Article

Developing a QSPR Model of Organic Carbon Normalized Sorption Coefficients of Perfluorinated and Polyfluoroalkyl Substances

Lan Jiang [1], Yue Xu [1,*], Xiaoyu Zhang [2,*], Bingfeng Xu [1], Ximeng Xu [1] and Yixing Ma [3]

[1] Faculty of Civil Engineering and Mechanics, Kunming University of Science and Technology, Kunming 650500, China
[2] Yunnan Research Academy of Eco-environmental Sciences, Kunming 650034, China
[3] Faculty of Environmental Science and Engineering, Kunming University of Science and Technology, Kunming 650500, China
* Correspondence: xuyue@kust.edu.cn (Y.X.); todzhang1999@hotmail.com (X.Z.)

Abstract: Perfluorinated and polyfluoroalkyl substances (PFASs) are known for their long-distance migration, bioaccumulation, and toxicity. The transport of PFASs in the environment has been a source of increasing concerned. The organic carbon normalized sorption coefficient (K_{oc}) is an important parameter from which to understand the distribution behavior of organic matter between solid and liquid phases. Currently, the theoretical prediction research on log K_{oc} of PFASs is extremely limited. The existing models have limitations such as restricted application fields and unsatisfactory prediction results for some substances. In this study, a quantitative structure–property relationship (QSPR) model was established to predict the log K_{oc} of PFASs, and the potential mechanism affecting the distribution of PFASs between two phases from the perspective of molecular structure was analyzed. The developed model had sufficient goodness of fit and robustness, satisfying the model application requirements. The molecular weight (*MW*) related to the hydrophobicity of the compound; lowest unoccupied molecular orbital energy (E_{LUMO}) and maximum average local ionization energy on the molecular surface ($ALIE_{max}$), both related to electrostatic properties; and the dipole moment (μ), related to the polarity of the compound; are the key structural variables that affect the distribution behavior of PFASs. This study carried out a standardized modeling process, and the model dataset covered a comprehensive variety of PFASs. The model can be used to predict the log K_{oc} of conventional and emerging PFASs effectively, filling the data gap of the log K_{oc} of uncommon PFASs. The explanation of the mechanism of the model has proven to be of great value for understanding the distribution behavior and migration trends of PFASs between sediment/soil and water, and for estimating the potential environmental risks generated by PFASs.

Keywords: perfluorinated and polyfluoroalkyl substances; organic carbon normalized sorption coefficient; quantitative structure–property relationship; distribution behavior

1. Introduction

Perfluorinated and polyfluoroalkyl substances (PFASs), as a class of synthetic aliphatic compounds [1,2], have be widely used in industrial production and daily consumer products because of their hydrophobicity, oleophobicity, thermal stability, and chemical stability [3,4]. Up to now, PFASs and their precursors have been found in water, atmosphere, soil, sediment, and other environmental media [5–8]. Generally, the concentration of PFAS in these environmental media is within the scale range of $ng \cdot L^{-1}$, but in some areas with serious pollution (i.e., around fluorine chemical plants), the concentration of PFAS in water can reach the scale range of $mg \cdot L^{-1}$ [9]. PFASs can enter the human body and accumulate through drinking water, the food chain, and in other ways [10,11], and when a

certain threshold is reached, they will produce corresponding toxic effects, such as neurological, reproductive, liver, and endocrine toxicity, which can seriously endanger human health [12–16].

Sediment, soil, and water are important sinks for PFASs [1,10,17]. The accurate measurement of the organic carbon normalized sorption coefficient (K_{oc}) of PFASs can reflect their distribution behavior between sediment/soil and water [18–20], which is crucial for their environmental fate and risk assessment. So far, many studies have been undertaken on the distribution behavior of traditional PFASs, such as perfluorinated carboxylic acids (PFCAs) and perfluoroalkyl sulfonic acids (PFSAs), but there are few studies on that of emerging PFASs. PFASs are composed of a carbon skeleton and hydrophilic groups, where the hydrogen atoms connected to the carbon skeleton are partially or completely replaced by fluorine atoms [1,2]. From a structural perspective, both conventional and emerging PFASs are dominated by the carbon skeleton. However, the substituents on the skeleton and the functional groups at the ends have an important impact on the environmental transport characteristics of the compounds. Previous studies have shown that the migration of PFASs in different environmental media is closely related to their structural factors, such as their carbon chain length, substituents, and functional groups [17,21]. For instance, for the two-phase medium of sediment/soil and water, there is a linear relationship between the K_{oc} of PFASs and the number of perfluorinated carbon (CF). In general, the K_{oc} increases with the number of CF, while PFASs with a sulfonic group have a larger K_{oc} than similar compounds containing a carboxyl group [21]. The lack of research on the environmental migration law of emerging PFASs results from the variety of PFASs, the fact that their derivatives appeared one after another, and the insufficient understanding of their physicochemical properties. At the same time, the experimental measurement of K_{oc} is not only cumbersome and costly but may also pose environmental pollution and human health risks in large-scale experiments. However, it is possible to quickly fill the data gap syrrounding the log K_{oc} of PFASs at low experimental cost by constructing a mathematical model based on the structural characteristics of PFASs to predict their K_{oc}.

The quantitative structure–property relationship (QSPR) model is a theoretical prediction tool with a rapid development and a wide application range. It establishes a functional relationship between the molecular structure of compounds and their properties to effectively predict the compounds' properties [22,23]. The QSPR model can be used to predict the partition coefficient of various organic pollutants with high efficiency, such as the partition coefficient of polycyclic aromatic hydrocarbons (PAHs) between polydimethyl-siloxane (PDMS) and water [24], the partition coefficient of polychlorinated biphenyls (PCBs) between low-density polyethylene and water [25], and the partition coefficient of PFASs between gas and particles [2]. To date, few studies have employed PFASs as a unique research object to construct QSPR models for predicting their log K_{oc} [26,27]. A previous study reported the log K_{oc} of 824 organic compounds predicted by a QSPR model, but only a few PFASs were included [26]. Due to the limitation of its data set, the application scope of the model was narrow for PFASs Another limitation of prior work on the model's establishment was that the modeling process did not fully follow the five guidelines for QSPR model construction [27,28]. Generally, for the log K_{oc} prediction of PFASs by QSPR, it is necessary to improve the applicability and accuracy of the model. Meanwhile, the standardization of modeling also need further investigation. Owing to the issues with the above models, there is still a knowledge gap in the overall analysis of PFASs distribution mechanism between sediment/soil and water at the molecular level.

This study developed an optimal log K_{oc} prediction model for PFASs based on the QSPR model construction guidelines. A comprehensive verification and evaluation of the model was undertaken to ensure the integrity and standardization of the modeling process and achieve a reasonable prediction of the log K_{oc} of PFASs. The model used 22 PFASs from eight different classes as the modeling dataset, equipped with strong pertinence which greatly improves the applicability of the model for PFASs. Molecular descriptors with clear definitions were included in the model, which identified the potential

mechanism of PFASs distribution between sediment/soil and water from the perspective of the molecular level quickly and accurately, facilitating a better understanding of the distribution behavior of PFASs between the two phases. This combination of effects has significant practical implications for enriching the migration theory of PFASs with different structures between sediment/soil and water, and provides a reference for predicting the deposition concentration of emerging PFASs in environmental media.

2. Results and Discussion

2.1. Model Construction and Validation

After stepwise linear regression, the optimal QSPR model (Equation (1)) was obtained. The model contained four molecular descriptors: molecular weight (*MW*), dipole moment (μ), lowest unoccupied molecular orbital energy (E_{LUMO}), and maximum average local ionization energy on the molecular surface ($ALIE_{max}$).

$$\log K_{oc} = 7.334 \times 10^{-3}\, MW - 1.705\, \mu - 0.956\, E_{LUMO} - 1.398\, ALIE_{max} + 24.10 \quad (1)$$

Table S4 in the supplementary materials lists the calculated values of the four molecular descriptors and the predicted values of log K_{oc}. The statistical parameters of the developed model are shown in Table 1. According to the evaluation criteria of the QSPR model summarized in previous studies [29], coefficient of determination (R^2) > 0.8, multiple correlation coefficient of leave-one-out cross-validation (Q^2_{LOO}) > 0.5, external validation indicators (Q^2_{F1}, Q^2_{F2}, and Q^2_{F3}) > 0.5, indicating that the model has sufficient goodness of fit, robustness, and predictive ability and meets the requirements of the QSPR model construction guidelines [28]. In addition, the $R^2 - Q^2_{LOO}$ value of this model is less than 0.3, indicating that there is no overfitting phenomenon in this model [30]. Q^2_{LOO}, Q^2_{F1}, Q^2_{F2}, and Q^2_{F3} were calculated according to the method in a previous study [29], and the calculation formulas of these parameters are presented in the supplementary materials; R^2 was obtained using SPSS 26 software (IBM SPSS Inc., Chicago, IL, USA).

Table 1. Statistical parameters of the optimal QSPR model.

| Training Set						Validation Set				
n	R^2	Q^2_{LOO}	*RMSE*	*F*	*p*	*n*	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	*RMSE*
18	0.962	0.920	0.212	82.269	<0.001	4	0.961	0.955	0.959	0.219

Notes: *n*: the number of data points; R^2: coefficient of determination; Q^2_{LOO}: multiple correlation coefficient of leave-one-out cross-validation; *RMSE*: root mean square error; *F*: variance ratio; *p*: significance index; when $p < 0.05$, this indicates that the model is significant; Q^2_{F1}, Q^2_{F2}, and Q^2_{F3}: external validation indicators.

Figure 1 shows the error distribution of the model prediction. The prediction errors of the PFASs were randomly distributed on both sides of the baseline, and there was no obvious regularity, indicating that the built model had no obvious systematic errors. Table 2 lists the significance index (*p*) and variance inflation coefficient (*VIF*) values of the molecular descriptors contained in the model. When $p < 0.05$, this indicates that the molecular descriptor was significant; when $VIF < 10$, this indicates that there was no multicollinearity among the molecular descriptors [31]. It can be seen from the table that all molecular descriptors in the model were key descriptors, and there was no collinearity among them. *p* and *VIF* were obtained using SPSS 26 software (IBM SPSS Inc., Chicago, IL, USA).

**Table 2.** Statistical parameters of different descriptors.

Description	*p*	*VIF*
MW	0.000	1.192
μ	0.002	1.239
E_{LUMO}	0.000	3.420
$ALIE_{max}$	0.003	3.216

Notes: *VIF*: variance inflation coefficient.
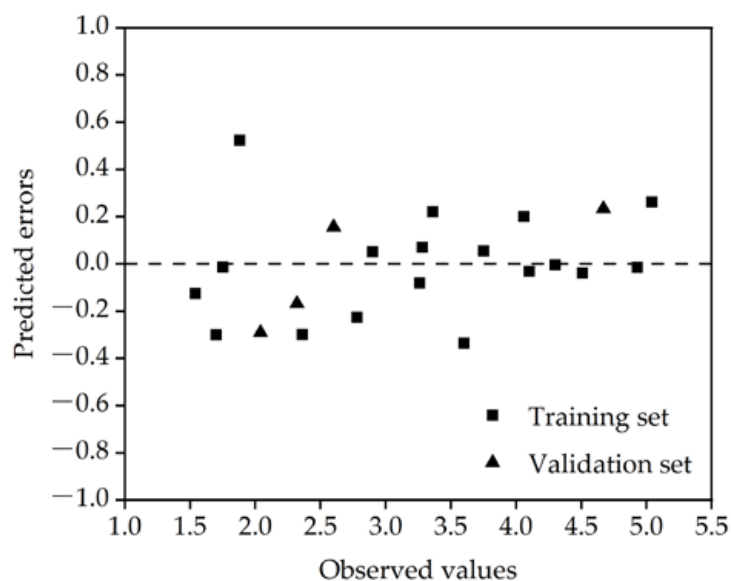
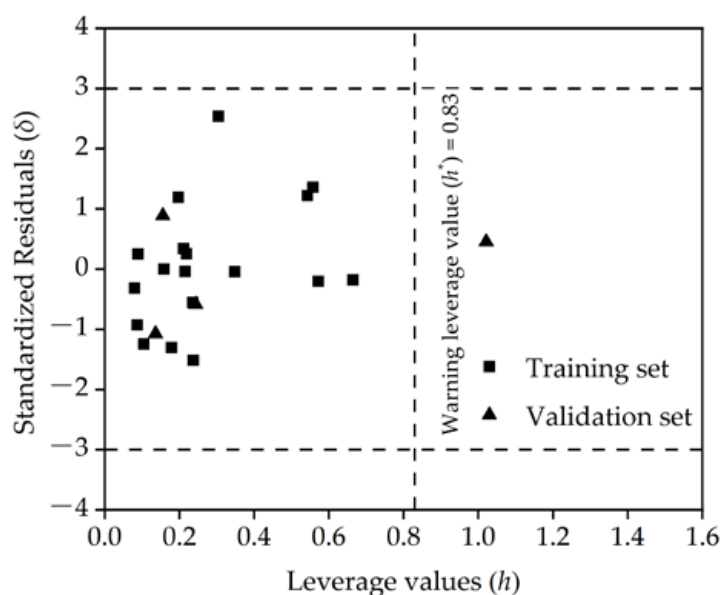**Figure 1.** Residual diagram of the optimal model.

As shown in Figure 2, the good consistency between the predicted value and the experimental value indicates that the established model has high prediction accuracy for the log K_{oc} of the PFASs.
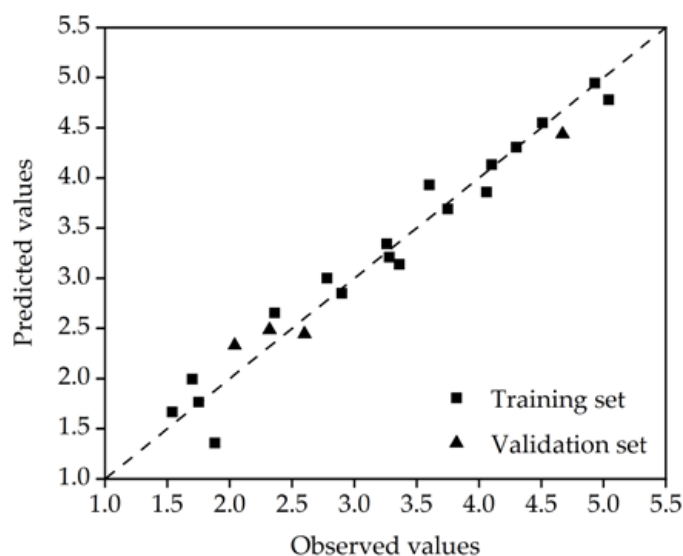
Figure 2. Observed and predicted values of the optimal model.

2.2. Application Domain

According to the Williams plot (Figure 3), none of the standardized residuals of the log K_{oc} of the PFASs obtained from the QSPR model exceeded the thresholds (-3, 3) [1]. In the test set, leverage values (h) > warning leverage value (h^*) ($h^* = 0.83$) of 6:2 chlorinated polyfluorinated ether sulfonate (6:2 Cl-PFAES), indicating that this substance was structurally quite different from most of the PFASs in the training set; 6:2 Cl-PFAES is an emerging PFAS that has been widely used in industry as a substitute for traditional PFASs (such as perfluorooctane sulfonic acid (PFOS)) [32,33]. The QSPR model predicts that the standardized residual of its log K_{oc} value is 0.454, which does not exceed the threshold. It can be seen that the QSPR model has a wide range of applications and strong generalization ability, which can successfully predict not only traditional PFASs but also emerging PFASs.

Figure 3. Williams diagram of the optimal model.

2.3. Mechanistic Interpretation of the Model

The standardized regression coefficient refers to the regression coefficient when all variables are expressed in standardized form. Since the same measurement unit is used, the independent variables are more comparable [34]. The standardized regression coefficients of *MW*, μ, E_{LUMO}, and $ALIE_{\text{max}}$ in the QSPR model were 1.048, −0.219, −0.495, and −0.362, respectively. Based on a comparison of their absolute values, it can be seen that the influence of the four molecular descriptors on log K_{oc} was $MW > E_{\text{LUMO}} > ALIE_{\text{max}} > \mu$.

MW is related to the molecular size and hydrophobicity and can reflect the effect of molecules on the formation and destruction of holes in water [35]. When *MW* increases, this increases the PFASs' molecular size and strengthens their hydrophobicity [36,37]. At this time, the energy of the adsorbate required for the formation of holes between water molecules will lead to stronger hydrophobic repulsion of water on the surface of the PFAS molecules [38], thus driving the adsorption of PFASs in solid media (such as sediment or soil) [17]. In this study, the QSPR model showed that *MW* was positively correlated with log K_{oc}. When *MW* increased, the log K_{oc} of PFSAs showed an increasing trend, and the log K_{oc} of PFCAs roughly showed an increasing trend (except for perfluorobutanoic acid (PFBA); the possible explanation is given below). This result is consistent with the previously reported results that the log K_{oc} of the same type of PFASs usually increases with the increase in the carbon chain [39]. PFASs of the same class have similar structures (with the same functional groups). With the increase in the carbon chain length, its *MW* and hydrophobicity increased, which promoted the adsorption of PFASs in solid medium and increased their log K_{oc}. In addition, the size of *MW* explained the change of log K_{oc} of different types of PFASs to a certain extent. For example, perfluorododecanoic acid (PFDoDA), 6:6 perfluoroalkyl phosphinic acid (6:6 PFPiA), and n-ethyl perfluorooctane sulfonamidoacetic acid (N-EtFOSAA) have the same carbon chain length (12 carbon atoms); their *MW* are 6:6 PFPiA > PFDoDA > N-EtFOSAA; their μ are N-EtFOSAA > PFDoDA > 6:6 PFPiA; their E_{LUMO} are PFDoDA > 6:6 PFPiA > N-EtFOSAA; their $ALIE_{\text{max}}$ are N-EtFOSAA > 6:6 PFPiA > PFDoDA, and their log K_{oc} are 6:6 PFPiA > PFDoDA > N-EtFOSAA, which shows the same ranking as *MW*.

E_{LUMO} reflects the ability of molecules to receive electrons [40]. When E_{LUMO} is more positive, it is more difficult for molecules to obtain electrons from the external environment. $ALIE_{\text{max}}$ is the average energy required to ionize electrons at any point in molecular space [41], which is related to their molecular electrostatic potential, electronegativity, hardness, and other properties [42]. When $ALIE_{\text{max}}$ is smaller, the electron activity is

stronger, and electrophilic and free radical reactions are more likely to occur [43]. According to the QSPR model, E_{LUMO} and $ALIE_{max}$ were negatively correlated with log K_{oc}, which reflects the electrostatic interaction between molecules. These two molecular descriptors were used to explain why the molecular weights of perfluoroheptanoic acid (PFPeA) and perfluorohexanoic acid (PFHxA) were both larger than that of PFBA but their log K_{oc} was smaller than that of PFBA. From the perspective of E_{LUMO}, it is more difficult for PFBA to obtain electrons from the external environment than PFPeA and PFHxA, but the difference between the compounds is small. From the perspective of $ALIE_{max}$, PFBA has stronger electronic activity than PFPeA and PFHxA and is more prone to electrophilic reactions, and the difference between the compounds is larger. Compared with PFPeA and PFHxA, PFBA may be more capable of electrostatic interaction with the external environment, increasing its log K_{oc}. This explanation is consistent with the results reported previously that electrostatic interactions may be the main factor affecting the adsorption of short-chain PFASs in solid-phase media [44].

μ is often used to describe the intermolecular dipole–dipole interactions in QSPR studies [45]. As μ increases, ionic compounds are more easily solvated in liquid phases (water) [46], and PFASs are adsorbed with greater difficulty in solid phases (aqueous aerosol) [1]. It can be seen from the QSPR model that μ was negatively correlated with log K_{oc}. The larger the μ value, the easier it is to solvate in water, so it is more difficult for PFASs to be adsorbed in solid media. For example, N-EtFOSAA is the precursor of perfluorooctane sulfonamide (PFOSA) [47]; the two have similar structures and they have the same carbon chain length; only the terminal functional groups are different. The *MW* of N-EtFOSAA is larger than that of PFOSA, and the $ALLE_{max}$ is smaller, but the log K_{oc} of PFOSA is much larger than that of N-EtFOSAA. In addition to the influence of E_{LUMO}, it is mainly due to the μ of N-EtFOSAA being relatively large (nearly threefold).

2.4. Model Comparison

The QSPR model with high R^2 was developed using the molar volume as a single molecule descriptor [27], but the robustness and external prediction ability of the model were not verified in the process of model construction, and the log K_{oc} prediction ability of PFASs with molar volume less than 160 $cm^3 \cdot mol^{-1}$ was limited. In this study, the robustness and external prediction ability of the developed QSPR model were verified through the Q^2_{LOO} and Q^2_{F1}, Q^2_{F2}, and Q^2_{F3}. The results show that the model has good robustness and external prediction ability. For example, the model built by Brusseau had prediction errors of 1.4 and 0.5 for PFBA and PFPeA (two PFASs with molar volumes of less than 160 $cm^3 \cdot mol^{-1}$) [27], while the QSPR model developed in this study had prediction errors of 0.52 and −0.13 for PFBA and PFPeA, respectively. The log K_{oc} prediction ability of this model is better for PFASs with a smaller molar volume.

A QSPR model was developed using nine molecular descriptors based on the structural properties of 824 compounds [26], but only six PFASs were included in these compounds. The QSPR model established by Wang et al. has a lower prediction accuracy of log K_{oc} for PFASs than the QSPR model, which only takes PFASs as the modeling data set [26]. The QSPR model developed in this study uses more PFAS modeling datasets and fewer molecular descriptors to obtain better model statistical parameters, and the log K_{oc} prediction accuracy for PFASs is higher. For example, the QSPR model developed by Wang et al. has prediction errors of −0.38 and 0.18 for traditional PFASs (such as perfluorooctanoic acid (PFOA) and PFOS) [26], while the prediction errors of this model for PFOA and PFOS are only −0.29 and 0.07, respectively.

To sum up, compared with the models developed in previous studies (Table 3), the data set used in this study has more types and numbers of PFASs, the developed QSPR model has higher prediction accuracy, and the model has a wider application field. It covers a large number of structurally diverse PFASs, including not only traditional PFASs but also emerging PFASs. In addition, the modeling process of this study is completely based on the standard framework of the QSPR model [28], which not only establishes a simple linear

relationship between the structural properties of PFASs and their log K_{oc} but also analyzes the distribution behavior of PFASs between sediment/soil and water in detail through the mechanical interpretation of the QSPR model.

Table 3. Comparisons of models in the current and earlier studies.

n	Chemicals	R^2	*RMSE*	Q^2_{LOO}	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	**Reference**
12	PFCAs, PFSAs	0.98	0.200	NR	NR	NR	NR	[27]
824 *	PFCAs, PFSAs	0.854	0.472	0.850	0.761	NR	NR	[26]
22	PFCAs, PFSAs, FOSAs, PFPiAs, and other emerging PFASs	0.962	0.212	0.920	0.961	0.955	0.959	This study

Notes: NR: not reported; *: the 824 compounds in the dataset contain only six PFASs; FOSAs: perfluoroalkane sulfonamide; PFPiAs: perfluoroalkyl phosphinic acid.

3. Materials and Methods

3.1. Data Collection and Processing

The log K_{oc} values were collected from the research literature on the adsorption of PFASs in sediments and soils [17,21,48–53], including 11 PFCAs, 5 PFSAs, 1 perfluoroalkane sulfonamide (FOSAS), 1 perfluoroalkyl phosphinic acid (PFPiAs), and 4 other PFASs. Since the experimental data in the previous studies were measured by different experimenters and under different experimental environments, in order to ensure the reliability of the data, this study first removed the outliers that obviously deviated from the overall data samples from the multiple log K_{oc} experimental values of the same PFASs collected and then calculated the average value to develop the QSPR model.

The log K_{oc} values of 22 PFASs ranged from 1.54 to 5.04, the span range was 3.50, the average value (mean) was 3.22, and the corresponding standard deviation (SD) was 1.11. All data fell within the interval of (mean −3SD, mean +3SD) and did not require further processing [54]. In total, 80% of the data in the dataset were randomly selected as the training set (18 PFASs) for developing the QSPR model; the remaining 20% of the data were the test set (4 PFASs) for external validation of the model. Details about PFASs, experimental values and references pertaining to the modelling and external validation sets are given in Table S1 in the Supplementary Materials.

3.2. Calculation of Molecular Descriptors

The B3LYP/6-31G* algorithm in the Gaussian program package (ver. G09W, Michael F, Wallingford, CT, USA) was used to optimize the molecular structure of the PFASs in the neutral electron ground state, and the stable molecular configuration of the PFASs with the lowest energy was obtained. The Multiwfn program (ver. 3.8, Lu T, Beijing, China) [55] calculated the optimized molecular structure of the PFASs and obtained 62 molecular descriptors, including the molecular structure features, orbital energy levels, electronegativity, atomic charge, polarity, and other physical and chemical information about the PFASs. The multiple physicochemical properties of the PFASs were successfully predicted using these molecular descriptors [1,56].

3.3. Model Development and Validation

Firstly, correlation analysis was performed between all molecular descriptors. For molecular descriptors with a correlation coefficient (R) higher than 0.9, only one molecular descriptor with a high correlation coefficient with log K_{oc} was retained. Secondly, based on SPSS 26 software (IBM SPSS Inc., Chicago, IL, USA), the retained molecular descriptors were taken as independent variables with log K_{oc} as the dependent variable to perform a stepwise linear regression to obtain the QSPR models containing different numbers of molecular descriptors. Lastly, the optimal QSPR model with the largest adjusted coefficient of determination (R^2_{adj}) and the smallest root mean square error (*RMSE*) was selected as the final model in this study. R, R^2_{adj} and *RMSE* were obtained by SPSS 26 software (IBM SPSS Inc., Chicago, IL, USA).

According to the QSPR model construction guidelines [28], the QSPR model should have sufficient goodness of fit, robustness, and predictive ability. In this study, the R^2 was used to evaluate the goodness of fit of the QSPR model, the Q^2_{LOO} was used to evaluate the robustness of the QSPR model, the test set was used to externally validate the QSPR model, and the Q^2_{F1}, Q^2_{F2}, Q^2_{F3} were used to evaluate the prediction ability of the QSPR model. In addition, in order to further verify the reliability of the developed model, the error distribution of the model prediction was used to evaluate whether the model had systematic errors; the *p* and *VIF* of the molecular descriptors contained in the QSPR model were used to determine whether each molecular descriptor was significant and whether there was multicollinearity among the molecular descriptors.

3.4. Application Domain

A Williams diagram [57] was used to characterize the application domain of the QSPR model, evaluate its scope of application, and determine whether there were outliers in the modeling samples. The composition of the Williams diagram and its calculation method are described in the Supplementary Materials.

4. Conclusions

In this study, we successfully developed an optimal QSPR model to predict the log K_{oc} of PFASs. The dataset of this model includes 22 PFASs in eight different categories, covering the common PFASs in current industrial production and daily life, and the model has a wide range of applicability. The comprehensive verification and evaluation of the model show that the developed model has sufficient goodness of fit, robustness, and predictive ability and can accurately predict the log K_{oc} of PFASs (within the application field defined by the model).

Through the mechanistic interpretation of the model, we found that the *MW*, E_{LUMO}, $ALIE_{\text{max}}$, and μ of PFASs are the main structural factors affecting the partitioning behavior between the solid and liquid phases, and the order of influence is $MW > E_{\text{LUMO}} > ALIE_{\text{max}} > \mu$. Specifically, *MW* reflects the hydrophobic property of the compound, μ reflects the polarity of the compound, while E_{LUMO} and $ALIE_{\text{max}}$ are related to the electrostatic interaction between molecules. The partitioning behavior of PFASs between the two phases is the result of the joint influence of multiple mechanisms. The hydrophobic interaction, electrostatic interaction, and dipole–dipole interaction play key roles in determining the partitioning of PFASs between the two phases. PFASs that can produce a strong hydrophobic interaction tend to be distributed in solid–phase media. The results of this study are of great significance to understanding the migration behavior and environmental fate of PFASs between sediment/soil and water, providing basic data for further environmental risk assessment.

In future research, the relationship between the log K_{oc} of PFASs and other partition coefficients can be compared and analyzed in order to illustrate the transport of PFASs in multiphase media as well. Moreover, new PFASs with less impact on the environment can be designed based on the structural factors that affect the distribution behavior of PFASs to reduce the environmental load caused by this kind of compound.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/molecules27175610/s1. 1. The five guidelines for QSPR construction; 2. Descriptor selection; 3. Model validation 4. The application domain; Table S1. The values of log K_{oc} for PFASs; Table S2. Description of the descriptors generated from Multiwfn; Table S3. Statistical parameters of QSPR model; Table S4. The observed and predicted log K_{oc} values of PFASs. The values of the descriptors used in the QSPR models.

References

1. Li, Y.; Yu, X.Y.; Chen, X.; Yin, J.; Zhong, W.J.; Zhu, L.Y. Underlying mechanisms for the impacts of molecular structures and water chemistry on the enrichment of poly/perfluoroalkyl substances in aqueous aerosol. *Sci. Total Environ.* **2022**, *803*, 150003. [CrossRef]
2. Yuan, Q.; Ma, G.C.; Xu, T.; Serge, B.; Yu, H.Y.; Chen, J.R.; Lin, H.J. Developing QSPR model of gas/particle partition coefficients of neutral poly-/perfluoroalkyl substances. *Atmos. Environ.* **2016**, *143*, 270–277. [CrossRef]
3. Ahrens, L.; Bundschuh, M. Fate and effects of poly-and perfluoroalkyl substances in the aquatic environment: A review. *Environ. Toxicol. Chem.* **2014**, *33*, 1921–1929. [CrossRef]
4. Zhang, W.; Zhang, Y.T.; Taniyasu, S.; Yeung, L.W.Y.; Lam, P.K.S.; Wang, J.S.; Li, X.H.; Yamashita, N.; Dai, J.Y. Distribution and fate of perfluoroalkyl substances in municipal wastewater treatment plants in economically developed areas of China. *Environ. Pollut.* **2013**, *176*, 10–17. [CrossRef]
5. Goodrow, S.M.; Ruppel, B.; Lippincott, R.L.; Post, G.B.; Procopio, N.A. Investigation of levels of perfluoroalkyl substances in surface water, sediment and fish tissue in New Jersey, USA. *Sci. Total Environ.* **2020**, *729*, 138839. [CrossRef]
6. Liu, J.; Zhao, X.R.; Liu, Y.; Qiao, X.C.; Wang, X.; Ma, M.Y.; Jin, X.L.; Liu, C.Y.; Zheng, B.H.; Shen, J.S.; et al. High contamination, bioaccumulation and risk assessment of perfluoroalkyl substances in multiple environmental media at the Baiyangdian Lake. *Ecotoxicol. Environ. Saf.* **2019**, *182*, 109454. [CrossRef]
7. Li, M.D. Numerical Simulation of Migration and Transformation of Perfluorinated Compounds in the Atmosphere of Shanghai. Master's Thesis, Shanghai Jiao Tong University, Shanghai, China, 2016.
8. Xiao, F.; Simcik, M.F.; Halbach, T.R.; Gulliver, J.S. Perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) in soils and groundwater of a U.S. metropolitan area: Migration and implications for human exposure. *Water Res.* **2015**, *72*, 64–74. [CrossRef]
9. Fardin, O.; Don, K.; Roland, W.; Alan, W. PFOS and PFC releases and associated pollution from a PFC production plant in Minnesota (USA). *Environ. Sci. Pollut. Res* **2013**, *20*, 1977–1992.
10. Gao, L.J.; Liu, J.L.; Bao, K.; Chen, N.N.; Meng, B. Multicompartment occurrence and partitioning of alternative and legacy per- and polyfluoroalkyl substances in an impacted river in China. *Sci. Total Environ.* **2020**, *729*, 138753. [CrossRef]
11. Ren, Y.Y.; Jin, Y.E.; Xu, H.H.; Qian, H.L.; Zheng, W.W.; Wu, C.; Guo, C.Y. Pollution status and risk assessment of perfluorinated compounds in drinking water in Shanghai. *J. Environ. Occup. Med.* **2020**, *37*, 1089–1094.
12. Xie, L.; Zhang, T.; Sun, H.W. The enrichment characteristics of perfluoroalkyl compounds in human liver and their relationship with liver injury. *Environ. Chem.* **2020**, *39*, 1479–1487.
13. Kar, S.; Sepulveda, M.S.; Roy, K.; Leszczynski, J. Endocrine-disrupting activity of per- and polyfluoroalkyl substances: Exploring combined approaches of ligand and structure based modeling. *Chemosphere* **2017**, *184*, 514–523. [CrossRef] [PubMed]
14. Mariussen, E. Neurotoxic effects of perfluoroalkylated compounds: Mechanisms of action and environmental relevance. *Arch. Toxicol.* **2012**, *86*, 1349–1367. [CrossRef]
15. Zhang, X.M.; Song, J.L.; Jin, Y.H.; Liu, W.; Liu, L.; Yu, H.Y. Effects of perfluorooctane sulfonic acid on reproductive toxicity of male quail. *Asian J. Ecotoxicol.* **2011**, *6*, 143–148.
16. Lau, C.; Anitole, K.; Hodes, C.; Lai, D.; Pfahles-Hutchens, A.; Seed, J. Perfluoroalkyl Acids: A Review of Monitoring and Toxicological Findings. *Toxicol. Sci.* **2007**, *99*, 366–394. [CrossRef]
17. Chen, M.; Wang, Q.; Shan, G.Q.; Zhu, L.Y.; Yang, L.P.; Liu, M.L. Occurrence, partitioning and bioaccumulation of emerging and legacy per- and polyfluoroalkyl substances in Taihu Lake, China. *Sci. Total Environ.* **2018**, *634*, 251–259. [CrossRef]
18. Pandey, S.K.; Roy, K. QSPR modeling of octanol-water partition coefficient and organic carbon normalized sorption coefficient of diverse organic chemicals using Extended Topochemical Atom (ETA) indices. *Ecotoxicol. Environ. Saf.* **2021**, *208*, 111411. [CrossRef]
19. Razzaque, M.M.; Grathwohl, P. Predicting organic carbon-water partitioning of hydrophobic organic chemicals in soils and sediments based on water solubility. *Water Res.* **2008**, *42*, 3775–3780. [CrossRef]

20. Li, D.Y.; Zhang, J.L.; Liang, C.S. Empirical Estimation Method of Soil-Water Partition Coefficient of Organic Pollutants in Underground Environment. In Proceedings of the 4th National Symposium on Young Geologists, The 4th National Symposium for Young Geologists, Beijing, China, 1 October 1999; Acta Geoscientia Sineca: Beijing, China, 1999; pp. 707–712.

21. Hu, H.M.; Zhang, Y.Y.; Zhao, N.; Xie, J.H.; Zhou, Y.Q.; Zhao, M.R.; Jin, H.B. Legacy and emerging poly- and perfluorochemicals in seawater and sediment from East China Sea. *Sci. Total Environ.* **2021**, *797*, 149052. [CrossRef]

22. Jiao, L. *Application of Topological Index: Quantitative Structure-Property Relationship between Hydrocarbons and Persistent Organic Pollutants*; China Petrochemical Press: Beijing, China, 2017.

23. Schultz, T.W.; Cronin, M.T.D.; Walker, J.D.; Aptula, A.O. Quantitative structure-activity relationships (QSARs) in toxicology: A historical perspective. *J. Mol. Struct. THEOCHEM* **2003**, *622*, 1–22. [CrossRef]

24. Zhu, T.Y.; Chen, W.X.; Cheng, H.M.; Wang, Y.J.; Singh, R.P. Prediction of polydimethylsiloxane-water partition coefficients based on the pp-LFER and QSAR models. *Ecotoxicol. Environ. Saf.* **2019**, *182*, 109374. [CrossRef] [PubMed]

25. Zhu, T.Y.; Chen, W.X.; Wu, J.; Singh, R.P.; Yan, B.P. Predicting low density polyethylene-water partition coefficients based on pp-LFER and QSPR models using molecular descriptors. *Fluid Phase Equilibria* **2020**, *506*, 112374. [CrossRef]

26. Wang, Y.; Chen, J.W.; Yang, X.H.; Lyakurwa, F.; Li, X.H.; Qiao, X.L. In silico model for predicting soil organic carbon normalized sorption coefficient (K_{OC}) of organic chemicals. *Chemosphere* **2015**, *119*, 438–444. [CrossRef]

27. Brusseau, M.L. Estimating the relative magnitudes of adsorption to solid-water and air/oil-water interfaces for per- and poly-fluoroalkyl substances. *Environ. Pollut.* **2019**, *254*, 113102. [CrossRef]

28. OECD. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q) SAR] Models*; Organisation for Economic Co-Operation and Development: Paris, France, 2007; Available online: http://www.OECD.Org/env/ehs/risk-assessment/guenvironment (accessed on 17 October 2020).

29. Kunal, R.; Supratik, K.; Rudra, N.D. *A Primer on QSAR/QSPR Modeling*; Springer: Berlin, Germany, 2015.

30. Kiralj, R.; Ferreira, M.M.C. Basic validation procedures for regression models in QSAR and QSPR studies: Theory and application. *J. Braz. Chem. Soc.* **2009**, *20*, 770–787. [CrossRef]

31. Lv, L.P.; Li, B.; He, S.H.; Li, H.; Xu, J.H. Prediction of azeotrope temperature of binary azeotropes containing low-carbon esters based on quantitative structure-activity relationship. *Chem. Eng.* **2019**, *47*, 44–49.

32. Zhou, X.J.; Wang, J.S.; Sheng, N.; Cui, R.N.; Deng, Y.Q.; Dai, J.Y. Subchronic reproductive effects of 6:2 chlorinated polyfluorinated ether sulfonate (6:2 Cl-PFAES), an alternative to PFOS, on adult male mice. *J. Hazard. Mater.* **2018**, *358*, 256–264. [CrossRef]

33. Gao, Y.X.; Deng, S.B.; Du, Z.W.; Liu, K.; Yu, G. Adsorptive removal of emerging polyfluoroalky substances F-53B and PFOS by anion-exchange resin: A comparative study. *J. Hazard. Mater.* **2017**, *323*, 550–557. [CrossRef]

34. Liu, Y.; Yu, Y.; Shi, B.Y.; Liu, S.M.; Wu, X. Analysis of the relative importance of factors affecting iron release in water supply network. *Environ. Sci.* **2017**, *38*, 5090–5096.

35. Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*; WILEY-VCH: Darmstadt, Germany, 2009.

36. Ding, G.H.; Shao, M.H.; Zhang, J.; Tang, J.Y.; Peijnenburg, W.J.G.M. Predictive models for estimating the vapor pressure of poly- and perfluorinated compounds at different temperatures. *Atmos. Environ.* **2013**, *75*, 147–152. [CrossRef]

37. Rayne, S.; Forest, K.; Friesen, K.J. Estimated bioconcentration factors (BCFs) for the C4 through C8 perfluorinated alkylsulfonic acid (PFSA) and alkylcarboxylic acid (PFCA) congeners. *J. Environ. Sci. Health Part A* **2009**, *44*, 598–604. [CrossRef]

38. Apul, O.G.; Perreault, F.; Ersan, G.; Karanfil, T. Linear solvation energy relationship development for adsorption of synthetic organic compounds by carbon nanomaterials: An overview of the last decade. *Environ. Sci. Water Res. Technol.* **2020**, *6*, 2949–2957. [CrossRef]

39. Guo, C.S.; Zhang, Y.; Zhao, X.; Du, P.; Liu, S.S.; Lv, J.P.; Xu, F.X.; Meng, W.; Xu, J. Distribution, source characterization and inventory of perfluoroalkyl substances in Taihu Lake, China. *Chemosphere* **2015**, *127*, 201–207. [CrossRef]

40. Cheng, Z.W.; Yang, B.W.; Chen, Q.C.; Ji, W.C.; Shen, Z.M. Characteristics and difference of oxidation and coagulation mechanisms for the removal of organic compounds by quantum parameter analysis. *Chem. Eng. J.* **2018**, *332*, 351–360. [CrossRef]

41. Sjoberg, P.; Murray, J.S.; Brinck, T.; Politzer, P. Average local ionization energies on the molecular surfaces of aromatic systems as guides to chemical reactivity. *Can. J. Chem.* **1990**, *68*, 1440. [CrossRef]

42. Politzer, P.; Murray, J.S.; Bulat, F.A. Average local ionization energy: A review. *J. Mol. Model.* **2010**, *16*, 1731–1742. [CrossRef]

43. Lu, T.; Chen, F.W. Quantitative analysis of molecular surface based on improved Marching Tetrahedra algorithm. *J. Mol. Graph. Model.* **2012**, *38*, 314–323. [CrossRef]

44. Zhao, L.X.; Zhu, L.Y.; Yang, L.P.; Liu, Z.T.; Zhang, Y.H. Distribution and desorption of perfluorinated compounds in fractionated sediments. *Chemosphere* **2012**, *88*, 1390–1397. [CrossRef]

45. Sang, P.; Zou, J.W.; Zhou, P.; Xu, L. QSPR modeling of bioconcentration factor of nonionic compounds using Gaussian processes and theoretical descriptors derived from electrostatic potentials on molecular surface. *Chemosphere* **2011**, *83*, 1045–1052.

46. Alvarez-Puebla, R.A.; Valenzuela-Calahorro, C.; Garrido, J.J. Theoretical study on fulvic acid structure, conformation and aggregation A molecular modelling approach. *Sci. Total Environ.* **2006**, *358*, 243–254. [CrossRef]

47. Zhao, S.Y.; Zhou, T.; Zhu, L.Y.; Wang, B.H.; Li, Z.; Yang, L.P.; Liu, L.F. Uptake, translocation and biotransformation of N-ethyl perfluorooctanesulfonamide (N-EtFOSA) by hydroponically grown plants. *Environ. Pollut.* **2018**, *235*, 404–410. [CrossRef]

48. Guelfo, J.L.; Higgins, C.P. Subsurface Transport Potential of Perfluoroalkyl Acids at Aqueous Film-Forming Foam (AFFF)-Impacted Sites. *Environ. Sci. Technol.* **2013**, *47*, 4164–4171. [CrossRef]

49. Pi, N.; Ng, J.Z.; Kelly, B.C. Uptake and elimination kinetics of perfluoroalkyl substances in submerged and free-floating aquatic macrophytes: Results of mesocosm experiments with Echinodorus horemanii and Eichhornia crassipes. *Water Res.* **2017**, *117*, 167–174. [CrossRef]

50. Chen, X.W.; Zhu, L.Y.; Pan, X.Y.; Fang, S.H.; Zhang, Y.F.; Yang, L.P. Isomeric specific partitioning behaviors of perfluoroalkyl substances in water dissolved phase, suspended particulate matters and sediments in Liao River Basin and Taihu Lake, China. *Water Res.* **2015**, *80*, 235–244. [CrossRef]

51. Labadie, P.; Chevreuil, M. Partitioning behaviour of perfluorinated alkyl contaminants between water, sediment and fish in the Orge River (nearby Paris, France). *Environ. Pollut.* **2011**, *159*, 1452–1453. [CrossRef]

52. Li, F.S.; Sun, H.W.; Hao, Z.N.; He, N.; Zhao, L.J.; Zhang, T.; Sun, T.H. Perfluorinated compounds in Haihe River and Dagu Drainage Canal in Tianjin, China. *Chemosphere.* **2011**, *84*, 265–271. [CrossRef]

53. Higgins, C.P.; Luthy, R.G. Sorption of perfluorinated surfactants on sediments. *Environ. Sci. Technol.* **2006**, *40*, 7251–7256. [CrossRef]

54. Zhang, L.; Song, X.; Wu, Y. *Theory, Methodology, Tools and Applications for Modeling and Simulation of Complex Systems*; Springer: Singapore, 2016.

55. Lu, T.; Chen, F.W. Multiwfn: A Multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592. [CrossRef]

56. Hu, C.; Wang, Z.Z.; Xu, W.B.; Xu, H.Y. QSPR Study on Vapor Pressure of Some Perfluorinated (and Polyfluorinated) Organic Compounds. *Chem. Res. Appl.* **2013**, *25*, 43–46.

57. Zhu, T.; Wu, J.; He, C.; Fu, D.; Wu, J. Development and evaluation of MTLSER and QSAR models for predicting polyethylene-water partition coefficients. *J. Environ. Manag.* **2018**, *223*, 600–606. [CrossRef]