

Article

Machine Learning-Driven Prediction of Organic Compound Adsorption onto Microplastics in Freshwater

Ya Wang , Peng Zhao, Honghong Yi and Xiaolong Tang *

School of Energy and Environmental Engineering, University of Science and Technology Beijing, Beijing 100083, China

* Correspondence: yaanne@126.com or wyanne@ustb.edu.cn (Y.W.); txiaolong@126.com (X.T.)

Abstract

Obtaining the adsorption equilibrium coefficient (K_d) of organic compounds on microplastics (MPs) is critical for understanding their environmental behaviors. Given the limited availability of these K_d values, it is imperative to develop predictive models for rapid acquisition of K_d values for different MPs. Herein, seven machine learning-based algorithms, i.e., MLR, RF, GBDT, XGBoost, CatBoost, LightGBM and SVM, were used to establish predictive models on the basis of 173 $\log K_d$ values in freshwater. The evaluation parameters, including R^2_t , $RMSE_t$, Q^2_v , $RMSE_v$ and Q^2 , indicate that the developed models have a satisfactory predictive capability. The developed MLR models can predict the $\log K_d$ values for chlorinated polyethylene (CPE), polybutylene succinate (PBS), polycaprolactone (PCL) and low-density polyethylene (LDPE) MPs. Given the limited performance of MLR in predicting adsorption on PE MPs, RF, GBDT, XGBoost, CatBoost, LightGBM and SVM were employed to develop predictive models, which significantly enhanced the predictive accuracy. The predictive models for PE MPs have a wider AD, covering organic compounds with different functional groups than previous models. Hydrogen bonding, hydrophobic, electrostatic and dispersion interactions may be involved in adsorption. The developed models can serve as efficient tools for estimating the K_d values for different MPs in freshwater, thereby providing the necessary data for evaluating the environmental risks of organic compounds and MPs.

Keywords: adsorption equilibrium coefficients (K_d); microplastics; organic pollutants; machine learning

1. Introduction

Microplastics (MPs) are small plastic fragments with a diameter less than 5 mm [1]. As a class of emerging pollutants, MPs have become ubiquitous in aquatic environments and can exert negative effects on human health and ecological systems [2–4]. Consequently, the environmental fate and ecological risks of MPs have attracted more and more attention [5–8]. Due to their hydrophobicity and large specific surface area, MPs are prone to adsorb organic pollutants in aquatic environments, thereby acting as a vector that can alter the transport and transformation behaviors of both organic pollutants and MPs [9–11]. For instance, the adsorption of organic pollutants such as polycyclic aromatic hydrocarbons (PAHs) onto microplastics can enhance the bioaccessibility of these PAHs in aquatic environments [12], which can also alter their ecological risks. Therefore, it is crucial to investigate the adsorption of organic compounds onto MPs so as to understand the environmental behaviors and ecological risks of both MPs and organic pollutants.



Academic Editor: Dimosthenis Giokas

Received: 30 December 2025

Revised: 27 January 2026

Accepted: 28 January 2026

Published: 1 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Some research, including studies on adsorption isotherms and kinetic experiments, have been conducted to investigate the adsorption of different organic compounds onto various MPs in seawater or freshwater [11,13–16]. The results have revealed that adsorption is promoted by different adsorption mechanisms such as van der Waals, hydrophobic, π - π , hydrogen bonding, and electrostatic interactions. These adsorption mechanisms can be influenced by the structures of both organic compounds and MPs [17,18]. For example, the adsorption of perfluorooctanesulfonate (PFOS) and perfluorooctanesulfonamide (FOSA) on polyethylene (PE), polystyrene (PS) and polyvinyl chloride (PVC) MPs are mainly driven by hydrophobic interactions [19]; the adsorption of bisphenol A onto polyamide MPs can be enhanced through hydrogen bonding [20]. In addition, some simulation methods, including molecular dynamics (MD) and density functional theory (DFT), have also been utilized to investigate the adsorption onto MPs at the molecular level [21,22]. For instance, Chen et al. [23] have explored the adsorption of polychlorinated biphenyls (PCBs) and hydroxy PCBs onto humic acid or MPs through MD and DFT methods. The results indicated that van der Waals interactions and hydrogen bonding promoted the adsorption of PCBs and hydroxy PCBs onto MPs. Su et al. [24] have probed the adsorption of 14 different organic compounds onto MPs with MD and machine learning methods and found that hydrophobic interactions played a dominant role in the adsorption.

However, previous research has primarily focused on marine environments, while research on microplastic adsorption in freshwater is extremely limited. In contrast to seawater, the distinct water chemistry condition of freshwater alters microplastic surface properties, making adsorption data from marine studies not directly transferable and leaving freshwater adsorption behaviors for many organic compounds onto MPs unclarified [25]. Given the time- and cost-intensive nature of investigating the adsorption of organic compounds onto different MPs one by one, it is of great necessity to develop predictive models so as to obtain adsorption data highly efficiently. To date, some predictive models have been developed to estimate the adsorption of organic compounds onto MPs in freshwater environments [6,10,24–31]. In terms of the predictive models which can be applied for adsorption onto chlorinated polyethylene (CPE), polybutylene succinate (PBS), polycaprolactone (PCL) or low-density polyethylene (LDPE) MPs [25,26], the descriptors in these models, such as $\log K_{OW}$, depend on experimental determination. As for organic compounds which lack these experimentally determined molecular descriptor values, these models cannot be used for predicting their adsorption. In addition, for the predictive models which are applicable to the adsorption onto polyethylene (PE) MPs [25,29–31], more efforts should focus on extending their application domains. Therefore, the new models for predicting adsorption onto CPE, PBS, PCL, LDPE and PE MPs should be established.

In this study, we collected the adsorption equilibrium coefficients (K_d) for various organic compounds onto CPE, PBS, PCL, LDPE and PE MPs. Based on the $\log K_d$ values and molecular structural descriptors, we established models for predicting the adsorption onto CPE, PBS, PCL and LDPE MPs through multiple linear regression analysis. Afterwards, we established prediction models for estimating the adsorption onto PE MPs with six machine learning algorithms, including random forest (RF), the gradient boosting decision tree (GBDT), extreme gradient boosting (XGBoost), categorical boosting (CatBoost), the light gradient boosting machine (LightGBM) and the support vector machine (SVM). Furthermore, the predictive performances of these established models were evaluated, and the adsorption mechanisms were also discussed.

2. Materials and Methods

2.1. Experimental Data Collection

The experimental $\log K_d$ values were collected from previous studies [10,14,16,17,25,26,32–45], and a dataset including 173 $\log K_d$ values was obtained, as shown in Table 1. The dataset covers 120 distinct organic compounds and five MPs, including CPE, PBS, PCL, LDPE, and PE MPs. For the CPE MPs, the dataset consists of 13 organic compounds; for the PBS and PCL MPs, the datasets comprise 18 organic compounds respectively; for the LDPE MPs, the dataset is composed of 19 organic compounds; for the PE MPs, the dataset consists of 105 organic compounds, which can be classified into 15 different categories in Table S1 of the Supplementary Materials. All the $\log K_d$ values were determined in freshwater (pH 6.8–7.5) at 298 ± 1 K and 101.3 kPa. The Chemical Abstracts Service (CAS) numbers, $\log K_d$ values, and molecular structure descriptors for these organic compounds are listed in Tables S1 and S2 of the Supplementary Materials.

Table 1. The numbers of $\log K_d$ values for different MPs in freshwater.

MPs	Numbers of $\log K_d$ Values
CPE	13
PBS	18
PCL	18
LDPE	19
PE	105

2.2. Molecular Structure Descriptors

The molecular structure descriptors of organic compounds are composed of Abraham descriptors [46–48] and theoretical molecular structure descriptors. The Abraham descriptors utilized in this study include E , S , A , B and V . E represents the excess molar refraction; S denotes the dipolarity/polarizability; A and B stand for the hydrogen bond donating and accepting abilities respectively; V is McGowan’s molar volume. All the E , S , A , B and V values for the 120 organic compounds were retrieved from the UFZ-LSER database (<https://www.ufz.de/index.php?en=31698>, accessed on 29 January 2025). In addition, on the basis of molecular structures for these organic compounds, 1444 theoretical molecular structure descriptors were calculated using the PaDEL software (v2.21) [49].

Then, we employed a two-stage feature optimization strategy to screen these molecular structural descriptors with the total dataset. We first used mutual information (MI) regression [50] (Equation (1)) to obtain the MI values for all molecular structure descriptors. According to the MI values, the top 200 molecular structure descriptors were retained. Subsequently, a lightweight least absolute shrinkage and selection operator (LASSO)–recursive feature elimination (RFE) model [51] was applied for recursive feature elimination, ultimately reducing the number of molecular structure descriptors to below 20.

$$I(X; Y) = \iint p(x_i, y) \log \frac{p(x_i, y)}{p(x_i)p(y)} dx_i dy \quad (1)$$

where $p(x_i, y)$ represents the joint probability distribution of X and Y and $p(x_i)$ and $p(y)$ denote the corresponding marginal probability distributions of X and Y .

2.3. Development of Predictive Models with Different Machine Learning-Based Algorithms

Z-score normalization was used for all the data. Afterwards, the datasets were randomly split into training sets and validation sets at a ratio of 4:1. For the CPE MPs, the training set consisted of 10 organic compounds and the validation set included 3 organic compounds; for the PBS and PCL MPs, the training sets were composed of 14 organic

compounds and the validation sets consisted of 4 organic compounds; for the LDPE MPs, the training set and validation set included 15 and 4 organic compounds, respectively; for the PE MPs, the training set and validation set were composed of 84 and 21 organic compounds, respectively. The training sets were utilized to develop predictive models and the validation sets were used for external validation.

Multiple linear regression was employed to build the linear predictive models for the adsorption on CPE, PBS, PCL and LDPE MPs. In addition, six different machine learning-based algorithms, i.e., RF, GBDT, XGBoost, CatBoost, LightGBM and SVM, were applied in the development of the nonlinear predictive models for organic compound adsorption onto PE MPs. The RF algorithm [52] is an ensemble learning method based on bagging, noted for its robustness and strong generalization performance. The GBDT algorithm [53] iteratively constructs decision trees so as to reduce errors and improve predictive ability. XGBoost [54], CatBoost [55] and LightGBM [56] are all algorithms derived from the GBDT framework, achieving some improvements, including predictive accuracy, robustness, and processing speed. SVM [57] is an algorithm that utilizes kernel functions to map data into a high-dimensional space, having good robustness and overfitting resistance. More details about the hyperparameter optimization can be found in Table S2.

2.4. Assessment of Predictive Models

The goodness of fit, robustness and predictive ability of the developed predictive models were evaluated with the following metrics: coefficients of determination (R^2), root mean square error (RMSE), leave-one out cross-validation (Q^2_{LOO}), 5-fold cross-validation Q^2 and external explained variance (Q^2_v) (more details can be found in the Supplementary Materials). In addition, the application domains (ADs) of the developed models were characterized using the Euclidean distance-based method and Williams plots [58]. The Euclidean distance-based method was performed through AmbitDiscovery-v0.04 (http://ambit.sourceforge.net/download_ambitdiscovery.html, 29 January 2025). Williams plots were shown on the basis of standardized residuals (δ^*) and leverage values (h). If one compound has a $|\delta^*|$ value larger than 3, it is an outlier. The h values can be calculated with the following equation:

$$h = x^T(X^T X)^{-1}x \quad (2)$$

where x denotes the descriptor vector of the organic compound, x^T is the transpose of x , X is the descriptor matrix and X^T is the transpose of X .

The warning leverage value (h^*) values can be calculated as follows:

$$h^* = 3(k + 1)/n \quad (3)$$

where n is the number of compounds in the training set and k represents the number of descriptors used in the predictive model.

2.5. Parameter Sensitivity Analysis

In terms of the predictive models based on the RF, GBDT, XGBoost, CatBoost, LightGBM and SVM algorithms, the Shapley method [59] was employed for analyzing the importance of molecular structure descriptors. The SHapley Additive exPlanations (SHAP) values were calculated on the basis of the TreeExplainer bridges theory, which can describe the contributions of different molecular structure descriptors to the predictions of models.

3. Results

3.1. $\log K_d$ Values for Organic Compounds on MPs

As shown in Figure 1, the $\log K_d$ (K_d , L/kg) values for the organic compounds on different MPs in freshwater are in the range of 0.42~8.78. In terms of the CPE MPs, the $\log K_d$ values range from 1.51 to 4.69; those for PBS MPs are in the range of 2.04~4.30; the $\log K_d$ values on PCL MPs range from 2.10 to 4.62; for the LDPE MPs, the $\log K_d$ values range from 0.72 to 3.91; as for the PE MPs, these are in the range of 0.42~8.78. It should be noted that the $\log K_d$ value for 1,2,3,4-tetrachlorobenzene is 4.69, which is an outlier. The same compound exhibits different $\log K_d$ values on various microplastics. For example, the $\log K_d$ values for naphthalene on these four microplastics are 3.45 (CPE), 3.25 (PCL), 3.01 (PBS), and 2.59 (LDPE). These results indicate that the Cl or O atoms of the MPs can enhance their adsorption capability for naphthalene. Moreover, the same MPs have shown different adsorption capabilities for distinct organic compounds. The organic compounds with more phenyl rings have larger $\log K_d$ values. For instance, the $\log K_d$ value for biphenyl on PE MPs is 3.25, which is larger than that for benzene (2.19); the $\log K_d$ value for pyrene on PE MPs is 4.70, which is larger than that for anthracene (4.30) and phenanthrene (4.30).

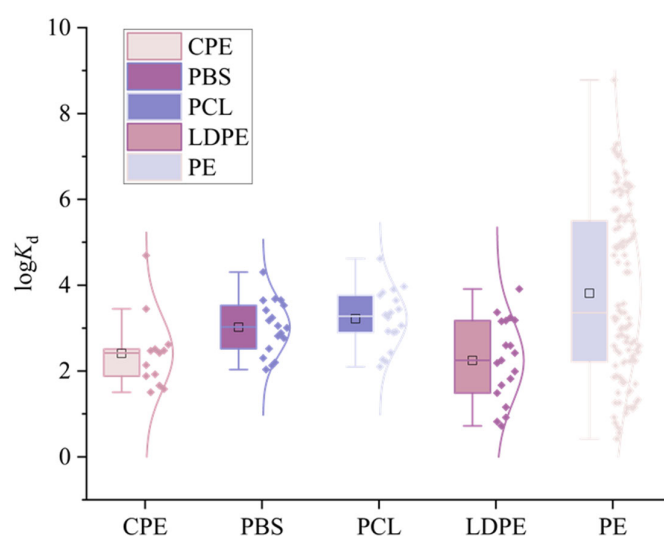


Figure 1. Box and whisker plots for the $\log K_d$ values for CPE, PBS, PCL, LDPE and PE MPs. (The lines of a box represent the lower quartile, median and upper quartile; the lower and upper whiskers represent the minimum and maximum values; the minimum value is defined as the lower quartile $- 1.5 \times (\text{upper quartile} - \text{lower quartile})$; the maximum value is defined as the upper quartile $+ 1.5 \times (\text{upper quartile} - \text{lower quartile})$).

3.2. Establishment and Evaluation of the Predictive Models for CPE, PBS, PCL, LDPE and PE MPs

(1) Linear models for CPE, PBS, PCL and LDPE MPs

The developed MLR models for predicting the adsorption of organic compounds onto CPE, PBS, PCL and LDPE MPs are shown below.

For CPE MPs:

$$\log K_d = -2.804 \times AATSC0i - 0.117 \times gmax + 18.852 \times AATSC6c + 7.700$$

$$n_t = 10, R^2_t = 0.976, RMSE_t = 0.124, n_v = 3, Q^2_v = 0.953, RMSE_v = 0.192, Q^2_{LOO} = 0.926 \quad (4)$$

For PBS MPs:

$$\log K_d = 0.370 \times LipoaffinityIndex + 73.540 \times AATSC3e + 0.0084 \times AATS4m + 0.7378$$

$$n_t = 14, R^2_t = 0.913, RMSE_t = 0.175, n_v = 4, Q^2_v = 0.958, RMSE_v = 0.135, Q^2_{LOO} = 0.836 \quad (5)$$

For PCL MPs:

$$\log K_d = 0.424 \times \text{LipoaffinityIndex} + 81.016 \times \text{AATSC3e} + 0.009 \times \text{AATS4m} + 0.648 \quad (6)$$

$$n_t = 14, R^2_t = 0.931, \text{RMSE}_t = 0.148, n_v = 4, Q^2_v = 0.993, \text{RMSE}_v = 0.077, Q^2_{\text{LOO}} = 0.856$$

For LDPE MPs:

$$\log K_d = 0.398 \times \text{LipoaffinityIndex} - 2.263 \times \text{SHBd} + 4.508 \times \text{ETA_BetaP_ns_d} + 0.461 \quad (7)$$

$$n_t = 15, R^2_t = 0.853, \text{RMSE}_t = 0.332, n_v = 4, Q^2_v = 0.948, \text{RMSE}_v = 0.238, Q^2_{\text{LOO}} = 0.729$$

where n_t is the number of organic compounds in the training set and n_v is the number of organic compounds in the validation set. The values for R^2_t , RMSE_t , Q^2_v , RMSE_v and Q^2_{LOO} were calculated to evaluate the goodness-of-fit, robustness and predictive capability of these four predictive models. Given the limited data available for adsorption on CPE, PBS, PCL and LDPE, leave-one out cross-validation (Q^2_{LOO}) was employed to assess the robustness of these models. All the R^2_t values exceeding 0.60 and Q^2_v/Q^2_{LOO} values above 0.50 [60] indicate that these four MLR predictive models demonstrate good performance. In addition, Figure 2 shows that the predicted $\log K_d$ values from the four linear models agree well with the experimental ones. Therefore, these linear models can be applied for predicting the $\log K_d$ values for different organic compounds on CPE, PBS, PCL and LDPE MPs.

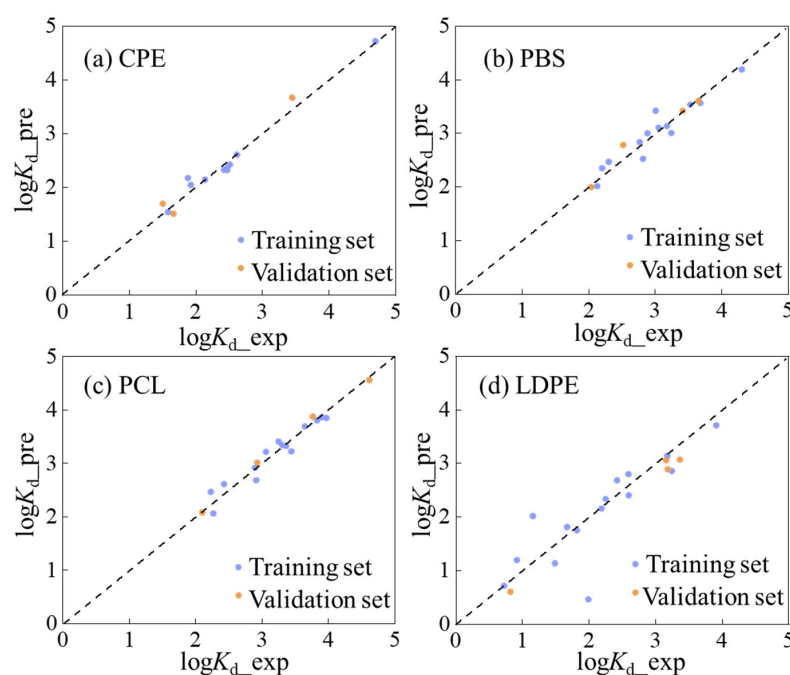


Figure 2. Predicted $\log K_d$ values from linear models ($\log K_{d_pre}$) versus experimental $\log K_d$ values ($\log K_{d_exp}$) for different MPs [(a) CPE, (b) PBS, (c) PCL and (d) LDPE].

(2) Nonlinear models for PE MPs

Due to the wide variety of organic compounds involved in the adsorption data on PE MPs, it is difficult to achieve satisfactory predictive performance using multiple linear regression analysis. Therefore, we employed six machine learning-based nonlinear algorithms, i.e., RF, GBDT, XGBoost, CatBoost, LightGBM and SVM, to construct the predictive models. Note that, prior to the model development, Pearson correlation analysis was conducted on $\log K_d$ values and the 14 molecular structural descriptors obtained through a two-stage feature optimization (Figure 3). As shown in Figure 3, these molecular structure descriptors, including *CrippenLogP*, *minaaCH*, *SpMax2_Bhm*, *AATSC3i*, *SpMax8_Bhi*,

ETA_dBeta, *SpMax6_Bhv*, *V* and *S*, are positively correlated with the $\log K_d$ values, whereas, the others, i.e., *B*, *A*, *SpMin8_Bhe*, *MATS7v* and *SpMax8_Bhp*, exhibit a negative correlation with $\log K_d$ values for PE MPs in freshwater. In addition, the histogram combined with the kernel density estimation (KDE) curve illustrates the distribution of the $\log K_d$ values and the 14 molecular structural descriptors.

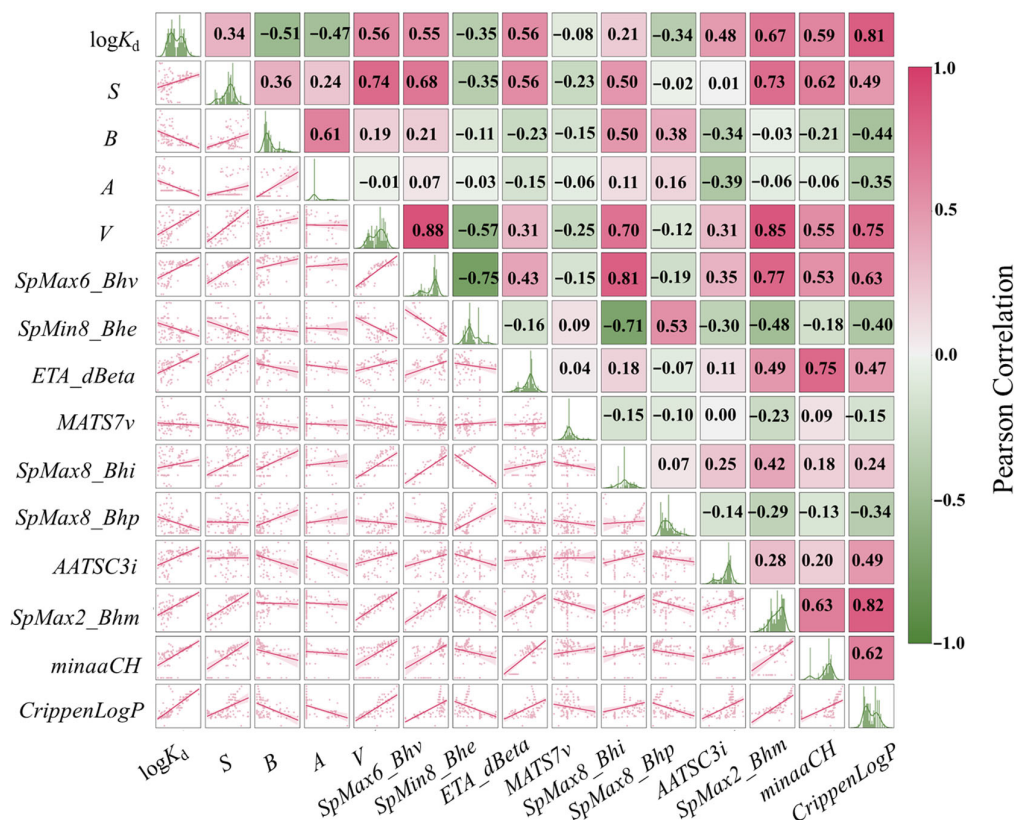


Figure 3. Heatmap of the Pearson correlation matrix for the $\log K_d$ values for PE MPs in freshwater and 14 molecular structural descriptors of organic compounds (The red line indicates the correlation between the two descriptors, while the green line on the diagonal represents the distribution of each descriptor).

Afterwards, six machine learning-based nonlinear algorithms, including RF, GBDT, XGBoost, CatBoost, LightGBM and SVM, were applied to develop the predictive models. The RF models was developed by utilizing 300 decision trees, with the maximum depth of each tree set to 6. The minimum number of samples required to split an internal node and that were required at a leaf node were set to 10 and 5, respectively. The optimal number of descriptors considered for each split was determined as the square root of the total number of features. The GBDT model was established with 200 weak learners, a learning rate of 0.05, and a maximum tree depth of 5. As for the CatBoost model, an L2 regularization coefficient was set to 1. For the LightGBM model, a learning rate of 0.03 was utilized. In addition, for the SVM model, the radial basis function was applied, and a baseline value for the parameter gamma was first obtained by setting gamma = 'scale', thereby obtaining the optimal values for the parameter C (0.5) and gamma (0.05).

The values for these evaluation parameters, i.e., R^2_t , $RMSE_t$, Q^2_v , $RMSE_v$ and Q^2 , listed in Table 2, imply that the six nonlinear predictive models also demonstrate satisfactory performance. Note that the CatBoost model demonstrates a notable performance gap between the training set ($R^2 = 0.99$) and the validation set ($Q^2 = 0.78$). This discrepancy is related not only to the sample size but also to the inherent characteristics of the CatBoost algorithm itself. To mitigate potential overfitting, we employed 5-fold cross-validation,

which yielded an average Q^2 of 0.78. This confirmed that the model maintains reasonable generalization capability under a robust validation framework. Therefore, the overall predictive performance of the CatBoost model remains reliable. Moreover, as exhibited in Figure 4, the predicted $\log K_d$ values from these six models are also in good agreement with those from experimental determinations. These nonlinear predictive models can serve as tools for quick acquisition of the $\log K_d$ values for PE MPs in freshwater.

Table 2. Different metric values for the predictive models of PE MPs.

Algorithms	n_t	R^2_t	$RMSE_t$	n_v	Q^2_v	$RMSE_v$	Q^2
RF	84	0.86	0.76	21	0.81	0.76	0.77
GBDT	84	0.97	0.37	21	0.86	0.65	0.79
XGBoost	84	0.95	0.45	21	0.86	0.64	0.80
CatBoost	84	0.99	0.11	21	0.84	0.70	0.78
LightGBM	84	0.87	0.74	21	0.79	0.80	0.77
SVM	84	0.79	0.93	21	0.78	0.83	0.71

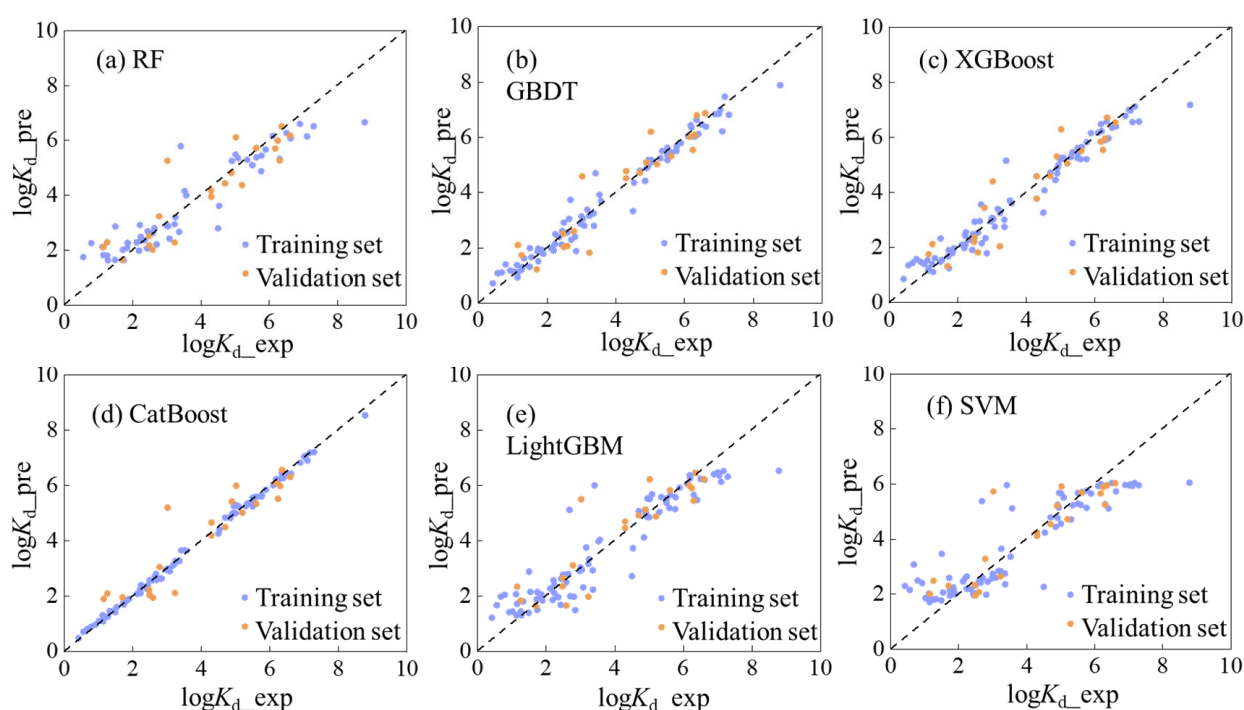


Figure 4. Predicted $\log K_d$ values from nonlinear models ($\log K_{d_pre}$) versus experimental $\log K_d$ values ($\log K_{d_exp}$) for PE MPs: (a) RF, (b) GBDT, (c) XGBoost, (d) CatBoost, (e) LightGBM and (f) SVM.

3.3. Application Domains for These Predictive Models

(1) Linear models for CPE, PBS, PCL and LDPE MPs

The application domains (ADs) for these four linear models of adsorption onto CPE, PBS, PCL and LDPE MPs were characterized with Williams plots (Figure 5) and Euclidean distance-based approach (Figure S1). The Williams plots (Figure 5) exhibit that all the organic compounds in the training sets are located in the ADs, and there are no outliers. However, as for the ADs characterized with the Euclidean distance-based approach (Figure S1), diethyl phthalate and phenanthrene are outliers for the PCL predictive model. This indicated that these two organic compounds differed significantly in structure from those in the training set for the PCL prediction model. In addition, atrazine is also an outlier for the LDPE model, implying that it exhibited structural differences from those in the training set. Note that the ADs of these models vary with the organic compounds

utilized for the development of predictive models. The ADs for the predictive model of CPE MPs cover the organic compounds with different functional groups, including $-\text{CH}_3$, $-\text{NO}_2$, $-\text{Cl}$, and $-\text{OH}$. The ADs for the models of PBS and PCL MPs are the same and cover the different organic compounds with distinct substituents, such as $-\text{CH}_3$, $-\text{CH}_2\text{CH}_2\text{CH}_3$, $-\text{C}_6\text{H}_5$, $-\text{NO}_2$, $-\text{Cl}$, $-\text{Br}$, and $-\text{OH}$. In addition, the AD for the predictive model of LDPE MPs is similar to those for PBS and PCL MPs.

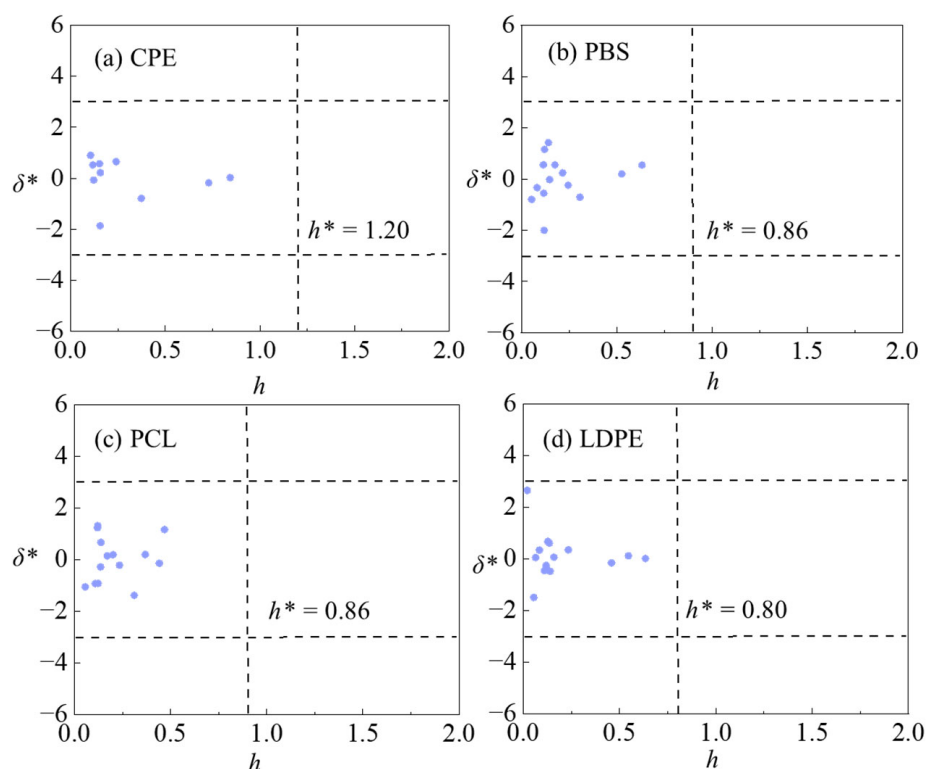


Figure 5. Williams plots of these linear models for (a) CPE, (b) PBS, (c) PCL and (d) LDPE MPs (The horizontal dashed lines represent -3 and 3 , and the vertical dashed line denotes the h^*).

(2) Nonlinear models for PE MPs

In Figure 6, the ADs for six nonlinear models, i.e., RF, GBDT, XGBoost, CatBoost, LightGBM and SVM, have been exhibited on the basis of h values and standardized residuals (δ^*). For the RF predictive model, there is no outlier; the $|\delta^*|$ value for trichlorfon is less than 3 , whereas its h value is slightly higher than h^* . This indicates that trichlorfon is very influential on the RF model. In terms of the CatBoost model, similar results have been found for diflubenzuron, which also implies that diflubenzuron exerts a significant influence on the CatBoost model. As shown in Figure 6b,c,e, the h values for 2,2',3,3',4,4',6-heptachlorobiphenyl are less than h^* , whereas the absolute δ^* values are larger than 3 . This exhibits that the 2,2',3,3',4,4',6-heptachlorobiphenyl is an outlier for the GBDT, XGBoost and LightGBM models. In addition, the compounds 3,3',4,4',5,5'-hexachlorobiphenyl in XGBoost and 2,3,4,5,6,2',3',4',5'-nonachlorobiphenyl in LightGBM also perform similarly, implying that they are outliers. The δ^* and h values for these outliers have been listed in Table S3. In addition, the Euclidean distance-based approach has also been applied to characterize the ADs in Figure S2, which indicated that no outliers were observed. All six nonlinear models can be used for estimating the adsorption on PE MPs of various organic compounds, including alcohols, alkanes, aromatic compounds, aromatic halogen compounds, carboxylic acids, cycloalkanes, esters, ethers, ketones, nitrogen heterocycles, oxygen heterocycles, phenols, phthalates, polycyclic aromatic hydrocarbons, biphenyls and chlorinated biphenyls.

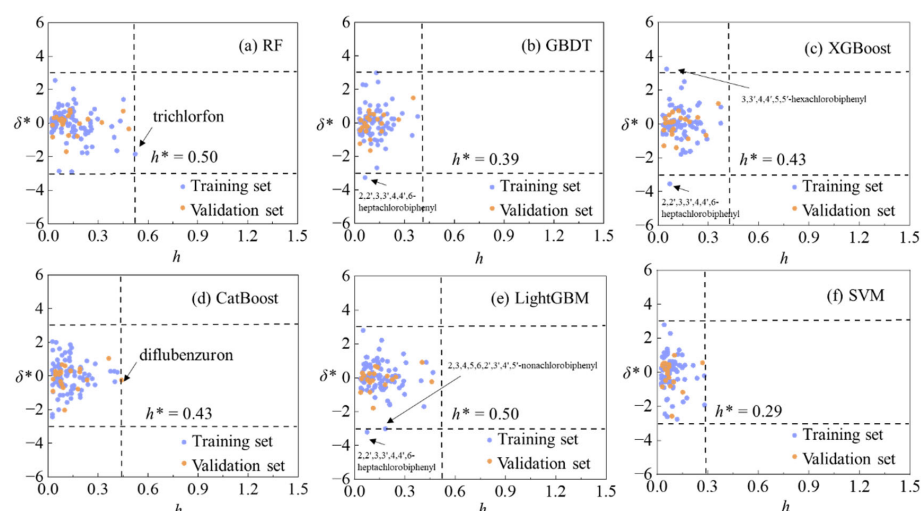


Figure 6. Williams plots of six nonlinear models with different algorithms for PE MPs: (a) RF, (b) GBDT, (c) XGBoost, (d) CatBoost, (e) LightGBM and (f) SVM (The horizontal dashed lines represent -3 and 3 , and the vertical dashed line denotes the h^*).

3.4. Adsorption Mechanisms

As for the linear predictive models of adsorption onto CPE, PBS, PCL and LDPE MPs, the utilized molecular structure descriptors have different coefficients (Table 3), implying that they play different roles in the prediction of adsorption onto MPs in freshwater.

Table 3. Molecular structure descriptors and their standardized coefficients for the linear models of CPE, PBS, PCL and LDPE MPs.

Descriptors	CPE	PBS	PCL	LDPE
<i>AATSC0i</i>	−0.704	−	−	−
<i>gmax</i>	−0.373	−	−	−
<i>AATSC6c</i>	0.279	−	−	−
<i>LipoaffinityIndex</i>	−	0.719	0.751	0.633
<i>AATSC3e</i>	−	0.509	0.545	−
<i>AATS4m</i>	−	0.253	0.360	−
<i>SHBd</i>	−	−	−	−0.912
<i>ETA_BetaP_ns_d</i>	−	−	−	0.176

In the predictive model of adsorption on CPE MPs, the descriptors *AATSC0i*, *gmax* and *AATSC6c* are utilized. The standardized coefficients of *AATSC0i* and *gmax* are negative, which demonstrates that these two descriptors make a negative contribution to the $\log K_d$ values. The descriptor *AATSC0i* [61] represents the average centered Broto–Moreau autocorrelation-lag 0/weighted by the first ionization potential, while *gmax* [62] is the maximum atom-level electrotopological state value in a molecule, related to the electronegative withdrawing group. In addition, the standardized coefficient of *AATSC6c* [49] is positive, which shows that the increase in *AATSC6c* will result in an increase in the $\log K_d$ value. *AATSC6c* is the average centered Broto–Moreau autocorrelation-lag6/weighted by charges, indicating that the electrostatic interactions related to the charges play a role in the adsorption on CPE MPs in freshwater.

In terms of the predictive models for PBS and PCL MPs, three different descriptors, i.e., *LipoaffinityIndex*, *AATSC3e*, and *AATS4m*, are used. All the standardized coefficients for these three molecular structure descriptors are positive. This indicates that the increase in these descriptors will lead to an increase in $\log K_d$ values. *LipoaffinityIndex* [63] is a parameter characterizing the lipophilicity of the organic compounds, implying that

the hydrophobic interactions are involved in the adsorption onto PE MPs in freshwater. Xu et al. [26] also found that the hydrophobic interactions played a dominant role in the adsorption onto MPs through poly-parameter linear free energy relationship models. *AATSC3e* [49] denotes the average centered Broto–Moreau autocorrelation-lag3/weighted by Sanderson electronegativities, indicating that the electrostatic interactions will influence adsorption. The other descriptor, *AATS4m* [49], represents the average Broto–Moreau autocorrelation-lag4/weighted by mass.

For the predictive model of LDPE MPs, the descriptor *LipoaffinityIndex* is also utilized. Analogous to the predictive models for PBS and PCL MPs, the standardized coefficient for *LipoaffinityIndex* is also positive. In addition, the parameters *SHBd* and *ETA_BetaP_ns_d* are also applied in the model for LDPE MPs. The standardized coefficient for *SHBd* is negative. This means that the increase in *SHBd* will result in a decrease in $\log K_d$ values. *SHBd* [64] is the sum of E-states for (strong) hydrogen bond donors. It demonstrates that hydrogen bonding can influence the adsorption of organic compounds on LDPE MPs in freshwater. In addition, the standardized coefficient for *ETA_BetaP_ns_d* is positive, implying that it has a positive relationship with the $\log K_d$ value. *ETA_BetaP_ns_d* [65] is a measure of lone electrons entering into resonance relative to molecular size. It shows that the electrostatic interactions related to the lone electrons can promote adsorption on LDPE MPs.

Moreover, we calculated the SHAP values for the molecular structure descriptors in the six nonlinear models so as to investigate the importance of different molecular structure descriptors utilized in these six nonlinear predictive models for adsorption onto PE MPs in freshwater. As shown in Figure 7, the same molecular structure descriptor has different effects on the prediction of $\log K_d$ values for distinct predictive models. For the predictive models, i.e., GBDT (Figure 7d), CatBoost (Figure 7h) and SVM (Figure 7l), the descriptor *CrippenLogP* has the most significant influence on the prediction. In addition, the SHAP values have shown that an organic compound with a larger *CrippenLogP* value will have a larger $\log K_d$ value. The descriptor *CrippenLogP* [66] represents Crippen's $\log P$, which is related to the hydrophobicity of the organic compound. This implies that the hydrophobic interactions play a key role in adsorption onto PE MPs in freshwater. However, in terms of the predictive models, i.e., RF and XGBoost, the descriptor *minaaCH* is the most influential factor. The descriptor *minaaCH* [67] is a kind of minimum atom-type E-state. Figure 7b,f exhibit that an increase in *minaaCH* will lead to an increase in $\log K_d$ values. In addition, the descriptor *B* has the most significant effects on the prediction of $\log K_d$ values by the LightGBM model. Figure 7j shows that one compound with a negative *B* value will have a larger $\log K_d$ value, indicating that the descriptor *B* makes a negative contribution to the predictions of the LightGBM model. *B* describes the hydrogen accepting ability of the organic compound. It demonstrates that the hydrogen bonding interactions play an important role in adsorption on PE MPs in freshwater.

Additionally, as for the six nonlinear prediction models, Figure 7b,d,f,h,j,l, show that the descriptors *CrippenLogP*, *ETA_dBeta*, *SpMax2_Bhm*, *SpMax8_Bhi*, and *V* are positively correlated with $\log K_d$ values, while the descriptor *B* shows a negative correlation with $\log K_d$ values. *ETA_dBeta* [68] is a measure of relative unsaturation content. This means that an organic compound with a higher relative unsaturation content is more liable to be adsorbed by PE MPs in adsorption. *SpMax2_Bhm* [49] represents the largest absolute eigenvalue of the Burden modified matrix-n2/weighted by relative mass; *SpMax8_Bhi* [49] is the largest absolute eigenvalue of the Burden modified matrix-n8/weighted by relative first ionization potential. The positive correlation indicates that one compound with a larger *SpMax2_Bhm* or *SpMax8_Bhi* value may have larger $\log K_d$ values. *V* is McGowan's molar volume, which can represent dispersion and hydrophobic interactions. Therefore, the positive correlation between the $\log K_d$ value and the descriptor *V* implies that dispersion

and hydrophobic interactions promote the adsorption of organic compounds onto PE MPs in freshwater. Wei et al. [30] also reported similar findings in adsorption mechanism analysis according to their predictive models.

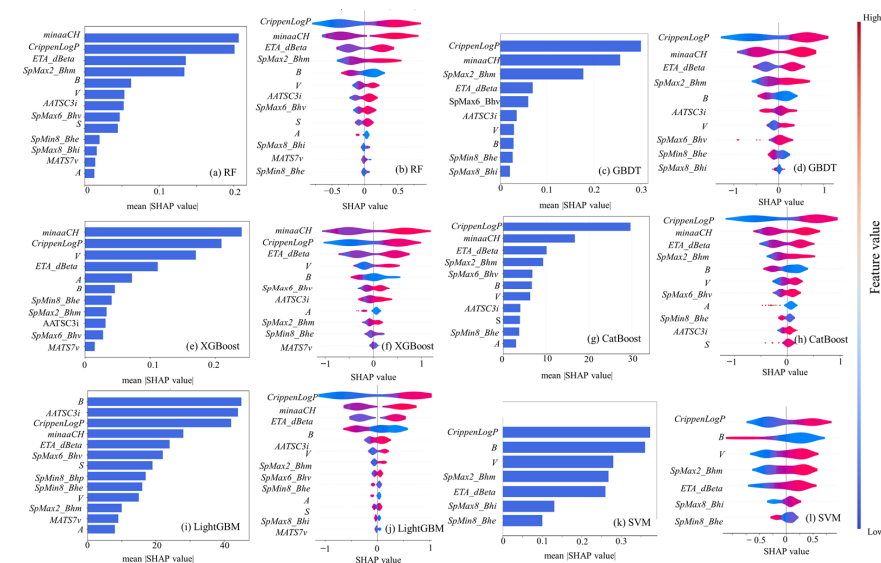


Figure 7. Mean |SHAP values| (a) RF, (c) GBDT, (e) XGBoost, (g) CatBoost, (i) LightGBM, (k) SVM and SHAP values (b) RF, (d) GBDT, (f) XGBoost, (h) CatBoost, (j) LightGBM, (l) SVM for the molecular structure descriptors across six nonlinear models for PE MPs.

In brief, different interactions, including hydrogen bonding, hydrophobic, electrostatic and dispersion interactions, are involved in the adsorption of organic compounds onto various MPs.

4. Comparisons with Previous Predictive Models

As shown in Table 4, some predictive models have been developed for predicting adsorption onto different MPs. Most of the molecular descriptors from these models for CPE, PBS, PCL and LDPE MPs were based on experimental determinations, while the descriptors used in the linear predictive models for these four MPs in the current study were independent of the experiments. Moreover, these linear predictive models have been evaluated through external validation, and their ADs have been characterized. In terms of the predictive models for the adsorption onto PE MPs, the six nonlinear predictive models developed in the current study have the widest ADs, covering 15 different categories. In addition, the external validation and ADs have been assessed for all these models.

Table 4. Previous prediction models for MPs in freshwater.

MPs	<i>n</i>	<i>k</i>	Algorithm	R^2_t	$RMSE_t$	Q^2_v	$RMSE_v$	Q^2_{Loo}	Q^2	AD	Ref.
PE	49	2	MLR	0.741	1.160	0.858	0.879	0.670		Y	[25]
CPE	13	2	MLR	0.930	0.236	0.913	0.361	0.749		Y	[25]
LDPE	18	5	MLR	0.909	0.288					N	[26]
PS	17	5	MLR	0.905	0.369					N	[26]
PBS	18	5	MLR	0.969	0.115					N	[26]
PCL	18	5	MLR	0.959	0.135					N	[26]
PE	23	2	MLR	0.909	0.909		0.608			Y	[30]
PE	24	1	MLR	0.903	0.903		0.686			Y	[31]
PE	24	7	RF	0.946	0.549	0.891	0.744			N	[29]
PE	24	7	SVM	0.953	0.536	0.893	0.770			N	[29]
PE	24	7	ANN	0.956	0.489	0.869	0.865			N	[29]

5. Conclusions

Seven different machine learning-based algorithms, including one linear (MLR) and six nonlinear (RF, GBDT, XGBoost, CatBoost, LightGBM and SVM) algorithms, were utilized to develop predictive models of adsorption onto various MPs in freshwater. The four MLR models developed are applicable to the prediction of adsorption onto CPE, PBS, PCL and LDPE MPs, while the established RF, GBDT, XGBoost, CatBoost, LightGBM and SVM can be used for predicting adsorption onto PE MPs in freshwater. All these 10 predictive models exhibit satisfactory goodness-of-fit, robustness and predictive capability. The results indicate that different adsorption mechanisms, including hydrogen bonding, hydrophobic, electrostatic and dispersion interactions, may be involved in the adsorption of various organic compounds onto CPE, PBS, PCL, LDPE and PE MPs. These 10 models have distinct ADs, among which the ADs for the predictive models of PE MPs exhibit the broadest coverage in terms of organic compound diversity. The predictive models for PE MPs can be used to predict the $\log K_d$ values of organic compounds with various functional groups, including $-\text{CH}_3$, $-\text{CH}_2\text{CH}_3$, $-\text{C}_6\text{H}_5$, $-\text{O}-$, $-\text{Cl}$, Br , $-\text{NO}_2$, $-\text{OH}$, NH_2 , $-\text{OCH}_3$, $-\text{C}(\text{CH}_3)_3$, and $-\text{C}(\text{O})\text{CH}_3$.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/separations13020050/s1>, Table S1: Categories of organic compounds used in PE predictive models; Table S2: Hyperparameter optimization for RF, GBDT, XGBoost, CatBoost, LightGBM and SVM; Table S3: Outliers and their standardized residuals (δ^*) and leverage values (h); Figure S1: Application domains with Euclidean distance-based approach for linear models of (a) CPE, (b) PBS, (c) PCL and (d) LDPE MPs; Figure S2: Application domains with Euclidean distance-based approach for six nonlinear models with different algorithms for PE MPs (a) RF, (b) GBDT, (c) XGBoost, (d) CatBoost, (e) LightGBM and (f) SVM.

Author Contributions: Conceptualization, Y.W., H.Y. and X.T.; methodology, Y.W. and P.Z.; software, Y.W. and P.Z.; validation, Y.W. and P.Z.; formal analysis, Y.W. and P.Z.; investigation, Y.W. and P.Z.; data curation, Y.W., H.Y., X.T. and P.Z.; writing—original draft preparation, Y.W. and P.Z.; writing—review and editing, Y.W., P.Z., H.Y. and X.T.; visualization, Y.W. and P.Z.; supervision, Y.W.; funding acquisition, Y.W., X.T. and H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 22206095; the Fundamental Research Funds for the Central Universities, grant numbers FRF-TP-22-087A1 and FRF-BD-25-003; and the National Key R&D Program of China, grant number 2023YFB3810800.

Data Availability Statement: Data is contained within the article or Supplementary Materials.

Conflicts of Interest: The authors declare no competing financial interests.

References

1. Collard, F.; Gasperi, J.; Gabrielsen, G.W.; Tassin, B. Plastic particle ingestion by wild freshwater fish: A critical review. *Environ. Sci. Technol.* **2019**, *53*, 12974–12988. [[CrossRef](#)] [[PubMed](#)]
2. Wang, Y.; Yang, Y.; Liu, X.; Zhao, J.; Liu, R.; Xing, B. Interaction of microplastics with antibiotics in aquatic environment: Distribution, sorption, and toxicity. *Environ. Sci. Technol.* **2021**, *55*, 15579–15595. [[CrossRef](#)] [[PubMed](#)]
3. Alimi, O.S.; Budarz, J.F.; Hernandez, L.M.; Tufenkji, N. Microplastics and nanoplastics in aquatic environments: Aggregation, deposition, and enhanced contaminant transport. *Environ. Sci. Technol.* **2018**, *52*, 1704–1724. [[CrossRef](#)] [[PubMed](#)]
4. Wang, T.; Wang, L.; Chen, Q.; Kalogerakis, N.; Ji, R.; Ma, Y. Interactions between microplastics and organic pollutants: Effects on toxicity, bioaccumulation, degradation, and transport. *Sci. Total Environ.* **2020**, *748*, 142427. [[CrossRef](#)]
5. Williams, T.; Walsh, C.; Murray, K.; Subir, M. Interactions of emerging contaminants with model colloidal microplastics, C_{60} fullerene, and natural organic matter—effect of surface functional group and adsorbate properties. *Environ. Sci.-Proc. Imp.* **2020**, *22*, 1190. [[CrossRef](#)]

6. Hatinoglu, M.D.; Perreault, F.; Apul, O.G. Modified linear solvation energy relationships for adsorption of perfluorocarboxylic acids by polystyrene microplastics. *Sci. Total Environ.* **2023**, *860*, 160524. [\[CrossRef\]](#)
7. Rochman, C.M.; Brookson, C.; Bikker, J.; Djuric, N.; Earn, A.; Bucci, K.; Athey, S.; Huntington, A.; McIlwraith, H.; Munno, K.; et al. Rethinking microplastics as a diverse contaminant suite. *Environ. Toxicol. Chem.* **2019**, *38*, 703–711. [\[CrossRef\]](#)
8. Galloway, T.S.; Cole, M.; Lewis, C. Interactions of microplastic debris throughout the marine ecosystem. *Nat. Ecol. Evol.* **2017**, *1*, 0116. [\[CrossRef\]](#)
9. Mizukawa, K.; Takada, H.; Ito, M.; Geok, Y.B.; Hosoda, J.; Yamashita, R.; Saha, M.; Suzuki, S.; Miguez, C.; Frias, J.; et al. Monitoring of a wide range of organic micropollutants on the Portuguese coast using plastic resin pellets. *Mar. Pollut. Bull.* **2013**, *70*, 296–302. [\[CrossRef\]](#)
10. Velzeboer, I.; Kwadijk, C.J.A.F.; Koelmans, A.A. Strong sorption of PCBs to nanoplastics, microplastics, carbon nanotubes, and fullerenes. *Environ. Sci. Technol.* **2014**, *48*, 4869–4876. [\[CrossRef\]](#)
11. Li, J.; Zhang, K.; Zhang, H. Adsorption of antibiotics on microplastics. *Environ. Pollut.* **2018**, *237*, 460–467. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Avio, C.G.; Gorbi, S.; Milan, M.; Benedetti, M.; Fattorini, D.; d’Errico, G.; Pauletto, M.; Bargelloni, L.; Regoli, F. Pollutants bioavailability and toxicological risk from microplastics to marine mussels. *Environ. Pollut.* **2015**, *198*, 211–222. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Zhang, P.; Huang, P.; Sun, H.; Ma, J.; Li, B. The structure of agricultural microplastics (PT, PU and UF) and their sorption capacities for PAHs and PHE derivatives under various salinity and oxidation treatments. *Environ. Pollut.* **2020**, *257*, 113525. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Liu, F.F.; Liu, G.Z.; Zhu, Z.L.; Wang, S.C.; Zhao, F.F. Interactions between microplastics and phthalate esters as affected by microplastics characteristics and solution chemistry. *Chemosphere* **2019**, *214*, 688–694. [\[CrossRef\]](#)
15. Chen, S.; Tan, Z.; Qi, Y.; Ouyang, C. Sorption of tri-n-butyl phosphate and tris(2-chloroethyl) phosphate on polyethylene and polyvinyl chloride microplastics in seawater. *Mar. Pollut. Bull.* **2019**, *149*, 110490. [\[CrossRef\]](#)
16. Guo, X.; Wang, J. Sorption of antibiotics onto aged microplastics in freshwater and seawater. *Mar. Pollut. Bull.* **2019**, *149*, 110511. [\[CrossRef\]](#)
17. Hüffer, T.; Hofmann, T. Sorption of non-polar organic compounds by micro-sized plastic particles in aqueous solution. *Environ. Pollut.* **2016**, *214*, 194–201. [\[CrossRef\]](#)
18. Yu, F.; Yang, C.; Zhu, Z.; Bai, X.; Ma, J. Adsorption behavior of organic pollutants and metals on micro/nanoplastics in the aquatic environment. *Sci. Total Environ.* **2019**, *694*, 133643. [\[CrossRef\]](#)
19. Wang, F.; Shih, K.M.; Li, X.Y. The partition behavior of perfluorooctanesulfonate (PFOS) and perfluorooctanesulfonamide (FOSA) on microplastics. *Chemosphere* **2015**, *119*, 841–847. [\[CrossRef\]](#)
20. Liu, X.; Shi, H.; Xie, B.; Dionysiou, D.D.; Zhao, Y. Microplastics as both a sink and a source of bisphenol A in the marine environment. *Environ. Sci. Technol.* **2019**, *53*, 10188–10196. [\[CrossRef\]](#)
21. Mo, Q.; Yang, X.; Wang, J.; Xu, H.; Li, W.; Fan, Q.; Gao, S.; Yang, W.; Gao, C.; Liao, D.; et al. Adsorption mechanism of two pesticides on polyethylene and polypropylene microplastics: DFT calculations and particle size effects. *Environ. Pollut.* **2021**, *291*, 118120. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Leng, Y.; Wang, W.; Cai, H.; Chang, F.; Xiong, W.; Wang, J. Sorption kinetics, isotherms and molecular dynamics simulation of 17 β -estradiol onto microplastics. *Sci. Total Environ.* **2023**, *858*, 159803. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Chen, Y.; Li, H.; Yin, Y.; Shan, S.; Huang, T.; Tang, H. Effect of microplastics on the adherence of coexisting background organic contaminants to natural organic matter in water. *Sci. Total Environ.* **2023**, *905*, 167175. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Su, L.; Wang, Z.; Xiao, Z.; Xia, D.; Wang, Y.; Chen, J. Rapidly predicting aqueous adsorption constants of organic pollutants onto polyethylene microplastics by combining molecular dynamics simulations and machine learning. *ACS EST Water* **2024**, *4*, 4184–4192. [\[CrossRef\]](#)
25. Gui, B.; Xu, X.; Zhang, S.; Wang, Y.; Li, C.; Zhang, D.; Su, L.; Zhao, Y. Prediction of organic compounds adsorbed by polyethylene and chlorinated polyethylene microplastics in freshwater using QSAR. *Environ. Res.* **2021**, *197*, 111001. [\[CrossRef\]](#)
26. Xu, J.; Wang, L.; Sun, H. Adsorption of neutral organic compounds on polar and nonpolar microplastics: Prediction and insight into mechanisms based on pp-LFERs. *J. Hazard. Mater.* **2021**, *408*, 124857. [\[CrossRef\]](#)
27. Qiu, Y.; Li, Z.; Zhang, T.; Zhang, P. Predicting aqueous sorption of organic pollutants on microplastics with machine learning. *Water Res.* **2023**, *244*, 120503. [\[CrossRef\]](#)
28. Yao, J.; Wen, J.; Li, H.; Yang, Y. Surface functional groups determine adsorption of pharmaceuticals and personal care products on polypropylene microplastics. *J. Hazard. Mater.* **2022**, *423*, 127131. [\[CrossRef\]](#)
29. Astray, G.; Soria-Lopez, A.; Barreiro, E.; Mejuto, J.C.; Cid-Samamed, A. Machine learning to predict the adsorption capacity of microplastics. *Nanomaterials* **2023**, *13*, 1061. [\[CrossRef\]](#)
30. Wei, X.; Li, M.; Wang, Y.; Jin, L.; Ma, G.; Yu, H. Developing predictive models for carrying ability of micro-plastics towards organic pollutants. *Molecules* **2019**, *24*, 1784. [\[CrossRef\]](#)

31. Li, M.; Yu, H.; Wang, Y.; Li, J.; Ma, G.; Wei, X. QSPR models for predicting the adsorption capacity for microplastics of polyethylene, polypropylene and polystyrene. *Sci. Rep.* **2020**, *10*, 14597. [\[CrossRef\]](#)
32. Wang, J.; Liu, X.; Liu, G. Sorption behaviors of phenanthrene, nitrobenzene, and naphthalene on mesoplastics and microplastics. *Environ. Sci. Pollut. Res.* **2019**, *26*, 12563–12573. [\[CrossRef\]](#)
33. Elizalde-Velázquez, A.; Subbiah, S.; Anderson, T.A.; Green, M.J.; Zhao, X.; Cañas-Carrell, J.E. Sorption of three common nonsteroidal anti-inflammatory drugs (NSAIDs) to microplastics. *Sci. Total Environ.* **2020**, *715*, 136974. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Lan, T.; Wang, T.; Cao, F.; Yu, C.; Chu, Q.; Wang, F. A comparative study on the adsorption behavior of pesticides by pristine and aged microplastics from agricultural polyethylene soil films. *Ecotox. Environ. Safe.* **2021**, *209*, 111781. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Fernandez, L.A.; Macfarlane, J.K.; Tcaciuc, A.P. Measurement of freely dissolved PAH concentrations in sediment beds using passive sampling with low-density polyethylene strips. *Environ. Sci. Technol.* **2009**, *43*, 1430–1436. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Hale, S.E.; Tomaszewski, J.E.; Luthy, R.G.; Werner, D. Sorption of dichlorodiphenyltrichloroethane (DDT) and its metabolites by activated carbon in clean water and sediment slurries. *Water Res.* **2009**, *43*, 4336–4346. [\[CrossRef\]](#)
37. Kong, X.; Zhou, A.; Chen, X.; Cheng, X.; Lai, Y.; Li, C.; Ji, Q.; Ji, Q.; Kong, J.; Ding, Y.; et al. Insight into the adsorption behaviors and bioaccessibility of three altered microplastics through three types of advanced oxidation processes. *Sci. Total Environ.* **2024**, *917*, 170420. [\[CrossRef\]](#)
38. Guo, X.; Liu, Y.; Wang, J. Sorption of sulfamethazine onto different types of microplastics: A combined experimental and molecular dynamics simulation study. *Mar. Pollut. Bull.* **2019**, *145*, 547–554. [\[CrossRef\]](#)
39. Pascall, M.A.; Zabik, M.E.; Zabik, M.J. Uptake of polychlorinated biphenyls (PCBs) from an aqueous medium by polyethylene, polyvinyl chloride, and polystyrene films. *J. Agric. Food Chem.* **2005**, *53*, 164–169. [\[CrossRef\]](#)
40. Razanajatovo, R.M.; Ding, J.N.; Zhang, S.S.; Jiang, H.; Zou, H. Sorption and desorption of selected pharmaceuticals by polyethylene microplastics. *Mar. Pollut. Bull.* **2018**, *136*, 516–523. [\[CrossRef\]](#)
41. Teuten, E.L.; Rowland, S.J.; Galloway, T.S. Potential for plastics to transport hydrophobic contaminants. *Environ. Sci. Technol.* **2007**, *41*, 7759–7764. [\[CrossRef\]](#)
42. Uber, T.H.; Hüffer, T.; Planitz, S.; Schmidt, T.C. Characterization of sorption properties of high-density polyethylene using the poly-parameter linearfree-energy relationships. *Environ. Pollut.* **2019**, *248*, 312–319. [\[CrossRef\]](#)
43. Wang, T.; Yu, C.; Chu, Q.; Wang, F.; Lan, T.; Wang, J. Adsorption behavior and mechanism of five pesticides on microplastics from agricultural polyethylene films. *Chemosphere* **2020**, *244*, 125491. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Wang, Z.; Chen, M.; Zhang, L.; Wang, K.; Yu, X.; Zheng, Z.; Zheng, R. Sorption behaviors of phenanthrene on the microplastics identified in a mariculture farm in Xiangshan Bay, southeastern China. *Sci. Total Environ.* **2018**, *628–629*, 1617–1626. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Wu, C.X.; Zhang, K.; Huang, X.L.; Liu, J.T. Sorption of pharmaceuticals and personal care products to polyethylene debris. *Environ. Sci. Pollut. Res.* **2016**, *23*, 8819–8826. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Abraham, M.H.; Grellier, P.L.; McGill, R.A. Determination of olive oil-gas and hexadecane-gas partition coefficients, and calculation of the corresponding olive oil-water and hexadecane-water partition coefficients. *J. Chem. Soc. Perkin Trans. 2* **1987**, *6*, 797–803. [\[CrossRef\]](#)
47. Abraham, M.H. Scales of solute hydrogen-bonding: Their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev.* **1993**, *22*, 73–83. [\[CrossRef\]](#)
48. Abraham, M.H.; McGowan, J.C. The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography. *Chromatographia* **1987**, *23*, 243–246. [\[CrossRef\]](#)
49. Yap, C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466–1474. [\[CrossRef\]](#)
50. Ye, H.; Jiang, S.; Yan, Y.; Zhao, B.; Grant, E.R.; Kitts, D.D.; Yada, R.Y.; Pratap-Singh, A.; Baldelli, A.; Yang, T. Integrating metal-phenolic networks-mediated separation and machine learning-aided surface-enhanced Raman spectroscopy for accurate nanoplastics quantification and classification. *ACS Nano* **2024**, *18*, 26281–26296. [\[CrossRef\]](#)
51. Ai, C.A. Method for cancer genomics feature selection based on LASSO-RFE. *Iran. J. Sci. Technol. Trans. Sci.* **2022**, *46*, 731–738. [\[CrossRef\]](#)
52. Meng, F.G.; Shi, Z.G.; Song, Y.X. The TPRF: A novel soft sensing method of alumina-silica ratio in red mud based on tpe and random forest algorithm. *Processes* **2024**, *12*, 663. [\[CrossRef\]](#)
53. Guo, C.; Yang, Z.; Yue, Y.C.; Li, W.X.; Wu, H.T. In-situ stresses ring hole measurement of concrete optimized based on finite element and GBDT algorithm. *Comput. Concr.* **2024**, *34*, 477–487.
54. Li, J.L.; Zhang, Z.S.; Wang, X.F. Performance-oriented road structure and material design method based on enhanced XGBoost algorithm. *Int. J. Pavement Eng.* **2024**, *25*, 2295899. [\[CrossRef\]](#)
55. Kuo, P.H.; Li, Y.H.; Yau, H.T. Development of feline infectious peritonitis diagnosis system by using CatBoost algorithm. *Comput. Biol. Chem.* **2024**, *113*, 108227. [\[CrossRef\]](#)

56. Yang, Y.D.; Wang, Y.J.; Zhang, X. Research on power system small signal stability analysis and correction based on LightGBM algorithm. *Electr. Eng.* **2024**, *106*, 4469–4486. [[CrossRef](#)]
57. Moradi, S.; Omar, A.; Zhou, Z.; Agostino, A.; Gandomkar, Z.; Bustamante, H.; Power, K.; Henderson, R.; Leslie, G. Forecasting and optimizing dual media filter performance via machine learning. *Water Res.* **2023**, *235*, 119874. [[CrossRef](#)]
58. Gramatica, P. Principles of QSAR models validation: Internal and external. *QSAR Combust. Sci.* **2007**, *26*, 694–701. [[CrossRef](#)]
59. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intel.* **2020**, *2*, 56–67. [[CrossRef](#)]
60. Golbraikh, A.; Shen, M.; Xiao, Z.Y.; Xiao, Y.D.; Lee, K.H.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aided Mol. Des.* **2003**, *17*, 241–253. [[CrossRef](#)]
61. Kobayashi, Y.; Miyake, Y.; Ishiwari, F.; Ishiwata, S.; Saeki, A. Machine learning of atomic force microscopy images of organic solar cells. *RSC Adv.* **2023**, *13*, 15107–15113. [[CrossRef](#)]
62. Votano, J.R.; Parham, M.; Hall, L.H.; Kier, L.B. New predictors for several ADME/Tox properties: Aqueous solubility, human oral absorption, and Ames genotoxicity using topological descriptors. *Mol. Divers.* **2004**, *8*, 379–391. [[CrossRef](#)]
63. Hou, S.; Wang, J.; Li, Z.; Wang, Y.; Wang, Y.; Yang, S.; Xu, J.; Zhu, W. Five-descriptor model to predict the chromatographic sequence of natural compounds. *J. Sep. Sci.* **2016**, *39*, 864–872. [[CrossRef](#)]
64. Przybyłek, M.; Jeliński, T.; Cysewski, P. Application of multivariate adaptive regression splines (MARSplines) for predicting Hansen solubility parameters based on 1D and 2D molecular descriptors computed from SMILES string. *J. Chem.* **2019**, *2019*, 9858371. [[CrossRef](#)]
65. Toropova, A.P.; Duchowicz, P.R.; Saavedra, L.M.; Castro, E.A.; Toropov, A.A. The use of the index of ideality of correlation to build up models for bioconcentration factor. *Mol. Inf.* **2020**, *39*, 1900070. [[CrossRef](#)]
66. Das, N.R.; Mishra, S.P.; Achary, P.G.R. Evaluation of molecular structure based descriptors for the prediction of pEC₅₀(M) for the selective adenosine A_{2A} receptor. *J. Mol. Struct.* **2021**, *1232*, 130080. [[CrossRef](#)]
67. Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2009.
68. De, P.; Roy, K. Greener chemicals for the future: QSAR modelling of the PBT index using ETA descriptors. *SAR QSAR Environ. Res.* **2018**, *29*, 319–337. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.