Supplementary Information

Molecular Connectivity Indices and Soil Properties to Predict Sorption of Per- and Polyfluoroalkyl Substances

Paulina Alulema-Pullupaxi[1], Fatih Evrendilek[2], Dilara Hatinoglu[1], Simin Moavenzadeh Ghaznavi[2], Kenneth Mensah[1], Manisha Choudhary[2], Sonora Ortiz[3], Onur Apul[1*]

[1] Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, PA 16802, USA

[2] Department of Civil and Environmental Engineering, University of Maine, Orono, ME 04469, United States.

[3] Department of Ecology and Environmental Sciences, University of Maine, Orono, ME 04469, United States.

[*]correspondence: oga5061@psu.edu

33 Pages, 15 Tables, 9 Figures

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
AICc	Corrected Akaike Information Criterion
ANN	Artificial neural network
C#	Number of carbon atoms in the molecule (e.g., C4, C8)
C_w	Concentration of a compound in the aqueous phase at equilibrium, (mg L^{-1})
C_s	Concentration of a compound in the solid matrix at equilibrium, (mg kg^{-1})
CEC	Cation exchange capacity
FTSs	Fluorotelomer sulfonates
f_{oc}	fraction of the solid that is organic carbon (kg OC per kg soil)
KNN	K-nearest neighbor
K_{oa}	Octanol–air partitioning coefficient
K_{ow}	Octanol–water partitioning coefficient
log K_d	Soil–water partitioning coefficient
LSER	Linear solvation energy relationship
MCIs	Molecular connectivity indices
ML	Machine Learning
N	Sample size
PFAS	Per- and polyfluoroalkyl substances
PFCAs	Perfluorocarboxylic acids
PFSAs	Perfluorosulfonic acids
QSPR	Quantitative structure–property relationship
QSAR	Quantitative structure–activity relationship
R^2_{adj}	Adjusted coefficient of determination on training dataset
R^2_{pred}	Predicted coefficient of determination on validation dataset
R^2_{ext}	Predicted coefficient of determination on independent external dataset
RF	Random forest
SHAP	Shapley additive explanations
SMILES	Simplified molecular-input line-entry system
SOC	Soil organic carbon
SP-3	Simple-path order 3
SVM	Support vector machine
VIF	Variance inflation factor
VP-7	Valence-path order 7
XGB	Extreme gradient boosting

Table S1. List of compounds for the PFAS dataset, data sources, and adsorption descriptors

PFAS Family	PFAS name	MW (mol/g)	pK_a Comptox	pK_a Chemicalize	logK_{ow}	logK_{oc} (3.4)	logK_{oc} (5.2)	logK_{oc} (7.2)	logK_{oc} (8.3)
PFCAs	PFBA	214.03	-0.21	1.07	2.14	2.00	1.90	1.70	1.80
	PFPeA	264.05	-0.8	0.34	2.81	2.00	1.80	1.70	1.70
	PFHxA	314.05	0.2	-0.78	3.48	2.00	1.90	1.70	1.70
	PFHpA	364.06	0.6	-2.29	4.15	2.50	2.20	2.10	2.10
	PFOA	414.07	0.34	-4.2	4.81	2.80	2.30	2.30	2.30
	PFNA	464.08	0.23	-6.51	5.48	3.00	2.70	2.70	2.70
	PFDA	514.086	0.4	-5.2	6.15	3.50	3.10	3.20	3.20
	PFUnDA	564.09	0.54	-5.2	6.82	4.30	4.00	3.80	3.80
	PFDoDA	614.1	0.54	-5.2	7.49	5.40	5.20	5.00	5.00
PFSAs	PFBS	300.1	-1.61	-3.31	1.82	2.00	1.80	1.70	1.70
	PFPeS	350.1	-1.77	-3.32	3.39	2.20	2.00	1.80	1.80
	PFHxS	400.11	-1.64	-3.32	3.16	2.60	2.70	2.10	2.10
	PFHpS	450.12	-1.81	-3.32	3.82	3.00	2.70	2.50	2.50
	PFOS	500	-1.64	-3.24	4.49	3.50	3.20	3.50	3.50
	PFNS	549.13	-1.68	-3.24	6.36	4.10	3.80	3.80	3.80
	PFDS	600.1	-1.54	-3.24	6.83	4.90	4.40	4.10	4.10
FTS	4:2 FTS	328.15	0.93	-2.64	2.04	2.00	1.90	1.70	1.70
	6:2 FTS	428.16	1.23	-2.72	4.54	2.50	2.30	2.10	2.10
	8:2 FTS	528.18	1.33	-2.61	5.58	3.70	3.50	3.30	3.30

- log K_{oc} was collected considering different pH conditions.
- Chemicalize[1]
- Comptox[2]

Table S2. List of 18 literature references considered for the collection of soil water partitioning coefficient (log K_d, L/kg), and soil properties (Soil organic carbon, SOC; Cation exchange capacity, CEC) in the training/validation datasets.

Soil Type	Country	References
Surface horizon Sphagnum peat uncontaminated PFAS soils	Sweden	Campos-Pereira, et al., 2022
Soils in this study were collected from under banana plantations in tropical Queensland.	Australia	Oliver et al., 2020
The surficial soils (0–20 cm) were collected in polyethylene (PE) self-locked packages from farmland and road greenbelt in Xiamen, China	China	Li et al., 2019
The sample (Risbergsh€ojden Oe) was collected in 2011 from a Spodosol in central Sweden , a site dominated by Scots Pine (*Pinus sylvestris*) vegetation.	Sweden	Campos Pereira et al., 2018
PFAS-contaminated soils, where PFAS in the soil came from the historic use of aqueous film-forming foams (AFFF).	Australia	Rayner et al., 2022
The sampling location was adjacent to a PFAS-contaminated site of interest.	Sweden	Niarchos et al., 2022
Samples were collected from AFFF-contaminated sites	Australia	Kabiri et al., 2022
Sandy soils collected after the last known AFFF application	United States	Schaefer et al., 2022
Four locally collected surface soils with varying properties were chosen for the present study. uncontaminated PFAS soils	Canada	Mejia-Avendaño et al., 2020
Ten soils designated as S1−S10 were chosen to represent a wide range of soil properties	Australia	Hong Nguyen et al., 2020
Eight surface soils (sampled at 0–20 cm in depth) varying in physicochemical properties were used.	Australia	Liu et al., 2020

Soil Type	Country	References
Samples collected from uncontaminated PFAS soils	Australia	Cai, et al., 2022
total of six uncontaminated soils (Soils 1−6) were selected from an archived soil collection at Oregon State University		Barzen-Hanson et al., 2017
Eleven temperate soils were chosen to represent a wide range of soil properties. uncontaminated PFAS soils	Sweden	Campos Pereira et al., 2023
Four soils were selected and sampled from different cities in China. uncontaminated PFAS soils	China	Yue-Rui Huang et al., 2023
Eight soils for spiking were collected from different uncontaminated areas (without native contamination of PFAS) with various physiochemical properties.	China	Pengfei Zhou, et al., 2024
Three natural clay minerals with different compositional and structural properties were used.	*	Qianqian Dong, et al., 2024
The three common clay minerals used in this study were kaolinite, illite, and montmorillonite.	*	Aamir Ahmad et al., 2024

Description of color patter used

Clay minerology
Uncontaminated samples
Organic soil samples
PFAS contaminated samples

Text S1. Molecular Connectivity Indices and Abraham Descriptors

Molecular connectivity is a method of molecular structure quantitation in which weighted counts of substructure fragments are incorporated into numerical indices. Structural features, such as size, branching, unsaturation, heteroatom content, and cyclicity, are encoded. In other words, this method describes molecular structure based solely on the molecule bonding and branching patterns. The calculation of the indices begins with the reduction of the molecule to the hydrogen-suppressed skeleton or graph. Each atom is assigned two atom descriptors based upon the count of sigma electrons or valence electrons present, other than those bonded to hydrogen atoms (Molconn Z tutorial). The molecular connectivity indices (MCI) or chi indices are symbolized as ${}^{n}\chi_{c}$.

The superscript 'n' represents the order of the index. Index order increases with the branching structure of a molecule. Substructures for a molecular skeleton are defined by the decomposition of the skeleton into fragments of:

a) atoms: zero order n = 0
b) one bond paths, first order, n = 1,
c) two bond fragments, second order n = 2, etc.)
d) three contiguous bond fragments (third order Path, m = 3, c = P)

The subscript 'c' represents the fragment configuration (p for path, c for cluster, ch for chain, pc for path-cluster).

a) cluster (three atoms attached to a central atom, m = 3, c = C)
b) the path/cluster (equivalent to the isopentane skeleton, m = 4, c = PC);
c) the chain fragment (ring) (cycles of 3, 4, 5 . . . atoms, m = 3, 4, 5. . ., c = CH). In the case of PFAS, CH indexes are zero because no rings are part of the molecule.

The second superscript 'v', in ${}^{n}\chi^{v}_{c}$, is calculated based on the molecule's valence electrons and represents the electronic information of a compound (Hall & Kier, 1991). Therefore, for each order and fragment type, a connectivity index may be calculated. In the present study, 56 connectivity descriptors were collected from PADEL software using PFAS molecules' SMILES information in ionic form. Those descriptors are described in Table S4. It is important to note that MCIs do not represent physical or chemical properties but are highly correlated with many properties like molar volume, solubility, and refraction.

a) Simple

The general formula for the simple chi connectivity indices ($^{m}\chi_t$ or xtm) is as follows:

$$^{m}\chi_t = \sum_{i=1}^{A} {}^{m}c_i$$

where m = order of the connectivity index, represents the number of bonds in a subgraph of type t (0 = atoms, 1 = fragments of one bond, 2 = fragments of two bonds, etc.), t = type of calculation (p = path, c = cluster, pc = path/cluster, ch = chain or cycle) and A = the number of non-hydrogen atoms in the molecule.

Example: m = 1 represents paths of length 1 i.e., bonds,

The bond contributions to the connectivity indices, named Cij by Kier and Hall, $^{m}C_i$ is calculated for all of the fragments of type t and path length m in the hydrogen-depleted graph of the molecule:

$$^{m}c_i = \prod_{k=1}^{m+1} (\delta_k)^{-0.5}$$

where k = the different atoms in the fragment and δ_k is the vertex degree of an atom given by

$$\delta_k = \sigma_k - h_k$$

Where σ_k is the number of electrons in sigma orbitals and h_k is the number of bonded hydrogen atoms (Simply – δ_k is the # of non-hydrogen atoms bonded to atom k).

a) Valence

The valence connectivity indices ($^{m}\chi^{v}_t$ or xvtm) are calculated in the same fashion as the simple connectivity indices except that the vertex degree is replaced by the valence vertex degree (δ^{v}_k) to give

$$^{m}c_i = \prod_{k=1}^{m+1} (\delta_k^v)^{-0.5}$$

where the valence vertex degree is given by

$$\delta_k^v = Z_k^v - h_k = \sigma_k + \pi_k + n_k - h_k$$

Where Z^v_k is the number of valence electrons, π_k = number of electrons in pi orbitals, n_k is the number of electrons in lone-pair orbitals, and h_k is the number of bonded hydrogen atoms.

For atoms of higher principal quantum levels, the valence vertex degree is given by

$$\delta_k^v = (Z_k^v - h_k)/(Z_k - Z_k^v - 1)$$

where Z_k is the number of electrons in atom k (the atomic number).

Table S3. Interpretation of molecular connectivity indices in the context of PFAS.

Class	Index Order (n)	Meaning of descriptors
SP - Simple path	n = 0 . . . n = 8	Single atoms and a group of 2 or more atoms connected by sigma bonds; these indices capture molecular size and chain structure. Example: SP-0: C, F, O SP-1: C-C, SP-3: F-C-C, C-C-O SP-4: F-C-C-C-O, etc.
SC - Simple Clusters	n = 3 . . . n = 6	A group of 4 or more atoms sharing a central atom and connected by sigma bonds, these indices partially capture functional groups and consider branches. Example: SC-3: CF_3, SC-6: CF_3-CF_2,
VP - Valence path	8	Single atoms and a group of 2 or more atoms connected by sigma and pi bonds; these indices include all valence electrons and capture molecular size and chain structure. Example: VP-0: :C:, :F: VP-3: F-C-C, F-C-C, C-COOH
VC - Valence cluster	n = 3 . . . n = 6	A group of 4 or more atoms sharing a central atom and connected by sigma bonds and pi bonds; these indices include all valence electrons and capture perfluorinated sections of the molecule and functional groups. Example: VC-3: CF_3, VC-6: CF_3-CF_2, -CF_2-COOH
SPC or VPC - Simple or average path cluster	n = 4 . . .	Combination of a simple or valence cluster with any order path; these indices consider the molecular structure itself. Example: SPC-3 or VPC-3: CF_3-C, C-COOH

Class	Index Order (n)	Meaning of descriptors
	n = 6	SC-6: CF_3-CF_2-C
ASP or AVP – Simple or Valence Average Simple Path	n = 0 . . . n = 7	These indices are calculated by dividing the total index value by the number of contributing paths. Gives a molecular size-independent metric. Example: PFBA: total 13 atoms including F, C, and O ASP-0 = SP-0/13 (13 atoms) ASP-1 = SP-1/12 (12 paths of 2 atoms) AVP-5 = VP-5/6 (6 different paths of 5 atoms)
Total	46	

Abraham Descriptors

Abraham's LSER approach describes solvation or related activities by the compounds' physicochemical properties. These models provide mechanistic insights because they reveal intermolecular interactions between adsorbents and adsorbates as well as quantifying their relative individual contribution to the adsorption. A typical LSER model that describes the adsorption of neutral organic compounds by using a set of predetermined independent descriptors is shown in Eq. (1):

$$Log\ k_d = c + eE + sS + aA + bB + vV \quad (1)$$

where log K_d is the partitioning coefficient between MPs and water under equilibrium conditions. E is the excess molar refraction in units of $(cm^3\ mol^{-1})/10$, representing non-specific van der Waals forces; S is the polarizability/dipolarity parameter, A and B are the hydrogen bond donating and accepting abilities, respectively; and 'V' is the molecular volume or McGowan's volume in units of $(cm^3\ mol^{-1})/100$. Lastly, c is the regression constant, and e, s, a, b, and v are the fitting coefficients indicating the contribution of each interaction on adsorption (Apul et al., 2020; Xu et al., 2021). Table S5 presents the Abraham descriptors at acidic and ionic states used to calculate the descriptors based on ionization percentage.

Table S4. Abraham descriptors for carboxylic PFAS reported by Hatinoglu et al. (2023).[3]

Compounds	E	E'	S	S'	A	A'	B	B'	V	V'	J-
PFBA, C4	-0.47	-0.32	0.1	1.27	0.46	0.02	0.33	2.80	0.87	0.85	1.95
PFPeA, C5	-0.62	-0.47	-0.02	1.12	0.46	0.04	0.33	2.88	1.05	1.03	2.00
PFHxA, C6	-0.77	-0.62	-0.15	0.95	0.46	0.06	0.33	2.97	1.22	1.2	2.0
PFHpA, C7	-0.7	-0.55	-0.27	1.00	0.46	0.08	0.33	3.05	1.40	1.38	2.15
PFOA, C8	-0.88	-0.73	-0.39	0.82	0.46	0.10	0.33	3.13	1.58	1.55	2.19
PFNA, C9	-1.03	-0.88	-0.51	0.66	0.46	0.12	0.33	3.21	1.75	1.73	2.23
PFDA, C10	-1.19	-1.04	-0.64	0.49	0.46	0.14	0.33	3.29	1.93	1.91	2.27
PFUnA, C11	-1.34	-1.19	-0.76	0.33	0.46	0.16	0.33	3.37	2.10	2.08	2.32
PFDoA, C12	-1.49	-1.34	-0.88	0.18	0.46	0.18	0.33	3.45	2.28	2.26	2.37

-C#: Number of C atoms in the molecule

MCI and Abraham values were corrected to reflect the dissociation state of PFAS during sorption experiments. Eq. (2) was used to calculate the ionization percentage of each compound, considering soil pH and pK_a.

$$\%dissociation = \frac{10^{soil\,pH - pka}}{1 + 10^{soil\,pH - pka}} * 100 \qquad (2)$$

Eq. (3) was used to adjust either MCI or Abraham predictors based on the ionic and acidic dissociation state of each compound.

$$Adjusted\,predictor = \frac{X_{acid} * P_{acid} + X_{ionic} * P_{ionic}}{100\%} \qquad (3)$$

where X_{acid} is the fraction of the compound at the acid state, X_{ionic} is the fraction of the compound at the ionic state. Considering that pK_a values are lower than 2, most of the compounds are almost 100% dissociated. P_{acid} is the predictor (MCI or Abraham) at acidic state, while P_{ionic} is the predictor at (MCI or Abraham) at ionic state.

Example: Dissociation percentage

PFHxA – Perfluorobutanoic Acid, $pk_a = 0.2$

Soil pH = 7.50

$$\%dissociation = \frac{10^{7.50 - 0.2}}{1 + 10^{7.50 - 0.2}} * 100$$

$$\%dissociation = 99.99\%$$

Soil pH = 2.80

$$\%dissociation = \frac{10^{2.80 - 0.2}}{1 + 10^{2.80 - 0.2}} * 100$$

$$\%dissociation = 99.74\%$$

Example: MCI adjusted based on dissociation percentage

VC-3 – Valence Cluster order 3

$VC\text{-}3_{Ion} = 0.465980$

$VC\text{-}3_{Acid} = 0.511623$

PFHxA fraction

Ion = 99.74%

Acid = 0.26%

VC-3 – Valence Cluster order 3 adjusted

$$Adjusted\ predictor = \frac{0.26 * 0.511623 + 99.74 * 0.465980}{100\%}$$

$$Adjusted\ predictor = 0.46609$$

Table S5. Descriptive statistics of the PFCAs dataset used to compare the predictive power of the Abraham descriptors versus MCIs. (N3M: Normal 3 mixture; N2M: Normal 2 mixture; and SHASH: Sinh–Arcsinh).

Data	Measure	SOC (%)	CEC (cmol/kg)	E'	VP-7	log K_d
PFCAs training	Distribution	SHASH	N3M	N3M	N3M	N3M
	N	245	207	245	245	245
	Mean	6.25	16.32	-0.80	0.06	0.56
	SD	14.44	12.04	0.31	0.05	1.13
	Min	0.00	0.50	-1.34	0.00	-1.37
	Max	53.60	41.40	-0.32	0.15	3.33
	IQR	1.90	14.50	0.49	0.10	1.62
	CV	231	74	-39	82	200
	Median	0.70	16.00	-0.73	0.06	0.30
PFCAs validation	Distribution	SHASH	N3M	N3M	N3M	N2M
	N	82	76	82	82	81
	Mean	4.27	15.67	-0.81	0.07	0.52
	SD	12.45	13.83	0.31	0.05	1.05
	Min	0.00	0.50	-1.34	0.00	-0.93
	Max	53.60	80.00	-0.32	0.15	3.10
	IQR	1.00	10.30	0.49	0.10	1.62
	CV	292	88	-38	78	203
	Median	0.40	8.00	-0.73	0.06	0.22

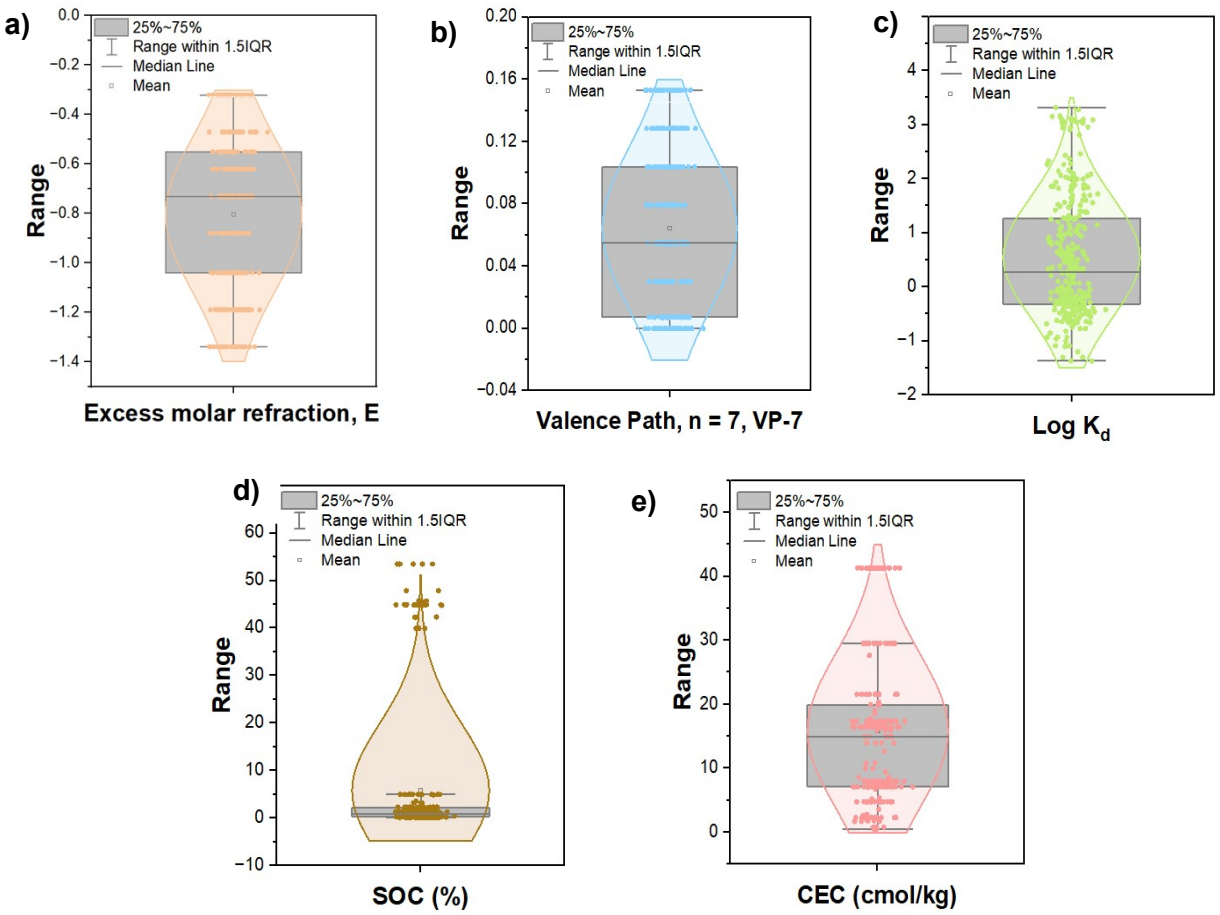

Figure S1. Combined violin and box-and-whisker plot of the most important molecular predictors a) Excess molar refraction, E', and b) Valence path, VP-7; soil properties c) soil organic carbon, SOC, and d) Cation exchange capacity, CEC; and e) log K_d values in the PFCAs dataset ($N = 327$). The violin shape (light color outline) indicates the distribution and frequency across the dataset. Wider sections of the violin correspond to more frequent values. The gray box shows the interquartile range (25–75%), with the black horizontal line representing the median. Whiskers extend to 1.5 times the interquartile range, and individual dots indicate observed data points.

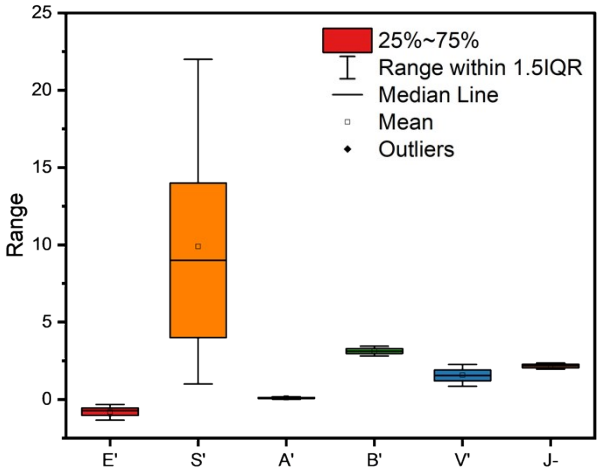

Figure S2. Abraham descriptors considered for model development. The light colors indicate the distribution and frequency across the database. The box shows the interquartile range (25–75%), with the black horizontal line representing the median. Whiskers extend to 1.5 times the interquartile range, and individual dots indicate observed data points.

Table S6. Descriptive statistics of the PFAS training/validation and external (generalization) datasets compiled in this study for log K_d (N3M: Normal 3 mixture; N2M: Normal 2 mixture; and SHASH: Sinh–Arcsinh).

Data	Measure	SOC	CEC	SP-3	ASP-0	ASP-1	AVP-0	log K_d
PFAS Training	Distribution	N3M	N3M	N3M	N3M	N3M	N3M	N2M
	N	482	480	524	524	524	524	523
	Mean	1.13	16.27	11.16	0.84	0.43	0.41	0.49
	SD	1.36	11.98	3.36	0.01	0.01	0.01	0.99
	Min	0.00	0.50	4.93	0.84	0.43	0.39	-1.37
	Max	7.70	80.00	17.43	0.85	0.45	0.43	3.33
	IQR	1.15	14.50	4.69	0.01	0.01	0.01	1.45
	CV	121	74	30	1	1	3	205
	Median	0.70	16.50	11.18	0.84	0.43	0.41	0.34
PFAS Validation	Distribution	N3M	N3M	N3M	N3M	N3M	N3M	N2M
	N	166	164	175	175	175	175	175
	Mean	1.09	14.32	11.18	0.84	0.43	0.41	0.53
	SD	1.35	11.57	3.39	0.01	0.01	0.01	1.03
	Min	0.00	0.50	4.93	0.84	0.43	0.39	-1.30
	Max	4.90	41.40	17.43	0.85	0.45	0.43	3.17
	IQR	1.00	12.78	4.69	0.01	0.01	0.01	1.45
	CV	124	81	30	1	1	3	195
	Median	0.40	8.00	11.18	0.84	0.43	0.41	0.32
Independent External Validation	Distribution	lognormal	SHASH	SHASH	SHASH	SHASH	SHASH	N3M
	N	628	658	658	658	658	658	628
	Mean	3.19	23.78	11.61	0.84	0.44	0.41	0.88
	SD	5.25	26.39	2.94	0.01	0.01	0.02	0.81
	Min	0.03	0.10	1.73	0.82	0.43	0.36	-1.15
	Max	37.60	140.00	18.99	0.87	0.49	0.51	3.57
	IQR	2.80	18.00	1.95	0.00	0.01	0.01	1.17
	CV	165	111	25	1	3	4	92
	Median	1.50	14.40	12.43	0.84	0.43	0.41	0.90

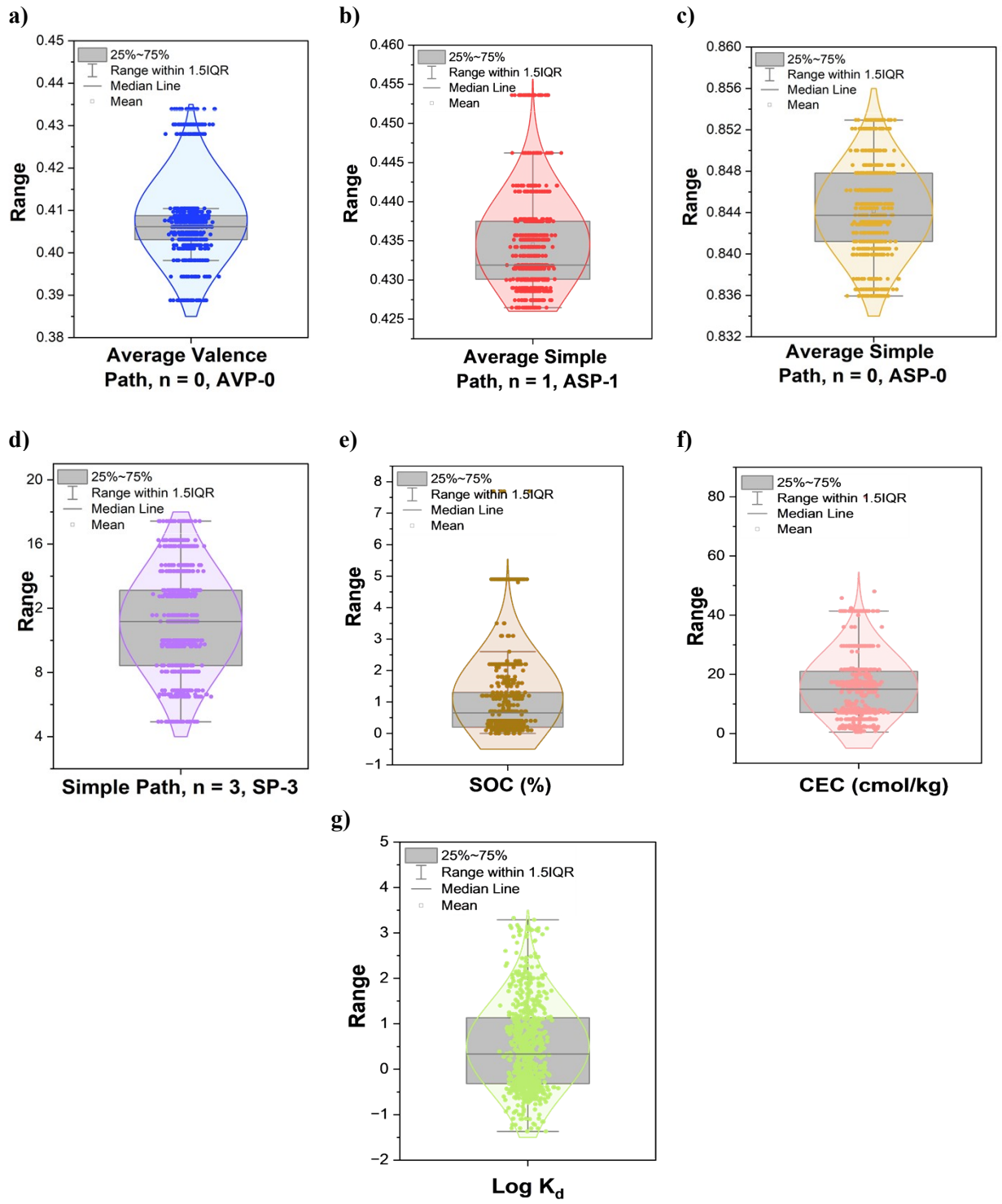

Figure S3. Combined violin and box-and-whisker plot of the most important molecular predictors a-c) Average simple and valence path index order 0 and 1, ASP-0, ASP-1, AVP-0; soil properties c) soil organic carbon, SOC, and d) cation exchange capacity, CEC; and e) log K_d values in the PFAS dataset (N = 699). The violin shape (light color outline) indicates the distribution and frequency across the dataset. Wider sections of the violin correspond to more frequent values. The gray box shows the interquartile range (25–75%), with the black horizontal line representing the

median. Whiskers extend to 1.5 times the interquartile range, and individual dots indicate observed data points.

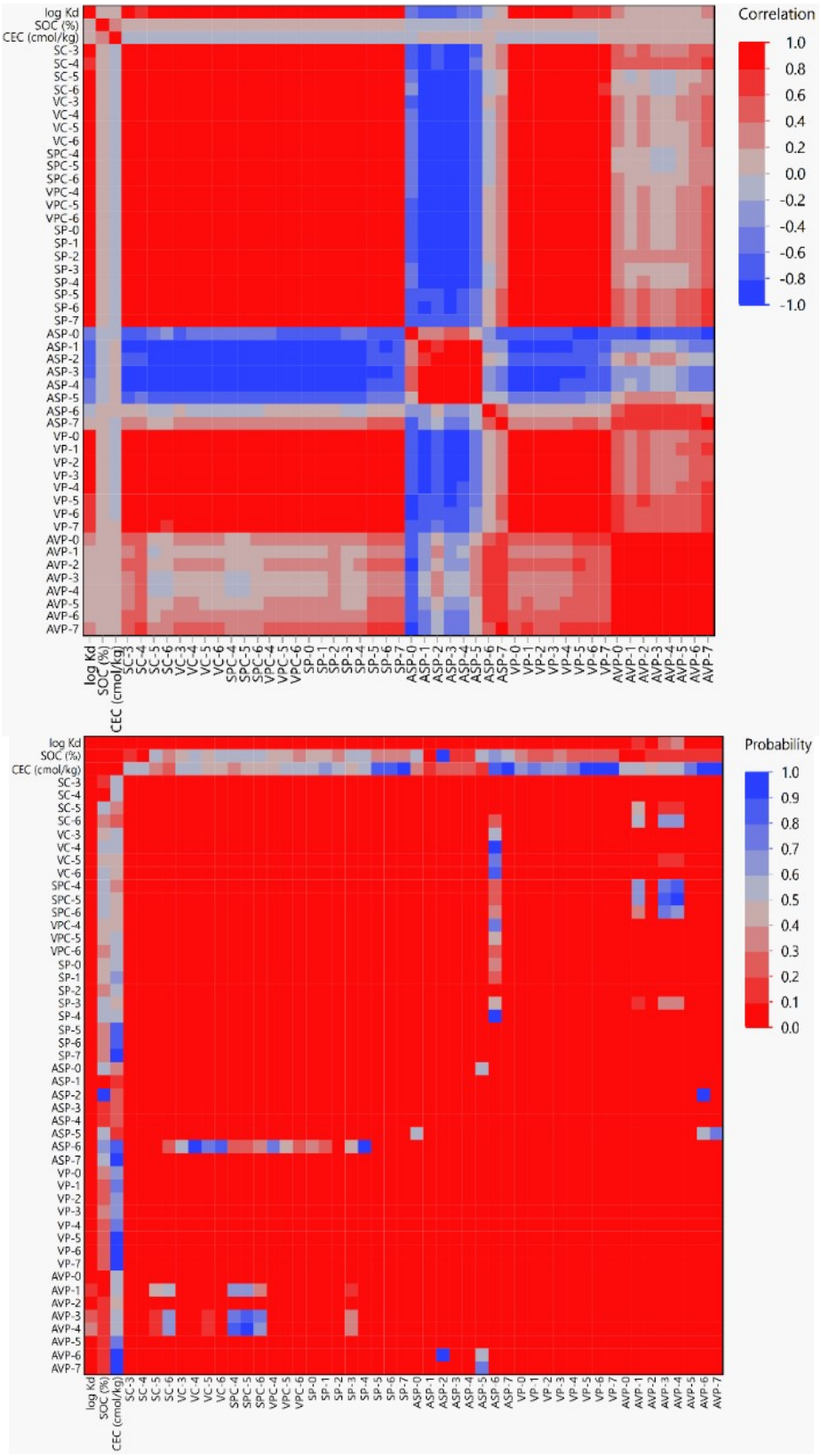

Figure S4. Pearson's correlation matrix analysis of log K_d, SOC, CEC, and 46 MCIs used for predictor-screening stage (the upper and lower parts refer to correlation coefficients and their associated *p*-values, respectively).

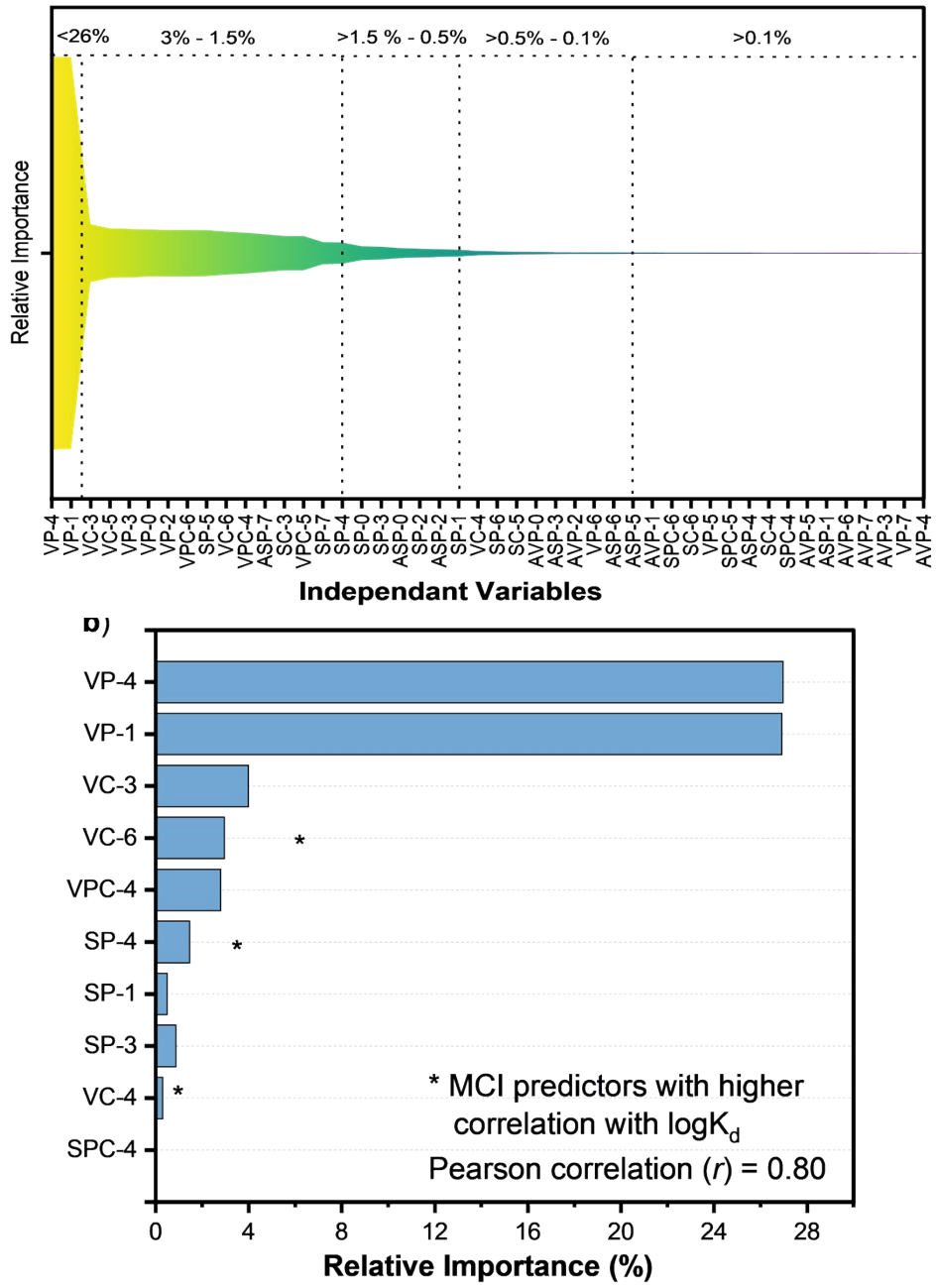

Figure S5. Screening results of individual MCIs based on (1) non-linear importance via RF (≥ 4%; N = one million trees), (2) Pearson's correlation with log K_d (absolute r value ≥ 0.85; N = 699), and (3) joint predictive power via preliminary MLR_{EN} models (including singular MCI and the soil properties) (R^2_{pred} ≥ 77%; N = 163) for log K_d

Table S7. Screening results of individual MCIs based on (1) non-linear importance via RF (≥ 4%; N = one million trees), (2) Pearson's correlation with log K_d (absolute r value ≥ 0.85; N = 699),

and (3) joint predictive power via preliminary MLR_{EN} models (including singular MCI and the soil properties) ($R^2_{pred} \geq 77\%$; $N = 163$) for log K_d.

ID	MCI	Relative importance (%)	*r*	R^2_{pred} (%)	ID	MCI	Relative importance (%)	*r*	R^2_{pred} (%)
1	VP-4	26.97	0.82	68.58	24	SP-6	0.24	0.81	67.02
2	VP-1	26.91	0.82	69.52	25	SC-5	0.19	0.83	76.02
3	VC-3	3.99	0.85	76.78	26	AVP-0	0.17	0.20	6.00
4	VC-5	3.42	0.85	77.58	27	ASP-3	0.13	-0.65	47.51
5	VP-3	3.33	0.84	72.71	28	AVP-2	0.12	0.18	5.31
6	VP-0	3.24	0.84	73.22	29	VP-6	0.12	0.79	61.67
7	VP-2	3.21	0.83	71.73	30	ASP-6	0.11	-0.11	6.41
8	VPC-6	3.20	0.85	75.89	31	ASP-5	0.10	-0.46	28.60
9	SP-5	3.16	0.81	66.87	32	AVP-1	0.09	0.06	4.15
10	VC-6	2.96	0.85	77.66	33	SPC-6	0.08	0.84	77.70
11	VPC-4	2.80	0.85	77.09	34	SC-6	0.08	0.81	73.81
12	ASP-7	2.60	0.09	4.37	35	VP-5	0.08	0.78	60.38
13	SC-3	2.37	0.82	70.88	36	SPC-5	0.08	0.84	77.70
14	VPC-5	2.36	0.85	76.53	37	ASP-4	0.07	-0.59	41.46
15	SP-7	1.52	0.81	66.75	38	SC-4	0.06	0.76	60.53
16	SP-4	1.47	0.85	77.53	39	SPC-4	0.06	0.84	77.57
17	SP-0	0.95	0.85	76.14	40	AVP-5	0.06	0.10	4.39
18	SP-3	0.88	0.85	77.91	41	ASP-1	0.06	-0.64	47.03
19	ASP-0	0.68	-0.47	20.89	42	AVP-6	0.05	0.16	5.48
20	SP-2	0.61	0.84	74.82	43	AVP-7	0.05	0.29	10.10
21	ASP-2	0.53	-0.68	53.68	44	AVP-3	0.03	0.04	4.19
22	SP-1	0.47	0.85	75.63	45	VP-7	0.03	0.79	62.92
23	VC-4	0.31	0.85	77.49	46	AVP-4	0.01	0.04	4.21

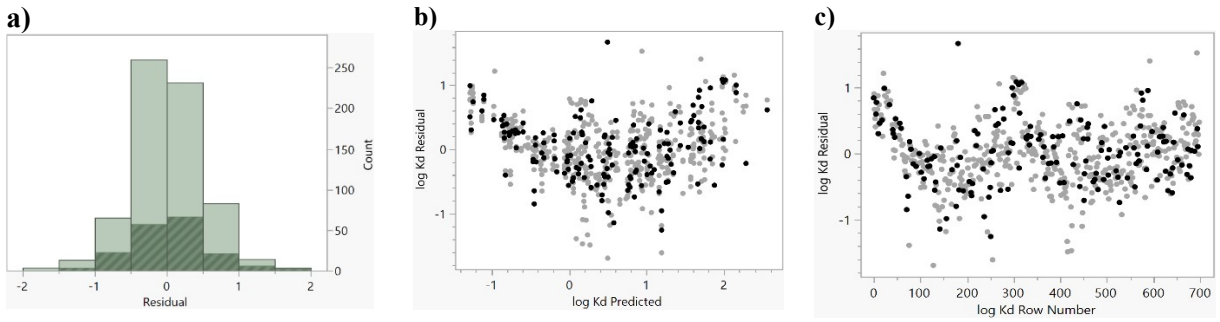

Figure S6. Diagnostic plots of a) residuals for the normality assumption, b) residuals by predicted values for the constant variance assumption, and c) residuals by order of data for the autocorrelation assumption based on training (75%) (gray dots; N = 484) and validation (25%) (black dots or patterned bars; $N = 154$) sub datasets for Model M7 (MLR_{EN} using SP-3 and the soil properties).

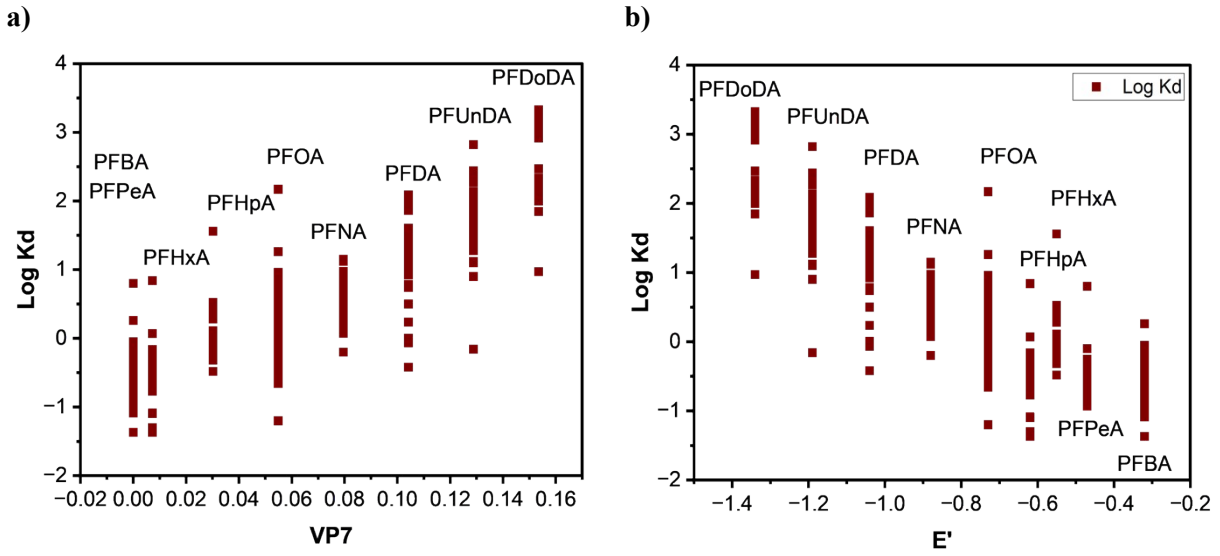

Figure S7. Distribution of soil-water partitioning coefficients (log K_d, L/kg) for 9 carboxylic PFAS (the PFCAs dataset) against the most important predictors. Seventh-order valence path (VP-7) predictor values increased with the C-chain, while excess molar refraction (E') decreased with longer C-chain in carboxylic PFAS.

Table S8. Parameters and performance metrics of the best-fit MLR_{EN} models for the Abraham descriptors and MCIs.

a) Model parameters

Predictor	Predictor	Intercept	Slope	SE	Wald χ^2	*P*	VIF
Abraham		-2.11		0.11	347	<0.0001	0.0
	E		-3.15	0.12	683	<0.0001	1.2
	SOC		0.072	0.04	4	0.0389	1.4
	CEC		0.004	0.003	2	0.159	1.0
MCI		-0.79		0.06	196	<0.0001	0.0
	VP-7		19.06	0.61	976	<0.0001	1.0
	SOC		0.07	0.03	7	0.0091	1.0
	CEC		0.003	0.002	3	0.102	0.7

- V: Validation; T: Training; Wald χ^2**:** Wald Chi-Square Test; P: p-value; VIF: Variance Inflation Factor; SE: Standard Error;

b) Model performance metrics

Predictor	T			V		
	R^2_{adj} (%)	RSME	AIC_c	R^2_{pred} (%)	RMSE	AIC_c
Abraham (E)	80.2	0.51	314	79.8	0.47	112
MCIs (VP-7)	82.9	0.47	295	84.8	0.40	81

- V: Validation; T: Training
- R^2_{pred}: goodness-of-fit on validation data; R^2_{adj}: goodness-of-fit on training data
- RMSE: Root mean squared error.
- AIC_c: Corrected Akaike Information Criterion
- Predictor Model with Abraham predictor, E: Samples size T, ($N = 207$), Sample size V, ($N = 76$)
- Predictive Model with MCI predictor, VP-7: Samples size T, ($N = 216$), Sample size V, ($N = 67$)

Table S9. Parameters and performance metrics of the best-fit simple linear regression (SLR) models (N = 523 for training and 175 for validation; $p < 0.0001$).

Model parameters						Performance metrics			
						T		V	
Model N.-	MCI Predictor	Intercept	Slope	SE	*t*-ratio	R^2 (%)	RSME	R^2_{Pred} (%)	RMSE
S1	VP-1	-2.578	0.608	0.093 0.018	-27.4 33.7	66.83	0.556	64.18	0.612
S2	VP-4	-1.651	1.921	0.068 0.057	-24.0 33.2	68.01	0.562	63.31	0.619
S3	VC-3	-2.284	2.391	0.078 0.065	-29.0 36.7	72.19	0.524	70.73	0.553
S4	VC-4	-2.209	22.269	0.075 0.598	-29.0 37.1	72.63	0.519	71.39	0.547
S5	VC-6	-1.833	21.250	0.067 0.578	-27.3 36.7	72.18	0.524	71.50	0.546
S6	VPC-4	-1.964	1.207	0.071 0.033	-27.6 36.5	71.89	0.526	70.96	0.551
S7	SP-3	-2.323	0.251	0.079 0.006	-29.0 36.6	72.08	0.525	71.73	0.543
S8	SP-4	-1.972	0.508	0.069 0.013	-28.1 37.1	72.61	0.520	71.43	0.546
S9	SPC-4	-2.282	0.153	0.082 0.004	-27.7 35.1	70.28	0.541	71.28	0.548

- V: Validation; T: Training; Wald χ^2: Wald Chi-Square Test; *P*: Significance Level; VIF: Variance Inflation Factor; SE: Standard Error; RMSE: Root Mean Squared Error.

Table S10. Parameters and performance metrics of the MLR_{EN} and ANN models ($N = 475$ for training and 163 for validation).

a) Model parameters

Model type	Model N.-	Predictor	Intercept	Slope	SE	Wald χ^2	*P*	VIF
MLR_{EN} with singular MCI	M1		-2.86		0.249	585	<0.0001	0
		VP-1		0.63	0.022	810	<0.0001	1.4
		SOC		0.09	0.016	34	<0.0001	0.9
		CEC		0.002	0.001	2	0.137	0.8
	M2		-1.89		0.088	465	<0.0001	0
		VP-4		1.99	0.070	793	<0.0001	1.4
		SOC		0.098	0.017	32	<0.0001	0.9
		CEC		0.002	0.001	2	0.151	0.8
	M3		-2.57		0.097	693	<0.0001	0
		VC-3		2.48	0.077	1124	<0.0001	1.4
		SOC		0.10	0.014	51	<0.0001	0.8
		CEC		0.003	0.001	4	0.044	0.8
	M4		-2.48		0. 089	771	<0.0001	0
		VC-4		23.0	0.678	1155	<0.0001	1.3
		SOC		0.11	0.014	55	<0.0001	0.8
		CEC		0.003	0.001	3	0.068	0.8
	M5		-2.09		0.080	677	<0.0001	0
		VC-6		22.0	0.659	1112	<0.0001	1.3
		SOC		0.11	0.014	52	<0.0001	0.8
		CEC		0.003	0.001	4	0.041	0.8
	M6		-2.24		0.088	646	<0.0001	0
		VPC-4		1.25	0.039	1122	<0.0001	1.4
		SOC		0.10	0.014	50	<0.0001	0.8
		CEC		0.003	0.001	4	0.031	0.8
	M7		-2.59		0.090	818	<0.0001	0
		SP-3		0.26	0.007	1190	<0.0001	1.2
		SOC		0.11	0.015	53	<0.0001	0.8
		CEC		0.003	0.001	3	0.050	0.8
	M8		-2.24		0.082	728	<0.0001	0
		SP-4		0.53	0.015	1153	<0.0001	1.3
		SOC		0.11	0.014	55	<0.0001	0.8
		CEC		0.003	0.001	3	0.063	0.8
	M9		-2.56		0.092	761	<0.0001	0
		SPC-4		0.16	0.004	1109	<0.0001	1.2
		SOC		0.11	0.015	45	<0.0001	0.9
		CEC		0.004	0.001	5	0.019	0.8

Continue in the next page

Model type	Model N.-	Predictor	Intercept	Slope	SE	Wald χ^2	*P*	VIF
			-183.13		14.08	168	<0.0001	0
MLR_{EN} with multiple MCIs	M10	SP-3		0.52	0.018	737	<0.0001	12
		ASP-1		123.7	7.521	270	<0.0001	6
		ASP-0		122.0	11.32	115	<0.0001	8
		AVP-0		51.5	4.495	131	<0.0001	6
		SOC		0.12	0.012	90	<0.0001	0.8
		CEC		0.002	0.001	3	0.060	0.9
ANN with multiple MCIs	A11	SP-3 ASP-1 AVP-0 ASP-0 SOC CEC						0

- Wald χ^2: Wald Chi-Square Test; *P*: Significance Level; VIF: Variance Inflation Factor; SE: Standard Error; RMSE: Root Mean Squared Error.

b) Model performance metrics

Model type	Model N.-	Predictor importance (highest→lowest)	T			V		
			R^2_{adj} (%)	RSME	AIC_c	R^2_{pred} (%)	RMSE	AIC_c
MLR_{EN} singular MCI	M1	VP-1, SOC, CEC	72.61	0.726	741	69.52	0.695	288
	M2	VP-4, SOC, CEC	71.88	0.529	754	68.57	0.573	309
	M3	VC-3, SOC, CEC	76.94	0.479	659	76.78	0.493	242
	M4	VC-4, SOC, CEC	77.62	0.472	645	77.48	0.485	237
	M5	VC-6, SOC, CEC	77.05	0.478	657	77.65	0.483	236
	M6	VPC-4, SOC, CEC	76.61	0.482	666	77.08	0.489	240
	M7	SP-3, SOC, CEC	77.06	0.478	657	77.91	0.481	234
	M8	SP-4, SOC, CEC	77.58	0.472	646	77.53	0.485	237
	M9	SPC-4, SOC, CEC	74.96	0.499	698	74.96	0.484	237
MLR_{EN} multiple MCIs	M10	SP-3, ASP-1, ASP-0, AVP-0, SOC, CEC	84.37	0.394	481	83.70	0.413	191
ANN multiple MCIs	A11	SP-3, ASP-1, AVP-0, ASP-0, SOC, CEC	86.30*	0.369	0.276**	84.89	0.397	0.302**

- V: Validation; T: Training
- *R^2: goodness-of-fit on training data in ANN, R^2_{adj}: goodness-of-fit on training data for MLR_{EN}, R^2_{pred}: goodness-of-fit on validation data
- RMSE: Root mean squared error.
- AIC_c: Corrected Akaike Information Criterion
- **MAD = Mean Absolute Deviation

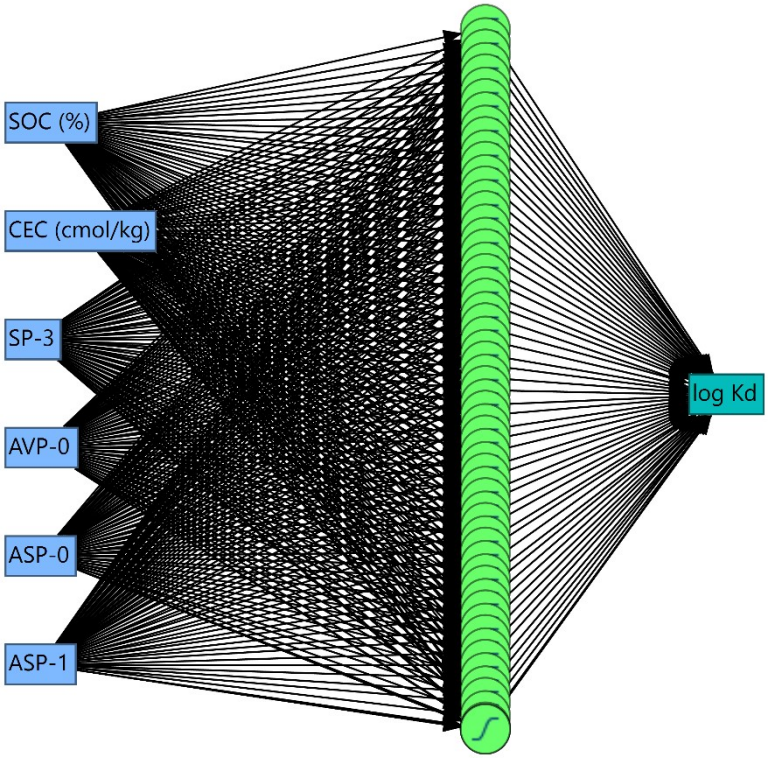

Figure S8. The architecture of the artificial neural network (ANN) model (Model A11) used in the study. Six inputs (blue boxes) supply the model with two key soil properties—soil organic carbon content (SOC, %), cation-exchange capacity (CEC, cmol kg^{-1})—and four molecular connectivity indices (SP-3, AVP-0, ASP-0, and ASP-1) connected by black lines to a single hidden layer comprising 50 neurons (green circles) that predict one output, log K_d. This configuration allows the network to capture complex, nonlinear interactions between molecular descriptors and soil properties that govern PFAS sorption behavior.

Table S11. Scenario analysis (for maximizing log K_d) results for Elastic Net-regularized multiple linear regression (MLR_{EN}) and ANN models as a function of the MCIs and soil properties.

Model type	Model N.-	Predictor	Optimal value	Maximum log K_d value	*D*
MLR_{EN} with singular MCI	M1	VP-1	7.3	2.7 (2.4-3.0)	0.830
		SOC	7.7		
		CEC	80		
	M2	VP-4	1.8	2.7 (2.4-3.0)	0.825
		SOC	7.7		
		CEC	80		
	M3	VC-3	1.7	2.9 (2.7-3.2)	0.876
		SOC	7.7		
		CEC	80		
	M4	VC-4	0.1	3.0 (2.7-3.2)	0.891
		SOC	7.7		
		CEC	80		
	M5	VC-6	0.1	3.0 (2.7-3.2)	0.886
		SOC	7.7		
		CEC	80		
	M6	VPC-4	3.2	2.9 (2.7-3.2)	0.876
		SOC	7.7		
		CEC	80		
	M7	SP-3	17.4	3.0 (2.8-3.3)	0.894
		SOC	7.7		
		CEC	80		
	M8	SP-4	7.9	3.0 (2.7-3.2)	0.890
		SOC	7.7		
		CEC	80		
	M9	SPC-4	27.8	3.0 (2.7-3.2)	0.885
		SOC	7.7		
		CEC	80		
MLR_{EN} with multiple MCIs	M10	SP-3	17.4	9.4 (8.5-10.3)	0.999
		ASP-1	0.4		
		ASP-0	0.8		
		AVP-0	0.4		
		SOC	7.7		
		CEC	80		
ANN with multiple MCIs	A11	SP-3	17.4	6.7	0.999
		ASP-1	0.4		
		AVP-0	0.4		
		ASP-0	0.8		
		SOC	7.7		
		CEC	80		

- D: Desirability function scale from 0 (completely undesirable) to 1 (fully desirable)

a) b)

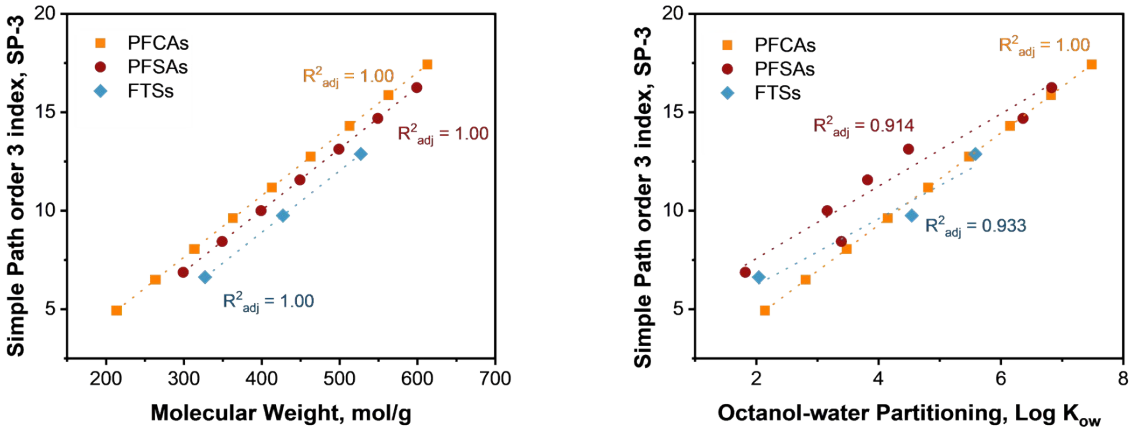

Figure S9. a) Molecular weight (mol/g) and octanol-water partitioning, log K_{ow} versus simple path order 3, SP-3 index for PFCAs, PFSAs, and FTSs.

Table S12. Uncertainty analysis results based on Monte Carlo simulations and Model M7 (estimates with 95% CIs).

Component	Model M7	Monte Carlo simulation
Low sorption (High PFAS mobility)	μ = –0.85 95% CI = –0.91, –0.79 σ = 0.26 π = 0.11	μ = –0.28 95% CI = –0.31, –0.26 σ = 0.50 π = 0.31
Mid sorption	μ = 0.54 95% CI = 0.48, 0.60 σ = 0.70 π = 0.79	μ = 0.69 95% CI = 0.67, 0.70 σ = 0.46 π = 0.51
High sorption (High PFAS retention)	μ = 1.80 95% CI = 1.75, 1.85 σ = 0.21 π = 0.11	μ = 1.63 95% CI = 1.60, 1.66 σ = 0.42 π = 0.18
N	638	5,000
Distribution	Normal 3 Mixture	Normal 3 Mixture
Mean [95% CI]	0.52 [0.45, 0.59]	0.56 [0.53, 0.58]
SD	0.88 [0.84, 0.94]	0.81 [0.79, 0.82]
Median	0.50	0.56
IQR	1.23	1.13
Prediction range	[–1.28, 2.56]	[–1.65, 2.91]
Probability of extremes (π1 + π3)	21.5%	49.2%

N: Samples size
SD: Standard deviation
IQR: Interquartile range
μ: component (low/mid/high)-specific mean sorption;
σ: component-specific standard deviation;
π: component-specific proportion; and
95% CI: 95% confidence interval for component-specific mean

Table S13. List of studies where 658 data points are derived for independent external validation dataset.

Author	Reference
Zhi, Y. & Liu, J., 2018[4]	10.1016/j.cej.2018.04.042
Xiao, F. et al., 2019[5]	10.1021/acs.est.9b05379
Liu, J. & Lee, L., 2007[6]	10.1021/es070228n
Liu, J. & Lee, L., 2005[7]	10.1021/es051125c
Richey, D. et al., 1997[8]	10.1021/es960649x
McLachlan, M. et al., 2019[9]	10.1016/j.chemosphere.2019.02.012
Guelfo, J. & Higgins, C., 2013[10]	10.1021/es3048043
Gredelj, A. et al., 2020[11]	10.1016/j.scitotenv.2019.134766
Cai, W. et al., 2022[12]	10.1016/j.scitotenv.2022.152975
Fabregat-Palau, J. et al., 2021[13]	10.1016/j.scitotenv.2021.149343
Enevoldsen, R. & Juhler, R., 2010[14]	10.1007/s00216-010-4066-0
Aly, Y., et al., 2019[15]	10.1039/c9ew00426b
Chen, Y. et al., 2013[16]	10.1080/19443994.2013.792145
Yin, C. et al., 2022[17]	10.1016/j.envpol.2022.118957
Higgins, C. & Luthy, R., 2006[18]	10.1021/es061000n
Jeon, J. et al., 2011[19]	10.1039/c0em00791a
Miao, Y. et al., 2017[20]	10.1016/j.ecoenv.2017.01.022
Knight, E. et al., 2019[21]	10.1016/j.scitotenv.2019.05.339
Oliver, D. et al., 2020[22]	10.1016/j.scitotenv.2020.137263
Milinovic, J. et al., 2015[23]	10.1016/j.scitotenv.2014.12.017
Chen, H. et al., 2016[24]	10.1016/j.chemosphere.2015.10.055
Xiang, L. et al., 2018[25]	10.1021/acs.jafc.8b03492
Ahrens, L. et al., 2011[26]	10.1016/j.chemosphere.2011.06.046
Li, C. et al., 2012[27]	10.1039/c2em30394a
Higgins, C. & Luthy, R., 2006[28]	10.1021/es061000n
Wang, W. et al., 2022[29]	10.1016/j.chemosphere.2021.133224
Chen, H. et al., 2012[30]	10.1016/j.marpolbul.2012.03.012
Jhonson, R. et al., 2007[31]	10.1021/je060285g
You, C. et al., 2010[32]	10.1016/j.envpol.2010.01.009
Wei, C. et al., 2017[33]	10.1016/j.ecoenv.2017.03.040
Kwadijk, C. et al., 2013[34]	10.1016/j.chemosphere.2012.08.041
Pan, G. et al., 2009[35]	10.1016/j.envpol.2008.06.035
Brusseau, M. et al., 2019[36]	10.1021/acs.est.9b02343
Chen, H. et al., 2009[37]	10.1016/j.chemosphere.2009.09.008
Umeh, A. et al., 2021[38]	10.1021/acs.est.0c07202
Chen, X. et al., 2020[39]	10.1016/j.ecoenv.2020.111105
Qian, J. et al., 2017[40]	10.1016/j.chemosphere.2016.11.114
Wei, C. et al., 2019[41]	10.1016/j.chemosphere.2018.10.098
Lee, H. & Mabury, S., 2017[42]	10.1021/acs.est.6b04395

Table S14. Reported modeling frameworks and metrics evaluating the performance of the most recent K_d prediction models reported in the literature.

Study	Training\Validation Dataset Description	Modeling Approach and Performance	External Validation Dataset and Model Performance	Descriptors / Predictors Used
Our Study	699 total K_d values (75% training/25% validation) Matrix: Soil PFAS compounds: 19	Multiple Linear Regression-Elastic Net (MLR_{EN}) R^2 = 77.9%; RMSE = 0.5 Artificial Neural Networks (ANN) R^2 = 84.9%; RMSE = 0.4	658 total K_d values Matrix: Soil & Sediments PFAS compounds: 35 R^2 = 52.4% (MLR_{EN}); 29.4% (ANN) RMSE = 0.6 (MLR_{EN}); 1.2 (ANN)	Molecular Connectivity Indices
Fabregat-Palau et al., 2025	1,274 total K_d values (80% training/20% validation) Matrix: Soil & Sediments PFAS compounds: 51	Two-layer ML model RPD = 3.16 NRMSE = 0.07	N.A.	Molecular weight; Log Kow; Charge density
Xie et al., 2024	2,148 total K_d values (80% training / 20% validation) Matrix: Soil PFAS compounds: 26	Random Forest (RF) R^2 = 93.0% RMSE = 0.86	N.A.	Molecular weight; Log Kow; ATSm; SpMax; Solubility
Fu et al., 2025	4,731 total K_d values (90% training / 10% validation) Matrix: Soil PFAS compounds: 42	LGBM model R^2 = 88.0% RMSE = 0.28	312 K_d values Matrix: Soil PFAS compounds: 6 R^2 = 72.0% RMSE = 0.36	Predictors collected from PaDEL and RDKit

Predictors Meaning: FilterLogS (LogS): The logarithm of the compound's solubility in water; ATSm8: Molecule's size and complexity; SpDiam: Maximum distance between any two atoms in a molecule; charge density: Indicative of electrostatic interactions between PFAS properties and charged soil surfaces; LogP and Log K_{ow}: Octanol-water partitioning coefficient.
Machine Learning (ML) Models: ANN: Artificial Neural Networks, RF: Random Forest, LGBM: Light Gradient Boosting Machine; two-layer framework ML model.
Model Performance: R^2: goodness-of-fit on validation data; RPD: Ratio of Performance Deviation; RSME: Root Mean Squared Error; NRMSE: Normalized Root Mean Squared Error
N.A.: Not Available

Table S15. Descriptive Statistics of Training/Validation Dataset used in our study and recent literature.

Study	SOC (%)					CEC (cmol/Kg)					Soil pH					Log K_d				
	Mean	**SD**	**Min**	**Max**	**Median**	**Mean**	**SD**	**Min**	**Max**	**Median**	**Mean**	**SD**	**Min**	**Max**	**Median**	**Mean**	**SD**	**Min**	**Max**	**Median**
Our Study	1.13	1.36	0	7.70	0.70	16.3	11.9	0.50	80.0	16.5	6.81	0.93	2.80	8.30	7.10	0.49	0.99	-1.40	3.33	0.34
Fabregat-Palau, et al., 2025	4.70	10.5	0.03	53.6	1.30	19.0	21.1	0.10	140	13.0	6.40	1.22	2.80	9.00	6.50	0.77	0.99	-1.40	3.94	0.78
Xie et al., 2024	2.01	4.41	0	41.0	1.10	16.2	17.3	0.20	140	14.0	5.99	1.59	3.10	9.00	6.00	0.77	0.96	-1.70	3.48	0.76
Fu et al., 2025	2.51	3.33	0	20.0	1.20	18.5	16.3	0.10	120	15.0	6.14	1.36	3.50	8.50	6.20	0.9	0.85	-1.50	3.5	0.75

SD: Standard Deviation; **Min:** Minimum value in the dataset; **Max:** Maximum values in the dataset.

References

(1) Nguyen, T. M. H.; Brä, J.; Thompson, K.; Thompson, J.; Kabiri, S.; Navarro, D. A.; Kookana, R. S.; Grimison, C.; Barnes, C. M.; Higgins, C. P.; Mclaughlin, M. J.; Mueller, J. F. Influences of Chemical Properties, Soil Properties, and Solution PH on Soil−Water Partitioning Coefficients of Per-and Polyfluoroalkyl Substances (PFASs). *Cite This: Environ. Sci. Technol* **2020**, *54*, 15883–15892. https://doi.org/10.1021/acs.est.0c05705.

(2) USEPA. *CompTox Chemicals Dashboard*. https://comptox.epa.gov/dashboard/ (accessed 2025-05-02).

(3) Hatinoglu, M. D.; Perreault, F.; Apul, O. G. Modified Linear Solvation Energy Relationships for Adsorption of Perfluorocarboxylic Acids by Polystyrene Microplastics. *Science of The Total Environment* **2023**, *860*, 160524. https://doi.org/10.1016/j.scitotenv.2022.160524.

(4) Zhi, Y.; Liu, J. Sorption and Desorption of Anionic, Cationic and Zwitterionic Polyfluoroalkyl Substances by Soil Organic Matter and Pyrogenic Carbonaceous Materials. *Chemical Engineering Journal* **2018**, *346*, 682–691. https://doi.org/10.1016/j.cej.2018.04.042.

(5) Xiao, F.; Jin, B.; Golovko, S. A.; Golovko, M. Y.; Xing, B. Sorption and Desorption Mechanisms of Cationic and Zwitterionic Per- and Polyfluoroalkyl Substances in Natural Soils: Thermodynamics and Hysteresis. *Environ Sci Technol* **2019**, *53* (20), 11818–11827. https://doi.org/10.1021/acs.est.9b05379.

(6) Liu, J.; Lee, L. S. Effect of Fluorotelomer Alcohol Chain Length on Aqueous Solubility and Sorption by Soils. *Environ Sci Technol* **2007**, *41* (15), 5357–5362. https://doi.org/10.1021/es070228n.

(7) Liu, J.; Lee, L. S. Solubility and Sorption by Soils of 8:2 Fluorotelomer Alcohol in Water and Cosolvent Systems. *Environ Sci Technol* **2005**, *39* (19), 7535–7540. https://doi.org/10.1021/es051125c.

(8) Richey, D. G.; Driscoll, C. T.; Likens, G. E. Soil Retention of Trifluoroacetate. *Environ Sci Technol* **1997**, *31* (6), 1723–1727. https://doi.org/10.1021/es960649x.

(9) McLachlan, M. S.; Felizeter, S.; Klein, M.; Kotthoff, M.; De Voogt, P. Fate of a Perfluoroalkyl Acid Mixture in an Agricultural Soil Studied in Lysimeters. *Chemosphere* **2019**, *223*, 180–187. https://doi.org/10.1016/j.chemosphere.2019.02.012.

(10) 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* (9), 4164–4171. https://doi.org/10.1021/es3048043.

(11) Gredelj, A.; Nicoletto, C.; Valsecchi, S.; Ferrario, C.; Polesello, S.; Lava, R.; Zanon, F.; Barausse, A.; Palmeri, L.; Guidolin, L.; Bonato, M. Uptake and Translocation of Perfluoroalkyl Acids (PFAA) in Red Chicory (Cichorium Intybus L.) under Various Treatments with Pre-Contaminated Soil and Irrigation Water. *Science of The Total Environment* **2020**, *708*, 134766. https://doi.org/10.1016/j.scitotenv.2019.134766.

(12) Cai, W.; Navarro, D. A.; Du, J.; Ying, G.; Yang, B.; McLaughlin, M. J.; Kookana, R. S. Increasing Ionic Strength and Valency of Cations Enhance Sorption through Hydrophobic Interactions of PFAS with Soil Surfaces. *Science of The Total Environment* **2022**, *817*, 152975. https://doi.org/10.1016/j.scitotenv.2022.152975.

(13) Fabregat-Palau, J.; Vidal, M.; Rigol, A. Modelling the Sorption Behaviour of Perfluoroalkyl Carboxylates and Perfluoroalkane Sulfonates in Soils. *Science of The Total Environment* **2021**, *801*, 149343. https://doi.org/10.1016/j.scitotenv.2021.149343.

(14) Enevoldsen, R.; Juhler, R. K. Perfluorinated Compounds (PFCs) in Groundwater and Aqueous Soil Extracts: Using Inline SPE-LC-MS/MS for Screening and Sorption Characterisation of Perfluorooctane Sulphonate and Related Compounds. *Anal Bioanal Chem* **2010**, *398* (3), 1161–1172. https://doi.org/10.1007/s00216-010-4066-0.

(15) Aly, Y. H.; McInnis, D. P.; Lombardo, S. M.; Arnold, W. A.; Pennell, K. D.; Hatton, J.; Simcik, M. F. Enhanced Adsorption of Perfluoro Alkyl Substances for *in Situ* Remediation. *Environ Sci (Camb)* **2019**, *5* (11), 1867–1875. https://doi.org/10.1039/C9EW00426B.

(16) Chen, Y.-C.; Lo, S.-L.; Li, N.-H.; Lee, Y.-C.; Kuo, J. Sorption of Perfluoroalkyl Substances (PFASs) onto Wetland Soils. *Desalination Water Treat* **2013**, *51* (40–42), 7469–7475. https://doi.org/10.1080/19443994.2013.792145.

(17) Yin, C.; Pan, C.-G.; Xiao, S.-K.; Wu, Q.; Tan, H.-M.; Yu, K. Insights into the Effects of Salinity on the Sorption and Desorption of Legacy and Emerging Per-and Polyfluoroalkyl Substances (PFASs) on Marine Sediments. *Environmental Pollution* **2022**, *300*, 118957. https://doi.org/10.1016/j.envpol.2022.118957.

(18) Higgins, C. P.; Luthy, R. G. Sorption of Perfluorinated Surfactants on Sediments. *Environ Sci Technol* **2006**, *40* (23), 7251–7256. https://doi.org/10.1021/es061000n.

(19) Jeon, J.; Kannan, K.; Lim, B. J.; An, K. G.; Kim, S. D. Effects of Salinity and Organic Matter on the Partitioning of Perfluoroalkyl Acid (PFAs) to Clay Particles. *Journal of Environmental Monitoring* **2011**, *13* (6), 1803. https://doi.org/10.1039/c0em00791a.

(20) Miao, Y.; Guo, X.; Dan Peng; Fan, T.; Yang, C. Rates and Equilibria of Perfluorooctanoate (PFOA) Sorption on Soils from Different Regions of China. *Ecotoxicol Environ Saf* **2017**, *139*, 102–108. https://doi.org/10.1016/j.ecoenv.2017.01.022.

(21) Knight, E. R.; Janik, L. J.; Navarro, D. A.; Kookana, R. S.; McLaughlin, M. J. Predicting Partitioning of Radiolabelled 14C-PFOA in a Range of Soils Using Diffuse Reflectance Infrared Spectroscopy. *Science of The Total Environment* **2019**, *686*, 505–513. https://doi.org/10.1016/j.scitotenv.2019.05.339.

(22) Oliver, D. P.; Navarro, D. A.; Baldock, J.; Simpson, S. L.; Kookana, R. S. Sorption Behaviour of Per- and Polyfluoroalkyl Substances (PFASs) as Affected by the Properties of Coastal Estuarine Sediments. *Science of The Total Environment* **2020**, *720*, 137263. https://doi.org/10.1016/j.scitotenv.2020.137263.

(23) Milinovic, J.; Lacorte, S.; Vidal, M.; Rigol, A. Sorption Behaviour of Perfluoroalkyl Substances in Soils. *Science of The Total Environment* **2015**, *511*, 63–71. https://doi.org/10.1016/j.scitotenv.2014.12.017.

(24) Chen, H.; Reinhard, M.; Nguyen, V. T.; Gin, K. Y.-H. Reversible and Irreversible Sorption of Perfluorinated Compounds (PFCs) by Sediments of an Urban Reservoir. *Chemosphere* **2016**, *144*, 1747–1753. https://doi.org/10.1016/j.chemosphere.2015.10.055.

(25) Xiang, L.; Xiao, T.; Yu, P.-F.; Zhao, H.-M.; Mo, C.-H.; Li, Y.-W.; Li, H.; Cai, Q.-Y.; Zhou, D.-M.; Wong, M.-H. Mechanism and Implication of the Sorption of Perfluorooctanoic Acid by Varying Soil Size Fractions. *J Agric Food Chem* **2018**, *66* (44), 11569–11579. https://doi.org/10.1021/acs.jafc.8b03492.

(26) Ahrens, L.; Yeung, L. W. Y.; Taniyasu, S.; Lam, P. K. S.; Yamashita, N. Partitioning of Perfluorooctanoate (PFOA), Perfluorooctane Sulfonate (PFOS) and Perfluorooctane Sulfonamide (PFOSA) between Water and Sediment. *Chemosphere* **2011**, *85* (5), 731–737. https://doi.org/10.1016/j.chemosphere.2011.06.046.

(27) Li, C.; Ji, R.; Schäffer, A.; Sequaris, J.-M.; Amelung, W.; Vereecken, H.; Klumpp, E. Sorption of a Branched Nonylphenol and Perfluorooctanoic Acid on Yangtze River Sediments and Their Model Components. *Journal of Environmental Monitoring* **2012**, *14* (10), 2653. https://doi.org/10.1039/c2em30394a.

(28) Higgins, C. P.; Luthy, R. G. Sorption of Perfluorinated Surfactants on Sediments. *Environ Sci Technol* **2006**, *40* (23), 7251–7256. https://doi.org/10.1021/es061000n.

(29) Wang, W.; Rhodes, G.; Zhang, W.; Yu, X.; Teppen, B. J.; Li, H. Implication of Cation-Bridging Interaction Contribution to Sorption of Perfluoroalkyl Carboxylic Acids by Soils. *Chemosphere* **2022**, *290*, 133224. https://doi.org/10.1016/j.chemosphere.2021.133224.

(30) Chen, H.; Zhang, C.; Yu, Y.; Han, J. Sorption of Perfluorooctane Sulfonate (PFOS) on Marine Sediments. *Mar Pollut Bull* **2012**, *64* (5), 902–906. https://doi.org/10.1016/j.marpolbul.2012.03.012.

(31) Johnson, R. L.; Anschutz, A. J.; Smolen, J. M.; Simcik, M. F.; Penn, R. L. The Adsorption of Perfluorooctane Sulfonate onto Sand, Clay, and Iron Oxide Surfaces. *J Chem Eng Data* **2007**, *52* (4), 1165–1170. https://doi.org/10.1021/je060285g.

(32) You, C.; Jia, C.; Pan, G. Effect of Salinity and Sediment Characteristics on the Sorption and Desorption of Perfluorooctane Sulfonate at Sediment-Water Interface. *Environmental Pollution* **2010**, *158* (5), 1343–1347. https://doi.org/10.1016/j.envpol.2010.01.009.

(33) Wei, C.; Song, X.; Wang, Q.; Hu, Z. Sorption Kinetics, Isotherms and Mechanisms of PFOS on Soils with Different Physicochemical Properties. *Ecotoxicol Environ Saf* **2017**, *142*, 40–50. https://doi.org/10.1016/j.ecoenv.2017.03.040.

(34) Kwadijk, C. J. A. F.; Velzeboer, I.; Koelmans, A. A. Sorption of Perfluorooctane Sulfonate to Carbon Nanotubes in Aquatic Sediments. *Chemosphere* **2013**, *90* (5), 1631–1636. https://doi.org/10.1016/j.chemosphere.2012.08.041.

(35) Pan, G.; Jia, C.; Zhao, D.; You, C.; Chen, H.; Jiang, G. Effect of Cationic and Anionic Surfactants on the Sorption and Desorption of Perfluorooctane Sulfonate (PFOS) on Natural Sediments. *Environmental Pollution* **2009**, *157* (1), 325–330. https://doi.org/10.1016/j.envpol.2008.06.035.

(36) Brusseau, M. L.; Khan, N.; Wang, Y.; Yan, N.; Van Glubt, S.; Carroll, K. C. Nonideal Transport and Extended Elution Tailing of PFOS in Soil. *Environ Sci Technol* **2019**, *53* (18), 10654–10664. https://doi.org/10.1021/acs.est.9b02343.

(37) Chen, H.; Chen, S.; Quan, X.; Zhao, Y.; Zhao, H. Sorption of Perfluorooctane Sulfonate (PFOS) on Oil and Oil-Derived Black Carbon: Influence of Solution PH and [Ca2+]. *Chemosphere* **2009**, *77* (10), 1406–1411. https://doi.org/10.1016/j.chemosphere.2009.09.008.

(38) Umeh, A. C.; Naidu, R.; Shilpi, S.; Boateng, E. B.; Rahman, A.; Cousins, I. T.; Chadalavada, S.; Lamb, D.; Bowman, M. Sorption of PFOS in 114 Well-Characterized Tropical and Temperate Soils: Application of Multivariate and Artificial Neural Network Analyses. *Environ Sci Technol* **2021**, *55* (3), 1779–1789. https://doi.org/10.1021/acs.est.0c07202.

(39) Chen, X.-T.; Yu, P.-F.; Xiang, L.; Zhao, H.-M.; Li, Y.-W.; Li, H.; Zhang, X.-Y.; Cai, Q.-Y.; Mo, C.-H.; Wong, M. H. Dynamics, Thermodynamics, and Mechanism of Perfluorooctane Sulfonate (PFOS) Sorption to Various Soil Particle-Size Fractions of Paddy Soil. *Ecotoxicol Environ Saf* **2020**, *206*, 111105. https://doi.org/10.1016/j.ecoenv.2020.111105.

(40) Qian, J.; Shen, M.; Wang, P.; Wang, C.; Hou, J.; Ao, Y.; Liu, J.; Li, K. Adsorption of Perfluorooctane Sulfonate on Soils: Effects of Soil Characteristics and Phosphate Competition. *Chemosphere* **2017**, *168*, 1383–1388. https://doi.org/10.1016/j.chemosphere.2016.11.114.

(41) Wei, C.; Song, X.; Wang, Q.; Liu, Y.; Lin, N. Influence of Coexisting Cr(VI) and Sulfate Anions and Cu(II) on the Sorption of F-53B to Soils. *Chemosphere* **2019**, *216*, 507–515. https://doi.org/10.1016/j.chemosphere.2018.10.098.

(42) Lee, H.; Mabury, S. A. Sorption of Perfluoroalkyl Phosphonates and Perfluoroalkyl Phosphinates in Soils. *Environ Sci Technol* **2017**, *51* (6), 3197–3205. https://doi.org/10.1021/acs.est.6b04395.