刘 冰,邹可可,刘 辉,等.基于可解释性机器学习的 PFAS 土壤中分配机制 [J]. 中国环境科学, 2026,46(3):1460−1474.
Liu B, Zou K K, Liu H, et al. Explainable machine learning for uncovering PFAS adsorption mechanisms in soils [J]. China Environmental Science, 2026,46(3): 1460−1474.

基于可解释性机器学习的 PFAS 土壤中分配机制

刘 冰[1*],邹可可[2],刘 辉[1],时凯歌[1],陈燕敏[1],卢 鑫[1],韩帅军[1],古 励[3] (1.郑州师范学院化学化工学院,河南 郑州 450044；2.河南省平顶山水文水资源测报分中心,河南 平顶山 467000；3.重庆大学环境与生态学院,重庆 400044)

摘要：基于机器学习(ML)模型预测全氟和多氟烷基物质(PFAS)的土壤−水分配系数(K_d),并阐释其在土壤中的分配行为.数据集涵盖 47 种 PFAS 共 1227 个样本,采用 16 个 PFAS 物化性质及土壤理化参数作为特征.结果表明: lgK_d 范围为−1.40~3.95;当 CF_2 官能团数>5 且分子量(MW)>400g/mol 时,lgK_d 与 CF_2 数及 MW 呈显著线性关系(R^2 分别为 0.96 和 0.94);土壤特性对 PFAS 吸附的影响因类型而异:非两性离子 PFAS 主要受有机碳质量分数和粉砂质量分数影响,而两性离子 PFAS 则主要受粉砂与砂质量分数影响;经参数优化,ML 模型预测性能良好,R^2 为 0.85,RMSE 为 0.35,MAE 为 0.26;其中 MW,有机碳质量分数,水溶性(lg$S$)和净电荷密度为关键特征,贡献率分别为 31.6%,29.8%,19.7%和 10.0%.三维相互作用分析表明,高有机碳含量,高阳离子交换能力(CEC)及低 pH 值土壤与大 MW PFAS 结合时,lgK_d>0.40,吸附性强;反之,当 CEC<8.00cmol/kg,有机质含量<1.00%,MW<380g/mol 时,lgK_d<0.40,吸附能力较弱,环境迁移潜力较高.研究结果为预测 PFAS 吸附行为提供了可靠工具.

关键词：全氟和多氟烷基物质；可解释性机器学习；土壤−水分配系数；预测
中图分类号：X53 **文献标识码**：A **文章编号**：1000–6923(2026)03−1460−15

Explainable machine learning for uncovering PFAS adsorption mechanisms in soils. LIU Bing[1*], ZOU Ke-ke[2], LIU Hui[1], SHI Kai-ge[1], CHEN Yan-min[1], LU Xin[1], HAN Shuai-jun[1], GU Li[3] (1.School of Chemistry and Chemical Engineering, Zhengzhou Normal University, Zhengzhou 450044, China；2.Pingdingshan Hydrology and Water Resources Reporting Subcenter of Henan Province, Pingdingshan 467000, China；3.College of Environment and Ecology, Chongqing University, Chongqing 400044, China). *China Environmental Science*, 2026,46(3)：1460~1474

Abstract：The soil-water partition coefficient (K_d) of per- and polyfluoroalkyl substances (PFAS) was predicted using machine learning (ML) models, and their partitioning behavior in soils was elucidated. A dataset comprising 1,227 samples of 47 PFAS was employed, with 16 PFAS physicochemical properties and soil parameters used as input features. It was demonstrated that the lgK_d values ranged from −1.40 to 3.95. When the number of CF_2 groups exceeded 5and the molecular weight (MW) was greater than 400g/mol, significant linear relationships were observed between lgKd and both the CF_2 number and MW, with R^2 values of 0.96 and 0.94, respectively. The influence of soil properties on PFAS adsorption was found to be dependent on the type of PFAS: non-zwitterionic PFAS were mainly affected by organic carbon content and silt content, whereas zwitterionic PFAS were primarily influenced by silt and sand content. After parameter optimization, the ML model exhibited good predictive performance, achieving an R^2 of 0.85, an RMSE of 0.35, and an MAE of 0.26. MW, organic carbon content, water solubility (lg$S$), and net charge density were identified as the most important features, with contribution rates of 31.6%, 29.8%, 19.7%, and 10.0%, respectively. Three-dimensional interaction analysis indicated that when soils with high organic carbon content, high cation exchange capacity (CEC), and low pH were combined with PFAS of large MW, lgK_d exceeded 0.40, reflecting strong adsorption. Conversely, under conditions where CEC was below 8.00 cmol/kg, organic matter content was less than 1.00%, and MW was below 380g/mol, lgK_d was lower than 0.40, indicating weaker adsorption and higher potential for environmental migration. The findings provide a reliable tool for predicting the adsorption behavior of PFAS.

Key words：per-and polyfluoroalkyl substances；explainable machine learning；soil-water partition coefficient；prediction

全氟和多氟烷基物质(PFAS)可通过大气干湿沉降、污泥农用及污水灌溉等多种途径在土壤中累积,其中消防训练场等特定区域污染尤为严重,全氟辛酸(PFOA)和全氟辛烷磺酸(PFOS)的平均浓度分别高达 101,103ng/g[1-2].由于 PFAS 分子中碳−氟键的高键能(110kcal/mol),其表现出极高的化学稳定性与

收稿日期：2025−08−03
基金项目：国家自然科学基金资助项目(51208448);河南省高等学校重点科研项目(24B610016);河南省科技攻关项目(262102320243);河南省 2025 年水利科技攻关项目(编号 51)
* 责任作者, 副教授, liubing_982002@163.com

环境持久性,可在生物体内长期蓄积[3-5].研究表明,土壤中的 PFOA 可经植物吸收进入食物链,人体摄入后代谢极其缓慢,平均生物半衰期达 2~4a[6].这些特性使得 PFAS 成为对生态环境和人类健康具有长期危害的典型持久性有机污染物[6-7].

理解 PFAS 的环境行为,特别是其在土壤-水系统中的分配特征,对于准确评估该类污染物在土壤环境中的迁移转化规律及其生态与健康风险具有关键科学意义[8].在土壤环境研究中,K_d 是表征 PFAS 环境行为的重要参数.有研究表明,约 12000 种 PFAS 化合物的 K_d 需通过实验测定,这一过程不仅耗时且成本高昂.因此,基于现有数据构建 K_d 预测模型具有重要的应用价值.但传统预测方法存在明显局限性,通常假设 K_d 与土壤有机碳含量、阳离子交换量(CEC)、MW 及辛醇-水分配系数(K_{OW})等参数间存在简单的线性关系[9-12].虽然最近通过结合光谱表征技术与回归分析方法在一定程度上提高了预测准确性,但仍未能充分纳入土壤异质性的复杂影响[13].值得注意的是,过度依赖这类简化模型可能导致对 PFAS 环境行为的系统性误判,而忽视 PFAS 与土壤各组分间可能存在的协同/拮抗作用机制.

PFAS 在环境中的分配机制存在研究结论不一致的问题.部分研究发现 PFAS 浓度与土壤有机碳含量呈显著正相关(R^2>0.7)[14-15],而另一些研究则未发现显著相关性[16].这表明仅靠有机碳含量不能完全解释 PFAS 的吸附行为.另外,pH 值和 CEC 等参数的变化也可能导致分配结果出现差异[17-19].因此,PFAS 在土壤-水系统中的吸附行为需要通过多参数进行系统评估[20].研究结果的差异主要源于多环境因子间协同/拮抗作用的复杂性,以及 PFAS 自身化学特性(如链长、离子类型)的影响.要准确理解其在环境界面中的行为,需要采用更综合的建模方法,同时对多个关键参数进行全面评估.

机器学习(ML)模型具有较强的非线性关系识别能力,已成为预测 PFAS 环境行为的新技术.最新研究表明,通过整合多维特征变量,ML 模型可显著提升预测准确性[21-22].例如,ML 模型不仅可用于分类识别具有特定生物相分配倾向的化合物(如高血浆结合性 PFAS)[23],还能通过整合多源异构特征(涵盖分子结构、环境参数及生物生理指标等)实现关键过程参数(如根系富集因子)的精准回归预测[24].此外,基于 ML 构建的预测框架展现了良好的跨介质或场景外推能力[25].这些 ML 模型在数据有限的条件下仍展现出优异的预测性能,为 PFAS 的初步风险评估和污染优先级排序提供了可靠工具.特别值得关注的是,可解释性 ML 方法(如 SHAP 分析)能够解析驱动 PFAS 环境行为的关键因子及其贡献度.

本研究选择了 5 种 ML 模型对 K_d 预测开展研究,并集成特征贡献率排列、shapley 加法解释(SHAP)、部分依赖图(PDP)和 3D 交互图,揭示土壤和 PFAS 物理化学属性之间的相互依赖关系和吸附机制.

1　材料与方法

1.1　数据集中 PFAS 类型

表 1　数据集中 47 种 PFAS 化合物

Table 1　47 PFAS compounds in the datasets

序号	类别	中文名称	化学分子式	缩写	CAS 编号	样本数
1	全氟烷基羧酸(PFCA)	三氟乙酸	$C_2HF_3O_2$	TFA	76-05-1	20
		全氟丁酸	$C_4HF_7O_2$	PFBA	375-22-4	24
		全氟戊酸	$C_5HF_9O_2$	PFPeA	2706-90-3	25
		全氟己酸	$C_6HF_{11}O_2$	PFHxA	307-24-4	26
		全氟庚酸	$C_7HF_{13}O2$	PFHpA	375-85-9	24
		全氟辛酸	$C_8HF_{15}O_2$	PFOA	335-67-1	238
		全氟壬酸	$C_9HF_{17}O_2$	PFNA	375-95-1	33
		全氟癸酸	$C_{10}HF_{19}O_2$	PFDA	335-76-2	49
		全氟十一烷酸	$C_{11}HF_{21}O_2$	PFUnA	2058-94-8	30
		全氟十二酸	$C_{12}HF_{23}O_2$	PFDoA	307-55-1	24
		全氟十三酸	$C_{13}HF_{25}O_2$	PFTrA	72629-94-8	3
		全氟代十四酸	$C_{14}HF_{27}O_2$	PFTeA	376-06-7	2
2	全氟烷基磺酸(PFSA)	全氟丁磺酸	$C_4HF_9O_3S$	PFBS	375-73-5	44
		全氟戊烷磺酸	$C_5HF_{11}O_3S$	PFPeS	2706-91-4	12

续表 1

序号	类别	中文名称	化学分子式	缩写	CAS 编号	样本数
2	全氟烷基磺酸(PFSA)	全氟己烷磺酸	$C_6F_{13}O_3SH$	PFHxS	355-46-4	85
		全氟庚基磺酸	$C_7HF_{15}O_3S$	PFHpS	375-92-8	12
		全氟辛烷磺酸	$C_8HF_{17}O_3S$	PFOS	1763-23-1	274
		全氟壬烷磺酸	$C_9HF_{19}O_3S$	PFNS	68259-12-1	12
		全氟癸烷磺酸	$C_{10}HF_{21}O_3S$	PFDS	335-77-3	16
3	全氟烷基膦酸(PFPA)	全氟己基膦酸	$C_6H_2F_{13}O_3P$	PFHxPA	40143-76-8	7
		全氟辛基膦酸	$C_8H_2F_{17}O_3P$	PFOPA	40143-78-0	7
		全氟癸基膦酸	$C_{10}H_2F_{21}O_3P$	PFDPA	52299-26-0	7
4	全氟烷基磺酰胺(FOSA)	全氟丁基磺酰胺	$C_4H_2F_9NO_2S$	FBSA	30334-69-1	10
		全氟正己基磺酰胺	$C_6H_2F_{13}NO_2S$	FHxSA	41997-13-1	10
		全氟辛基磺酰胺	$C_8H_2F_{17}NO_2S$	PFOSA	754-91-6	25
		氟虫胺	$C_{10}H_6F_{17}NO_2S$	EtFOSA	4151-50-2	3
5	N-甲基/乙基全氟辛烷磺酰胺基乙酸(FOSAA)	2-(N-甲基全氟辛烷磺酰氨基)乙酸	$C_{11}H_6F_{17}NO_4S$	N-MeFOSAA	2355-31-9	5
		2-(N-乙基全氟辛烷磺酰氨基)乙酸	$C_{12}H_8F_{17}NO_4S$	N-EtFOSAA	2991-50-6	5
6	全氟烷基醚羧酸(PFECA)	全氟(2-甲基-3-氧杂己酸)	$C_6HF_{11}O_3$	GenX	13252-13-6	9
		4.8-二氧杂-3H-全氟壬酸	$C_7H_2F_{12}O_4$	ADONA	919005-14-4	10
7	氯化多氟烷基醚磺酸盐(PFAES)	乙磺酸杂质 7	$C_8HClF_{16}O_4S$	8:2Cl-PFAES	756426-58-1	15
8	氟调聚物醇(FTOH)	全氟丁基乙基醇	$C_6H_5F_9O$	4:2FTOH	2043-47-2	1
		全氟己基乙醇	$C_8H_5F_{13}O$	6:2FTOH	647-42-7	3
		全氟辛基乙基醇	$C_{10}H_5F_{17}O$	8:2FTOH	678-39-7	5
		1,1,2,2-四氢全氟十二醇	$C_{12}H_5F_{21}O$	10:2FTOH	865-86-1	3
9	氟调聚物磺酸盐(FTS)	1H,1H,2H,2H-全氟己烷磺酸	$C_6H_5F_9O_3S$	4:2FTS	757124-72-4	10
		全氟己基乙基磺酸	$C_8H_5F_{13}O_3S$	6:2FTS	27619-97-2	27
		1H,1H,2H,2H-全氟癸磺酸	$C_{10}H_5F_{17}O_3S$	8:2FTS	39108-34-4	26
10	阳离子型全氟辛烷磺酸	全氟辛酰胺基丙胺	$C_{14}H_{16}F_{15}N_2O^+$	PFOAAmS	45305-66-6	11
		N-[3-(二甲氨基)丙基]-全氟己烷磺酰胺	$C_{11}H_{13}F_{13}N_2O_2S$	AmPr-FHxSA	50598-28-2	10
		N-[3-氟-4-[(甲基氨基)羰基]苯基]-2-甲基丙氨酸	$C_{12}H_{15}F_{13}N_2O_2S$	TAmPr-FHxSA	38850-51-0	10
11	两性离子全氟辛烷磺酸	全氟烷基甜菜碱	$C_{15}H_{15}F_{17}N_2O_4S$	PFOSB	75046-16-1	5
		N-[3-[二(2-羟乙基)氨基]丙基]-2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-十五氟辛酰胺	$C_{15}H_{15}F_{15}N_2O_3$	PFOAB	90179-39-8	11
		N-[羧甲基-N,N-二甲基-3-[(3,3,4,4,5,5,6,6,7,7,8,8,8-十三氟代辛基)磺酰胺基]丙基]铵内盐	$C_{15}H_{19}F_{13}N_2O_4S$	6:2FtSaB	34455-29-3	22
		N-[羧甲基-N,N-二甲基-3-[(3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,10-十七氟代癸基)磺酰胺基]丙基]铵内盐	$C_{17}H_{19}F_{17}N_2O_4S$	8:2FtSaB	34455-21-5	5
		N-[羧甲基-N,N-二甲基-3-[(3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,11,11,12,12,12-二十一氟代十二烷基)磺酰胺基]丙基]铵内盐	$C_{19}H_{19}F_{21}N_2O_4S$	10:2FtSaB	34455-35-1	2

本研究数据集整合了 1997~2023 年发表的 48 篇文献数据,涵盖 11 类 47 种 PFAS 化合物,共 1227 个样本(表 1).(1)C_2~C_{14} 全氟烷基羧酸(PFCA)和(2)C_4~C_{10} 全氟烷基磺酸(PFSA),这两类在土壤和水体中分布广泛,浓度跨度从痕量至显著污染水平,且难以被微生物降解[26-28];(3)C_6~C_{10} 全氟烷基膦酸(PFPA)、(4)C_4~C_{10} 全氟烷基磺酰胺(FOSA)及(5)N-烷基全氟辛烷磺酰胺基乙酸(FOSAA),其水体浓度在 2.5~5.8ng/L[29];(6)全氟烷基醚羧酸(PFECA)和(7)氯化多氟烷基醚磺酸盐(PFAES)在工业排放水体中浓度较高,其中 GenX 和 8:2Cl-PFAES 可达 5 和 112μg/L[27];(8)氟调聚物醇(FTOH)和(9)氟调聚物磺酸盐(FTS)分别在大气(0.3μg/m³)和土壤(4ng/g)中检出[26,30];(10)阳离子及(11)两性离子全氟辛烷磺酸是轻水泡沫灭火剂的主要组份,主要存在于消防泡沫污染场地[31-32].各化合物环境行为差异显著,如三氟乙酸(TFA)易挥发且土壤吸附性弱,而长链 PFAS 更易在土壤中蓄积[28].

1.2　K_d 数据集构建

文献中实验批量数据必须符合以下标准.(1)排除纯有机相(如腐殖酸)或纯矿物相(如针铁矿)测得的吸附数据;(2)批量吸附实验遵循官方指南(Test No.106,OECD),在室温(20~25°C)条件下进行,达到吸附平衡(饱和条件);(3)如根据不同浓度的等温线数据得到多个 K_d 值,优先选择在低浓度范围下计算出的 K_d 值,以更好地代表实际环境中的 PFAS 浓度,或者选择在最低离子强度下测得的 K_d 值.

使用 3 种方法获得数据集中 K_d 值.(1)直接采用文献中报告的 K_d 值,参考文献中已计算出低单浓度实验的 K_d 值或由吸附等温线低浓度区间计算得出的 K_d 值;(2)文献报道了 K_{OC} 值,根据式(1)换算得到 K_d 值;(3)文献只报告了等温线模型(如 Freundlich 或 Langmuir)的拟合参数,从等温线拟合参数中推导 K_d 值(式(2)、(3)和(4)).

$$K_d = K_{OC} \cdot f_{OC} \tag{1}$$

式中:f_{OC} 为土壤或沉积物中的有机碳质量分数.

$$K_d = C_s/C_w \tag{2}$$

式中:C_S 为污染物在固相中的吸附浓度,mg/kg;C_W 为污染物在水相中未吸附浓度,mg/kg.

Freundlich 模型:$K_d=KF\cdot C_W^{(1/n-1)}$ (3)

式中:KF 为 Freundlich 系数,(mg/kg)/(mg/L)$^{1/n}$;1/n 为 Freundlich 系数.

Langmuir 模型:$K_d=[(KL\cdot Q_{max}\cdot C_W)/(1+KL\cdot C_W)]/C_W$ (4)

式中:KL 为 Langmuir 常数,L/mg;Q_{max} 为 Langmuir 最大吸附容量,mg/kg.

1.3　ML 模型中特征工程

ML 模型特征变量(表 2)主要包括两个部分,(1)PFAS 的物理化学性质包括:MW、lgK_{OW}、lg$S$、pKa、净电荷密度、C 原子数、F 原子数、C/F 质量比、F 质量分数和 CF_2 官能团数;(2)土壤理化参数包括:pH 值、有机碳质量分数、CEC、土壤质地(砂质量分数、粉砂质量分数、黏土质量分数).

表 2　特征变量及其分类

Table 2　Feature variables and classification

特征变量	参数	个数
PFAS 的物理化学特性	分子量(MW)、辛醇−水分配系数(lgK_{OW})、水溶性(lg$S$)、酸离解常数(pKa)、净电荷密度、C 原子数、F 原子数、C/F 质量比、F 质量分数、CF_2 官能团数	10
土壤理化参数	pH 值、有机碳质量分数、阳离子交换容量(CEC)、砂质量分数、粉砂质量分数、黏土质量分数	6

1.4　数据填充和标准化处理

数据集缺失的数据主要集中在 pH 值、CEC 和土壤质地(砂粒含量、淤泥含量、黏土含量),缺失占比分别 4.01%、20.36%和 8.13%,使用 K 最近邻模型插补缺失数据.为消除数据集中不同特征之间量纲的影响,研究中采用 Z−score 标准化处理方法.将原始数据的均值(mean)和标准差(standard deviation)进行数据的标准化,标准化公式见式(5).

$$Z_i = \frac{(X_i - \mu)}{\sigma} \tag{5}$$

式中:Z_i 是标准化后的数据;X_i 为需要标准化的数据;μ 是数据的均值;σ 是数据的标准差.

1.5　ML 模型建立

本研究采用常用的 5 种机器学习模型:多元线性回归(MLR)、支持向量回归(SVR)、随机森林(RF)、梯度提升回归树(GBRT)和极限梯度提升(XGBoost),这些模型在基本原理、训练方法和应用场景方面各具特点.

SVR 作为一种基于核方法的非线性回归模型,通过最大化间隔并引入 ε−不敏感损失函数,在高维特征空间中实现鲁棒的回归拟合.RF 则通过聚合多个决策树来提高预测精度和稳定性.GBRT 作为一种集成学习方法,通过迭代地添加决策树来修正前序树的残差误差.XGBoost 作为一种高效的梯度提升框架,通过引入正则化项和二阶泰勒展开优化损失函数,显著提升了模型的泛化能力和收敛速度.此外,MLR 模型与其他模型不同,其假设自变量与因变量之间存在线性关系[33].

为提高模型性能,采用了 5 折交叉验证和贝叶斯优化方法.交叉验证可增强模型鲁棒性,而贝叶斯

优化则用于超参数调优.超参数作为模型外部设定的关键参数,直接影响学习过程,对优化模型性能至关重要[34].与训练过程中获得的模型参数不同,超参数需在训练前预先设定,包括学习率、集成方法中决策树数量等关键设置[35].

1.6 ML 模型评估与验证

ML 模型评估指标包括相关系数(R^2)、均方根误差(RMSE)和平均绝对误差(MAE).R^2越接近 1,表明模型拟合度越好,而较低的 RMSE 和 MAE 值则反映更高的预测精度.此外,还通过交叉验证计算的预测决定系数(Q^2)评估模型泛化能力,基于训练集的交叉验证结果,通过比较各折预测值与实际值获得,Q^2接近 1 表示 ML 模型具有优秀的预测能力[36].模型训练前,数据集按 8:2 比例随机划分为训练集和测试集[37].内部验证采用 5 折交叉验证法,将数据集均分为 5 份,每次取 1 份作为验证集,其余 4 份作为训练集,循环五次确保每个数据点都参与训练和验证[38].外部验证则使用未参与任何训练环节的 20%独立测试集进行,进一步确保模型的泛化能力[38].

1.7 ML 模型解释

首先采用特征贡献率排列,通过随机打乱特征值并观测模型性能变化来评估各特征的重要性[39].其次应用 SHAP 值解析单个特征对特定预测的贡献度,揭示排除某特征后预测值的变化规律[40].接着采用部分依赖图(PDP)可视化特征与预测值的关系,表示出特征在整个数据集上的平均影响,提供宏观层面的特征作用规律[41].最后,通过三维交互作用图解析两个特征的协同效应对 ML 模型预测的影响.

2 结果与讨论

2.1 数据集分布范围

对数据集中变量特征和预测目标进行描述性统计分析,结果如表 3 所示.变异系数(CV)范围为 19.05%~306.80%,F 质量分数和 pH 值变异系数(CV)相对较小(9.21%和 19.05%).但是,总体数据差异显著,分异程度较高.47 种 PFAS 的 lgK_d 范围为−1.40~3.95;MW 范围 113.93~770.07g/mol;C 原子和 CF_2 官能团数分别为 2~19 和 1~13;F 质量分数在 43.33%~71.85%之间;净电荷密度在−854~188C/g 之间;土壤 pH 值从 2.76~8.96;土壤中有机碳质量分数范围为 0.03%~53.60%,但大多数土壤样本属于有机碳含量较低类型(<2.00%).

表 3 数据集中预测目标和特征统计值

Table 3 Statistical values of targets and features in the datasets

指标	最小值	最大值	均值	标准差	变异系数(%)
MW(g/mol)	113.93	770.07	446.45	98.76	22.12
lgK_{OW}	0.90	9.00	4.65	1.24	26.64
lg$S$	−2.69	7.99	2.68	1.62	60.69
pKa	−0.20	14.20	0.72	2.20	306.80
净电荷密度(C/g)	−853.9	187.56	−221.42	118.99	53.74
C 原子数	2	19	8.07	2.46	30.41
F 原子数	3	27	15.04	3.62	24.06
CF_2 官能团数	1	13	7.03	1.81	25.76
C/F 质量比	0.28	0.73	0.34	0.08	23.56
F 质量分数(%)	43.33	71.85	63.86	5.88	9.21
pH 值	2.76	8.96	6.40	1.22	19.05
CEC(cmol/kg)	0.05	140.00	19.03	21.13	110.99
有机碳质量分数(%)	0.03	53.60	4.71	10.51	223.05
砂质量分数(%)	0.00	100.00	46.21	22.97	49.72
粉砂质量分数(%)	0.00	100.00	27.62	19.64	71.11
黏土质量分数(%)	0.00	69.30	25.59	16.87	65.90
K_d 对数值(lgK_d)	−1.40	3.95	0.77	0.92	119.03

2.2 碳链、MW 和分子结构对 K_d 的影响

2.2.1 碳链对 K_d 的影响 数据集中 47 种 PFAS 化合物 lgK_d 统计值如图 1 所示.根据碳链数,47 种 PFAS 化合物分为 3 组:短链 PFAS(C 原子数≤6)、长链 PFAS(C 原子数≥7)和两性离子 PFAS.另外,lgK_d 统计值随着碳链数和 CF_2 官能团数变化趋势如图 2 所示.

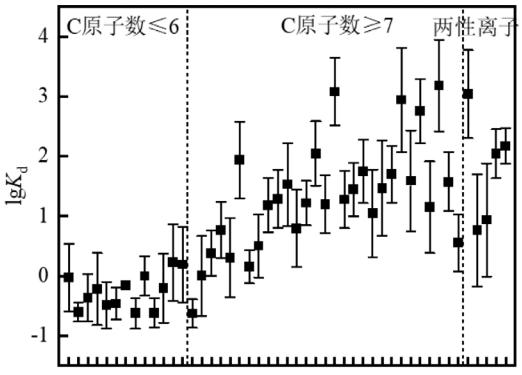


图 1 数据集中 PFAS 的 lgK_d 分布

Fig.1 The distribution of lgK_d values for PFAS in the datasets

横刻度值从左到右分别为 TFA、FBSA、PFBA、PFBS、PFPeA、PFPeS、4:2FTOH、4:2FTS、FHxSA、GenX、PFHxA、PFHxPA、PFHxS、ADONA、PFHpA、PFHpS、6:2FTOH、6:2FTS、8:2Cl-PFAES、PFEtCHxS、PFOA、PFOPA、PFOS、PFOSA、PFNA、PFNS、8:2FTOH、8:2FTS、EtFOSA、PFDA、PFDPA、PFDS、AmPr-FHxSA、N-MeFOSAA、PFUnA、10:2FTOH、N-EtFOSAA、PFDOA、TAmPr-FHxSA、PFTrA、PFOAAmS、PFOAB、PFTeA、6:2FtSaB、PFOSB、8:2FtSaB 和 10:2FtSaB

从图 2(a)可以知道,短链 PFAS 的 lgK_d 值与碳链数无相关性,这表明碳链长度并不能决定土壤对短链 PFAS 的吸附性能,lgK_d 值差异可能更多归因于短链 PFAS 离子基团与土壤表面相互作用[41]相.对于长链 PFAS,短链 PFAS 的 lgK_d 值为−0.48~0.003,优先与土壤中的水相结合,具有高度流动性,受土壤性质的影响较小,这对 FBSA 和 ADONA 来说尤为明显.长链 PFAS 的 lgK_d 值与碳链数具有较强正相关性(R^2=0.93),平均增加 1 个碳链数 lgK_d 值提高 0.38 对数单位.这表明对于长链 PFAS 疏水性是吸附的主要驱动力,这与之前的大多数吸附研究一致[42-43].此外,对于两性离子 PFAS,lgK_d 值与碳链数具有正相关性(R^2=0.82),平均增加 1 个碳链数 lgK_d 值提高 0.34 对数单位.

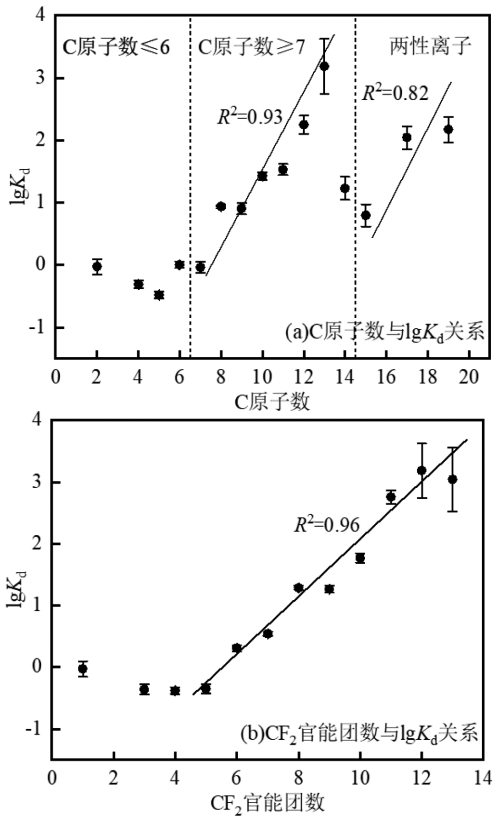


图 2　*C* 原子和 CF_2 官能团数对 lgK_d 值的影响

Fig.2　Impact of C atoms and CF_2 groups on lgK_d values

从图 2(b)可以知道,lgK_d 值随着 CF_2 官能团数的增加而上升.当 CF_2 官能团数小于 6,lgK_d 值相近.当 CF_2 官能团数≥6 时,lgK_d 值与 CF_2 官能团数具有较强正相关性(R^2=0.96),平均增加 1 个 CF_2 官能团数 lgK_d 值提高 0.48 对数单位.此外,该趋势也适用于两性离子 PFAS.总的来说,对于 CF_2 官能团数≥6 的 PFAS 化合物,其与土壤有机物质的疏水相互作用更加有效[44-45].

2.2.2　MW 对 K_d 的影响　从图 3 可知,根据 MW 分为 7 组:100~200, 200~300, 300~400, 400~500, 500~600, 600~0.57, 700~800g/mol.对于小 MW 的 PFAS (MW<400g/mol 的短链 PFAS、PFHpA、6:2FTOH 和 ADONA),lgK_d 值与 MW 之间无线性关系.但对于大 MW 的 PFAS(MW>400g/mol),MW 和 lgK_d 值具有较强线性正相关关系(R^2=0.94),这主要是因为碳链越长、MW 越大的 PFAS 具有更强疏水性;而且大 MW 的 PFAS 还可通过多点接触机制增强其在土壤界面的物理吸附作用[46].另外,从图 4 也可以看出,对于大 MW 的 PFAS,当 CF_2 官能团数为 6、8 和 10 时,组内两性离子 PFAS(6:2FtSaB、8:2FtSaB 和 10:2FtSaB)的 MW 最大,lgK_d 平均值也是最大(FTOH 除外,其 lgK_d 值与分子结构相关).

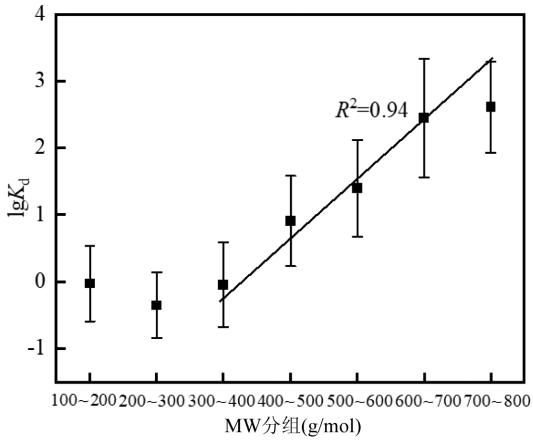


图 3　MW 对 lgK_d 值的影响

Fig.3　Impact of molecular weight on lgK_d values

根据 Kleber 等[47]提出的表面活性剂类物质多区吸附模型,长链 PFAS 在土壤矿物表面的吸附行为可描述为 3 个动力学区域:(1)接触区.PFAS 亲水头基(如−SO_3^-/−COO^-)与土壤矿物表面羟基发生配体交换,形成初级定向吸附层;(2)疏水区.在界面疏水区域,熵驱动效应通过降低体系自由能,促使已吸附的 PFAS 疏水碳链与其他两亲分子发生疏水相互作用主导的自组装行为;(3)动力学区.外层 PFAS 亲水基团通过弱氢键或范德华力实现动态结合,形成松散外围结构.另外,该模型得到 Maimaiti 等研究的支持[48].因此,MW 增大可显著提升土壤吸附作用,其机

制包括(i)增强的疏水作用力;(ii)多点接触带来的结合能累积;(iii)分子构象灵活性降低导致土壤吸附有机物的解吸阻力上升.

2.2.3 分子结构对 K_d 的影响 从图 4 和图 1 可知,ADONA 和 GenX 的 lgK_d 值(−0.63±0.24)和(−0.62±0.24),远低于其前体物 PFOA 的 lgK_d 值(0.50±0.53),与其他短链 PFAS(C_3−C_5)lgK_d 值相近.在 GenX(单醚键)和 ADONA(双醚键)的分子结构中,醚键(−O−)取代−CF2−基团,导致其 MW 比 PFOA 小.MW 减少导致形成水分子溶剂孔穴所需能量降低、增加其亲水性[49];另外醚键(−O−)氧原子提供孤电子对,增强与水分子形成氢键的能力[50].

4:2FTOH、6:2FTOH、8:2FTOH 和 10:2FTOH 的 lgK_d 平均值分别(−0.15±0.24)、(0.77±0.47)、(2.05±0.54)和(2.94±0.0.87),MW 分别是 264,364,464,564g/mol.从图 4 可以看出,当 CF_2 官能团数(4、6、8 和 10)相同时,4 个组内 PFAS 平均 MW 分别是 29,461,525,620g/mol,FTOH 的 MW 都是组内最小,而 lgK_d 平均值最大.FTOH 分子更易通过疏水作用吸附于有机质表面,这主要归因于其末端中性羟基(−OH)的非电离特性、全氟烷基链的强疏水性,以及乙基桥链结构赋予的构象柔性和疏水相互作用增强效应[50].

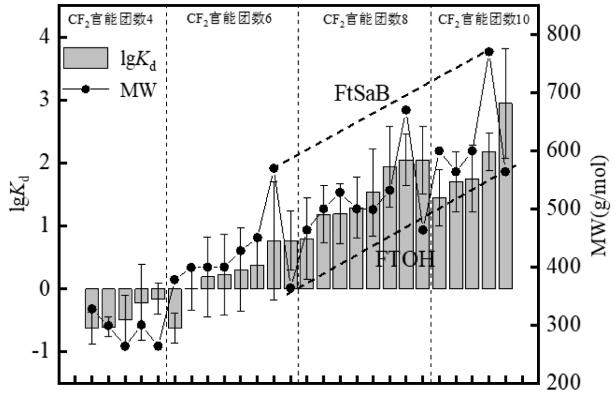


图 4 不同 CF_2 官能团数 PFAS 的 lgK_d 统计值和 MW

Fig.4 Statistical values of lgK_d and MW for PFAS with different numbers of CF_2 moieties

横刻度值从左到右分别为 4:2FTS、FBSA、PFPeA、PFBS、4:2FTOH、ADONA、FHxSA、PFHxS、PFHxPA、6:2FTS、PFHpS、6:2FtSaB、6:2FTOH、PFNA、PFOPA、8:2FTS、PFOS、PFOSA、8:2Cl-PFAES、8:2FtSaB、8:2FTOH、PFDPA、PFUnA、PFDS、10:2FtSaB 和 10:2FTOH

2.3 土壤性质参数对 K_d 的影响

采用偏最小二乘回归(PLSR)方法,分析 6 种关键土壤理化参数(有机碳质量分数、砂质量分数、粉砂质量分数、黏土质量分数、pH 值和 CEC)与 lgK_d 值的定量关系,考察不同参数对 K_d 的特征重要性.在对 PFAS 化合物、PFAS 化合物(除两性离子)和两性离子 PFAS 化合物 3 种情况分析后,如图 5(a)和 5(b)所示,对 PFAS 化合物和 PFAS 化合物(除两性离子),有机碳质量分数和粉砂质量分数的 VIP 值大于 1,为主要影响特征,表现为正向影响.有机碳质量分数越大,越有利于土壤吸附 PFAS.这可能是因为 PFAS 的碳氟链(−CF_2−)具有强疏水性,而土壤有机质中的非极性烷基链、芳香环等疏水微域可与之发生疏水分配作用[51-52].粉砂对 PFAS 主要影响机制为金属氧化物配位作用、阳离子架桥和疏水作用[52-54].(1)粉砂表面的 Fe/Al 氧化物羟基(≡Fe−OH、≡Al−OH)与 PFAS 的阴离子基团(如−SO_3^-或−COO^-)发生配位(配体交换),特别对长链 PFAS(PFOA/PFOS)亲和力更高;(2)金属阳离子(如 Ca^{2+}和 Mg^{2+})中和粉砂表面负电荷形成阳离子桥,可显著提高 PFAS 吸附量;(3)吸附在粉砂表面的天然有机质微域可提供疏水吸附位点,PFAS 疏水效应可增强其在氧化物表面滞留时间[55]总.体来说,对于 PFAS 化合物,土壤中有机质与 PFAS 之间的疏水相互作用占主导机制.

两性离子 PFAS 包括正电荷和负电荷,其吸附行为与其他阴离子和非离子 PFAS 不同.如图 5(c)所示,对于两性离子 PFAS,粉砂质量分数和砂质量分数 VIP 值大于 1(1.73 和 1.47),是主要影响特征,分别表现为正、负向影响.土壤中粉砂对两性离子 PFAS 吸附具有正向主导作用,这可能是(1)两性离子 PFAS 同时具有正电荷基团(如−$N^+(CH_3)_3$)和负电荷基团(如−SO_3^-或−COO^-),粉砂颗粒表面通常带负电荷,正电荷基团与粉砂表面的负电荷位点发生静电吸引.负电荷基团可能通过阳离子桥接(如 Ca^{2+}、Mg^{2+})与粉砂颗粒表面连接[56-57].Xiao 等[57]通过分子静电势分析表明,两性离子 PFOSB 的极性端电荷强度显著高于 PFOA,其分子表面最大静电势差值达 114.6kcal/mol.(2)粉砂中的硅氧烷基(≡Si−O−Si≡)或金属氧化物表面的羟基(≡Fe−OH,≡Al−OH)可以与两性离子 PFAS 的磺酸基、羧酸基和氨基形成氢键.(3)

粉砂中金属氧化物形成活性位点(≡Fe−OH、≡Al−OH),再与两性离子 PFAS 的负电荷基团发生配体交换形成内圈配合物(式(6))[56-57].

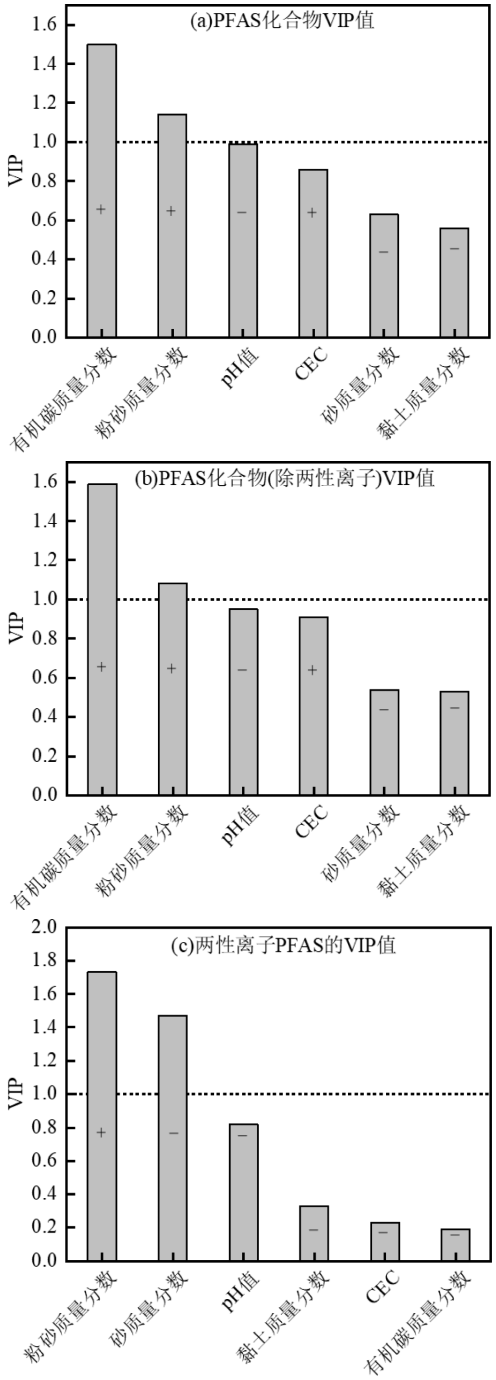


图 5　PLSR 模型分析得出特征变量 VIP 得分

Fig.5　VIP scores of features derived from PLSR model

另外,砂粒表现为负向影响,特别对两性离子型 PFAS 表现更加明显(图 5).这可能主要是砂粒通过比表面积限制、电荷位点缺失和有机质稀释三重机制,抑制两性离子 PFAS 的吸附能力[58-59,69].(1)相对粉砂和黏土,砂粒比表面积更小(0.01~0.1m^2/g),导致 PFAS 吸附效果降低.(2)砂粒表面主要成分为石英(SiO_2),表面均匀带负电,缺乏金属氧化物(Fe/Al 氧化物<0.5%),不能与阴离子 PFAS 化合物发生配位.另外,也不能与两性离子 PFAS 的负电荷、正电荷基团发生静电吸引.(3)砂粒表面有机质覆盖率<10%,无法形成“电荷锚定−疏水捕获”协同结构.

最后,从图 5 可以看出,pH 值的 VIP 值接近 1(0.99、0.95 和 0.82)和排名第三,对 PFAS 吸附表现为负向影响.从图 6 也可以看出,lgK_d 值随溶液 pH 值的增加而降低.这与前期研究结果一致,Higgins 与 Luthy 报道土壤/沉积物对阴离子 PFAS 及有机酸的吸附系数,pH 值降低 1 个单位时 lgK_d 值增加 0.37 对数单位[51].pH 值导致 lgK_d 值变化可能主要归因于土壤表面电荷、阳离子桥接作用和疏水协同效应[60-62].(1)当 pH 值下降时,土壤表面铁/铝氧化物和有机质的质子化程度更高(式(7)~(10)),通过促进疏水性和静电吸引来提高吸附效果.(2)H^+通过静电引力优先占据土壤中阳离子交换位点,置换并释放出吸附阳离子态(如 Ca^{2+}/Al^{3+}),与 PFAS 形成桥接复合物(≡Soil−H^+−Ca^{2+}−OOC−PFAS).(3)土壤中有机质基团(如−COOH)质子化增强其疏水性,同时降低 PFAS 分子脱水能垒,双路径协同增强固−液界面疏水作用.Gu 等研究报道,当 pH 值发生变化时,土壤表面饱和阳离子的变化会影响 PFAS 分配系数[61].

$$\equiv \text{M-OH} + \text{PFOS}^- \rightarrow \equiv \text{M-O-SO}_2\text{-PFOS} + \text{OH}^- \quad (6)$$

$$\equiv \text{Fe-OH} + \text{H}^+ \rightarrow \equiv \text{Fe-OH}_2^+ \quad (7)$$

$$\equiv \text{Al-OH} + \text{H}^+ \rightarrow \equiv \text{Al-OH}_2^+ \quad (8)$$

$$\text{R-COOH} + \text{H}^+ \rightarrow \text{R-COOH}_2^+ \quad (9)$$

$$\text{R-NH}_2 + \text{H}^+ \rightarrow \text{R-NH}_2^+ \quad (10)$$

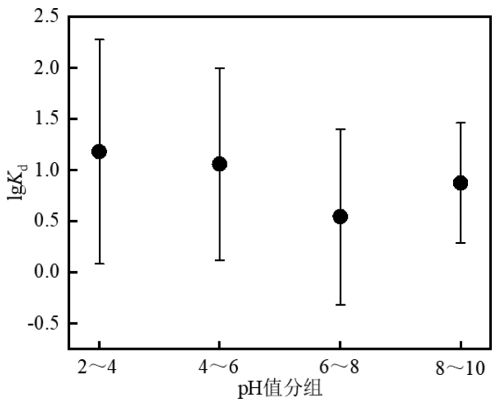


图 6　pH 值对 lgK_d 的影响

Fig.6　Impact of pH values on lgK_d

2.4 ML 模型对 K_d 预测

2.4.1 ML 模型性能 经过参数优化后,在预测 lgK_d 值时 MLR、SVR、RF、XGBoost 和 GBRT 模型训练集和测试集结果如表 4 和图 7 所示.从 5 种 ML 模型性能指标可以看出,所研究的 PFAS 尽管存在不同形态(阴离子、阳离子和两性离子等),RF、XGBoost 和 GBRT 模型表现出最佳性能,突显了高预测精度和可靠性.RF、XGBoost 和 GBRT 模型的优异性能归功于其基于决策树,自动识别特征间的复杂交互和非线性效应,在性能、鲁棒性和易用性方面具有显著优势[63].相比之下,GBRT 模型的 Q^2 最高(0.85),具有较高泛化能力.SVR 模型表现尚可,R^2 值为 0.89 和 0.84,Q^2 为 0.81,但其对数据维度和噪声的敏感性限制了其可靠性.MLR 模型表现最弱,R^2 值为 0.74 和 0.74,Q^2 为 0.73,MAE 值为 0.38 和 0.37,RMSE 值为 0.47 和 0.47,反映了其在捕捉变量间非线性关系方面的局限性.总之,传统线性模型(MLR)难以准确预测 K_d 值,而 RF、XGBoost 和 GBRT 模型在处理环境数据集的复杂性和异质性方面表现出色,这体现出先进 ML 模型在处理复杂环境数据中的效用.其中因 GBRT 模型卓越的准确性和泛化能力,后续选择 GBRT 模型进行 PFAS 在土壤吸附机制进一步分析.

表 4 ML 模型预测性能指标

Table 4 Performance metrics of ML models prediction

ML 模型	训练集 R^2	训练集 RMSE	训练集 MAE	测试集 R^2	测试集 RMSE	测试集 MAE	Q^2
MLR	0.74	0.47	0.38	0.74	0.47	0.37	0.73
SVR	0.89	0.31	0.22	0.84	0.37	0.28	0.81
RF	0.98	0.14	0.10	0.84	0.35	0.26	0.83
XGBoost	0.99	0.09	0.07	0.86	0.36	0.27	0.85
GBRT	0.99	0.09	0.07	0.86	0.35	0.26	0.85

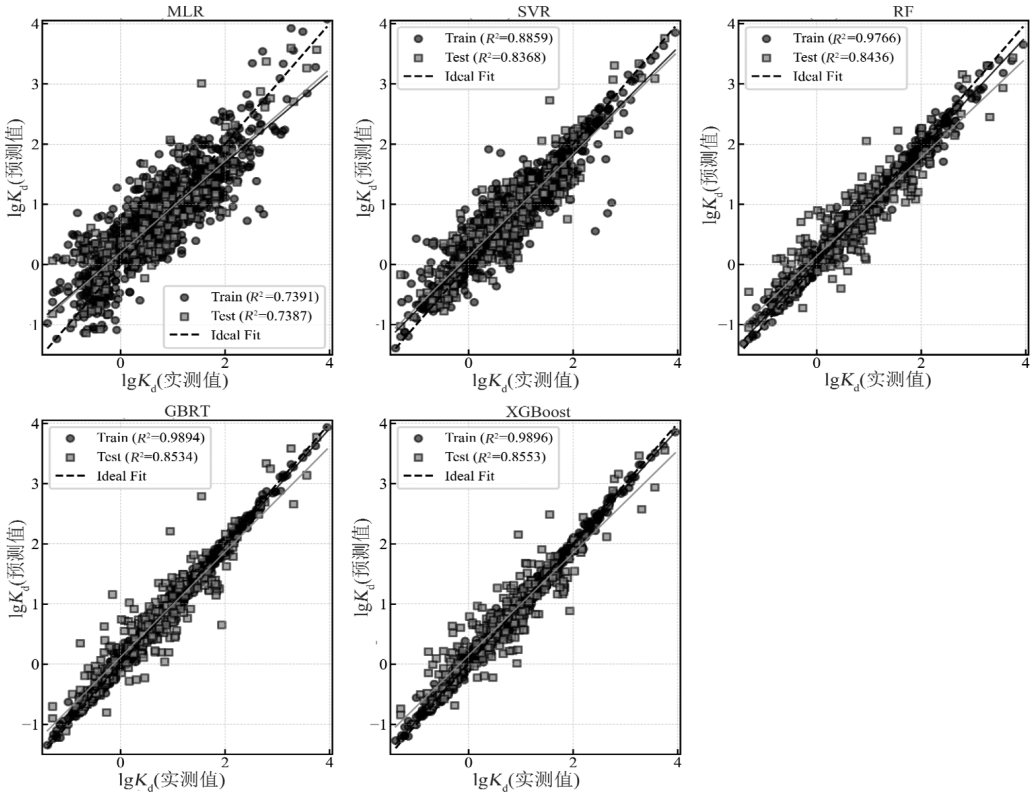


图 7 lgK_d 实测值与 ML 模型预测值散点图

Fig.7 Scatter plots of measured lgK_d values and ML model predicted values

2.4.2 特征重要性 图 8(a)的特征贡献率排列量化了每个特征的相对贡献.特征贡献率越高,表明对模型性能的影响越大.在所有特征中,MW、有机碳质量分数、lg$S$ 和净电荷密度特征贡献率之和大于 90%,为最具影响力的 4 个特征,MW 和有机碳质量分数特征贡献率之和大于 60%,结果反映了 MW 在决定 PFAS 土壤分布的关键作用,而有机碳质量分数则反映了有利于 PFAS 吸附的土壤环境条件.相比之下,

其变量特征如 lgK_{ow} 和 pKa,以及土壤特性如 pH 值、CEC、砂质量分数和粉砂质量分数,表现出较低的特征贡献率,在 0.1%~3.5%范围之间.虽然这些特征对 ML 模型的预测也有贡献,但其影响远不如 MW 和有机碳质量分数.土壤特性(pH 值、CEC、砂质量分数和粉砂质量分数)的特征贡献率较低,这表明可能数据集中的土壤类型数据缺失,造成土壤多样性存在一定局限,这进一步说明需考虑更广泛土壤类型来提高 ML 模型对土壤特定效应的敏感性.

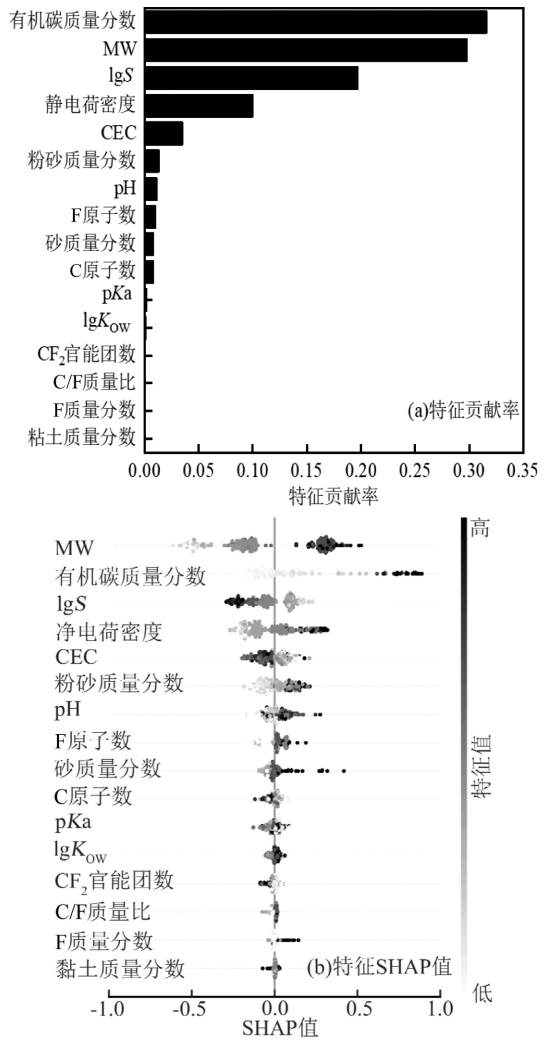


图 8　(a)GBRT 模型特征贡献率排列及(b)GBRT 模型特征 SHAP 值

Fig.8　(a) Feature contribution ranking for the GBRT model and (b) SHAP values distribution for the GBRT model

图 8(b)中 SHAP 值提供了每个特征对模型预测个体贡献的另一个视角,再次确认了 MW 的主导影响,显示出与预测 lgK_d 值的强正相关性.此外,SHAP 值揭示了土壤有机质对 PFAS 行为的显著影响.CEC 和土壤质地成分的影响相对较小,这可能与这些土壤组分与 PFAS 分子之间复杂的相互作用有关.特征贡献率排列和 SHAP 值的结合使用为理解影响 PFAS 土壤吸附的多重因素提供了一种方法.

2.4.3　PFAS 在土壤中分配的依赖性　图 9(a)显示,MW 与 lgK_d 值之间的非线性关系,当 MW 小于 350g/mol 时,PFAS 的分配行为相对稳定,这与图 3 一致,较小 MW PFAS 的 lgK_d 更相近.MW 大于 430g/mol 时 lgK_d 急剧上升,之后效应趋于平稳.这表明 MW 对 PFAS 分配的影响存在一个限度,超过该限度后,PFAS 与土壤基质之间的相互作用变得更加复杂.Umeh 等[64]报道,较低 MW 的 PFAS 更有可能从饱和土壤迁移到水中,而长链 PFAS 则倾向于被土壤强烈吸附.较高 MW 的 PFAS 通常表现出更强的疏水性,主要通过疏水相互作用以及与土壤颗粒间的 π–π 堆叠或层状结构作用等机制,显著增强其与土壤有机质之间的相互作用[65].

如图 9(b)所示,表现出土壤有机质在结合 PFAS 中的重要性.随着土壤中有机碳质量分数增加,PFAS 的 lgK_d 呈现指数上升趋势,说明较高的有机质水平能提高土壤滞留 PFAS 的能力.首先,土壤有机质作为 PFAS 的主要吸附位点,其含量的上升直接增加了可供分配的介质体积[66].PFAS(特别是长链疏水性 PFAS)会优先分配到有机质中,有机质的增加显著增强了这种分配能力[67].有研究表明,当有机质质量分数从 0.1%增至 1%时,PFOS 的 K_d 值可提高 10~100 倍.其次,丰富的有机化合物吸附阳离子,形成一层屏障,减少了土壤颗粒与 PFAS 之间的静电斥力,斥力的减少增强了 PFAS 与土壤中疏水和非带电基团的相互作用[68].Hubert 等研究发现,K_d 值随着土壤有机质含量增加而增加,并且与 PFAS 的 MW 和头基特性密切相关[68].当在有机质质量分数大于 10%时, lgK_d 会逐渐趋于平缓,表明吸附位点接近饱和[69].

从图 9(c)可以看出,lg$S$(溶解度对数)对 PFAS 的吸附行为(以 lgK_d 表示)具有显著影响.当 lg$S$ 值较高时,lgK_d 较低,表明 PFAS 的疏水性增强(水溶性降低)会促进其在固相中的吸附.具体表现为:超低溶解度区间(lg$S$<0)lgK_d 保持高值且稳定,主要对应长链 PFAS(如 PFOS、PFOA)的超疏水吸附机制.此时,PFAS 倾向于分配至土壤有机质中,溶解度变化

对其吸附行为影响较小[70,74].过渡区间(lgS=0~4.0)lgK_d 急剧下降,反映吸附机制从疏水主导向疏水-静电协同作用转变.随着 PFAS 水溶性增加(如中链或含极性基团变体),静电作用(如阳离子桥接)、氢键等逐渐增强,形成混合吸附机制[70-73].高溶解度区间(lg$S$>4)lgK_d 趋于稳定,对应短链 PFAS(如 PFBA、PFBS)的吸附行为.此时,PFAS 的亲水性较强,吸附主要依赖与土壤表面极性位点(如金属氧化物、有机质含氧官能团)的特异性作用,而溶解度影响减弱[73-74].此外,短链 PFAS 的小分子尺寸降低了空间位阻,使其更易进入微孔结构,而高溶解度下的吸附位点饱和效应也进一步稳定 lgK_d[73].

净电荷密度对 lgK_d 的影响如图 9(d)所示,当净电荷密度在−260~−140C/g 时,lgK_d 呈显著增加,吸附增强可归因 PFAS 与铁和铝氧化物中羧酸基团和金属阳离子之间形成表面络合物,或静电相互作用.相关研究证实,这两种机制可能单独或共同起作用[75].当净电荷密度小于−240C/g(负电性更强)时,PFAS 分子与带负电的土壤表面(如有机质、黏土矿物)之间的强静电排斥作用占主导地位,尽管可能存在阳离子桥接效应,但仍限制了吸附作用[73].反之,当电荷密度接近中性(>−140C/g)时,疏水作用和范德华力成为主要吸附作用力,导致 lgK_d 相对稳定[73].此外,图中平台区的出现表明存在饱和效应,即进一步的电荷改变不会显著改变界面相互作用能,这可能是因为其他因素(如分子尺寸、土壤孔隙结构)开始成为限制因素,这说明 PFAS 环境行为的静电作用与疏水作用之间复杂的协同关系.

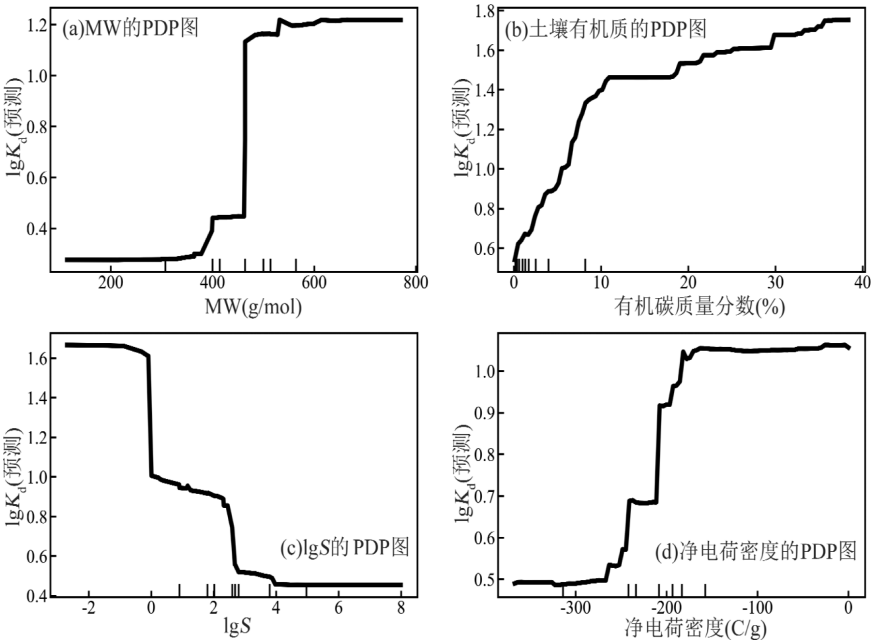


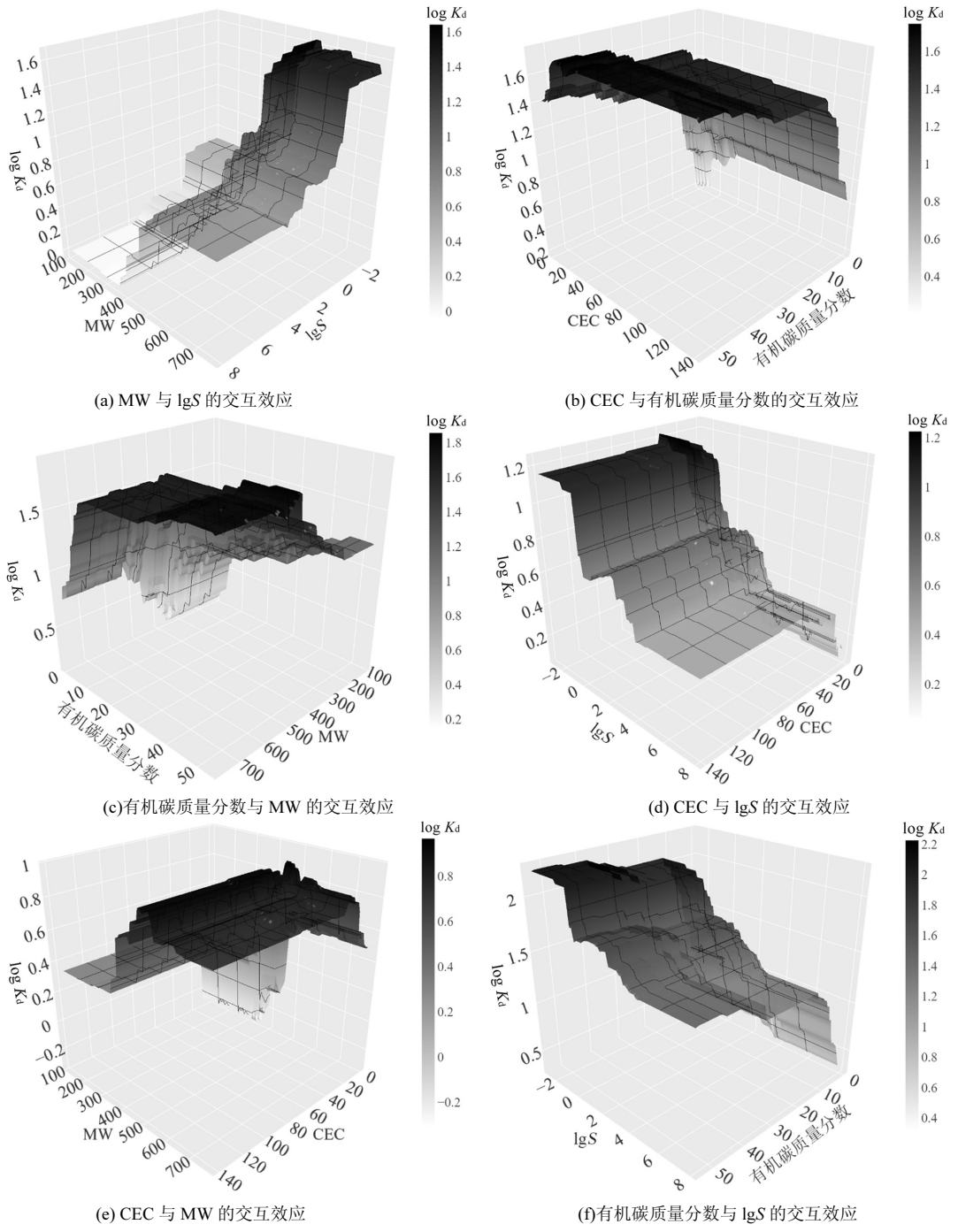
图 9 GBRT 模型部分依赖图(PDPs)

Fig.9 The partial dependence plots (PDPs) for the GBRT model

2.4.4 PFAS 性质与土壤属性之间的相互作用 PFAS 物理化学特征(MW、lg$S$)与土壤特性(有机碳质量分数和 CEC)之间的多维相互作用如图 10 所示,这些特征在之前的分析中被认为重要特征(图 8). MW 与 lg$S$ 的交互效应分析结果(图 10(a))表明,MW 是影响 lgK_d 变化的主导因子(平均偏依赖变化=0.03),这一结果证实了长链 PFAS(MW>380g/mol)由于其疏水性增强导致 lgK_d 上升特征.有机碳质量分数与 CEC 的交互作用分析(图 10(b))显示,有机质对吸附过程的影响更为显著(平均偏依赖变化=0.01).研究表明,有机质含量每增加 1%,CEC 可提升 2.3~3.5cmol/kg,这主要归因于有机质中的羧基(−COOH)和酚羟基(−OH)等官能团提供了额外的阳离子交换位点[68,75].

图 10(c)~(f)表明,较大 PFAS 分子提高其疏水性和空间结构,能够在土壤有机质内形成更多的吸附位点.富含有机质的土壤提供了丰富的官能团和疏水基质,促进了疏水相互作用和氢键结合.此外,

高 CEC 可以降低土壤颗粒上的负电荷密度,从而减少土壤颗粒负电荷密度与阴离子型 PFAS 之间的静电斥力.斥力的减少促进了更紧密的相互作用,特别是在较低 pH 值下,此时土壤组分的质子化程度增加增强了 PFAS 吸附作用.这种相互作用对于大分子的 PFAS 尤为突出,它们可以通过与土壤有机质的堆叠或层状相互作用形成更稳定的带状结构.因此,高有机质含量(>1%)和 CEC(>8cmol/kg)与低 pH 值的协同作用显著提高了 PFAS 的吸附能力($\lg K_d$ >0.4),特别是对长全氟烷基链(低 lg$S$ 值)和高 MW(>380g/mol)的 PFAS 化合物.相反,当 CEC(<8cmol/kg)、有机质含量(<1%)和 MW(<380g/mol)较低时,PFAS 的吸附能力显著减弱($\lg K_d$<0.4),表现出更高的环境迁移性.



(a) MW 与 lg$S$ 的交互效应　(b) CEC 与有机碳质量分数的交互效应

(c)有机碳质量分数与 MW 的交互效应　(d) CEC 与 lg$S$ 的交互效应

(e) CEC 与 MW 的交互效应　(f)有机碳质量分数与 lg$S$ 的交互效应

图 10　关键土壤性质与分子特征间的交互作用

Fig.10　The interaction between pairs of key soil and molecular features

3 结论

3.1 采用 16 个特征变量和参数优化后 GBRT 模型对 $\lg K_d$ 的预测表现出良好性能,其 R^2 约为 0.85, RMSE、MAE 和 Q^2 分别达到 0.35、0.26 和 0.85.MW、有机碳质量分数、$\lg S$ 和净电荷密度被识别为主要影响特征,累计贡献率超过 90%.

3.2 具有较高 MW 和更强疏水性的 PFAS 在富含有机质且 CEC 较高的土壤中表现出更强的吸附能力,这类环境条件促进了 PFAS 与土壤基质之间的稳定相互作用.特别在低 pH 值条件下,质子化作用减弱静电斥力,进一步增强了疏水性 PFAS 与土壤颗粒的结合能力.GBRT 模型能够有效适应土壤类型和 PFAS 化合物的复杂性,展现出了良好的应用潜力.

参考文献:

[1] Adu O, Ma X, Sharma V K. Bioavailability, phytotoxicity and plant uptake of per–and polyfluoroalkyl substances (PFAS): a review [J]. Journal of Hazardous Materials, 2023,447:130805.

[2] Evich M G, Davis M J B, McCord J P, et al. Per– and polyfluoroalkyl substances in the environment [J]. Science, 2022,375:658–660.

[3] Wang Y, Munir U, Huang Q. Occurrence of per– and polyfluoroalkyl substances (PFAS) in soil: sources, fate, and remediation [J]. Soil & Environmental Health, 2023,1:100004.

[4] Mei W, Sun H, Song M, et al. Per– and polyfluoroalkyl substances (PFASs) in the soil–plant system: sorption, root uptake, and translocation [J]. Environment International, 2021,156:106642.

[5] Trang B, Li Y, Xue X S, et al. Low–temperature mineralization of perfluorocarboxylic acids [J]. Science, 2022,377:839–845.

[6] [6] Ehrlich V, Bil W, Vandebriel R, et al. Consideration of pathways for immunotoxicity of per– and polyfluoroalkyl substances (PFAS) [J]. Environmental Health, 2023,22:1910.

[7] Zahm S, Bonde J P, Chiu W A, et al. Carcinogenicity of perfluorooctanoic acid and perfluorooctanesulfonic acid [J]. Lancet Oncology, 2024,25:16–17.

[8] Mohona T M, Ye Z, Dai N, et al. Adsorption behavior of long–chain perfluoroalkyl substances on hydrophobic surface: a combined molecular characterization and simulation study [J]. Water Research, 2023,239:120074.

[9] Lyu X, Xiao F, Shen C, et al. Perand polyfluoroalkyl substances (PFAS) in subsurface environments: occurrence, fate, transport, and research prospect [J]. Reviews of Geophysics, 2022,60:8755–1209.

[10] Ahmad A, Tian K, Tanyu B, et al. Effect of clay mineralogy on the partition coefficients of perfluoroalkyl substances [J]. ACS ES&T Water, 2023,3:2899–2909.

[11] Fabregat–Palau J, Vidal M, Rigol A. Modelling the sorption behaviour of perfluoroalkyl carboxylates and perfluoroalkane sulfonates in soils [J]. Science of the Total Environment, 2021,801:149343.

[12] Wanzek T, Stults J F, Johnson M G, et al. Role of mineral–organic interactions in PFAS retention by AFFF–impacted soil [J]. Environmental Science & Technology, 2023,57:5231–5242.

[13] Knight E R, Janik L J, Navarro D A, et al. Predicting partitioning of radiolabelled 14C–PFOA in a range of soils using diffuse reflectance infrared spectroscopy [J]. Science of the Total Environment, 2019, 686:505–513.

[14] 陈 新,孙 霄,单国强,等.全氟与多氟化合物在场地土壤中的垂直迁移行为 [J]. 环境化学, 2025,44(1):140–148.
Chen X, Sun X, Shan GQ, et al. Vertical transportation behavior of per– and polyfluoroalkyl substances (PFAS) in soil columnaround a fluorochemical industrial park [J]. Environmental Chemistry, 2025, 44(1):140–148.

[15] Nguyen T M H, Braunig J, Thompson K, et al. Influences of chemical properties, soil properties, and solution pH on soil–water partitioning coefficients of per– and polyfluoroalkyl substances (PFASs) [J]. Environmental Science & Technology, 2020,54:15883–15892.

[16] Oliver D P, Navarro D A, Baldock J, et al. Sorption behaviour of per– and polyfluoroalkyl substances (PFASs) as affected by the properties of coastal estuarine sediments [J]. Science of the Total Environment, 2020,720:137263.

[17] Du Z, Deng S, Bei Y, et al. Adsorption behavior and mechanism of perfluorinated compounds on various adsorbents—a review [J]. Journal of Hazardous Materials, 2014,274:443–454.

[18] Qi Y, Cao H, Pan W, et al. The role of dissolved organic matter during per– and polyfluorinated substance (PFAS) adsorption, degradation, and plant uptake: a review [J]. Journal of Hazardous Materials, 2022, 436:129139.

[19] Wang W, Rhodes G, Zhang W, et al. Implication of cation–bridging interaction contribution to sorption of perfluoroalkyl carboxylic acids by soils [J]. Chemosphere, 2022,290:133224.

[20] Li Y, Oliver D P, Kookana R S. A critical analysis of published data to discern the role of soil and sediment properties in determining sorption of per and polyfluoroalkyl substances (PFASs) [J]. Science of the Total Environment, 2018,628:110–120.

[21] Dou B, Zhu Z, Merkurjev E, et al. Machine learning methods for small data challenges in molecular science [J]. Chemical Reviews, 2023, 123:8736–8780.

[22] Gupta S, Aga D, Pruden A, et al. Data analytics for environmental science and engineering research [J]. Environmental Science & Technology, 2021,55:10895–10907.

[23] Cao H, Peng J, Zhou Z, et al. Investigation of the binding fraction of PFAS in human plasma and underlying mechanisms based on machine learning and molecular dynamics simulation [J]. Environmental Science & Technology, 2023,57:17762–17773.

[24] Xiang L, Qiu J, Chen Q Q, et al. Development, evaluation, and application of machine learning models for accurate prediction of root uptake of per– and polyfluoroalkyl substances [J]. Environmental Science & Technology, 2023,57:18317–18328.

[25] Cao H, Peng J, Zhou Z, et al. Insight into the defluorination ability of per– and polyfluoroalkyl substances based on machine learning and quantum chemical computations[J]. Science of the Total Environment, 2022,807:151018.

[26] Brusseau M L, Anderson R H, Guo B. PFAS concentrations in soils: background levels versus contaminated sites [J]. Science of the Total Environment, 2020,740:140017.

[27] Xiao F. Emerging poly‐ and perfluoroalkyl substances in the aquatic environment: A review of current literature [J]. Water Research, 2017, 124:482–495.

[28] Xie G, Cui J, Zhai Z, et al. Distribution characteristics of trifluoroacetic acid in the environments surrounding fluorochemical production plants in Jinan, China [J]. Environmental Science and Pollution Research, 2020,27(1):983–991.

[29] Huang Y, Lu M, Li H, et al. Sensitive determination of perfluoroalkane sulfonamides in water and urine samples by multiple monolithic fiber solid‐phase microextraction and liquid chromatography tandemmass spectrometry [J]. Talanta, 2019,192:24–31.

[30] Schellenberger S, Jönsson C, Mellin P, et al. Release of side‐chain fluorinated polymer‐containing microplastic fibers from functional textiles during washing and first estimates of perfluoroalkyl acid emissions [J]. Environmental Science & Technology, 2019,53(24): 14329–14338.

[31] Schüßler M, Capitain C, Bugsel B, et al. Non‐target screening reveals 124PFAS at an AFFF‐impacted field site in Germany specified by novel systematic terminology [J]. Analytical & Bioanalytical Chemistry, 2025,417:6049–6064.

[32] Liu M, Munoz G, Vo Duy S, et al. Per‐ and polyfluoroalkyl substances in contaminated soil and groundwater at airports: a canadian case study [J]. Environmental Science & Technology, 2022,56(2):885–895.

[33] Nimon K F, Oswald F L. Understanding the results of multiple linear regression: beyond standardized regression coefficients [J]. Organizational Research Methods, 2013,16:650–674.

[34] Gao F, Shen Y, Sallach J B, et al. Direct prediction of bioaccumulation of organic contaminants in plant roots from soils with machine learning models based on molecular structures [J]. Environmental Science & Technology, 2021,55:16358–16368.

[35] Roostaei J, Colley S, Mulhern R, et al. Predicting the risk of GenX contamination in private well water using a machine‐learned bayesian network model [J]. Journal of Hazardous Materials, 2021,411:125075.

[36] Yang C, Dong H, Chen Y, et al. New insights on the formation of nucleation mode particles in a coastal city based on a machine learning approach [J]. Environmental Science & Technology, 2023,58(2):1187–1198.

[37] Zhu J J, Yang M, Ren ZJ. Machine learning in environmental research: common pitfalls and best practices [J]. Environmental Science & Technology, 2023,57:17671–17689.

[38] 于林堂,陈咚咚,陶翠翠,等.基于可解释性机器学习的芬顿工艺降解有机污染物速率的研究 [J]. 中国环境科学, 2025,45(8):4294–4302.
Yu L T, Chen D D, Tao C C, et al. Research on the degradation rate of organic pollutants in fenton process based on explainable machine learning [J]. China Environmetal Science, 2025,45(8):4294–4302.

[39] Mi X, Zou B, Zou F, et al. Permutation‐based identification of important biomarkers for complex diseases via machine learning models [J]. Nature Communications, 2021,12:3008.

[40] Rodríguez‐Perez R, Bajorath J. Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values [J]. Journal of Medicinal Chemistry, 2020,63:8761–8777.

[41] 姜明星,王斯坦,许端平.基于机器学习的金属有机框架吸附水中重金属性能预测 [J]. 中国环境科学, 2023,43(5):2319–2327.
Jiang M X, Wang S T, Xu D P. Prediction of adsorption performance of MOFs for heavy metals in water based on machine learnin [J]. China Environmental Science, 2023,43(5):2319–2327.

[42] 蔡绮思,李星颖,张　芳,等.扩散梯度薄膜探针原位分析全/多氟烷基物质在水‐沉积物系统中的分布特征 [J]. 分析化学, 2025,53(2): 300–318.
Cai Q S, Li X Y, Zhang F, et al. In situ analysis of per‐ and polyfluoroalkyl substances distribution characteristics in water‐sediment system using diffusive gradients in thin films probe [J]. Chinese Journal of Analytical Chemistry, 2025,53(2):300–318.

[43] Sorengard M, Kleja D B, Ahrens L. Stabilization of per‐ and polyfluoroalkyl substances (PFASs) with colloidal activated carbon (PlumeStop) as a function of soil clay and organic matter content [J]. Journal of Environmental Management, 2019,249:109345.

[44] Munoz G, Ray P, Mejia‐Avendaño S, et al. Optimization of extraction methods for comprehensive profiling of perfluoroalkyl and polyfluoroalkyl substances in firefighting foam impacted soils [J]. Analytica Chimica Acta, 2018,1034:74–84.

[45] Backe W J, Day T C, Field J A. Zwitterionic, cationic, and anionic fluorinated chemicals in aqueous film forming foam formulations and groundwater from U.S. military bases by nonaqueous large‐volume injection HPLC‐MS/MS [J]. Environmental Science & Technology, 2013,47:5226–5234.

[46] Wiseman C L S, Püttmann W. Interactions between mineral phases in the preservation of soil organic matter [J]. Geoderma, 2006,134:109–118.

[47] Kleber M, Sollins P, Sutton R. A conceptual model of organo‐mineral interactions in soils: self‐assembly of organic molecular fragments into zonal structures on mineral surfaces [J]. Biogeochemistry, 2007, 85:9–24.

[48] Maimaiti A, Deng S, Meng P, et al. Competitive adsorption of perfluoroalkyl substances on anion exchange resins in simulated AFFF‐impacted groundwater [J]. Chemical Engineering Journal, 2018, 348:494–502.

[49] Joerss H, Apel C, Ebinghaus R. Emerging per‐ and polyfluoroalkyl substances (PFASs) in surface water and sediment of the North and Baltic Seas [J]. Science of the Total Environment, 2019,686:360–369.

[50] Narasimhan L, Angela K W. Adsorption, structure, and dynamics of short‐ and long‐chain PFAS molecules in kaolinite: molecular‐level insights [J]. Environmental Science & Technology, 2022,56(12): 8043–8052.

[51] Higgins C P, Luthy R G. Sorption of perfluorinated surfactants on sediments [J]. Environmental Science & Technology, 2006,40:7251–7256.

[52] 苗　育.全氟辛酸在土壤介质上吸附平衡的研究 [D]. 淮南:安徽理工大学, 2017.
Miao Y. Adsorption equilibrium of perfluorooctanoic acid on soil and its compostions [D]. Huainan: Anhui University of Science and Technology, 2017.

[53] Maja S H, Sarah J, Arwel V H, et al. Sorption of perfluoroalkyl substances to two types of minerals [J]. Chemosphere, 2016,159: 385–391.

[54] Mejia-avendano S, Zhi Y, Yan B, et al. Sorption of polyfluoroalkyl surfactants on surface soils: Effect of molecular structures, soil properties, and solution chemistry [J]. Environmental Science & Technology, 2020,54(3):1513–1521.

[55] Zhao L X, Bian J N, Zhang Y H, et al. Comparison of the sorption behaviors and mechanisms of perfluorosulfonates and perfluorocarboxylic acids on three kinds of clay minerals [J]. Chemosphere, 2014,114:51–58.

[56] Cogorno J, Ahmadi N, Muniruzzaman M, et al. Electrostatic and chemical interplay between air–water and mineral–water interfaces during PFOS transport in unsaturated porous media [J]. Environmental Science & Technology, 2025,59:7634–7645.

[57] Xiao F, Jin B S, Golovko S A, et al. Golovko, and Baoshan Xing. Sorption and desorption mechanisms of cationic and zwitterionic Per- and Polyfluoroalkyl substances in natural soils: thermodynamics and hysteresis [J]. Environmental Science & Technology, 2019,53:11818–11827.

[58] Lv X Y, Sun Y Y, Ji R, et al. Physicochemical factors controlling the retention and transport of perfluorooctanoic acid (PFOA) in saturated sand and limestone porous media [J]. Water research: A journal of the international water association, 2018,141:251–258.

[59] Wang Y K, Khan N, Huang D D, et al. Transport of PFOS in aquifer sediment: transport behavior and a distributed-sorption model [J]. Science of the Total Environment, 2021,779:146444.

[60] Tang C Y, Fu Q S, Gao D W, et al. Effect of solution chemistry on the adsorption of perfluorooctane sulfonate onto mineral surfaces [J]. Water Research, 2010,44(8):2654–2662.

[61] Gu B, Schmitt J, Chen Z, et al. Adsorption and desorption of different organic matter fractions on iron oxide [J]. Geochimica Et Cosmochimica Acta, 1995,59:219–229.

[62] Johnson R L, Anschutz A J, Smolen J M, et al. The adsorption of perfluorooctane sulfonate onto sand, clay, and iron oxide surfaces [J]. Journal of Chemical & Engineering Data, 2007,52:1165–1170.

[63] Zhou X, Lu P, Zheng Z, et al. Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree [J]. Reliability Engineering & System Safety, 2020,200:106931.

[64] Umeh A C, Naidu R, Olisa E, et al. A systematic investigation of single solute, binary and ternary PFAS transport in water-saturated soil using batch and 1-dimensional column studies: focus on mixture effects [J]. Journal of Hazardous Materials, 2024,461:132688.

[65] 赵丽霞.典型全氟化合物在沉积物中的分配行为与微观机制研究 [D]. 天津:南开大学, 2013.
Zhao L X. The study about partitioning behavior and mechanisms of typical perfluoroalkyl substances in sediment [D]. Tianjin:Nankai University, 2013.

[66] Campos-Pereira H, Kleja D B, Sjostedt C, et al. The adsorption of per- and polyfluoroalkyl substances (PFASs) onto ferrihydrite is governed by surface charge [J]. Environmental Science & Technology, 2020,54:15722–15730.

[67] Gagliano E, Sgroi M, Falciglia P P, et al. Removal of poly- and perfluoroalkyl substances (PFAS) from water by adsorption: role of PFAS chain length, effect of organic matter and challenges in adsorbent regeneration [J]. Water Research, 2020,171:115381.

[68] Hubert M, Arp H P H, Hansen M C, et al. Influence of grain size, organic carbon and organic matter residue content on the sorption of per- and polyfluoroalkyl substances in aqueous film forming foam contaminated soils-implications for remediation using soil washing [J]. Science of the Total Environment, 2023,875:162668.

[69] Hellsing M S, Hellsing S, Josefsson A V, et al. Sorption of perfluoroalkyl substances to two types of minerals [J]. Chemosphere, 2016,159:385-391.

[70] 郭羽彤,牛世欣,吕宏虹,等.农田土壤中全氟化合物的迁移转化与污染控制技术研究进展 [J]. 农业环境科学学报, 2025,44(3):566–579.
Guo Y T, Niu S X, Lv H H, et al. Advances in migration, transformation,and pollution control technologies of perfluoroalkyl substances in farmland soils: a review [J]. Journal of Agro-Environment Science, 2025,44(3):566–579.

[71] Umeh A C, Naidu R, Shilpi S, et al. Sorption of PFOS in 114well-characterized tropical and temperate soils: application of multivariate and artificial neural network analyses [J]. Environmental Science & Technology, 2021,55:1779–1789.

[72] Campos-Pereira H, Makselon J, Kleja D B, et al. Binding of per- and polyfluoroalkyl substances (PFASs) by organic soil materials with different structural composition- charge- and concentration-dependent sorption behavior [J]. Chemosphere, 2022,297:134167.

[73] Campos-Pereira H, Kleja D B, Ahrens L, et al. Effect of pH, surface charge and soil properties on the solid-solution partitioning of perfluoroalkyl substances (PFASs) in a wide range of temperate soils [J]. Chemosphere, 2023,321:138133.

[74] 程 静,梁光愉,冯雯凤,等.水环境中短链全氟及多氟烷基化合物污染水平及其处理技术研究进展 [J]. 环境化学, 2024,43(12):4022–4043.
Cheng J, Liang G Y, Feng W F, et al. Advances in pollution levels and treatment technologies of short chain per- and polyfluoroalkyl substances in aquatic environment [J]. Environmental Chemistry, 2024,43(12):4022–4043.

[75] 黄丹丹.复合污染情况下典型全氟化合物在孔隙介质中的迁移特性研究 [D]. 北京:中国地质大学(北京), 2021.
Huang D D. The retardation and transport of typical PFAS in porous media under co-contaminated conditions [D]. Beijing: China University of Geosciences (Beijing), 2021.

作者简介： 刘 冰(1978-),男,河南许昌人,副教授,博士,主要研究方向为水体中含氮消毒副产物及前体物识别和控制城市污水处理原理与技术.发表论文 30 余篇.liubing_982002@163.com.