1=CC2=C(C=CC=CC3=C(C=C1)C3=C4C=CC(C=C4O2)=[N+](CC)CC)C(O)=O	81-88-9	479.01

[a]Samples in the external test set at pH = 3

[b]Samples in the external test set at pH = 7

[c]Samples in the external test set at pH = 10

The scanning record is in the scope of 400–4000 cm^{-1} with the scanning speed being 10 cm^{-1}/min.

All the experiments and measurements above were performed at room temperature (298K).

2.2. Computation details

Density functional theory (DFT) is a computational modelling means, based on Hohenberg-Kohn theorems, to explore the electronic structures of many-body systems. The elementary principle of DFT is the quantities which evaluate the feedback of the energy in light of diverse perturbations may be regard as descriptors, constructing the relationship between electronic structure and the affinity for other reaction molecules [29].

The structure of each organic molecule was optimized in Gaussian 09 by B3LYP function with 6-311G basis set. Then total energy ($E_{(B3LYP)}$) of the optimized structure was computed by using the same method. Finally, other 8 quantum descriptors were achieved from computation results. μ stands for the dipole moment in a vacuum, reflecting the asymmetry of a molecule. Based on natural bond orbital (NBO) theory, qH^+ refers to the maximal value of positive NBO charge on a hydrogen atom (qH^+); $q(C^-)_n$ and $q(C^-)_x$ refer to minimal and maximal values of negative NBO charge on a carbon atom; $q(CH^+)_n$ and $q(CH^+)_x$ refer to minimum and maximum positive NBO charge on a hydrogen atom linked to a carbon atom. These 5 descriptors reflect the distribution of charge on different atoms of a molecule. E_{HOMO} represents the energy of highest occupied molecular orbital and E_{LUMO} represents the energy of the lowest unoccupied molecular orbital. E_{GAP} is the difference between E_{LUMO} and E_{HOMO}, reflecting the difficulty of a molecule to be excited.

Two Bond order descriptors and six Fukui indices were calculated by another software Material Studio 6.1 with the set of DMol^3 code, gradient approximation (GGA) method and double numerical plus polarization (DNP) 3.5 basis. In molecular orbital theory, bond order is defined as half the difference between the number of bonding electrons and the number of antibonding electrons as the equation below. Bond order is also an index of bond strength which gives an indication of the stability of a bond. Usually a molecule has the tendency to be more stable if its BO increase when BO is less than 4 [30]. BO_x and BO_n are defined as the maximum and the minimum bond order of a molecule.

$$\mathrm{BO} = \frac{\text{number of boding electrons} - \text{number of antibonding electrons}}{2} \quad (4)$$

Fukui indices are particularly useful as they reflect the change of energy with a perturbation in the number of electrons [31].

$$f(+) = \rho N+1(r) - \rho N(r) \quad (5)$$

$$f(-) = \rho N(r) - \rho N-1(r) \quad (6)$$

$$f(0) = 1/2[\rho N+1(r) - \rho N-1(r)] \quad (7)$$

where ρN+1(r), ρN(r) and ρN–1(r) are the electron densities of the N + 1, N and N –1 electron system, and f(+), f(–) and f(0) quantitatively reflect the affinity with nucleophilic attack, electrophilic attack and ·OH radical attack. It has been reported that Fukui indices were used in QSAR analyses to build up models for the removal of organic compounds in ozone oxidation process [25,30] and in the Fenton oxidation process [32].

2.3. Construction method

Multiple linear regression (MLR) is an approach of analysis for measuring the relevance between a series of independent variables, and a dependent variable [33]. In this study, multiple linear regression was analyzed by statistical software of SPSS (17.0) in order to figure out the relationship between adsorption activity of nascent state MnO_2 and quantum structures of organic compounds. The statistical analyses in SPSS usually output a series of potential models, which requires evaluation, validation and screening.

2.4. Tests and validations

The quality of a QSAR model can be evaluated in accord with squared correlation coefficient (R^2). Based on Tropsha's test [34,35], the whole data set should be partitioned into training set involved in the construction of the models and test set not concerned in the training practice. The following statistical indices were adopted to reckon the quality of the models.

Squared correlation coefficient R^2:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - \tilde{y}_{tr}\right)^2} \quad (8)$$

$$R_{EXT}{}^2 = 1 - \frac{\sum_{i=1}^{ntest}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{ntest}\left(y_i - \tilde{y}_{tr}\right)^2} \quad (9)$$

$$k = \frac{\sum_{i=1}^{ntest} y_i \tilde{y}_i}{\sum_{i=1}^{ntest} \tilde{y}_i^2} \quad (10)$$

$$R_0{}^2 = 1 - \frac{\sum_{i=1}^{ntest}(\tilde{y}_i - \tilde{y}_i^{r0})^2}{\sum_{i=1}^{ntest}\left(\tilde{y}_i - \bar{\tilde{y}}\right)^2} \quad (11)$$

where *n* and *ntest* are the numbers of data divided into the training set and the test set, $\tilde{y}_{tr}$ is the average of the dependent variable in the training set, y_i and $\tilde{y}_i$ are the observed values and the predicted values of the dependent variable, $\bar{\tilde{y}}$ is the average of all in test set (I =1,…, *ntest*) and $\tilde{y}_i^{r0} = ky_i$ (I = 1,…, *ntest*).

Tropsha [36] made the criteria of a qualified QSAR model as follows: $R^2 > 0.5$, $R_{EXT}{}^2 > 0.6$, $\frac{\left(R^2 - R_0{}^2\right)}{R^2} < 0.1$ and $0.85 \le k \le 1.15$

Validations are indispensable to verify the robustness and predictability of the models [37]. The internal validation ($Q_{INT}{}^2$) employed leave-one-out (LOO) cross-validation method as well as the external validation ($Q_{EXT}{}^2$) were adopted in this study. The models are robust, predictive and acceptable only if $Q_{INT}{}^2 > 0.5$ and $Q_{EXT}{}^2 > 0.5$ [38].

Other common statistical indices, such as standard deviation (SD) and P-value, F-value, were also calculated during the evaluation of the models. t test and colinearity diagnostics were applied to check the significance and interdependency of each independent variables in models.

Applicability domain (APD) is an unnegligible factor that reflects the limitation of reliable predictions [39]. The predictions derived from models should be regarded reliable only if they are within applicability domain. The distance of a test compound to its nearest neighbor (d_{NN}) in the training set was compared to the APD. The predictions were unreliable if their d_{NN} excess APD. The calculation of APD was in accord with the equation below:

$$APD = < d > + Z\sigma \quad (12)$$

< d > and σ was computed based on the following steps: (1) the average of Euclidean distances between all pairs of training data was computed; (2) the set of distances less than the average was remained to form a new set; (3) < d > and σ were computed as the average and standard deviation of all distances in this new set. In this study, Z was set as 0.5 [35].

Y-randomization test is a powerful procedure to check if the QSAR model is robust and statistically significant. In this study, the dependent variable in training set is randomly permuted for 10 rounds, and the new models were generated based on the data of original independent variable and randomly permuted dependent variable. If a new model gave much lower R^2 and Q_{INT}^2 than the original one, then the model had statistically significance and robustness without the risk of chance correlation [35,40]. We calculated the metric ${}^cR_p^2$ by using the following formula [41,42]:

$${}^cR_p^2 = \sqrt{R \times (R^2 - R_r^2)} \quad (13)$$

R_r^2 is the average of R^2s for the randomized models. For an acceptable model, the value of ${}^cR_p^2$ should be more than 0.5.

Similarly, we defined the indicator ${}^cQ_p^2$.

$${}^cQ_p^2 = Q_{INT} \times \sqrt{Q_{INT}^2 - Q_r^2} \quad (14)$$

where Q_r^2 is the average of Q_{INT}^2s for the randomized models, whose value should excess 0.5.

Analysis of the possible unusual observations, namely outliers, high leverage observations, and influential observations, is carried out based on the following statistics. Outliers are identified by studentized residuals (SRE). An observation with a studentized residual that is larger than 3 (in absolute value) is generally deemed an outlier [43]. High leverage observations are identified by leverages (LEV).An observation is regarded as high leverage if the leverage value is 3 times larger than the mean leverage value:

$$LEV > 3\left(\frac{p}{n}\right) \quad (15)$$

Influential observations are identified by Difference in fits (DFFITS). An observation is deemed influential if the absolute value of its DFFITS value excesses a threshold.

$$DFFITS > 2\sqrt{\frac{p+1}{n-p-1}} \quad (16)$$

where n = the number of observations and p = the number of parameters including the intercept.

3. Result and discussion

3.1. XRD and FT-IR spectra of nascent state manganese dioxide

The XRD spectrum of nascent state manganese dioxide is shown in Fig. 1. The pattern of manganese dioxide has peaks (001), (002) and (111) at about 2θ = 2.5°, 25° and 37°, indexed to δMnO_2. According to JCPDS 80–1098, it belongs to C2/m group of monoclinic system with a = 5.15 Å, b = 2.84 Å, c = 7.17 Å. The crystal structure of δMnO_2 is constructed from layers of edge-sharing MnO_6 octahedra containing a specific amount of H_2O molecules and alien cations between them. The spacing between the two layers is about 7.13 Å, larger than the tunnel size of other kinds [44]. Hence δMnO_2 needs more H_2O or other foreign cations to stabilize the structure. On the basis of the XRD pattern, the (001) plane δMnO_2has much higher diffraction intensity than that of (002) and (111) plane, indicating that δ-MnO_2 might expose the (001) plane.

The FT-IR spectrum of nascent state manganese dioxide is shown in Fig. 2. The strongest adsorption band at 582 cm^{-1} is attributed to Mn-O bond. The adsorption band at about 1000 cm^{-1} correspond to surface hydroxyl groups(≡Mn-OH) of δMnO_2. The adsorption band at about 1400 cm^{-1} is due to the flexural vibration of H-O-H bond, the direct proof of the existence of bound moisture on surface. The strong and wide adsorption band at around 3400 cm^{-1} is related to the stretching vibration of H-O-H bond and OH group.

Combined the information obtained from XRD and FT-IR spectra, the nascent state manganese dioxide is in the pattern of δMnO_2, with large surface area and it is formed of layers of MnO_6 octahedra. Many water molecules and hydroxyls existing on the surface can promote the adsorption as it can generate hydrogen bond with organic molecules. It explains why nascent state MnO_2 exhibits much

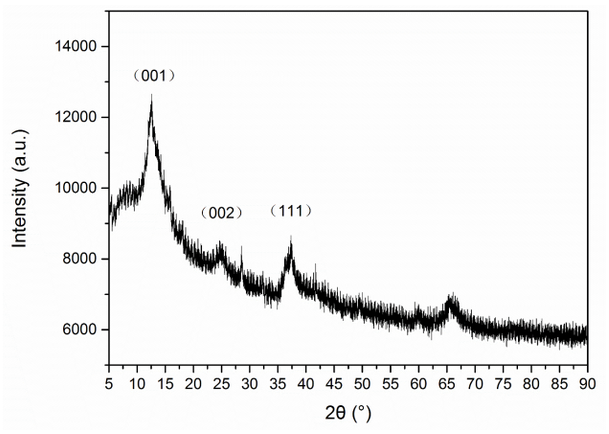

Fig. 1. XRD spectrum of nascent state manganese dioxide.

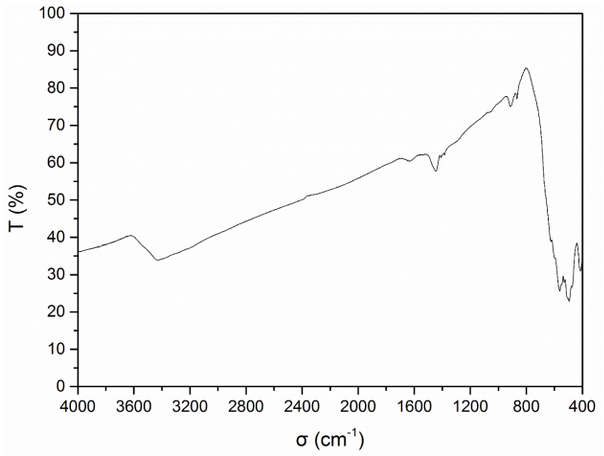

Fig. 2. FT-IR spectrum of nascent state manganese dioxide.

stronger adsorption ability than commercially available MnO_2 does.

3.2. Adsorption experiment results

The experimental removal rate (R_{exp}) of 29 organic contaminants absorbed by δMnO_2 was calculated by the equation below:

$$R_{exp} = (A_0 - A_f)/A_0 * 100\% \quad (17)$$

where A_0 and A_f mean the absorbance of original solution and final filtrate at the maximal absorption wavelength (λ_{max}) of each organic contaminant.

The values of R_{exp} under different pH conditions are listed in Table S1–S3. The averages of R_{exp} at pH = 3, pH = 7 and pH = 10 are 77.29%, 32.96% and 23.58%. It indicates that the experimental removal efficiency is highest under acid condition and lowest under alkaline condition. Therefore, in most cases, the optimum condition for δMnO_2 to treat organic pollution is acid condition. Seen from Table S1–S3, the experimental removal rates of these organic compounds vary a lot. Under acid condition, for Nitrobenzene, R_{exp} is just 17.66%, less than 20%, while for Rhodamine B, R_{exp} is 100.00%, completely removed. However, the removal rates of Rhodamine B decrease to 51.35% and 7.06% with the pH value changed to 7 and 10. Accordingly, δMnO_2 is not efficient for treatment of all organic compounds. R_{exp} of most organic compounds have obvious decrease tendencies when pH value increases, which is shown in Fig. S1. While, some organic compounds display "V" shape curve, lowest at neutral pH surroundings, such as Acid chrome blue K, Bromocresol green and Cresol red.

3.3. Quantum descriptors computation results

18 quantum descriptors we mentioned in Section 2.2 were used as structure descriptors of organic compounds. They were computed by Gaussian 09 and Material Studio 6.1, which are shown in Table 2. These theoretical descriptors are of great significance to observe which factors are essential to affect the adsorption by δMnO_2.

3.4. Construction, validation and selection of QSAR models

29 organic contaminants were separated into training set for model construction and internal validation, and test set for external validation. According to the previous researches by us, an optimum QSAR model included three or four independent variables in general [25,30,32]. Topliss and P. [45] considered that the proportion of the amount of training contaminants to the amount of independent variables ought to be over 5:1. The training set thus should consist of more than 20 contaminants. However, we had to divide enough contaminants into test set for external validation. Therefore, we divided 23 contaminants into training set and 6 contaminants into test set. Since R_{exp} of investigated contaminants varied a lot, especially under acid condition, the training and test contaminants ought to cover wide enough scope. This study did not investigate massive organic contaminants, so the randomized grouping of organic contaminants was likely to cause the two data sets to scatter in a narrow scope. On this account, we did not employ randomization method. Under each pH condition, the contaminants were ranked from small to big according to the values of R_{exp}. We picked out the contaminants placed at 3, 8, 13, 18, 23 and 28. The test set compounds at pH = 3 is marked by [a] before their number and marked by [b] and [c] at pH = 7 and pH = 10.

The MLR analysis of training set performed by SPSS output a series of QSAR models at each pH. We selected those models with the amount of variables from 2 to 5. Those chosen models as well as the corresponding statistical indices such as R^2, R_{EXT}^2, k, R_0^2 are listed in Table 3 and Table 4. We adopted model validation to assess if the models have fine stability and concise predictability. The criteria of acceptable models are $R^2 > 0.6$, $R_{EXT}^2 > 0.5$, 0.85k, $(R^2R_0^2)/R^2 < 0.1$ $Q_{INT}^2 > 0.5$, $Q_{EXT}^2 > 0.5$, relative small SD, $P < 0.05$ and large enough F (Define N is the number of variables in the model. If N = 2, F > 3.49; N = 3, F > 3.13; N = 4, F > 2.93; N = 5, F > 2.81.)

The predictions for removal rates (R_{pre}) of both training contaminants and test contaminants and the differences between R_{pre} and R_{exp} (Diff.) are listed in Table S1–S3. For a more intuitive understanding, the plots of R_{pre} vs. R_{exp} are displayed in Fig. 3–5. The nearer the points to the regression curve, the more precise the prediction is.

The predefined applicability domain (APD) threshold and the distances of test contaminants to their nearest neighbors (d_{NN}) in the training set are displayed in Table 5. All the values of d_{NN} should not be larger than APD if the predictions by models have good reliability.

The outcomes of Y-randomization test with 10 iterations of shuffles of dependent variable (R_{exp}) are displayed in Table 6. To be an acceptable model, the series of new R^2 and Q^2 should be lower than the original one. The averages of new R^2/Q^2 and ${}^cR_p^2/{}^cQ_p^2$ are also available in Table 6. The definitions of ${}^cR_p^2/{}^cQ_p^2$ were in Section 2.4., and the values of ${}^cR_p^2/{}^cQ_p^2$ should be over 0.5 for an acceptable model.

T test along with collinearity diagnostics were performed by SPSS. The results of t test and collinearity diagnostics are available in Table S4–S6. The threshold of t test

Table 2
18 quantum parameters of 29 organic molecules

No.	$E_{(B3LYP)}$ (a. u)	μ (Debye)	qH^+ (e)	$q(CH^+)_n$ (e)	$q(CH^+)_x$ (e)	$q(C^-)_n$ (e)	$q(C^-)_x$ (e)	E_{LUMO} (eV)	E_{HOMO} (eV)	E_{GAP} (eV)	BO_n -	BO_x -	$f(+)_n$ (e)	$f(+)_x$ (e)	$f(-)_n$ (e)	$f(-)_x$ (e)	$f(0)_n$ (e)	$f(0)_x$ (e)
1[c]	–571.741	3.198	0.207	0.207	0.180	–0.236	0.193	–1.661	–6.479	4.818	1.101	1.570	0.022	0.078	0.018	0.050	0.020	0.063
2	–590.560	3.933	0.489	0.226	0.205	–0.198	0.437	–2.885	–6.152	3.267	1.169	1.547	0.020	0.068	0.013	0.052	0.021	0.060
3	–1226.790	1.141	0.476	0.235	0.223	–0.243	0.306	–1.116	–6.615	5.499	1.259	1.411	0.041	0.119	0.049	0.079	0.060	0.092
4[bc]	–752.152	7.859	0.419	0.247	0.280	–0.148	0.485	–3.242	–7.122	3.880	0.954	1.607	–0.005	0.071	0.013	0.105	0.001	0.068
5[c]	–1206.920	5.034	0.381	0.219	0.204	–0.279	0.201	–0.735	–5.853	5.118	1.272	1.425	0.035	0.113	0.038	0.081	0.050	0.090
6[c]	–3309.310	14.878	0.494	0.199	0.244	–0.285	0.386	–0.375	–6.374	5.999	0.899	1.452	–0.009	0.078	–0.009	0.130	0.001	0.096
7	–1587.500	8.051	0.474	0.232	0.201	–0.248	0.341	–2.423	–5.662	3.239	1.162	1.522	–0.023	0.042	–0.007	0.041	–0.015	0.041
8[b]	–1597.480	14.245	0.384	0.226	0.166	–0.296	0.477	–0.436	–7.595	7.159	1.045	1.579	0.014	0.034	0.012	0.033	0.012	0.033
9[a]	–937.930	8.116	0.383	0.205	0.166	–0.591	0.203	–2.368	–4.982	2.613	1.006	1.620	0.006	0.073	0.008	0.046	0.008	0.056
10[ab]	–11879.780	6.139	0.477	0.249	0.212	–0.598	0.294	–1.797	–6.696	4.900	0.954	1.419	–0.012	0.071	–0.009	0.035	0.002	0.056
11[a]	–11801.150	7.656	0.479	0.245	0.212	–0.257	0.290	–1.851	–6.860	5.009	0.950	1.415	–0.008	0.095	–0.010	0.039	0.002	0.053
12	–2448.380	15.057	0.491	0.236	0.197	–0.292	0.396	–1.987	–5.335	3.348	1.190	1.443	0.014	0.056	0.003	0.041	0.014	0.048
13	–1225.480	5.981	0.484	0.224	0.197	–0.607	0.470	–2.777	–6.070	3.294	0.987	1.395	0.047	0.120	0.054	0.126	0.068	0.095
14	–1595.455	14.763	0.271	0.271	0.186	–0.383	0.261	–2.749	–4.110	1.361	1.103	1.503	0.002	0.053	0.007	0.050	0.007	0.051
15[ab]	–2020.900	2.915	0.477	0.241	0.201	–0.266	0.355	–3.321	–5.825	2.504	1.098	1.513	–0.010	0.050	–0.002	0.024	–0.002	0.035
16	–1807.770	7.110	0.497	0.248	0.176	–0.271	0.453	–0.245	–7.513	7.268	1.187	1.532	0.001	0.046	0.005	0.039	0.007	0.046
17	–875.880	5.235	0.388	0.219	0.203	–0.257	0.511	–2.695	–5.553	2.858	0.964	1.421	0.013	0.058	0.012	0.050	0.016	0.044
18[c]	–513.200	4.622	0.419	0.219	0.206	–0.254	0.217	–2.858	–6.778	3.920	0.878	1.392	0.026	0.119	0.017	0.076	0.025	0.096
19	–1461.410	13.575	0.239	0.239	0.187	–0.368	0.302	–2.940	–4.709	1.769	1.068	1.514	0.002	0.056	0.003	0.056	0.002	0.056
20	–458.041	2.686	0.462	0.228	0.193	–0.411	0.376	0.272	–6.043	6.315	1.310	1.336	0.087	0.090	0.058	0.098	0.074	0.102
21	–1645.010	3.686	0.382	0.241	0.199	–0.248	0.197	–1.960	–5.335	3.375	1.296	1.468	0.001	0.048	0.004	0.042	0.003	0.042
22[b]	–1485.780	8.801	0.217	0.217	0.136	–0.287	0.254	–0.245	–7.731	7.486	1.312	1.459	0.015	0.046	0.016	0.032	0.015	0.038
23	–895.524	4.269	0.474	0.223	0.186	–0.351	0.211	–0.054	–6.560	6.506	0.977	1.430	0.012	0.056	0.006	0.042	0.009	0.052
24[c]	–1643.410	12.083	0.239	0.239	0.191	–0.366	0.256	–3.457	–4.709	1.252	1.038	1.418	0.009	0.037	0.010	0.037	0.012	0.036
25[a]	–5388.740	16.502	0.286	0.286	0.259	–0.222	0.429	–0.381	–6.370	5.989	0.638	1.619	–0.050	0.056	–0.030	0.057	–0.060	0.057
26[b]	–436.865	4.541	0.238	0.238	0.208	–0.191	0.061	–3.838	–7.513	3.675	1.323	1.390	0.035	0.094	–0.002	0.060	0.016	0.077
27	–2373.120	3.932	0.458	0.262	0.196	–0.265	0.350	–1.933	–5.308	3.375	1.198	1.514	0.010	0.063	0.005	0.047	0.006	0.055
28[a]	–479.670	6.427	0.218	0.218	0.095	–0.351	0.215	–1.279	–5.716	4.437	1.055	1.457	0.018	0.143	0.029	0.080	0.023	0.095
29	–1881.430	8.788	0.482	0.280	0.190	–0.581	0.438	–0.327	–6.234	5.907	0.962	1.530	–0.005	0.063	–0.007	0.038	–0.006	0.062

[a] Samples in the external test set at pH = 3; [b] Samples in the external test set at pH = 7; [c] Samples in the external test set at pH = 10.

Table 3
QSAR models

pH	Model No.	Models
3	1	$R_{pre}=0.900-6.309f(+)_x+0.950q(C^-)_x$
	2	$R_{pre}=1.437-5.627f(+)_x+1.018q(C^-)_x+0.099E_{HOMO}$
	3	$R_{pre}=0.632-4.899f(+)_x+0.941q(C^-)_x+0.106E_{HOMO}+0.560BO_x$
	4	$R_{pre}=0.735-5.404f(+)_x+0.981q(C^-)_x+0.113E_{HOMO}+0.569BO_x-0.007\mu$
7	1	$R_{pre}=-4.152+2.595BO_x+0.595BO_n$
	2	$R_{pre}=-4.153+2.632BO_x+0.688BO_n-0.696q(C^-)_n$
	3	$R_{pre}=-3.470+2.110BO_x+0.574BO_n-0.772q(C^-)_n-2.613f(+)_x$
	4	$R_{pre}=-3.650+2.428BO_x+0.376BO_n-0.636q(C^-)_n-3.407f(+)_x+2.483f(+)_n$
10	1	$R_{pre}=-0.759+0.032\mu+0.682BO_n$
	2	$R_{pre}=-0.671+0.037\mu+0.543BO_n+2.226f(0)_n$
	3	$R_{pre}=-0.057+0.031\mu+0.243BO_n+4.976f(0)_n-3.938f(+)_x$
	4	$R_{pre}=0.191+0.029\mu+0.185BO_n+4.924f(0)_n-4.162f(+)_x-0.817q(CH^+)_n$

Table 4
Statistical indices of QSAR models

pH	Model No.	R^2	R_{EXT}^2	k	R_0^2	$(R^2-R_0^2)/R^2$	Q_{INT}^2	Q_{EXT}^2	SD	P	F
3	1	0.7702	0.5390	1.1225	0.8779	−0.1398	0.6829	0.8795	0.1277	0.0000	33.5204
	2	0.9113	0.7272	1.1272	0.8407	0.0775	0.8771	0.8902	0.0914	0.0000	65.0801
	3	0.9273	0.8461	1.0786	0.9377	−0.0112	0.8826	0.9184	0.0757	0.0000	57.3940
	4	0.9380	0.8182	1.0851	0.9337	0.0046	0.8816	0.9153	0.0719	0.0000	51.4495
7	1	0.6280	0.7499	1.0183	0.9986	−0.5901	0.4481	0.7532	0.1527	0.0001	16.8837
	2	0.7459	0.6273	1.0920	0.9495	−0.2730	0.5767	0.6738	0.1295	0.0000	18.5941
	3	0.8271	0.5872	1.0109	0.9992	−0.2081	0.6243	0.6149	0.1097	0.0000	21.5287
	4	0.8460	0.8045	1.0029	1.0000	−0.1820	0.6398	0.8370	0.1066	0.0000	18.6804
10	1	0.6249	0.5321	1.3427	0.6175	0.0118	0.4623	0.6733	0.1102	0.0001	16.6569
	2	0.7034	0.5807	1.1966	0.8617	−0.2250	0.5601	0.6275	0.1005	0.0000	15.0178
	3	0.9213	0.8200	1.1445	0.8522	0.0750	0.8552	0.9248	0.0532	0.0000	52.6434
	4	0.9398	0.7284	1.3257	0.7739	0.1765	0.8931	0.9199	0.0479	0.0000	53.0586

in this study (23 samples with 0.05 significance) is 2.074. Accordingly, the absolute t value of each variable ought to excess 2.074 and values of Sig. ought to be under 0.05. In addition, we could measure the correlation degree of each independent variable by Variance inflation factor (VIF). When VIF is equal to 1.0, there is no interdependency between each variable; when VIF is in the scope of 1.0 to 5.0, the collinearity exists but is not serious; when VIF is higher than 10.0 (the cut of value), the collinearity is strong enough to require the reconstruction of the model [46].

The standards to select optimum models at different pH are: (1) qualified in all the criteria mentioned above (to be an acceptable model); (2) the predictive ability to be better than other acceptable models if more than one model satisfy the criteria.

Firstly, we compare the models at pH = 3. For Model 1 and Model 2, all the criteria are satisfied, indicating that Model 1 and Model 2 are both acceptable models. However, for Model 3 and Model 4, the t test and significance test of the variables in the models are not passed. The absolute t value for BO_x in Model 3 is only 1.989, smaller than the threshold value 2.074, and its Sig. value is 0.062, larger than 0.05. The absolute t value for μ in Model 4 is only 1.7.24 (<2.074), and its Sig. value is 0.105 (>0.05). Consequently, Model 3 and Model 4 would not to be considered as acceptable models. The optimum model at pH = 3 is supposed to be selected between Model 1 and Model 2. According to the standards we mentioned above, the optimum model ought to have better predictive ability. R^2 of Model 2 is 0.9113, larger than that of Model 1 (0.7702). Apart from that, R_{EXT}^2, Q_{INT}^2, and Q_{EXT}^2 of Model 2 are lower than those of Model 1. Model 2 has smaller SD (0.0914), compared with Model 1 (0.1277). Seen from Fig. 3. (a)(b), the points of Model 2 more closely and uniformly spread along the regression curve than those of Model 1. All these comparisons lead to the conclusion that Model 2 have better predictive ability and is the optimum at pH = 3. Therefore, the optimum QSAR model (pH = 3) is:

$$R_{pre} = 1.437 - 5.627f(+)_x + 1.018q(C^-)_x + 0.099E_{HOMO} (R^2 = 0.9273) \quad (18)$$

Secondly, the discussion of the 4 models under neutral condition is carried out. For Model 1, Q_{INT}^2 is 0.4481,

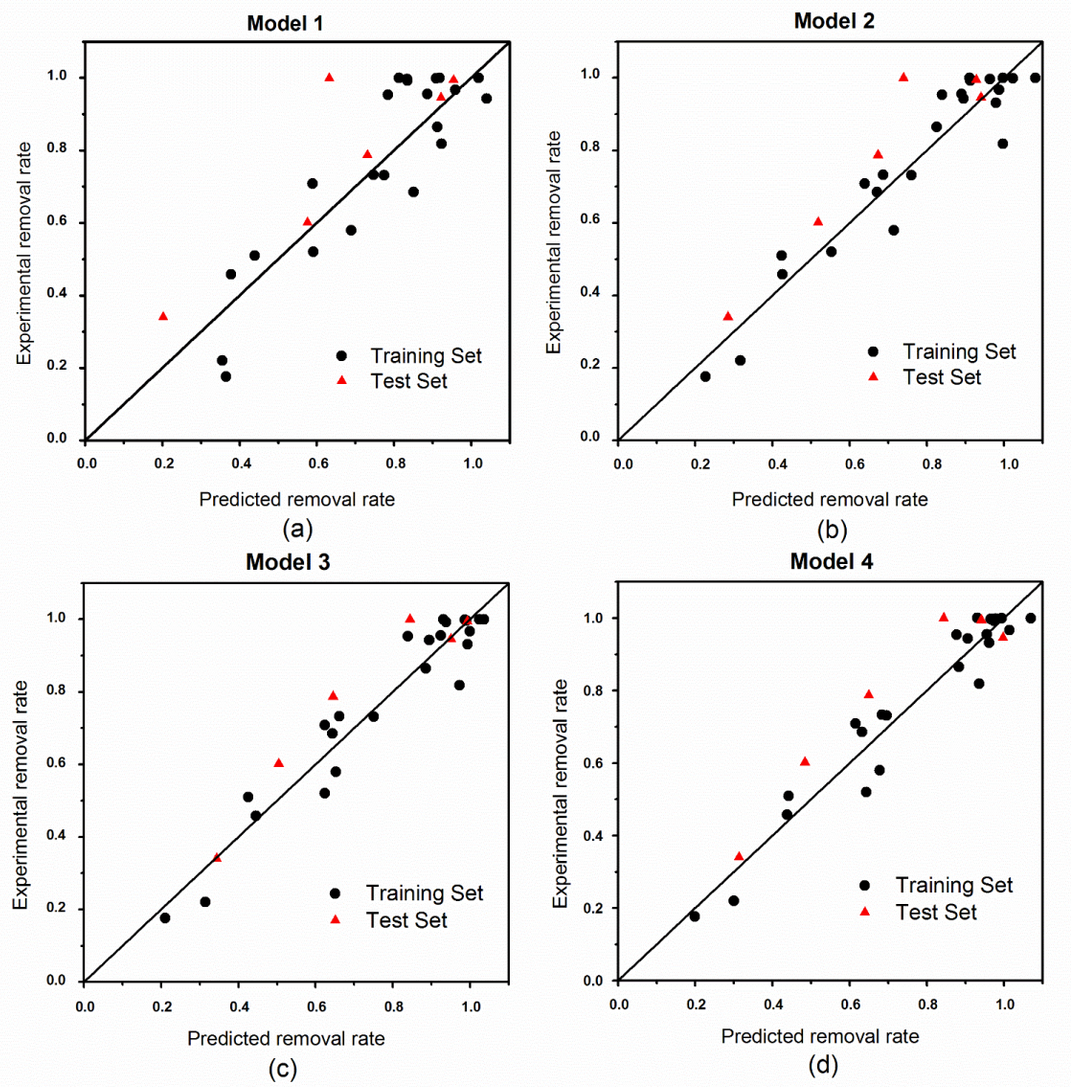

Fig. 3. The plots of predicted removal rate versus experimental removal rate derived from models at pH = 3.

slightly less than 0.5, $^{c}Q_{p}^{2}$ in Y-randomization test is 0.3779, less than 0.5, so Model 1 is excluded. Model 2 along with Model 3 satisfy all the criteria, which makes them acceptable. The applicability domain (APD) threshold of Model 4 is 0.2397, while the d_{NN} of No.8 test compound (AzureI) is 0.2433, which means that the prediction for this test compounds is unreliable. In addition, based on Table S5, the variable $f(+)_n$ does not pass t test and significances test since t = 1.445 (<2.074), Sig. = 0.167 (>0.05). Therefore Model 4 is excluded from acceptable models. Then the choice of the optimum model at pH = 7 is between Model 2 and Model 3. R^2 as well as Q_{INT}^2 of Model 2 is 0.7459 and 0.5767, smaller than those of Model 3 (0.8271, 0.6243). On the other hand, R_{EXT}^2 and Q_{EXT}^2 of Model 2 excess those of Model 3. The SD of Model 2 is larger than that of Model 3. According to Figs. 4b, c, in Model 2, the dots of compounds in training set as a whole are closer to the regression curve than Model 3. Only one compound in the test set is far away from the regression curve both in Model 2 and Model 3. For Model 3 the deviation is larger, which cause the lower value of R_{EXT}^2 and Q_{EXT}^2. The experimental removal rate of this test compound (Bromocresol green) is merely 1.51%, but the predictions derived by Model 2 and Model 3 are 27.53% and 34.78% with the differences as high as 26.02%and 33.27%. It is likely that the experimental result of this compound is not accurate. So it is necessary to repeat the adsorption experimental to check the accuracy of the result. Model 3 performs better in internal validation, while Model 2 performs better in external validation. Synthetically, taking all these factors into account, we choose Model 2 as the optimum QSAR model at pH = 7, though the external validation performance is not good, in need of improvement. The optimum model is:

$$R_{pre} = -4.153 + 2.632BO_x + 0.688BO_n - 0.696q(C^-)_n (R^2 = 0.7459) \quad (19)$$

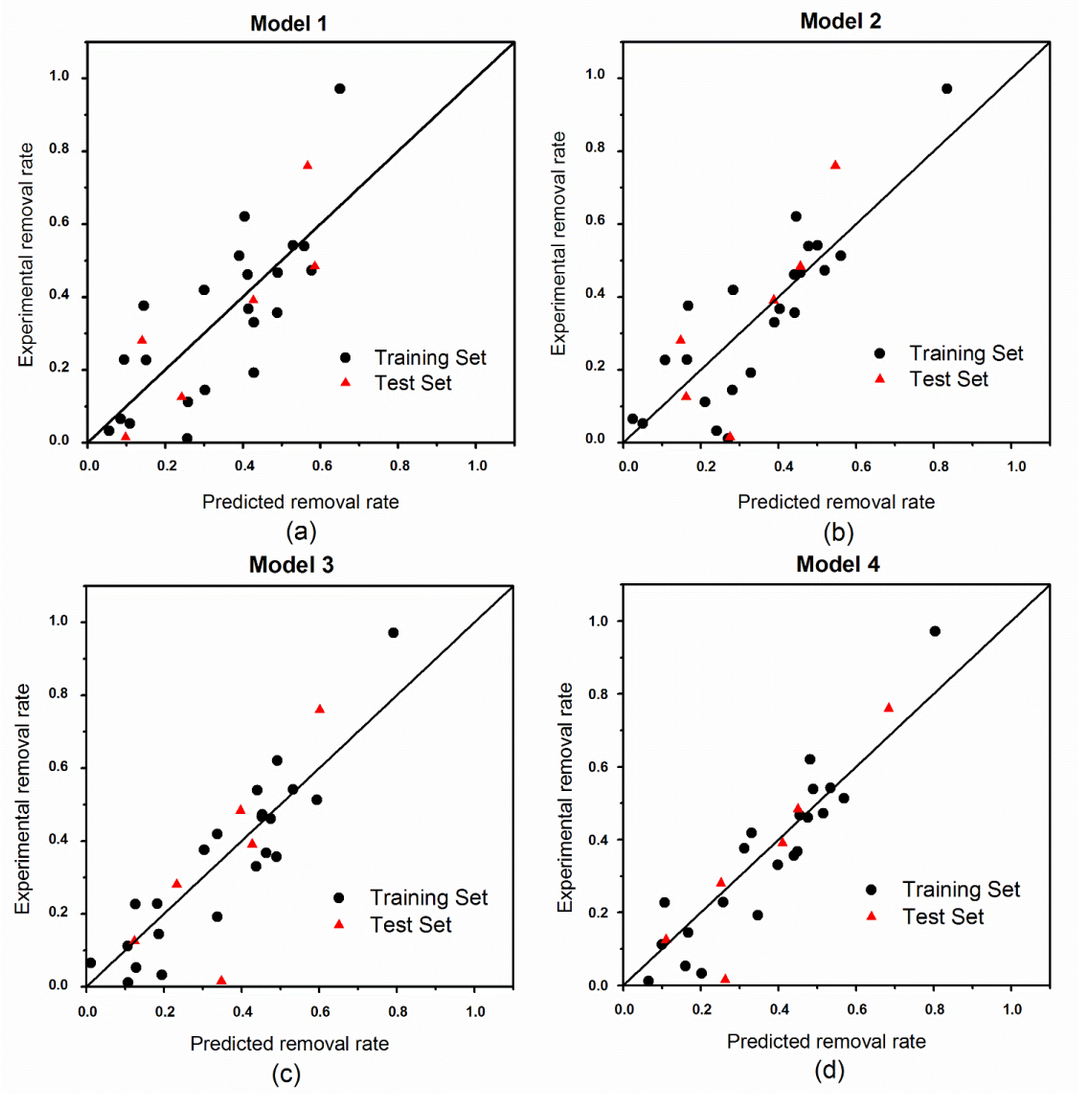

Fig. 4. The plots of predicted removal rate versus experimental removal rate derived from models at pH = 7.

Finally, we analyze the models at pH = 10 to find out the optimum. For Model 1, k = 1.3427 (>1.15), Q_{INT}^2 = 0.4623 (<0.5) and $^cQ_p^2$ = 0.4043 (<0.5), hence Model 1 is not qualified. Model 2 is also not qualified as k value (1.1966) is not within the range from 0.85 to 1.15. For Model 4, k = 1.3257 (>1.15), $(R^2–R_0^2)/R^2$ = 0.1765 (>0.1), and Sig. (BO_n) = 0.053 (>0.05). Therefore Model 4 is not acceptable. Among these 4 models at pH = 10, Model 3 is the only one that satisfies all the criteria. It can be seen from Fig. 5 c that most of the points of compounds in training set and test set distribute along the regression curve within the range from 0.0 to 0.6 The recommended QSAR model at pH = 10 is:

$$R_{pre} = –0.057 + 0.031\mu + 0.243BO_n + 4.976f(0)_n – 3.938f(+)_x (R^2 = 0.9213) \quad (20)$$

The analysis of the possible unusual observations in three optimum models was carried out by SPSS and the results are displayed in Table S7. The threshold for SRE, LEV and DIFFITS are 3, 0.2069 and 0.6794. In the optimum model at pH = 3, there are no outliers and high leverage observations but two influential observations (Chromotropic acid and Isatin). The point stands for Chromotropic acid has rather high SRE value (–2.2828) but still within the threshold. The high SRE value is the main reason that results in the high DIFFITS value. We retain this point but would retest the experimental removal rate in the future study. The point of Isatin has relatively high SRE value (–1.6077) and LEV value (0.1148), and both of them lead to the point to be an influential observation. This point are also kept as it is not far away from the regression line and obeys the trend. AT pH = 7, one high leverage observation (Basic fuchsin) is identified. The high LEV value (0.2410) of Basic fuchsin which excesses the threshold results from the fact that its removal rate is much higher than other points. As the point is not far away from the regression

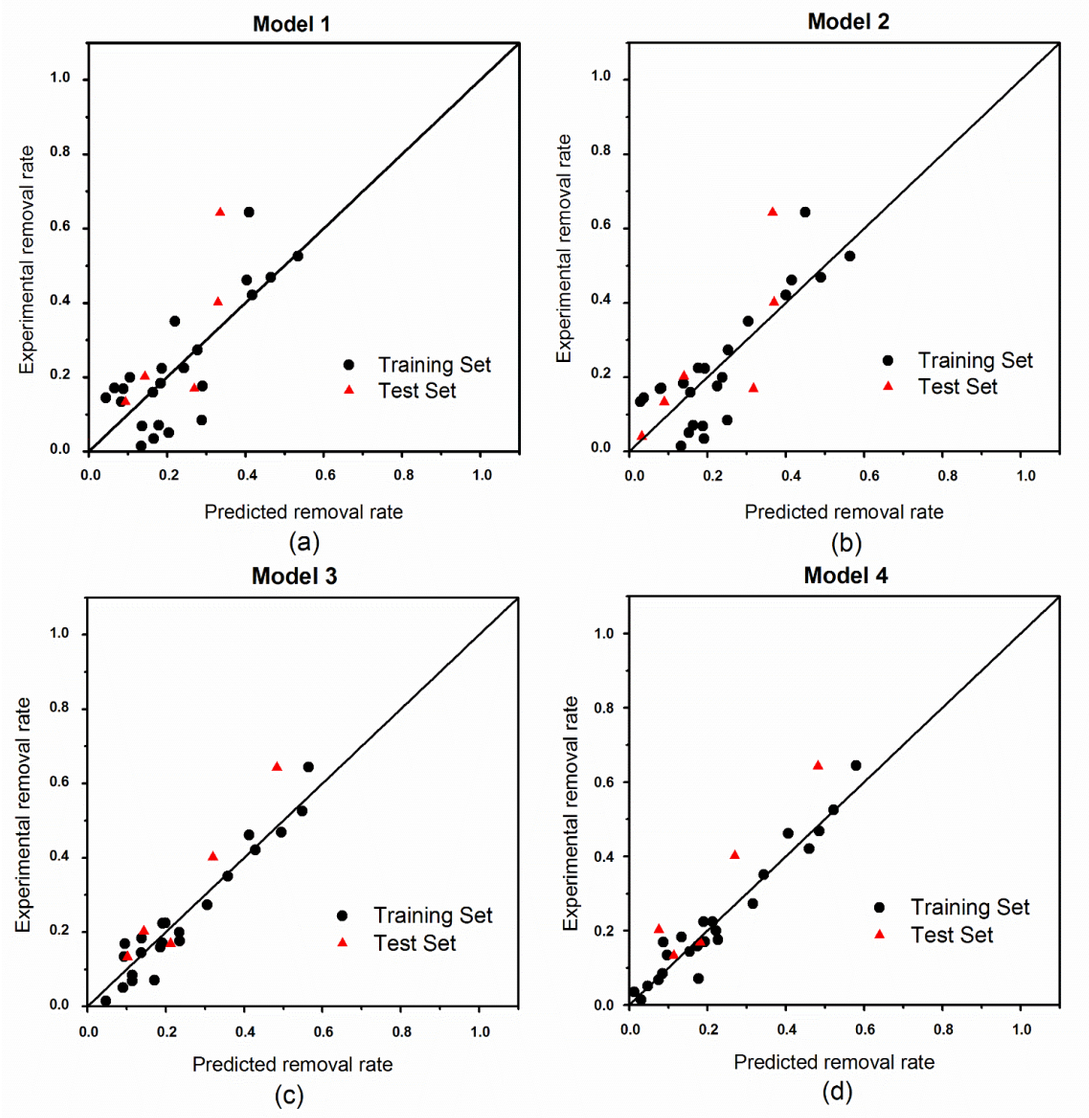

Fig. 5. The plots of predicted removal rate versus experimental removal rate derived from models at pH = 10.

line, it should be kept but rechecked in the future study. At pH = 10, the point of Methylene blue is identified as an influential observation. This point belongs to the test set, not being used to construct the model, so its influential property does not affect the modelling step. However, its SRE value (2.5961) is so close to the threshold that it should be rechecked in the future study. These unusual observations indicate that though the optimum models pass all the tests and validations, they still have some potential limits and problems, in need of further development in the upcoming study.

Fig. S2–S4 shows the distribution of the training set and test set over the quantum parameters which are selected as the independent variables in the optimum models at different pH conditions. From these plots, we can see that the training set and test set scatter over the whole range of the considered space.

3.5 Analysis and discussion of optimum models

QSAR models build up the relationship between the molecule structures in micro scale with physical or chemical properties in macro scale. Apart from making predictions for properties of unknown materials, QSAR models are able to explain the microcosmic mechanisms of macrocosmic activities.

According to Section 3.4, optimum model at pH = 3 contains three variables ($f(+)_x$, $q(C^-)_x$, E_{HOMO}). optimum model at pH = 7 contains three variables (BO_x, BO_n, $q(C^-)_n$). optimum model at pH = 10 contains four variables (μ, BO_n, $f(0)_n$, $f(+)_x$). These variables are bound up with adsorption mechanisms by nascent state MnO_2.

Two optimum models have one common variable: $f(+)_x$. So $f(+)_x$ is the vital variables in the adsorption process by δMnO_2. $f(+)_x$ is a fukui index quantifying the affinity of nuc-

Table 5
The distances of test compounds to their nearest neighbors (d_{NN}) in the training set and predefined threshold (APD)

pH	Model No.	APD	No. of test compounds					
			9	10	11	15	25	28
3	1	0.2563	0.1588	0.0563	0.0830	0.0369	0.0001	0.1949
	2	0.2483	0.1096	0.0556	0.0887	0.0417	0.0000	0.1244
	3	0.2337	0.0462	0.0810	0.1212	0.0283	0.0000	0.1238
	4	0.2273	0.0569	0.0640	0.1016	0.0272	0.0001	0.1209
pH	Model No.	APD	No. of test compounds					
			4	8	10	15	22	26
7	1	0.2534	0.0001	0.2207	0.0391	0.0266	0.0546	0.0213
	2	0.2527	0.0000	0.1720	0.0057	0.0280	0.0553	0.0506
	3	0.2369	0.0006	0.1773	0.1538	0.0611	0.0724	0.0222
	4	0.2397	0.0001	0.2433	0.0631	0.0449	0.0529	0.0170
pH	Model No.	APD	No. of test compounds					
			1	4	5	6	18	24
10	1	0.1702	0.0112	0.0376	0.0225	0.0017	0.1183	0.0740
	2	0.1633	0.0376	0.0187	0.0930	0.0006	0.0945	0.0838
	3	0.1621	0.0088	0.0195	0.0239	0.0019	0.0075	0.0806
	4	0.1672	0.0178	0.0347	0.0113	0.0037	0.0662	0.0969

APD applicability domain.

leophilic attack. Under normal circumstances a molecule is more attractive to nucleophilic reagent if the value of $f(+)_x$ tends to be larger. The regression coefficients of $f(+)_x$ are negative, indicating that removal rate will decline with the rise of $f(+)_x$. It can be concluded that the affinity of an organic compound to nucleophilic reagent impedes the adsorption process. This may be because in the solution H_2O is a stronger nucleophilic reagent than δMnO_2. So the compounds with higher values of $f(+)_x$ tend to have stronger affinity to water molecules than δMnO_2 and is less possible to be absorbed onto the surface of manganese dioxide.

$q(C^-)_x$ and $q(C^-)_n$ is defined as the maximum and minimum negative partial charge on carbon atom. This pair of quantum descriptors appear in the optimum models at pH = 3 and at pH = 7 respectively. The regression coefficient of $q(C^-)_x$ is positive, but that of $q(C^-)_n$ is negative. Several previous studies pointed out that hydrous metal oxides were of pH-dependent surface charge [11], and the point of zero charge (PZC) of MnO_2 was in the range of 2.5 to 4.8 [47,48]. The surface charge of MnO_2 is positive charged when pH is smaller than PZC, and the surface charge is negative when pH is larger than PZC. It is possible that at pH = 3, the surface of δMnO_2 is positively charged and at pH = 7 and 10, the surfaces are negatively charged. The surface reactions of δMnO_2 may be as follows:

At pH = 3,

$$MnO\text{-}(OH)_2(H_2O)_n + H^+ \rightarrow [Mn\text{-}(OH)_3]^+(H_2O)_n \quad (21)$$

At pH = 7 and 10,

$$MnO\text{-}(OH)_2(H_2O)_n + OH^- \rightarrow [MnO_2\text{-}(OH)]^-(H_2O)_n + H_2O \quad (22)$$

$$MnO\text{-}(OH)_2(H_2O)_n + 2OH^- \rightarrow [MnO_3]^{2-} + 2H_2O \quad (23)$$

Therefore, under acid condition, the electrostatic attraction between δMnO_2, with positively charged surface, and an organic compound become stronger if $q(C^-)_x$ increase since there is more negative partial distributed on a carbon atom. That's why the coefficient of $q(C^-)_x$ is positive at pH = 3. On the contrary, under neural condition, the surface of δMnO_2 is negative charged and δMnO_2 is more repulsive to the organic molecule if $q(C^-)_n$ is larger. So the removal rate is negative related to $q(C^-)_n$. Accordingly, removal rate is negatively correlated with $q(C^-)_n$ at pH = 7. From Eqs. (21)–(23), we know that the increase of pH results in a decrease in the number of hydroxyls release from the surface, a decline of adsorption capacity by nascent state manganese dioxide [11]. That explains why the removal rates of most organic compounds are highest at pH = 3 and lowest at pH = 10.

The third variable in optimum model at pH = 3 is $E_{HOMO.}$ E_{HOMO} stands for energy of highest occupied molecular orbital reflecting the ability of a molecule to donate electrons [49]. At pH = 3, the adsorbent, having positively charged surface, is a good electron acceptor. The adsorption will be promoted if E_{HOMO} rises because the molecule is more easy to donate electrons.

The first two variables, BO_x and BO_n, in the optimum model at pH = 7 both reflect the strength of a chemical bond. BO_n is also the second important variables at pH = 10. When BO is under 4, the stability of a molecule will improve with the increase of BO. According to Table 2, all the values of BO_x and BO_n are less than 4. Since both BO_x and BO_n are positively related to removal rate, the improvement of stability is beneficial to the adsorption under neutral and alkaline condition. According to Table 3, BO_x is also included in Model 3 and Model 4 at pH = 3, having positive regression coefficients, which indicates that the speculation above is also established under acid condition. But the influence of

Table 6
Y-randomization test of QSAR models

pH	Iteration	Model 1		Model 2		Model 3		Model 4	
		R^2	Q_{INT}^2	R^2	Q_{INT}^2	R^2	Q_{INT}^2	R^2	Q_{INT}^2
3	Original	0.7702	0.6820	0.9113	0.8771	0.9273	0.8826	0.9380	0.8816
	1	0.2336	0.0156	0.2621	0.0111	0.2634	0.0040	0.2859	0.0028
	2	0.0270	0.1267	0.0518	0.0979	0.0627	0.1294	0.0677	0.2323
	3	0.0857	0.0158	0.1289	0.0270	0.1329	0.0650	0.2162	0.0113
	4	0.0896	0.0729	0.2150	0.0004	0.2193	0.0023	0.3315	0.0107
	5	0.1772	0.0252	0.1788	0.0000	0.2520	0.0044	0.2523	0.0005
	6	0.2122	0.0113	0.2989	0.0473	0.3257	0.0484	0.3524	0.0400
	7	0.0340	0.1740	0.0436	0.1605	0.0500	0.2351	0.0530	0.3263
	8	0.1284	0.0001	0.1444	0.0196	0.1532	0.0522	0.1998	0.0438
	9	0.0886	0.0658	0.1239	0.0774	0.1331	0.0900	0.1371	0.1225
	10	0.2325	0.0294	0.2435	0.0081	0.2468	0.0011	0.2502	0.0008
	Average	0.1309	0.0537	0.1691	0.0449	0.1839	0.0632	0.2146	0.0791
	$^cR_p^2/{}^cQ_p^2$	0.7017	0.6546	0.8224	0.8543	0.8303	0.8504	0.8237	0.8411
7	Original	0.6280	0.4481	0.7459	0.5767	0.8271	0.6243	0.8460	0.6398
	1	0.0489	0.0394	0.0495	0.0473	0.0853	0.0627	0.0870	0.1142
	2	0.0266	0.1054	0.0266	0.2491	0.1632	0.0035	0.3158	0.0234
	3	0.0130	0.1917	0.0325	0.1806	0.0379	0.2509	0.1542	0.0337
	4	0.0055	0.2786	0.0528	0.0838	0.3112	0.0517	0.3113	0.0123
	5	0.0127	0.2931	0.0195	0.3481	0.1737	0.0605	0.1970	0.0790
	6	0.3239	0.0953	0.4645	0.2182	0.2909	0.2079	0.5090	0.1928
	7	0.2240	0.0437	0.2269	0.0099	0.2386	0.0000	0.2649	0.0011
	8	0.0414	0.1080	0.1580	0.0000	0.1622	0.0076	0.2675	0.0017
	9	0.0478	0.0968	0.1137	0.0140	0.1138	0.0379	0.1661	0.0352
	10	0.0481	0.0428	0.0493	0.0821	0.0662	0.2943	0.2251	0.0283
	Average	0.0792	0.1295	0.1193	0.1233	0.1643	0.0977	0.2498	0.0522
	$^cR_p^2/{}^cQ_p^2$	0.5871	0.3779	0.6836	0.5113	0.7404	0.5734	0.7102	0.6132
10	Original	0.6249	0.4623	0.7034	0.5601	0.9213	0.8552	0.9398	0.8931
	1	0.1242	0.0007	0.1683	0.0062	0.1684	0.0000	0.2215	0.0039
	2	0.0853	0.0108	0.1034	0.0375	0.1128	0.0745	0.1770	0.0460
	3	0.0505	0.0792	0.1120	0.1411	0.1178	0.1650	0.1266	0.2197
	4	0.0668	0.0501	0.0013	0.0407	0.1657	0.0000	0.1823	0.0001
	5	0.3764	0.2097	0.3998	0.2162	0.4444	0.1888	0.4444	0.0402
	6	0.0237	0.3338	0.1187	0.0032	0.2623	0.0197	0.2634	0.0547
	7	0.0223	0.0652	0.1458	0.0704	0.1636	0.0786	0.1999	0.0279
	8	0.0331	0.2131	0.1173	0.0430	0.1598	0.0101	0.2184	0.0001
	9	0.0780	0.0701	0.0784	0.1545	0.0800	0.1921	0.0890	0.1308
	10	0.0313	0.0549	0.0920	0.0129	0.2161	0.0265	0.2197	0.0064
	Average	0.0892	0.1088	0.1337	0.0726	0.1891	0.0755	0.2142	0.0530
	$^cR_p^2/{}^cQ_p^2$	0.5786	0.4043	0.6330	0.5226	0.8213	0.8166	0.8258	0.8662

bond order is less significant compared with the other two conditions.

The key variable at pH = 10 is dipole moment (μ), which used to measure the polarity of a molecule. The larger the value of dipole moment, the stronger the polarity is. The regression coefficient of dipole moment is positive at pH = 10 while is negative at pH = 3 based on Model 4. This maybe because at pH = 3, the polarity of the solvent (H_2O) is stronger than that of adsorbent (δMnO_2), so the increase of dipole moment will cause more organic molecule to stay in the solution rather than be adsorbed on δMnO_2. At pH = 10, the polarity of δMnO_2 excesses water molecule, resulting in the opposite relationship between dipole moment and removal rate.

The last quantum descriptor we will discuss about is $f(0)_n$, the third variable in the optimum at pH = 10. $f(0)_n$ measures the affinity with ·OH radical attack. $f(0)_n$ is positively related with removal rate and it may be because

at pH = 10 some ·OH radical is formed on the surface of δMnO_2. The rise of $f(0)_n$ prompts the organic molecule to be more attractive to the surface with ·OH radical, so the coefficient of $f(0)_n$ is positive.

In total, at pH = 3, with positive charge surface of δMnO_2, the important factors that influence the adsorption are 1) affinity with solvent, 2) charge distribution and 3) ability to donate electrons; at pH = 7, with slightly negative charged surface, the important factors are 1) stability of a molecule, 2) charge distribution; at pH = 10, with negative charged surface, the important factors are 1) polarity of a molecule, 2) stability of a molecule, 3) attraction with ·OH radical and 4) affinity with solvent.

3.6. Application of QSAR models

There are two main applications of QSAR models. One is to predict the removal rate of an organic compound at a certain pH. The other one is to select the best pH condition of one or more organic pollutants based on the prediction of the removal rates at three pH. The basic flow of the applications of QSAR models is:

1. Select a certain kind of organic pollutants;
2. Compute the quantum descriptors of the organic pollutants;
3. Input the quantum descriptors into three QSAR models;
4. Output the predictions of removal rates at different pH;
5. Select the best pH condition of the treatment.

4. Conclusion

The nascent state MnO_2 in our experiment is δMnO_2. According to the XRD and FT-IR spectra, δMnO_2 has large surface area and many water molecules and hydroxyls on the surface.

R_{exp} of 29 organic contaminants were tested, at three pH conditions (pH = 3,7,10), adsorbed by nascent state MnO_2. In total, R_{exp} decreases with the increase of pH values. It is because the increase of pH results in a decrease in the number of hydroxyls released from the surface and a decline of adsorption capacity by MnO_2.

By MLR analysis of R_{exp} and 18 quantum descriptors computed by DFT method, a series of different QSAR models were given at different pH. After validation, three optimum models at different pH were recommended. In the light of the three optimum models, the affinity with solvent, charge distribution and stability of a molecule were main molecular characters influencing removal effect.

QSAR models in this study can be applied to predict the removal rate of an organic compound at a certain pH. They can also be used to select the best treatment pH condition of one or more organic pollutants basing on the prediction of the removal rates at three pH.

In future, the QSAR models should be developed by rechecking the unusual observations, introducing more quantum parameters related to adsorption activity and more different kinds of organic compounds in training and test sets to make the models more accurate, stable and reliable.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Project No. 21537002,No. 41401562), the program for New Century Excellent Talents in Shanghai Jiao Tong University, Young Talents Program of University of Hebei (No.BJ2014041) and National Science Foundation for Post-doctoral Scientists of China (No.2014M561197).

References

[1] S. Dash, S. Patel, B.K. Mishra, Oxidation by permanganate: synthetic and mechanistic aspects, Tetrahedron, 65 (2009) 707–739.
[2] B. Sun, X. Guan, J. Fang, P.G. Tratnyek, Activation of Manganese oxidants with Bisulfite for enhanced oxidation of organic contaminants: the involvement of Mn(III), Environ. Sci. Technol., 49 (2015) 12414–12421.
[3] E.S. Lee, F.W. Schwartz, Characteristics and applications of controlled-release $KMnO_4$ for groundwater remediation, Chemosphere, 66 (2007) 2058–2066.
[4] N. Li, M. Fan, J. Van Leeuwen, B. Saha, H. Yang, C.P. Huang, Oxidation of As(III) by potassium permanganate, J. Environ. Sci., 19 (2007) 783–786.
[5] K.-C. Huang, G.E. Hoag, P. Chheda, B.A. Woody, G.M. Dobbs, Kinetics and mechanism of oxidation of tetrachloroethylene with permanganate, Chemosphere, 46 (2002) 815–825.
[6] T. Rodriguez-Alvarez, R. Rodil, J.B. Quintana, S. Trinanes, R. Cela, Oxidation of non-steroidal anti-inflammatory drugs with aqueous permanganate, Water Res., 47 (2013) 3220–3230.
[7] L. Hu, H.M. Martin, S.T. J., Oxidation kinetics of anitibiotics during water treatment with potassium permanganate, Envrion. Sci. Technol., 44 (2010) 6416–6433.
[8] A. Bruchet, J.P. Duguet, I.H. Suffet, Role of oxidants and disinfectants on the removal, masking and generation of tastes and odours, Rev. Environ. Sci. Biotechnol., 3 (2004) 33–41.
[9] J.J. Chen, H.H. Yeh, The mechanisms of potassium permanganate on algae removal, Water Res., 39 (2005) 4420–4428.
[10] X. Guan, D. He, J. Ma, G. Chen, Application of permanganate in the oxidation of micropollutants: a mini review, Front. Environ. Sci. Engin. China, 4 (2010) 405–413.
[11] S. Mustafa, M.I. Zaman, S. Khan, pH effect on phosphate sorption by crystalline MnO(2), J. Colloid. Interface. Sci., 301 (2006) 370–375.
[12] R. Liu, H. Liu, X. Zhao, J. Qu, R. Zhang, Treatment of dye wastewater with permanganate oxidation and in situ formed manganese dioxides adsorption: cation blue as model pollutant, J. Hazard Mater., 176 (2010) 926–931.
[13] L. Zhang, J. Ma, M. Yu, The microtopography of manganese dioxide formed in situ and its adsorptive properties for organic micropollutants, Solid State Sci., 10 (2008) 148–153.
[14] M. Xu, H. Wang, D. Lei, D. Qu, Y. Zhai, Y. Wang, Removal of Pb(II) from aqueous solution by hydrous manganese dioxide: Adsorption behavior and mechanism, J. Environ. Sci., 25 (2013) 479–486.
[15] F. Qu, Z. Yan, W. Liu, S. Shao, X. Ren, N. Ren, G. Li, H. Liang, Effects of manganese dioxides on the ultrafiltration membrane fouling by algal extracellular organic matter, Sep. and Purif. Technol., 153 (2015) 29–36.
[16] Z. Aksu, Application of biosorption for the removal of organic pollutants: a review, Process Biochem., 40 (2005) 997–1026.
[17] B.D. Blair, J.P. Crago, C.J. Hedman, R.J. Treguer, C. Magruder, L.S. Royer, R.D. Klaper, Evaluation of a model for the removal of pharmaceuticals, personal care products, and hormones from wastewater, Sci. Total Environ., 444 (2013) 515–521.
[18] K.Z. Myint, X.Q. Xie, Recent advances in fragment-based QSAR and multi-dimensional QSAR methods, Int. J. Mol. Sci., 11 (2010) 3846–3866.

[19] C.H. Andrade, K.F. Pasqualoto, E.I. Ferreira, A.J. Hopfinger, 4D-QSAR: perspectives in drug design, Molecules, 15 (2010) 3281–3294.
[20] V. Aruoja, M. Moosus, A. Kahru, M. Sihtmae, U. Maran, Measurement of baseline toxicity and QSAR analysis of 50 non-polar and 58 polar narcotic chemicals for the alga Pseudokirchneriella subcapitata, Chemosphere, 96 (2014) 23–32.
[21] K. Mansouri, V. Consonni, M.K. Durjava, B. Kolar, T. Oberg, R. Todeschini, Assessing bioaccumulation of polybrominated diphenyl ethers for aquatic species by QSAR modeling, Chemosphere, 89 (2012) 433–444.
[22] I.A. Aguayo-Villarreal, V. Hernández-Montoya, N.A. Rangel-Vázquez, M.A. Montes-Morán, Determination of QSAR properties of textile dyes and their adsorption on novel carbonaceous adsorbents, J. Mol. Liq., 196 (2014) 326–333.
[23] Y. Zhang, D. Wei, R. Huang, M. Yang, S. Zhang, X. Dou, D. Wang, V. Vimonses, Binding mechanisms and QSAR modeling of aromatic pollutant biosorption on Penicillium oxalicum biomass, Chem. Eng. J., 166 (2011) 624–630.
[24] A.M. Redding, F.S. Cannon, S.A. Snyder, B.J. Vanderford, A QSAR-like analysis of the adsorption of endocrine disrupting compounds, pharmaceuticals, and personal care products on modified activated carbons, Water Res., 43 (2009) 3849–3861.
[25] H. Zhu, Z. Shen, Q. Tang, W. Ji, L. Jia, Degradation mechanism study of organic pollutants in ozonation process by QSAR analysis, Chem. Eng. J., 255 (2014) 431–436.
[26] H. Kusic, B. Rasulev, D. Leszczynska, J. Leszczynski, N. Koprivanac, Prediction of rate constants for radical degradation of aromatic pollutants in water matrix: a QSAR study, Chemosphere, 75 (2009) 1128–1134.
[27] S. Sudhakaran, G.L. Amy, QSAR models for oxidation of organic micropollutants in water based on ozone and hydroxyl radical rate constants and their chemical classification, Water Res., 47 (2013) 1111–1122.
[28] R. Xiao, T. Ye, Z. Wei, S. Luo, Z. Yang, R. Spinney, Quantitative structure-activity relationship (QSAR) for the oxidation of trace organic contaminants by sulfate radical, Environ. Sci. Technol., 49 (2015) 13394–13402.
[29] P.W. Ayers, R.C. Morrison, R.K. Roy, Variational principles for describing chemical reactions: Condensed reactivity indices, J. Chem. Phys., 116 (2002) 8731.
[30] H. Zhu, W. Guo, Z. Shen, Q. Tang, W. Ji, L. Jia, QSAR models for degradation of organic pollutants in ozonation process under acidic condition, Chemosphere, 119 (2015) 65–71.
[31] R. Balawender, P. Geerlings, Nuclear Fukui function from coupled perturbed Hartree–Fock equations, J. Chem. Phys., 114 (2001) 682.
[32] L. Jia, Z. Shen, W. Guo, Y. Zhang, H. Zhu, W. Ji, M. Fan, QSAR models for oxidative degradation of organic pollutants in the Fenton process, J. Taiwan Inst. Chem. E., 46 (2015) 140–147.
[33] V. Yangali-Quintanilla, A. Sadmani, M. McConville, M. Kennedy, G. Amy, A QSAR model for predicting rejection of emerging contaminants (pharmaceuticals, endocrine disruptors) by nanofiltration membranes, Water Res., 44 (2010) 373–384.
[34] A. Golbrakih, A. Tropsha, Beware of q2 !, J. Mol. Graph. Mod, 20 (2002) 269–276.
[35] G. Melagraki, A. Afantitis, Enalos KNIME nodes: Exploring corrosion inhibition of steel in acidic medium, Chemom. Intell. Lab. Syst., 123 (2013) 9–14.
[36] A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, Mol. Inform., 29 (2010) 476–488.
[37] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci., 26 (2007) 694–701.
[38] B. Ma, H. Chen, M. Xu, T. Hayat, Y. He, J. Xu, Quantitative structure-activity relationship (QSAR) models for polycyclic aromatic hydrocarbons (PAHs) dissipation in rhizosphere based on molecular structure and effect size, Environ. Pollut., 158 (2010) 2773–2777.
[39] G. Melagraki, A. Afantitis, Enalos InSilico Nano platform: an online decision support tool for the design and virtual screening of nanoparticles, RSC Adv., 4 (2014) 50713–50725.
[40] S. Zhang, G. Alexander, O. Scott, T. Alexander, A novel automated lazy learning QSAR (all-QSAR) approach: method development, application and virtual screening of chemical databases using validated all-QSAR models, J. Chem. Inf. Model., 46 (2006) 1984–1995.
[41] I. Mitra, A. Saha, K. Roy, Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants, Mol. Simulat., 36 (2010) 1067–1079.
[42] S. Kar, O. Deeb, K. Roy, Development of classification and regression based QSAR models to predict rodent carcinogenic potency using oral slope factor, Ecotoxicol. Environ. Saf., 82 (2012) 85–95.
[43] R. Lehmann, 3σ-Rule for outlier detection from the viewpoint of geodetic adjustment, J. Surv. Eng., 139 (2013) 157–165.
[44] Y. Dai, J.-h. lI, Y. Peng, X.-f. Tang, Effects of MnO_2 crystal structure and surface property on the NH_3-SCR reaction at low temperature, Acta Phys. -Chim. Sin., 28 (2012) 1771–1776.
[45] J.G. Topliss, E.R. P., Chance factors in studies of quantitative structure-activity relatiionships, J. Med. Chem., 22 (1979) 1238–1244.
[46] T.A. Craney, J.G. Surles, Model-dependent variance inflation factor cutoff values, Qual. Eng., 14 (2002) 391–403.
[47] A.T. Stone, J.J. Morgan, Reduction and dissolution of managese (IV) oxides by organics: 2. Survey of the reactivity of organids, Envrion. Sci. Technol., 18 (1984) 617–624.
[48] X.D. Li, F.W. Schwartz, DNAPL remediation with in situ chemical oxidation using potassium permanganate, J. Contam. Hydrol., 68 (2004) 39–53.
[49] N.O. Eddy, E.E. Ebenso, Adsorption and quantum chemical studies on cloxacillin and halides for the corrosion of mild steel in acidic medium, Int. J. Electrochem. Sci., 5 (2010) 731–750.

Supplementary Information

Table S1
Experimental and predicted removal rate of organic compounds at pH = 3

No.	Organic compound	R_{exp}	Model 1		Model 2		Model 3		Model 4	
			R_{pre}	Diff.	R_{pre}	Diff.	R_{pre}	Diff.	R_{pre}	Diff.
1	1,10-Phenanthroline monohydrate	52.00%	59.12%	7.13%	55.32%	3.32%	62.39%	10.40%	64.37%	12.37%
2	1-Nitroso-2-naphthol	95.55%	88.61%	–6.94%	89.02%	–6.54%	92.43%	–3.12%	95.58%	0.02%
3	2,4-Dichlorophenol	50.95%	43.99%	–6.96%	42.40%	–8.55%	42.60%	–8.35%	44.15%	–6.80%
4	2,4-Dinitrophenylhydrazine	86.52%	91.28%	4.76%	82.61%	–3.91%	88.55%	2.03%	88.36%	1.84%
5	3,4-Dichloroaniline	45.79%	37.80%	–7.99%	42.64%	–3.15%	44.52%	–1.27%	43.78%	–2.01%
6	Acid chrome blue K	73.12%	77.46%	4.34%	76.00%	2.88%	75.07%	1.95%	69.60%	–3.52%
7	Acid orange	96.71%	95.90%	–0.81%	98.73%	2.02%	99.93%	3.22%	101.44%	4.73%
8	Azure I	93.15%	113.86%	20.71%	97.94%	4.79%	99.35%	6.20%	96.17%	3.02%
9[a]	Basic fuchsin	99.98%	63.23%	–36.75%	73.97%	–26.00%	84.46%	–15.52%	84.37%	–15.60%
10[a]	Bromocresol green	78.68%	73.14%	–5.54%	67.38%	–11.30%	64.56%	–14.12%	64.95%	–13.73%
11[a]	Bromophenol blue	60.16%	57.61%	–2.55%	51.85%	–8.31%	50.47%	–9.69%	48.45%	–11.71%
12	Chromotropic acid	81.84%	92.29%	10.45%	99.68%	17.84%	97.28%	15.44%	93.56%	11.72%
13	Cresol red	70.87%	58.94%	–11.93%	63.93%	–6.95%	62.41%	–8.46%	61.55%	–9.32%
14	Crystal violet	99.96%	81.36%	–18.61%	99.75%	–0.21%	102.39%	2.43%	99.40%	–0.56%
15[a]	Eriochrome black T	94.60%	92.18%	–2.42%	94.03%	–0.57%	95.09%	0.49%	99.73%	5.13%
16	Eriochrome blue black R	94.30%	104.01%	9.71%	89.55%	–4.75%	89.44%	–4.85%	90.58%	–3.72%
17	Indigo	99.97%	101.95%	1.98%	108.11%	8.14%	103.58%	3.61%	106.93%	6.95%
18	Isatin	22.00%	35.54%	13.53%	31.73%	9.72%	31.43%	9.42%	30.06%	8.05%
19	Malachite green	99.72%	83.36%	–16.36%	96.31%	–3.41%	99.05%	–0.67%	96.49%	–3.22%
20	m-Cresol purple	57.99%	68.94%	10.95%	71.51%	13.52%	65.25%	7.26%	67.80%	9.81%
21	Metanil yellow	95.39%	78.43%	–16.96%	83.92%	–11.47%	83.88%	–11.52%	87.75%	–7.65%
23	Methyl orange	68.57%	85.11%	16.54%	67.14%	–1.43%	64.32%	–4.25%	63.26%	–5.31%
22	Methyl red	73.28%	74.71%	1.43%	68.72%	–4.56%	66.16%	–7.12%	68.38%	–4.90%
24	Methylene blue	99.84%	90.98%	–8.87%	102.32%	2.48%	98.65%	–1.19%	97.83%	–2.01%
25[a]	Naphthol green B	99.42%	95.42%	–4.00%	92.80%	–6.62%	99.28%	–0.14%	94.11%	–5.31%
26	Nitrobenzene	17.66%	36.49%	18.83%	22.64%	4.98%	21.09%	3.43%	19.90%	2.24%
27	Orange G	99.24%	83.50%	–15.74%	91.33%	–7.92%	93.79%	–5.46%	97.40%	–1.84%
28[a]	p-dimethylaminobenzaldehyde	34.03%	20.21%	–13.82%	28.53%	–5.50%	34.37%	0.34%	31.32%	–2.71%
29	Rhodamine B	100.00%	91.86%	–8.13%	91.12%	–8.87%	93.15%	–6.84%	93.09%	–6.91%

[a] Samples in the external test set at pH = 3.

Table S2
Experimental and predicted removal rate of organic compounds at pH = 7

No.	Materials	R_{exp}	Model 1		Model 2		Model 3		Model 4	
			R_{pre}	Diff.	R_{pre}	Diff.	R_{pre}	Diff.	R_{pre}	Diff.
1	1,10-Phenanthroline monohydrate	47.26%	57.72%	10.46%	51.90%	4.63%	45.31%	–1.96%	51.49%	4.23%
2	1-Nitroso-2-naphthol	53.92%	55.80%	1.88%	47.74%	–6.18%	44.03%	–9.89%	48.96%	–4.97%
3	2,4-Dichlorophenol	11.18%	25.87%	14.69%	21.09%	9.91%	10.65%	–0.53%	10.02%	–1.16%
4[b]	2,4-Dinitrophenylhydrazine	48.35%	58.55%	10.20%	45.66%	–2.69%	39.68%	–8.66%	45.00%	–3.34%
5	3,4-Dichloroaniline	14.46%	30.27%	15.82%	28.15%	13.69%	18.70%	4.24%	16.75%	2.30%
6	Acid chrome blue K	22.72%	15.12%	–7.60%	10.79%	–11.93%	12.62%	–10.10%	10.70%	–12.02%
7	Acid orange	35.63%	48.90%	13.27%	44.17%	8.54%	49.01%	13.38%	43.99%	8.36%
8[b]	Azure I	75.97%	56.73%	–19.25%	54.70%	–21.27%	60.12%	–15.86%	68.39%	–7.58%
9	Basic fuchsin	97.15%	65.05%	–32.10%	83.42%	–13.73%	79.11%	–18.04%	80.37%	–16.78%
10[b]	Bromocresol green	1.51%	9.79%	8.28%	27.53%	26.02%	34.78%	33.27%	26.27%	24.76%
11	Bromophenol blue	6.48%	8.52%	2.04%	2.48%	–4.01%	1.11%	–5.37%	–3.73%	–10.21%
12	Chromotropic acid	41.89%	30.06%	–11.83%	28.31%	–13.58%	33.69%	–8.21%	33.07%	–8.82%

(*Continued*)

Table S2 *(Continued)*

13	Cresol red	3.27%	5.53%	2.26%	24.04%	20.77%	19.50%	16.23%	20.21%	16.94%
14	Crystal violet	62.03%	40.46%	–21.57%	44.63%	–17.40%	49.16%	–12.87%	48.20%	–13.83%
15[b]	Eriochrome black T	39.09%	42.75%	3.66%	38.78%	–0.31%	42.74%	3.65%	41.04%	1.95%
16	Eriochrome blue black R	54.18%	52.98%	–1.20%	50.08%	–4.10%	53.29%	–0.89%	53.41%	–0.77%
17	Indigo	5.26%	10.91%	5.65%	4.99%	–0.27%	12.85%	7.59%	16.08%	10.82%
18	Isatin	0.07%	–1.74%	–1.81%	–8.60%	–8.67%	–14.38%	–14.45%	–11.94%	–12.02%
19	Malachite green	46.09%	41.23%	–4.86%	44.14%	–1.95%	47.53%	1.44%	47.58%	1.49%
20	m-Cresol purple	22.80%	9.44%	–13.36%	16.45%	–6.35%	18.30%	–4.50%	25.72%	2.92%
21	Metanil yellow	33.07%	42.86%	9.79%	38.91%	5.84%	43.74%	10.67%	39.83%	6.76%
23	Methyl orange	36.76%	41.47%	4.71%	40.33%	3.57%	46.29%	9.53%	44.88%	8.12%
22[b]	Methyl red	28.07%	14.02%	–14.05%	14.77%	–13.30%	23.27%	–4.79%	25.16%	–2.90%
24	Methylene blue	37.59%	14.53%	–23.05%	16.73%	–20.86%	30.37%	–7.22%	31.23%	–6.36%
25	Naphthol green B	19.23%	42.89%	23.66%	32.89%	13.66%	33.74%	14.50%	34.71%	15.47%
26[b]	Nitrobenzene	12.54%	24.22%	11.69%	16.22%	3.68%	12.41%	–0.12%	11.05%	–1.49%
27	Orange G	46.68%	48.96%	2.28%	45.66%	–1.03%	45.22%	–1.47%	45.52%	–1.17%
28	p-dimethylaminobenzaldehyde	1.15%	25.66%	24.52%	27.09%	25.94%	10.72%	9.57%	6.50%	5.35%
29	Rhodamine B	51.35%	39.07%	–12.27%	56.10%	4.75%	59.44%	8.09%	56.90%	5.55%

[b] Samples in the external test set at pH = 7.

Table S3
Experimental and predicted removal rate of organic compounds at pH = 10

No.	Materials	r_{exp}	Model 1		Model 2		Model 3		Model 4	
			r_{pre}	Diff.	r_{pre}	Diff.	r_{pre}	Diff.	r_{pre}	Diff.
1[c]	1,10-Phenanthroline monohydrate	13.31%	9.42%	–3.89%	8.97%	–4.34%	0.1020	–3.10%	11.42%	–1.89%
2	1-Nitroso-2-naphthol	15.91%	16.41%	0.50%	15.60%	–0.31%	0.1857	2.66%	17.42%	1.51%
3	2,4-Dichlorophenol	6.79%	13.62%	6.82%	18.84%	12.05%	0.1142	4.63%	7.50%	0.70%
4[c]	2,4-Dinitrophenylhydrazine	20.20%	14.30%	–5.90%	14.00%	–6.20%	0.1438	–5.82%	7.60%	–12.60%
5[c]	3,4-Dichloroaniline	16.92%	26.96%	10.04%	31.73%	14.80%	0.2120	4.27%	18.15%	1.23%
6[c]	Acid chrome blue K	40.13%	33.00%	–7.13%	36.97%	–3.16%	0.3204	–8.09%	26.97%	–13.17%
7	Acid orange	17.58%	29.11%	11.53%	22.45%	4.87%	0.2349	5.91%	22.66%	5.08%
8	Azure I	64.44%	40.95%	–23.49%	45.02%	–19.42%	0.5644	–8.00%	57.94%	–6.50%
9	Basic fuchsin	22.41%	18.68%	–3.73%	19.34%	–3.08%	0.1914	–3.28%	21.24%	–1.17%
10	Bromocresol green	16.90%	8.81%	–8.09%	7.86%	–9.04%	0.0955	–7.35%	8.67%	–8.23%
11	Bromophenol blue	1.41%	13.39%	11.98%	13.26%	11.85%	0.0470	3.30%	3.00%	1.60%
12	Chromotropic acid	52.54%	53.44%	0.90%	56.34%	3.80%	0.5481	2.26%	52.27%	–0.27%
13	Cresol red	19.94%	10.55%	–9.39%	23.76%	3.82%	0.2341	3.46%	22.15%	2.20%
14	Crystal violet	46.82%	46.57%	–0.26%	48.97%	2.15%	0.4948	2.66%	48.51%	1.69%
15	Eriochrome black T	13.43%	8.31%	–5.12%	2.86%	–10.57%	0.0933	–4.10%	9.65%	–3.78%
16	Eriochrome blue black R	27.31%	27.81%	0.49%	25.22%	–2.09%	0.3055	3.24%	31.60%	4.29%
17	Indigo	17.07%	6.60%	–10.47%	8.18%	–8.89%	0.1907	2.01%	19.27%	2.20%
18[c]	Isatin	3.99%	–1.23%	–5.22%	3.24%	–0.75%	–0.0446	–8.45%	–5.30%	–9.29%
19	Malachite green	46.16%	40.38%	–5.78%	41.57%	–4.60%	0.4128	–4.88%	40.63%	–5.54%
20	m-Cresol purple	35.09%	22.04%	–13.05%	30.44%	–4.64%	0.3584	0.75%	34.34%	–0.75%
21	Metanil yellow	22.45%	24.28%	1.83%	17.58%	–4.88%	0.1981	–2.65%	19.01%	–3.45%
23	Methyl orange	42.09%	41.74%	–0.34%	40.04%	–2.04%	0.4281	0.73%	46.02%	3.94%
22	Methyl red	14.40%	4.39%	–10.01%	3.75%	–10.65%	0.1370	–0.70%	15.48%	1.08%
24[c]	Methylene blue	64.25%	33.56%	–30.69%	36.64%	–27.61%	0.4838	–15.87%	48.25%	–16.00%
25	Naphthol green B	5.05%	20.42%	15.37%	15.24%	10.19%	0.0905	4.00%	4.75%	–0.30%
26	Nitrobenzene	8.46%	28.86%	20.40%	25.10%	16.64%	0.1147	3.01%	8.51%	0.05%
27	Orange G	18.34%	18.39%	0.04%	13.84%	–4.51%	0.1378	–4.56%	13.39%	–4.95%
28	p-dimethylaminobenzaldehyde	3.49%	16.62%	13.13%	19.09%	15.60%	–0.0501	–8.50%	1.30%	–2.19%
29	Rhodamine B	7.06%	17.83%	10.77%	16.32%	9.25%	0.1712	10.06%	17.68%	10.62%

[c] Samples in the external test set at pH = 10.

Table S4
Checking statistical data of variables in models at pH = 3

Model No.	Variables	Coefficients	t	Sig.	VIF
1	$f(+)_x$	–6.309	–6.172	0.000	1.038
	$q(C^-)_x$	0.950	4.106	0.001	1.038
2	$f(+)_x$	–5.627	–8.484	0.000	1.075
	$q(C^-)_x$	1.018	6.881	0.000	1.045
	E_{HOMO}	0.099	5.498	0.000	1.039
3	$f(+)_x$	–4.899	–6.832	0.000	1.453
	$q(C^-)_x$	0.941	6.581	0.000	1.128
	E_{HOMO}	0.106	6.198	0.000	1.088
	BO_x	0.560	1.989	0.062	1.535
4	$f(+)_x$	–5.404	–7.280	0.000	1.724
	$q(C^-)_x$	0.981	7.118	0.000	1.162
	E_{HOMO}	0.113	6.742	0.000	1.148
	BO_x	0.569	2.127	0.048	1.536
	μ	–0.007	–1.714	0.105	1.459

VIF: variance inflation factor.

Table S5
Checking statistical data of models at pH = 7

Model No.	Variables	Coefficients	t	Sig.	VIF
1	BO_x	2.595	5.640	0.000	1.083
	BO_n	2.595	2.910	0.009	1.083
2	BO_x	2.632	6.743	0.000	1.085
	BO_n	0.668	3.813	0.001	1.105
	$q(C^-)_n$	–0.696	–2.969	0.008	1.020
3	BO_x	2.110	5.607	0.000	1.404
	BO_n	0.574	3.782	0.001	1.157
	$q(C^-)_n$	–0.772	–3.854	0.001	1.038
	$f(+)_x$	–2.613	–2.907	0.009	1.325
4	BO_x	2.428	5.690	0.000	1.917
	BO_n	0.377	1.862	0.080	2.164
	$q(C^-)_n$	–0.636	–2.941	0.009	1.282
	$f(+)_x$	–3.407	–3.303	0.004	1.851
	$f(–)_n$	2.483	1.445	0.167	3.985

VIF: variance inflation factor.

Table S6
Checking statistical data of models at pH = 10

Model No.	Variables	Coefficients	t	Sig.	VIF
1	μ	0.032	5.390	0.000	1.237
	BO_n	0.682	4.218	0.000	1.237
2	μ	0.037	6.317	0.000	1.440
	BO_n	0.543	3.393	0.003	1.456
	$f(0)_n$	2.226	2.243	0.037	1.605
3	μ	0.031	9.682	0.000	1.545
	BO_n	0.243	2.560	0.020	1.823
	$f(0)_n$	4.976	7.607	0.000	2.488
	$f(+)_x$	–3.938	–7.057	0.000	1.901
4	μ	0.003	9.878	0.000	1.642
	BO_n	0.089	2.079	0.053	1.983
	$f(0)_n$	0.589	8.359	0.000	2.490
	$f(+)_x$	0.512	–8.135	0.000	1.973
	$q(CH^+)_x$	0.357	–2.287	0.035	1.187

VIF: variance inflation factor.

Table S7
The analysis of the possible unusual observations in three optimum models.

No.	Material	pH = 3			pH = 7			pH = 10		
		SRE	LEV	DIFFIT	SRE	LEV	DIFFIT	SRE	LEV	DIFFIT
1[c]	1,10-Phenanthroline monohydrate	–0.6988	0.0240	–0.1725	–0.4576	0.0345	–0.1227	0.3959	0.0195	0.0931
2	1-Nitroso-2-naphthol	0.5640	0.0116	0.1224	0.4233	0.0211	0.1011	–0.6445	0.0020	–0.1240
3	2,4-Dichlorophenol	0.6728	0.0654	0.2218	–0.7654	0.0127	–0.1690	–0.9691	0.0158	–0.2228
4[bc]	2,4-Dinitrophenylhydrazine	0.2364	0.0033	0.0460	0.1516	0.0157	0.0342	0.8526	0.0085	0.1797
5[c]	3,4-Dichloroaniline	0.0233	0.0644	0.0076	–1.0838	0.0020	–0.2115	–0.9358	0.0002	–0.1770
6[c]	Acid chrome blue K	–0.5742	0.0000	–0.1072	1.0223	0.0451	0.3009	1.1874	0.0119	0.2639
7	Acid orange	–0.4102	0.0336	–0.1092	–0.7355	0.0123	–0.1615	–1.2281	0.0001	–0.2349
8[b]	Azure I	–0.7397	0.0314	–0.1948	1.6243	0.0453	0.4940	1.1703	0.1497	0.5601
9[a]	Basic fuchsin	2.7785	0.0001	0.6106	1.0310	0.2410	0.6365	0.3875	0.0014	0.0736
10[ab]	Bromocresol green	1.0495	0.0037	0.2096	–2.0562	0.0025	–0.4308	1.1458	0.0217	0.2811
11[a]	Bromophenol blue	0.6628	0.0331	0.1766	0.4197	0.0858	0.1528	–0.7186	0.0410	–0.2034
12	Chromotropic acid	–2.2828	0.0364	–0.6887	1.0697	0.0018	0.2082	–0.7633	0.1357	–0.3429
13	Cresol red	0.5319	0.0077	0.1102	–1.6334	0.0071	–0.3516	–0.8026	0.0001	–0.1509
14	Crystal violet	–0.1433	0.0366	–0.0389	1.3238	0.0133	0.3009	–0.7963	0.0947	–0.3045
15[ab]	Eriochrome black T	–0.1179	0.0215	–0.0282	–0.0141	0.0034	–0.0028	0.5742	0.0224	0.1392
16	Eriochrome blue black R	0.3577	0.0125	0.0781	0.2503	0.0282	0.0636	–0.7919	0.0085	–0.1666
17	Indigo	–1.1242	0.0658	–0.3772	0.1016	0.0722	0.0344	–0.5311	0.0015	–0.1012
18[c]	Isatin	–1.6077	0.1148	–0.6951	0.8834	0.1602	0.4325	1.4478	0.0941	0.5682
19	Malachite green	0.2239	0.0271	0.0563	0.0977	0.0122	0.0212	0.6017	0.0459	0.1758
20	m-Cresol purple	–1.8260	0.0008	–0.3662	0.5456	0.0249	0.1352	–0.3784	0.0232	–0.0921
21	Metanil yellow	1.1217	0.0046	0.2274	–0.5008	0.0036	–0.0982	0.2752	0.0009	0.0518
22[b]	Methyl orange	–0.1010	0.0040	–0.0198	–0.3264	0.0054	–0.0654	–0.4054	0.0537	–0.1241
23	Methyl red	0.2684	0.0026	0.0517	1.1102	0.0303	0.2934	–0.0410	0.0100	–0.0087
24[c]	Methylene blue	–0.4560	0.0447	–0.1317	1.7039	0.0240	0.4411	2.5961	0.0871	1.0942
25[a]	Naphthol green B	0.5887	0.0188	0.1380	–1.0964	0.0000	–0.2080	–0.8531	0.0234	–0.2103
26[b]	Nitrobenzene	–1.0840	0.1679	–0.5478	–0.2556	0.0256	–0.0635	–0.6846	0.0157	–0.1558
27	Orange G	0.7344	0.0158	0.1675	0.0187	0.0157	0.0042	0.6347	0.0098	0.1351
28[a]	p-dimethylaminobenzaldehyde	0.2709	0.1324	0.1191	–2.0489	0.0030	–0.4317	1.4624	0.0980	0.5843
29	Rhodamine B	0.8472	0.0154	0.1931	–0.4853	0.0512	–0.1465	–1.9316	0.0037	–0.4069
	Threshold	3	0.2069	0.6794	3	0.2069	0.6794	3	0.2069	0.6794

[a]Samples in the external test set at pH = 3
[b]Samples in the external test set at pH = 7
[c]Samples in the external test set at pH = 10

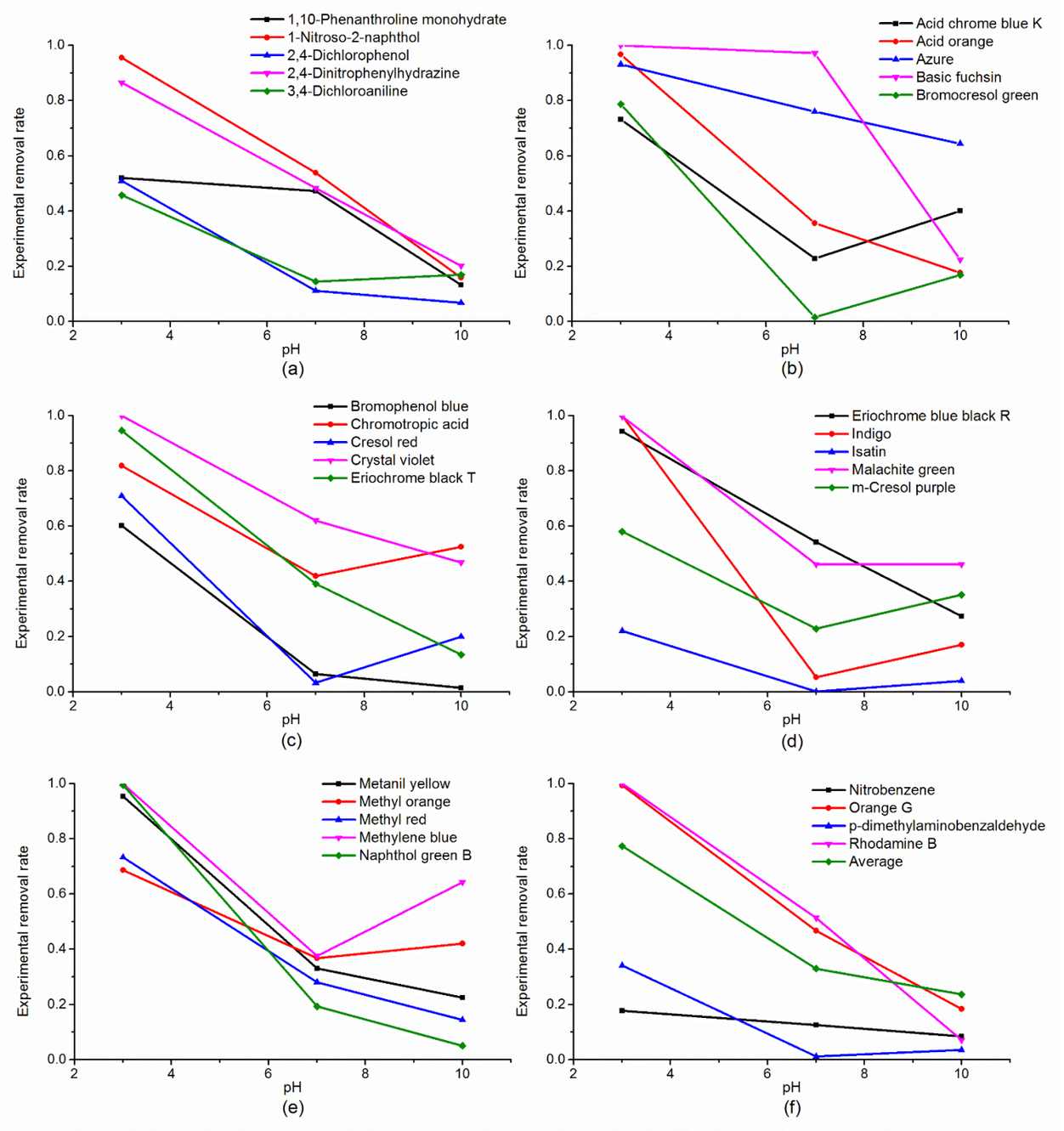

Fig. S1. The experimental removal rate of 29 organic compounds and average under different pH conditions.

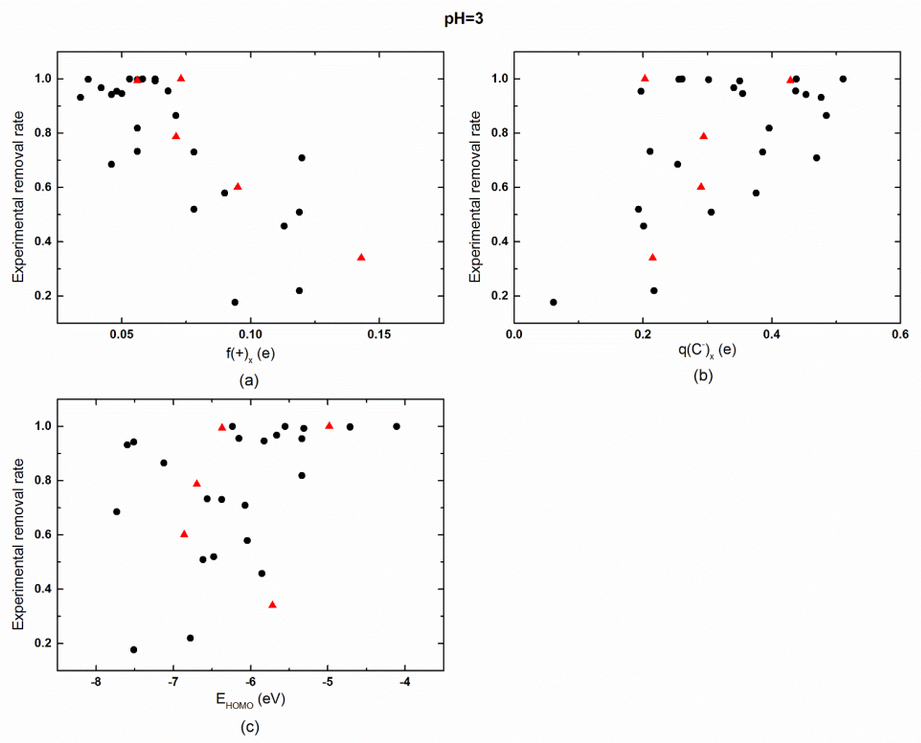

Fig. S2. The distribution of the training set and test set over independent variables in the optimum model at pH = 3. The training set is marked as (●) and the test set is marked as (▲).

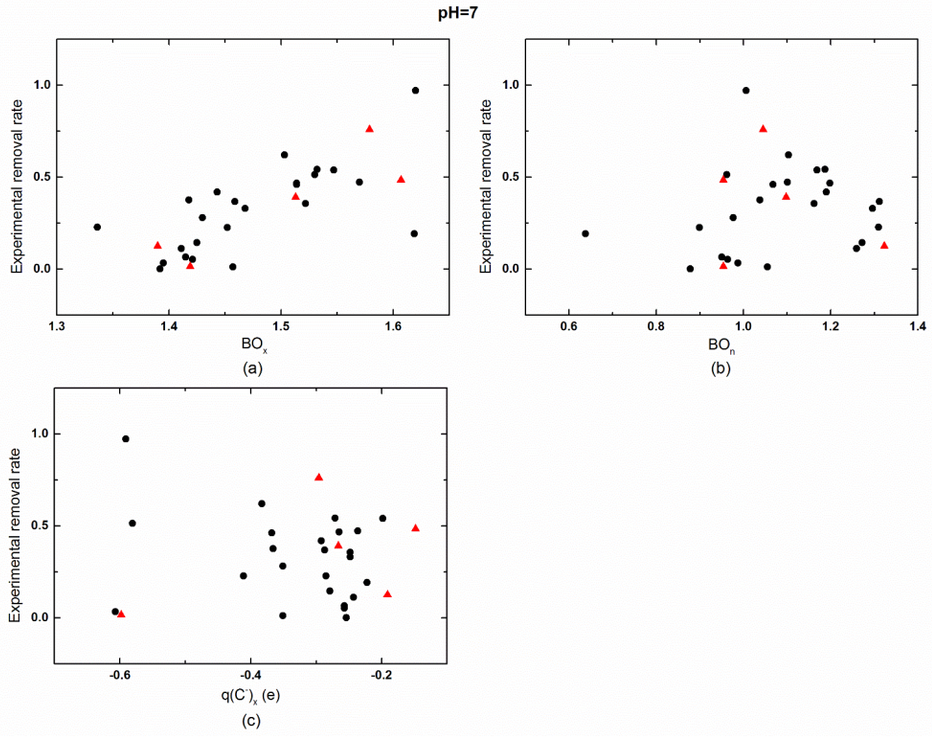

Fig. S3. The distribution of the training set and test set over independent variables in the optimum model at pH = 7. The training set is marked as (●) and the test set is marked as (▲).

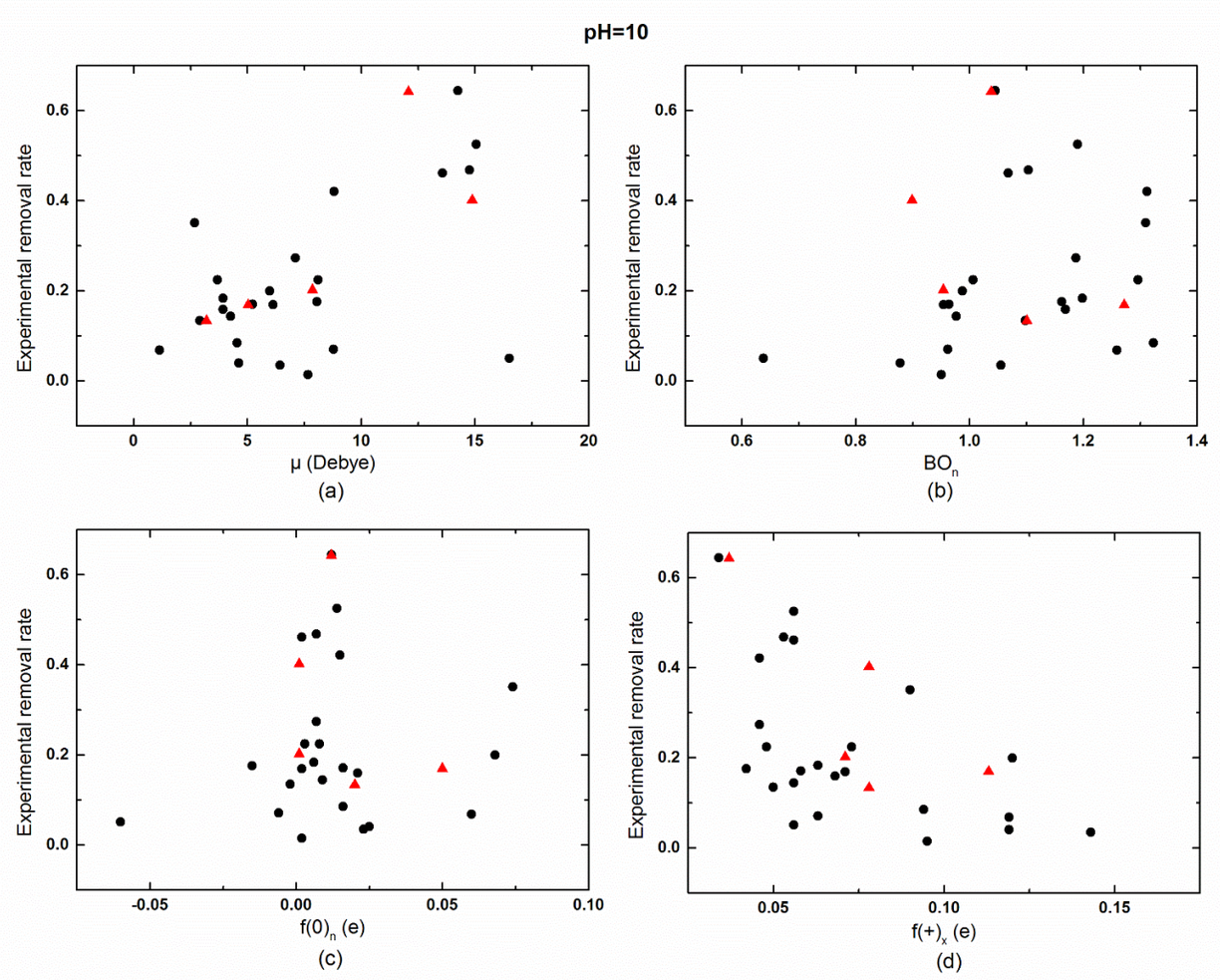

Fig. S4. The distribution of the training set and test set over independent variables in the optimum model at pH = 10. The training set is marked as (●) and the test set is marked as (▲).