Modeling sorption of environmental organic chemicals from water to soils

Zhizhen Zhang [a], Shenghong Wang [a], Trevor N. Brown [b], Alessandro Sangion [b], Jon A. Arnot [b,c,d], Li Li [a,*]

[a] *School of Public Health, University of Nevada, 1664, N. Virginia Street, Reno, NV 89557-274, United States*
[b] *ARC Arnot Research & Consulting, Toronto, Ontario M4M 1W4, Canada*
[c] *Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario M1C 1A4, Canada*
[d] *Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario M5S 1A8, Canada*

ABSTRACT

Reliable estimation of chemical sorption from water to solid phases is an essential prerequisite for reasonable assessments of chemical hazards and risks. However, current fate and exposure models mostly rely on algorithms that lack the capability to quantify chemical sorption resulting from interactions with multiple soil constituents, including amorphous organic matter, carbonaceous organic matter, and mineral matter. Here, we introduce a novel, generic approach that explicitly combines the gravimetric composition of various solid constituents and poly-parameter linear free energy relationships to calculate the solid-water sorption coefficient (K_d) for non-ionizable or predominantly neutral organic chemicals with diverse properties in a neutral environment. Our approach demonstrates an overall statistical uncertainty of approximately 0.9 log units associated with predictions for different types of soil. By applying this approach to estimate the sorption of 70 diverse chemicals from water to two types of soils, we uncover that different chemicals predominantly exhibit sorption onto different soil constituents. Moreover, we provide mechanistic insights into the limitation of relying solely on organic carbon normalized sorption coefficient (K_{OC}) in chemical hazard assessment, as the measured K_{OC} can vary significantly across different soil types, and therefore, a universal cut-off threshold may not be appropriate. This research highlights the importance of considering chemical properties and multiple solid constituents in sorption modeling and offers a valuable theoretical approach for improved chemical hazard and exposure assessments.

1. Introduction

The sorption of organic chemicals from water to solid phases in the environment, such as those in soils and sediments, is a critical process governing their transport and transformation, hence determining their environmental persistence, mobility, and bioavailability (Karickhoff, 1981; Kickham et al., 2012; Reichenberg and Mayer, 2006; Reid et al., 2000; Webster et al., 1998). Gaining insights into the sorption behavior of chemicals is essential for understanding and quantifying their environmental fate as well as potential ecological and human exposures.

Mineral matter (MM) and organic matter (OM) are two major components controlling the sorption of organic chemicals from water. More specifically, OM includes (i) amorphous organic matter (AOM) (e.g., humic acids and fulvic acids) and (ii) carbonaceous organic matter (COM) (e.g., black carbon, coal, natural biochar, and other pyrogenic carbonaceous matter). The sorption of chemicals to these constituents is

governed by different mechanisms. For example, AOM is fluid, rubbery, and rich in oxygen-containing functional groups, where neutral chemicals (non-ionizable organics or the neutral species of ionizable organics) can be sorbed mostly through the hydrophobic effect and hydrogen bonding (Pignatello and Xing, 1996). By contrast, COM is condensed, rigid, and rich in aromatic structures, where neutral chemicals can be sorbed additionally through π-bond interactions and pore sorption (Cornelissen et al., 2005; Kah et al., 2017; Pignatello et al., 2017). Since the solid phases, notably, MM, are often negatively charged, the sorption of cations additionally involves electrostatic interactions, including cation exchange, cation bridging, and electron donor-acceptor interactions (MacKay and Vasudevan, 2012). Other mechanisms, such as pore-filling, also contribute to sorption of chemicals to these constituents (Schwarzenbach et al., 2017).

Since the measurement of OM typically involves combustion, where both AOM and COM are converted into CO_2, OM is commonly quantified

* Corresponding author.
E-mail address: lili@unr.edu (L. Li).

using total organic carbon (TOC) (Nelson and Sommers, 1996). As such, most current chemical assessment paradigms evaluate sorption using the sorption coefficient (K_d; the ratio of sorbed and dissolved concentrations) normalized by the TOC fraction, known as K_{OC}. For example, the lowest K_{OC} at a pH range between 4 and 9 is currently used as a conservative criterion for screening chemicals for their mobility in drinking water protection (European Union, 2023; Neumann and Schliebner, 2019). However, the use of K_{OC} implies two simplifying assumptions: (1) sorption is controlled predominantly by OM, with less consideration given to MM, and (2) both AOM and COM exhibit similar sorption affinities for chemicals. While these simplifying assumptions have been widely accepted for decades, they have recently been challenged. First, although it is reasonable to assume sorption by OM to be the primary sorbent for neutral chemicals (Schwarzenbach et al., 2017), MM plays a crucial role in the sorption of ionizable chemicals (Droge and Goss, 2013b, 2013a), which constitute approximately half of the chemicals undergoing evaluations (Franco et al., 2010; Zhang et al., 2024). Second, the sorption affinities of chemicals can vary between AOM and COM (Cornelissen et al., 2005), and therefore, varying soil composition can impact the sorption of different chemicals (Spark and Swift, 2002). For these reasons, efforts have been made to separately consider sorption to different soil constituents (Droge and Goss, 2013a; Han et al., 2016; Plata et al., 2015). Unfortunately, these approaches have gained limited application in holistic environmental fate and exposure modeling. Most existing models, such as the well-known examples RAIDAR (Arnot et al., 2006; Arnot and Mackay, 2008), USEtox (Fantke et al., 2017), PROTEX (Li et al., 2018a, 2018b), and EUSES (European Commission, 2004), still rely on the traditional K_{OC} approach for sorption quantification.

To better characterize the sorption of non-ionizable or predominantly neutral organic chemicals (i.e., >99% of the mass present in water being neutral species at pH=7) from water, here we propose a novel, general composition-based modeling approach that integrates (i) poly-parameter linear free energy relationships (pp-LFERs) for sorption of non-ionizable organic chemicals to AOM, COM, and MM and (ii) the relative composition of the three solid constituents in soil. This modeling approach facilitates cost-effective screening-level predictions of chemical-specific and soil-specific K_d for the sorption of *non-ionizable or predominantly neutral organic chemicals* from *neutral* (pH close to 7) water at the *standard ambient temperature* (15 to 25 °C). This modeling approach helps us understand the relative contributions of various soil constituents to chemical sorption; it also allows comprehensive predictions of overall sorption based on multiple soil compositions, thereby facilitating tailored and targeted assessments for chemical screening and prioritization. Since inputs to pp-LFERs can be either experimentally determined or theoretically predicted from the chemical structure, this approach applies to a broad spectrum of non-ionizable or predominantly neutral organic chemicals, including those in the design or pre-manufacturing stage. This versatility renders the approach well-suited for precautionary chemical assessments.

2. The composition-based model

2.1. Composition-based model for estimating sorption coefficient (K_d)

In this approach, we discriminate between three solid constituents, namely AOM (with an equivalent carbon content of AOC, f_{AOC} in% weight), COM (with an equivalent carbon content of COC, f_{COC} in% weight, where AOC + COC = TOC), and MM (with a mineral content, f_{MM} in% weight). Table 1 illustrates the example gravimetric compositions of these three constituents across various "reference soil groups", i. e., soil types with distinctive physical, chemical, and biological attributes.

The sorption of chemicals on each constituent from water can be described using a sorption coefficient ($K_{AOM\text{-}water}$ in L_{water}/kg_{AOM}, $K_{COM\text{-}water}$ in L_{water}/kg_{COM}, and $K_{MM\text{-}water}$ in L_{water}/kg_{MM}), defined as the ratio

Table 1

Example gravimetric compositions of amorphous organic carbon (AOC), carbonaceous organic carbon (COC), and mineral matter (MM), as well as cation exchange capacity of MM (CEC$_{MM}$) in various reference soil groups.

Reference soil group	Representative sites and corresponding soil mapping units (SMUs)	AOC (%)[a]	COC (%)[a]	MM (%)	CEC$_{MM}$ (cmol$_C$/kg)[b]
Luvisol	Western U.S. arid areas (SMU 5091)	0.63	0.01	16	64
Fluvisol	Cairo, Egypt (SMU 16,414)	0.68	0.01	20	71
Retisol	Belgium, Brussel (SMU 9383)	1.28	0.06	20	29
Ferralsol	Amazon rainforest, Brazil (SMU 12,767)	1.34	0.06	35	8
Podzol	Baltic region (SMU 10,078)	6.37	0.85	6	74
	Great Toronto Area, Canada (SMU 4952)				
Gleysol	Great Lakes, U.S. and Canada (SMU 4701)	3.74	0.44	25	20
Histosol	Northeastern Canada (SMU 4933)	40.3	6.18	13	57
Urban soil	England, UK	2.52[c]	1.68[c]	9[d]	16

Notes:.
[a] For soil types other than the urban soil, the COC content is extrapolated from TOC content reported by the United Nations Food and Agriculture Organization (UN FAO) Harmonized World Soils Database version 2.0, based on an empirical relationship between TOC and COC contents as reported by (Glaser and Amelung, 2003). The AOC content is calculated to be the difference between the reported TOC and calculated COC content.
[b] Represented by the CEC of clay (CEC$_{clay}$), taken from the United Nations Food and Agriculture Organization's Harmonized World Soils Database version 2.0, which calculated CEC$_{clay}$ by correctting for the contribution of AOC's CEC (CEC$_{AOC}$; a typical mean of 350 cmol$_C$/kg$_{AOC}$ was used) from the measured soil's CEC (CEC$_{soil}$).
[c] Taken from the medians of Edmondson et al.(Edmondson et al., 2015).
[d] Taken from the UN FAO Harmonized World Soils Database version 2.0 based on soil type in London (SMU 7001).

of the chemical concentration in each constituent to its concentration in water (Eqs. 1 through 3). According to classic sorption theories, these partition coefficients are expressed as products of sorption coefficients (K_{AOC} in L_{water}/kg_{AOC}, K_{COC} in L_{water}/kg_{COC}, and K_{MM} in L_{water}/kg_{MM}) and corresponding constituent content (f_{AOC}, f_{COC}, and f_{MM}) (Eqs. 1 through 3).

$$K_{AOM-water} = C_{AOM} : C_{water} = K_{AOC} \times f_{AOC} \tag{1}$$

$$K_{COM-water} = C_{COM} : C_{water} = K_{COC} \times f_{COC} \tag{2}$$

$$K_{MM-water} = C_{MM} : C_{water} = K_{MM} \times f_{MM} \tag{3}$$

Therefore, the total sorption coefficient is:

$$K_d = K_{AOM-water} + K_{COM-water} + K_{MM-water}, \text{namely}$$
$$K_d = K_{AOC} \times f_{AOC} + K_{COC} \times f_{COC} + K_{MM} \times f_{MM} \tag{4}$$

Since we seek to develop a *screening-level* approach for sorption prediction, we base the above equations on the following simplifying assumptions:

(1) There is a linear, concentration-independent distribution of chemicals between each soil constituent and water. This assumption has been shown to hold for low-concentration chemicals found in the environment (Blum et al., 1994; Schwarzenbach et al., 2017); it is also a recommended assumption by the Organisation for Economic Co-operation and Development (OECD) Guidelines for the Testing of Chemicals #106 (OECD, 2000).

(2) A chemical maintains equilibrium partitioning between individual soil constituents and water. This assumption is widely

accepted in multimedia environmental modeling that involves multiple phases within an environmental compartment (Diamond et al., 1992; Mackay et al., 1992).

(3) We focus on the central-tendency estimates of sorption coefficients for the same chemical across different types of AOM (e. g., humic acids and fulvic acids), COM (e.g., black carbon generated at different combustion temperatures), and MM (e.g.,

et al., 2017) for many well-studied organic chemicals.

We apply Eqs. (6) (Bronner and Goss, 2011), (7) (Lu et al., 2016), and (8) to predict the sorption of non-ionizable or predominantly neutral organic chemicals onto AOC, COC, and MM in *neutral water-solid systems*, which represents either natural soil and sediment or the starting test conditions (pH, ionic strength, ionic type, etc.) recommended by the OECD Guidelines for the Testing of Chemicals #106 (OECD, 2000):

$$\log_{10} K_{\text{AOC}} = (0.81 \pm 0.08) \times E + (-0.61 \pm 0.11) \times S + (-0.21 \pm 0.14) \times A + (-3.44 \pm 0.18) \times B + (2.99 \pm 0.11) \times V + (-0.29 \pm 0.12) \tag{6}$$

$$\log_{10} K_{\text{COC}} = [(-0.35 \pm 0.02)\log\gamma] \times E + (-0.62 \pm 0.10) \times A + (-3.35 \pm 0.11) \times B + (3.74 \pm 0.11) \times V + (-1.45 \pm 0.09) \tag{7}$$

clays saturated with different cations), and we do not consider variability in sorption affinity caused by variations in sorbent structure or composition. This is because such within-chemical variability is usually less significant than variability in sorption coefficients among diverse chemicals (Grathwohl, 1990; Sigmund et al., 2020). Also, we ignore the impact of soil composition heterogeneity, i.e., the relative abundance of and resulting competitive sorption by the three soil constituents, on the non-linearity of sorption (Chiou and Kile, 1998).

(4) AOM, COM, and MM do not share sorption sites, and sorption to these constituents does not compete. This assumption does not consider the formation of organomineral complexes due to interactions between AOM and MM (Celis et al., 2006) or the changed physical conformation of AOM due to the co-occurrence of MM (Bonin and Simpson, 2007).

2.2. Predictions of sorption coefficients with pp-LFER

Based on the free energy theory, a pp-LFER calculates the distribution of a chemical between two phases as a linear combination of (i) Abraham solute descriptors (capital letters in Eq. (5)) of the neutral form of the chemical and (ii) system constants (lowercase letters in Eq. (5)) describing the water-sorbent system:

$$\log_{10} K = c + e \times E + s \times S + a \times A + b \times B + v \times V \tag{5}$$

where,

E: Excess molar refraction with a unit of $(\text{cm}^3 \text{ mol}^{-1})/10$, which correlates with a part of the dispersion interaction (induced dipole-induced dipole interaction) and Debye forces (dipole-induced dipole interaction) that constitute van der Waals interactions.
S: Polarizability/dipolarity parameter (Arey et al., 2005).
A: Solute H-bond acidity, which correlates with interactions (e.g., hydrogen bonding or π-bond interaction) between H-donating chemical and H-accepting sorbent (i.e., AOM or COM in this case) (Pignatello et al., 2017).
B: Solute H-bond basicity, which correlates with interactions (e.g., hydrogen bonding or π-bond interaction) between H-accepting chemical and H-donating solvent (i.e., water in this case) (Pignatello et al., 2017).
V: McGowan molar volume with a unit of $(\text{cm}^3 \text{ mol}^{-1})/100$, which correlates with a part of the dispersion interaction and the cavity formation during a chemical's transfer from one condensed phase to another.

The chemical-specific Abraham solute descriptors can be either experimentally determined or theoretically predicted. They have also been compiled and documented in databases such as UFZ-LSER (Ulrich
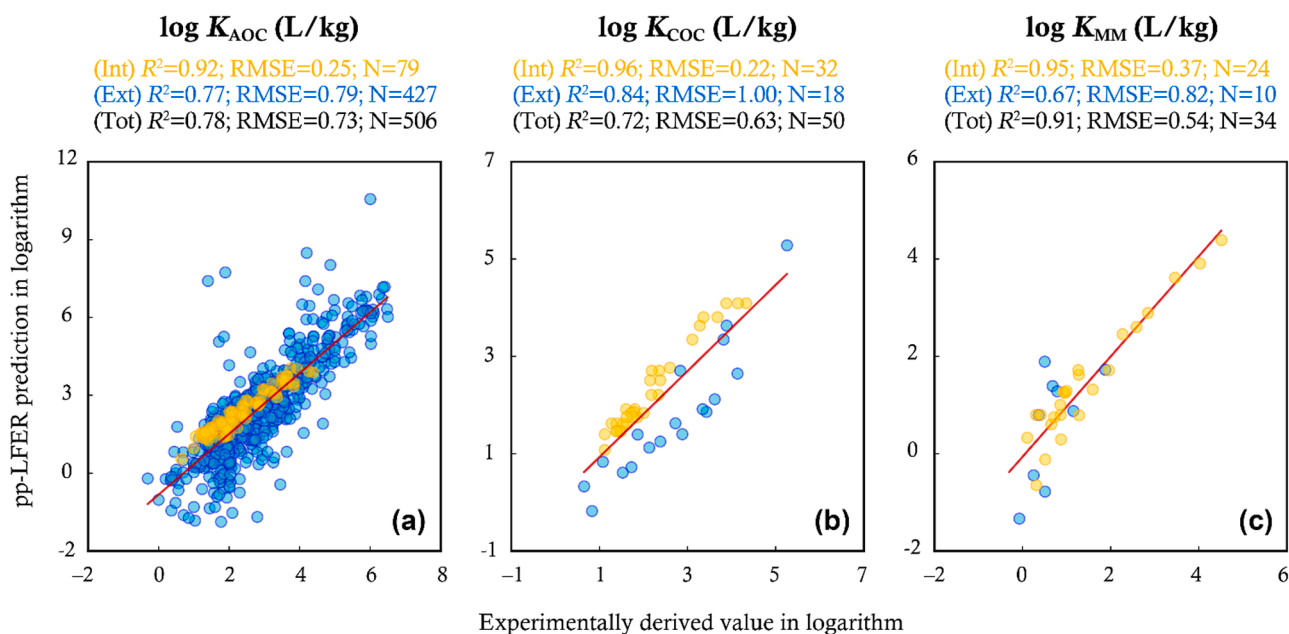
where, γ is chemical activity. While readers are encouraged to use environment-specific γ, we also recommend a conservative generic value of 0.001 be assigned for *screening-level* studies where environment-specific values are unavailable. Earlier studies show that most chemical contaminants exhibit a chemical activity in the range from 0.0001 to 0.01 in environmental samples (Gobas et al., 2018; Mayer et al., 2009), and this range is below a threshold of 0.01 that is likely to lead to narcosis or baseline toxicity (Gobas et al., 2018; Mackay and Arnot, 2011). We select the generic value of 0.001 because it gives a better agreement with the smallest root mean square deviation (RMSE) between predicted and measured K_{COC} of 50 chemicals, compared to chemical activities of 0.0001 and 0.01 (for details, see **Supplementary Materials Table S1**).

$$\log_{10} K_{\text{MM}} = 0.32 \times E - 2.55 \times S - 0.83 \times A - 0.65 \times B + 3.43 \times V - 0.68 \tag{8}$$

Supplementary Materials Text S1 details the derivation of Eq. (8) and underlying considerations.

3. Results and discussion

3.1. Evaluation of the proposed approach's performance

In earlier studies (Abraham et al., 1994; Bronner and Goss, 2011; Goss et al., 2003; Lu et al., 2016), the K_{AOC}, K_{COC}, and K_{MM} pp-LFERs (Eqs. 6 through 8) have undergone thorough *internal validation* to assess the goodness of fit between their predictions and the experimentally derived values of training chemicals (**Supplementary Materials Table S2**). Here, we additionally perform an *external validation* to assess their predictive ability for experimental measurements that are new to these pp-LFERs. Specifically, we first collect literature-reported experimentally derived K_{AOC}, K_{COC}, and K_{MM} for non-ionizable or predominantly neutral organic chemicals that were not used to develop these pp-LFERs (**Supplementary Materials Text S2**). We then use the above pp-LFERs to make predictions for these chemicals, using Abraham solute descriptors (with a priority given to experimentally determined values) and parameters obtained from the UFZ-LSER database (Ulrich et al., 2017). Subsequently, we assess the statistical agreement between the collected experimentally derived values and the predictions.

Fig. 1 shows the performance of the adopted pp-LFERs in internal and external validations. The internal validation results are obtained from corresponding literature, which reported coefficients of determination (R^2) of 0.92, 0.96, and 0.95 for K_{AOC}, K_{COC}, and K_{MM}, respectively (Abraham et al., 1994; Bronner and Goss, 2011; Goss et al., 2003; Lu et al., 2016) (Fig. 1). These values indicate that the developed pp-LFERs can explain more than 90% of the variance observed in the training set. The RMSEs were reported to be 0.25, 0.22, and 0.37 log units for K_{AOC},

log K_{AOC} (L/kg)

(Int) R^2=0.92; RMSE=0.25; N=79
(Ext) R^2=0.77; RMSE=0.79; N=427
(Tot) R^2=0.78; RMSE=0.73; N=506

log K_{COC} (L/kg)

(Int) R^2=0.96; RMSE=0.22; N=32
(Ext) R^2=0.84; RMSE=1.00; N=18
(Tot) R^2=0.72; RMSE=0.63; N=50

log K_{MM} (L/kg)

(Int) R^2=0.95; RMSE=0.37; N=24
(Ext) R^2=0.67; RMSE=0.82; N=10
(Tot) R^2=0.91; RMSE=0.54; N=34

Fig. 1. Comparison between pp-LFER predictions of K_{AOC}, K_{COC}, and K_{MM} and literature-reported experimentally derived values. Orange dots represent experimental measurements used for training the pp-LFERs as reported in the original papers (internal validation), whereas blue dots represent experimental measurements new to the pp-LFERs (external validation). Shown in the figures are coefficients of determination (R^2), root-mean-square deviations (RMSE), and numbers of chemicals (N) for internal validation ("Int"), external validation ("Ext"), and combined datasets ("Tot"). For details on the derivation of these datasets, see Supplementary Materials Text S3. Note that K_{AOC}, K_{COC}, and K_{MM} have been measured for different chemicals in the literature: 23 chemicals have measurements of both K_{AOC} and K_{COC} reported, and none of these chemicals have measurements of all of K_{AOC}, K_{COC}, and K_{MM} reported (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

K_{COC}, and K_{MM}, respectively (Abraham et al., 1994; Bronner and Goss, 2011; Goss et al., 2003; Lu et al., 2016) (Fig. 1). These values indicate that predictions deviate from experimentally derived values by generally less than a factor of 3. In the external validation, R^2 values are 0.77, 0.84, and 0.67 for K_{AOC}, K_{COC}, and K_{MM}, respectively, whereas the RMSEs were 0.79, 1.0, and 0.82 log units, respectively (Fig. 1). These results suggest that the pp-LFERs may be more uncertain for predicting experimental measurements that were new to the models; however, the overall performance is still reasonable and sufficient for screening-level assessments. In addition, using K_{COC} of chemicals in the validation sets as an example, **Supplementary Materials Fig. S1** shows that the regression residuals (i.e., the difference between measured and predicted values) across a diverse array of chemicals do not correlate with Abraham solute descriptors, indicating the absence of systematic bias in the prediction.

Combining the internal and external validation datasets, we calculate that RMSEs are 0.73, 0.63, and 0.54 log units, respectively, for K_{AOC}, K_{COC}, and K_{MM} predictions (Fig. 1). These numbers indicate that the pp-LFER predictions generally agree with experimentally derived values, with a deviation by a factor of less than 6. Since K_d is a linear combination of $K_{AOM-water}$, $K_{COM-water}$, and $K_{MM-water}$ (Eq. (4)), Eq. (9) allows calculating the overall uncertainty (RMSE$_{overall}$) by propagating these uncertainties:

$$RMSE_{overall} = \sqrt{f_{AOC}^2 RMSE_{AOC}^2 + f_{COC}^2 RMSE_{COC}^2 + f_{MM}^2 RMSE_{MM}^2} \qquad (9)$$

For illustration, we calculate RMSE$_{overall}$ to be approximately 0.9 log units for both Podzol (characterized by high AOC and COC content and low MM content; Table 1) and Ferralsol (characterized by high MM content and low AOC and COC content; Table 1). Therefore, from a statistical standpoint, a K_d prediction may likely suffer an uncertainty of less than a factor of 8 if no experimentally derived value is available for groundtruthing the prediction.

3.2. Relative contributions of AOM, COM, and MM in chemical sorption from water

Using the K_{AOC}, K_{COC}, and K_{MM} pp-LFERs paired with chemical-specific inputs, we assess the roles of AOM, COM, and MM in the sorption of 70 non-ionizable organic chemicals drawn from the substances registered in the REACH registration database (Fig. 2). These chemicals include commonly studied environmental contaminants, e.g., polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs), as well as chemicals that have recently been identified to be "mobile" or "very mobile" substances, based on a conservative criterion of log K_{OC} lower than 4 (Arp et al., 2017; Arp and Hale, 2022; Neumann and Schliebner, 2019). For detailed information on these chemicals, please refer to **Supplementary Materials Table S3**. The comparison between Fig. 2a and b illustrates the impact of soil composition on the relative importance of chemical sorption onto AOM, COM, and MM. Overall, AOM and COM are the main sorbents for most investigated chemicals to Podzol, which aligns with Podzol's higher TOC content, whereas MM is the main sorbent for most investigated chemicals to Ferralsol.

The sorption of certain chemicals is dominated by a single sorbent, regardless of soil composition. For example, our approach predicts that AOM is the dominant sorbent in both types of soil for the sorption of chlorobenzenes and chlorotoluenes (the "HArHCs" category in Fig. 2a and b). This agrees with experimental observations where the K_d of chlorobenzenes was strongly linearly correlated with soil AOC content; notably, the regression slope, which represents AOM's contribution, was 200 to 1500 times greater than the regression intercept, which represents combined contributions of sorbents other than AOM (Paya-Perez et al., 1991). Likewise, PAHs and some PCBs are predicted to be sorbed mainly onto COM in both soil types, due to their planar structures and the interaction between delocalized π electron clouds and π electron-rich surface of COM. This aligns with earlier observations that COM sorbs 91% to 99% of the PAH mass in soil (Cornelissen et al., 2006). By contrast, our approach predicts that the sorption of non-ionizable per-fluorinated and organosilicon chemicals (the "OF & OSi" category in
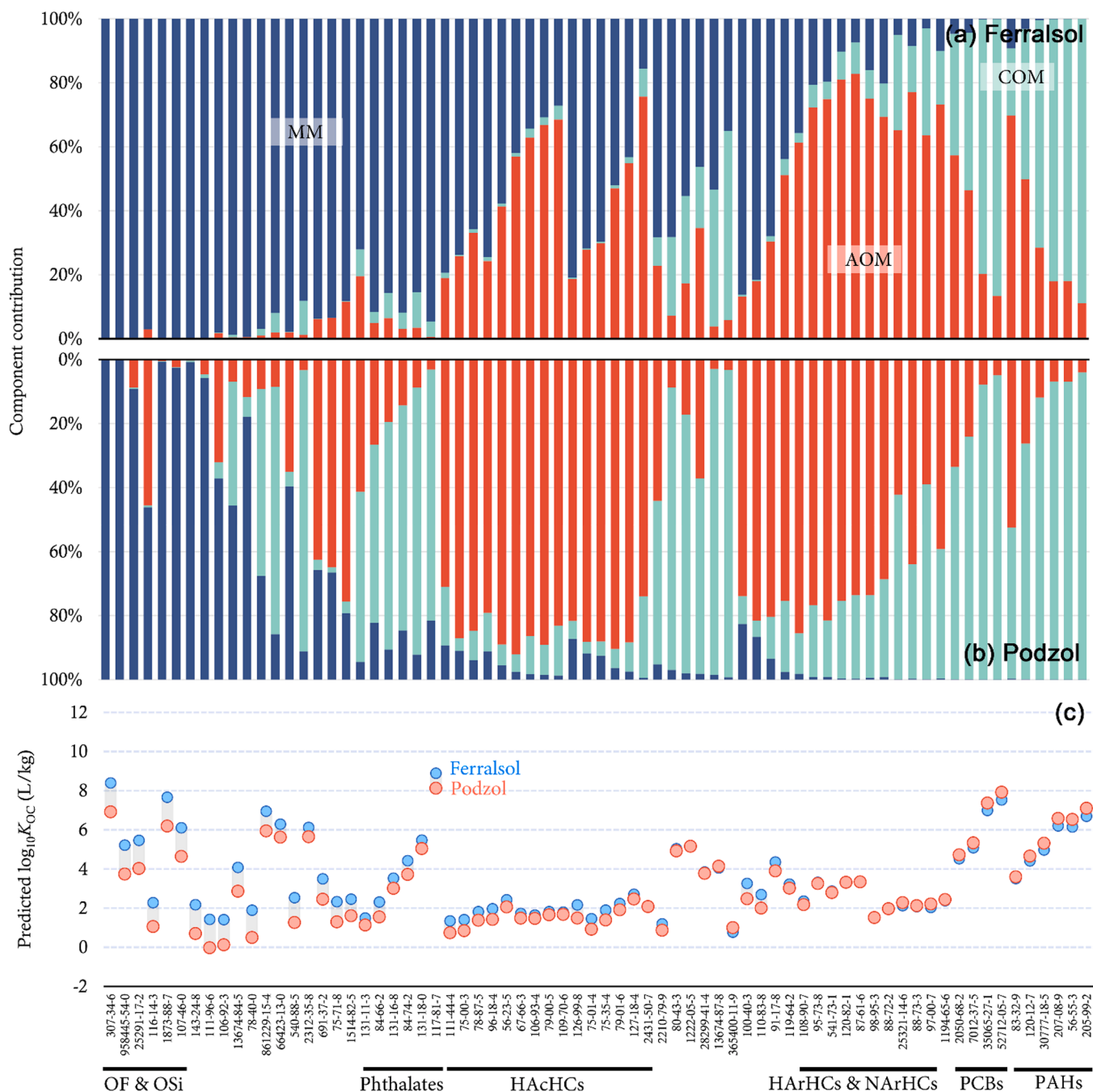
Fig. 2. The relative importance of mineral matter (MM), amorphous organic matter (AOM), and carbonaceous organic matter (COM) in the sorption of 70 organic chemicals (identified by their CAS Registry Numbers) to two types of soil, Ferralsol (Panel a) and Podzol (Panel b), and the calculated K_{OC} of the 70 organic chemicals in the two types of soil (Panel c). Highlighted are six categories of chemicals: Organofluorine (OF) and organosilicon (OSi) compounds, phthalates, halogenated acyclic hydrocarbons (HAcHCs), halogenated and/or nitro aromatic hydrocarbons (HArHCs), polychlorinated biphenyls (PCBs), and polycyclic aromatic hydrocarbons (PAHs).

Fig. 2a and b) is significantly influenced by MM in both types of soil. Although studies directly comparing the sorption of *non-ionizable* perfluorinated and organosilicon chemicals onto MM and other sorbents are currently lacking, a previous comprehensive review suggests that for *ionizable* long-chain PFAS, K_d correlates more strongly with soil clay content than with soil AOC content (Li et al., 2018). The reduced affinity of perfluorinated and organosilicon chemicals toward AOM likely stems from their much weaker van der Waals interactions with the "bulk" condensed phases (e.g., AOM or water phase) and lower polarizability/dipolarity, relative to compounds of similar molecular size (Goss and Bronner, 2006). This is evidenced by their considerably lower

(negative or near-zero) values of descriptors *E* and *S*, respectively, compared to other chemicals investigated here (**Supplementary Materials Table S3**). Such weaker interactions diminish their propensity to remain in the "bulk" condensed phases, leading to a preference for sorption on the MM surface (Goss and Bronner, 2006).

The dominant sorbent for chemical sorption may vary depending on the type of soil. For instance, our approach predicts that phthalates are mainly sorbed to MM in Ferralsol but to COM in Podzol (Fig. 2a and b). Conversely, MM and AOM are the primary sorbents for the sorption of halogenated acyclic hydrocarbons (e.g., chloroethane and chloropropane) to Ferralsol and Podzol, respectively (the "HAcHCs" category

in Fig. 2a and b). These two cases highlight the importance of investigating chemical sorption onto various soil constituents and considering variabilities in chemical properties and soil characteristics, if one seeks a thorough understanding of chemical sorption to the soil.

3.3. Variability in measured K_{OC} among soils with different compositions

In many current practices, sorption is quantified using K_{OC}, defined as K_d normalized by the TOC fraction (f_{TOC})(Eq. (10), where the symbol "≡" means "defined as"). However, from a mechanistic perspective, the measured K_d reflects simultaneous sorption to AOM, COM, and MM, and the measurement of f_{TOC} does not distinguish between AOC and COC (OECD, 2000). Therefore, a mechanistic analysis of the experimentally derived K_{OC} by this study gives Eq. (10):

$$K_{OC} \equiv \frac{K_d}{f_{TOC}} = \frac{K_{AOC} \times f_{AOC} + K_{COC} \times f_{COC} + K_{MM} \times f_{MM}}{f_{AOC} + f_{COC}} \quad (10)$$

Eq. (10) explains that K_{OC} is inherently variable across measurements in different environments, which results from variations in the environment-specific soil composition. We here quantify such an inter-environment variability using the difference between K_{OC} values calculated for Ferralsol and Podzol (Fig. 2c). As shown in Fig. 2c, among the 70 investigated chemicals, although the Ferralsol and Podzol K_{OC} values generally agree with each other, the Ferralsol K_{OC} may exceed Podzol K_{OC} by an order of magnitude for 16 chemicals. While our approach introduces an overall uncertainty of 0.9 orders of magnitude in K_d predictions, this degree of uncertainty is still less than the discrepancy in K_{OC} between the two types of soil. Therefore, it remains worthwhile to consider multiple geosorbents in hazard or risk assessments. In general, these 16 chemicals are primarily sorbed by MM, with sorption to MM accounting for over 90% of the sorption in Ferralsol and

et al., 2010; Zhang et al., 2024), there is a need to extend pp-LFERs to cover cations and anions. This extension involves the use of (i) ionic-species Abraham solute descriptors and (ii) system constants that are recalibrated specifically for cations and anions. It is important to note that, for the same organic chemical, ionic-species and neutral-species Abraham solute descriptors can differ substantially. For four categories of anionic (carboxylates and phenoxides) and cationic chemicals (amines and pyridines), Abraham and collaborators (Abraham, 2011; Abraham and Acree Jr, 2010; Abraham and Acree, 2010a, 2010b) have developed equations that allow for the expression of ionic-species Abraham solute descriptors as a linear combination of neutral-species Abraham solute descriptors and other parameters (**Supplementary Materials Text S3**). However, for chemicals falling outside of these four categories, conversion equations for neutral- to ionic-species descriptors are not currently available. Furthermore, as these conversion equations rely on neutral-species Abraham solute descriptors, they cannot be applied to permanently charged chemicals like quaternary ammonium compounds, for which neutral-species do not exist.

To develop K_{AOC} pp-LFERs for cations and anions, we collect experimentally determined K_{AOC} at a pH of approximately 7 for 55 cationic and 28 anionic organic chemicals (**Supplementary Materials Table S4**). We select cationic chemicals with a pKa (i.e., the dissociation constant for the conjugate base) greater than 7.5 and anionic chemicals with a pKa lower than 6.5, ensuring at least 70% of the mass present in water being ionic species at pH=7. For these chemicals, we convert the neutral-species Abraham solute descriptors (E, S, A, B, and V) to corresponding ionic-species Abraham solute descriptors (i.e., E_i, S_i, A_i, B_i, V_i, and J^+, or J^-)(**Supplementary Materials Table S4**). The K_{AOC} and ionic-species Abraham solute descriptors are then correlated using multivariant linear regression (Eqs. (11) and (12) for anions and cations, respectively):

$$\log K_{AOC,anion} = (0.07 \pm 0.11) \times E_i - (0.24 \pm 0.07) \times S_i + (3.40 \pm 1.78) \times A_i - (1.02 \pm 0.27) \times B_i + (1.33 \pm 0.28) \\ \times V_i + (2.07 \pm 0.34) \times J^- - (0.80 \pm 0.90) \quad (11)$$

$$\log K_{AOC,cation} = (0.60 \pm 0.80) \times E_i - (0.84 \pm 0.37) \times S_i + (0.82 \pm 0.67) \times A_i + (0.58 \pm 0.19) \times V_i + (0.21 \pm 0.92) \\ \times J^+ + (1.87 \pm 0.65) \quad (12)$$

over 40% in Podzol. It is particularly interesting to note that this subset of chemicals includes non-ionizable perfluorinated and organosilicon chemicals (**Supplementary Materials Table S3**). By contrast, chemicals for which AOM or COM serves as the dominant sorbents exhibit Ferralsol and Podzol K_{OC} values that differ by less than a factor of 5. This contrast indicates that sorption onto MM could significantly influence the measurement of K_d.

Such inherent variability in K_{OC} may impact the use of K_{OC} in chemical hazard assessment. For example, Fig. 2c shows that 3 chemicals [methyl perfluoro-3-(3-methoxypropoxy)−3H-propanoate (CASRN 958445−54−0), tris(2-chloroisopropyl)phosphate (CASRN 13674−84−5), and decalin (CASRN 91−17−8)] may have a logK_{OC} lower than 4 in Podzol but greater than 4 in Ferralsol. These discrepancies may challenge the use of K_{OC} alone in chemical mobility assessments because they constitute potential "false negatives" or "false positives" if only one type of soil is used in the experimental determination of K_{OC}. This highlights that K_{OC} is soil-specific, instead of intrinsic to a chemical.

4. Expansion of the approach to ionizable organic chemicals

In the above case, we have applied Eqs. 6 through 8 to non-ionizable or predominantly neutral organic chemicals. Given that the majority of the organic chemicals undergoing evaluations are ionizable (Franco

These pp-LFERs show satisfactory predictive ability for K_{AOC} of both anions and cations (respective R^2 of 0.93 and 0.49, and respective RMSE of 0.33 and 0.47 log units).

To our knowledge, there is currently a lack of pp-LFER calculating K_{COC} for ionizable organic chemicals, nor a comprehensive dataset of experimentally determined values that could support developing such a model. We collect experimentally determined K_{COC} at a pH of approximately 7 for 11 anionic and cationic organic chemicals (**Supplementary Materials Table S5**). We then evaluate whether Eq. (7), supplied with the neutral-species Abraham descriptors, can reasonably predict K_{COC} for these chemicals. Interestingly, our results indicate that Eq. (7) seems to perform comparably for both ionizable (R^2= 0.71; RMSE= 0.91 log units; N = 11) and non-ionizable (R^2= 0.72; RMSE= 0.63 log units; N = 50, combined internal and external evaluation datasets; Fig. 1b) organic chemicals. However, it is important to acknowledge that this comparison is made using a limited dataset. Further research is warranted to assess the applicability of this approach more thoroughly for ionizable organic chemicals.

Sorption of ionizable organic chemicals onto MM is believed to be controlled mainly by the ion exchange mechanism. For example, Droge and Goss (Droge and Goss, 2013a, 2013b) have demonstrated that $K_{MM-water}$ of cations is strongly correlated with the cation exchange

capacity (CEC$_{MM}$ in cmol$_C$/kg) (Eq. (13)):

$$K_{MM-water} = K_{CEC} \times CEC_{MM} \qquad (13)$$

In their research (Droge and Goss, 2013a, 2013b), Droge and Goss have also developed a pp-LFER for the prediction of K_{CEC} for sorption of cationic amines sharing a C$_x$H$_y$N$^+$ structure onto phyllosilicate minerals. However, there remains a significant need for additional research to understand and quantify the $K_{MM-water}$ of both cations and anions onto various types of MM, such as clay minerals and pedogenic metal oxides, so as to facilitate developing reliable pp-LFERs in the future.

4. Conclusions and perspectives

We present a generic composition-based approach to calculate K_d for sorption of non-ionizable organic chemicals from water. Statistically, this approach exhibits an overall uncertainty of 0.9 log units for predictions. With this approach, we demonstrate that different soil constituents may exert a dominant influence on the sorption of different chemicals, resulting in variability in K_{OC} across different types of soil. Therefore, in chemical hazard assessment, it may not be appropriate to view K_{OC} as a chemical's intrinsic property that is independent of specific environments.

While our proposed approach shows promising performance, the models developed herein are preliminary because the datasets are quite limited and for K_{MM} the data had to be derived from other experimental parameters. The applicability domain of the models is presently hard to define for applications to thousands of chemicals and they should therefore be used cautiously. Nonetheless, the framework provides an Integrated Approach for Testing and Assessment (OECD, 2020) that can be used to systematically improve the mechanistic understanding of these processes and efficiently guide experimental resources for subsequent revision. Further theoretical and experimental investigations are still necessary to enhance our understanding and characterization of the sorption processes between diverse chemicals and various soil components. For example, future experimental studies can provide valuable insights to examine whether the K_{MM} pp-LFER derived based on a thermodynamic circle is mechanistically sound (for detailed discussions, see **Supplementary Materials Text S3**). In addition, our proposed approach relies on a distinction between AOM and COM, which currently lacks clear theoretical definition and faces analytical challenges in wet labs. The current methods for determining COM, such as thermogravimetry, density separation, and the detection of marker substances like benzenecarboxylic acids, are complex and imprecise. Future research should focus on developing analytical methods for COM determination and generating datasets of reliable, reproducible, and consistent K_{COC} measurements to facilitate refining the model.

CRediT authorship contribution statement

Zhizhen Zhang: Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Shenghong Wang:** Investigation, Data curation, Writing – original draft. **Trevor N. Brown:** Methodology, Writing – review & editing. **Alessandro Sangion:** Methodology, Writing – review & editing. **Jon A. Arnot:** Methodology, Writing – review & editing, Funding acquisition. **Li Li:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.wroa.2024.100219.

References

Abraham, M.H., 2011. The permeation of neutral molecules, ions, and ionic species through membranes: brain permeation as an example. J. Pharm. Sci. 100 (5), 1690–1701.

Abraham, M.H., Acree Jr, W.E., 2010. The transfer of neutral molecules, ions and ionic species from water to ethylene glycol and to propylene carbonate; descriptors for pyridinium cations. New J. Chem. 34 (10), 2298–2305.

Abraham, M.H., Acree, W.E.J., 2010a. Equations for the transfer of neutral molecules and ionic species from water to organic phases. J. Org. Chem. 75 (4), 1006–1015.

Abraham, M.H., Acree, W.E.J., 2010b. Solute descriptors for phenoxide anions and their use to establish correlations of rates of reaction of anions with iodomethane. J. Org. Chem. 75 (9), 3021–3026.

Abraham, M.H., Andonian-Haftvan, J., Whiting, G.S., Leo, A., Taft, R.S., 1994. Hydrogen bonding. Part 34. The factors that influence the solubility of gases and vapours in water at 298K, and a new method for its determination. J. Chem. Soc., Perkin Trans. 2 (8), 1777–1791.

Arey, J.S., Green, W.H., Gschwend, P.M., 2005. The electrostatic origin of Abraham's solute polarity parameter. J. Phys. Chem. B 109 (15), 7564–7573.

Arnot, J.A., Mackay, D., 2008. Policies for chemical hazard and risk priority setting: Can toxicity, and quantity information be combined? Environ. Sci. Technol. 42 (13), 4648–4654.

Arnot, J.A., Mackay, D., Webster, E., Southwood, J.M., 2006. Screening level risk assessment model for chemical fate and effects in the environment. Environ. Sci. Technol. 40 (7), 2316–2323.

Arp, H.P.H., Brown, T.N., Berger, U., Hale, S.E., 2017. Ranking REACH registered neutral, ionizable and ionic organic chemicals based on their aquatic persistency and mobility. Environ. Sci. Process. Impacts. 19, 939–955.

Arp, H.P.H., Hale, S.E., 2022. Assessing the persistence and mobility of organic substances to protect freshwater resources. ACS. Environ. Au 6 (2), 482–509.

Blum, D.J.W., Suffet, I.H., Duguet, J.P., 1994. Quantitative structure-activity relationship using molecular connectivity for the activated carbon adsorption of organic chemicals in water. Water Res. 28 (3), 687–699.

Bonin, J.L., Simpson, M.J., 2007. Variation in phenanthrene sorption coefficients with soil organic matter fractionation: the result of structure or conformation? Environ. Sci. Technol. 41 (1), 153–159.

Bronner, G., Goss, K.U., 2011. Predicting sorption of pesticides and other multifunctional organic chemicals to soil organic carbon. Environ. Sci. Technol. 45 (4), 1313–1319.

Celis, R., De Jonge, H., De Jonge, L.W., Real, M., Hermosin, M.C., Cornejo, J., 2006. The role of mineral and organic components in phenanthrene and dibenzofuran sorption by soil. Eur. J. Soil. Sci. 57 (3), 308–319.

Chiou, C.T., Kile, D.E., 1998. Deviations from sorption linearity on soils of polar and nonpolar organic compounds at low relative concentrations. Environ. Sci. Technol. 32 (3), 338–343.

Cornelissen, G., Breedveld, G.D., Kalaitzidis, S., Christanis, K., Kibsgaard, A., Oen, A.M. P., 2006. Strong sorption of native PAHs to pyrogenic and unburned carbonaceous geosorbents in sediments. Environ. Sci. Technol. 40 (4), 1197–1203.

Cornelissen, G., Gustafsson, Ö., Bucheli, T.D., Jonker, M.T.O., Koelmans, A.A., van Noort, P.C.M., 2005. Extensive sorption of organic compounds to black carbon, coal, and kerogen in sediments and soils: mechanisms and consequences for distribution, bioaccumulation, and biodegradation. Environ. Sci. Technol. 39 (18), 6881–6895.

Diamond, M.L., Mackay, D., Welbourn, P.M., 1992. Models of multi-media partitioning of multi-species chemicals: the fugacity/aquivalence approach. Chemosphere 25 (12), 1907–1921.

Droge, S.T.J., Goss, K.U., 2013a. Development and evaluation of a new sorption model for organic cations in soil: contributions from organic matter and clay minerals. Environ. Sci. Technol. 47 (24), 14233–14241.

Droge, S.T.J., Goss, K.U., 2013b. Sorption of organic cations to phyllosilicate clay minerals: CEC-normalization, salt dependency, and the role of electrostatic and hydrophobic effects. Environ. Sci. Technol. 47 (24), 14224–14232.

Edmondson, J.L., Stott, I., Potter, J., Lopez-Capel, E., Manning, D.A.C., Gaston, K.J., Leake, J.R., 2015. Black carbon contribution to organic carbon stocks in urban soil. Environ. Sci. Technol. 49 (14), 8339–8346.

European Commission, 2004. European Union System for the Evaluation of Substances 2.0 (EUSES 2.0) Background Report. Prepared For the European Chemicals Bureau

by the National Institute of Public Health and the Environment (RIVM) (RIVM Report no. 601900005). European Chemicals Bureau. Bilthoven, The Netherlands. Available from http://ecb.jrc.it (Accessed on December 1, 2022).

European Union, 2023. Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on Classification, Labelling and Packaging of Substances and Mixtures, Amending and Repealing Directives 67/548/EEC and 1999/45/EC, and Amending Regulation (EC) No 1907/2006 [Amended on December 19, 2022; Commission Delegated Regulation (EU) 2023/707].

Fantke, P., Bijster, M., Guignard, C., Hauschild, M., Huijbregts, M., Jolliet, O., Kounina, A., Magaud, V., Margni, M., McKone, T., Posthuma, L., Rosenbaum, R., van de Meent, D., van Zelm, R., 2017. USEtox® 2.0 Documentation (version 1.1).

Franco, A., Ferranti, A., Davidsen, C., Trapp, S., 2010. An unexpected challenge: ionizable compounds in the REACH chemical space. Int. J. Life Cycle Assess. 15 (4), 321–325.

Glaser, B., Amelung, W., 2003. Pyrogenic carbon in native grassland soils along a climosequence in North America. Global. Biogeochem. Cycles. 17 (2), 1064.

Gobas, F.A.P.C., Mayer, P., Parkerton, T.F., Burgess, R.M., Van de Meent, D., Gouin, T., 2018. A chemical activity approach to exposure and risk assessment of chemicals. Environ. Toxicol. Chem. 37 (5), 1235–1251.

Goss, K.U., Bronner, G., 2006. What is so special about the sorption behavior of highly fluorinated compounds? J. Phys. Chem. A 110 (30), 9518–9522.

Goss, K.U., Buschmann, J., Schwarzenbach, R.P., 2003. Determination of the surface sorption properties of talc, different salts, and clay minerals at various relative humidities using adsorption data of a diverse set of organic vapors. Environ. Toxicol. Chem. 22 (11), 2667–2672.

Grathwohl, P., 1990. Influence of organic matter from soils and sediments from various origins on the sorption of some chlorinated aliphatic hydrocarbons: implications on Koc correlations. Environ. Sci. Technol. 24 (11), 1687–1693.

Han, L., Sun, K., Jin, J., Xing, B., 2016. Some concepts of soil organic carbon characteristics and mineral interaction from a review of literature. Soil Biol. Biochem. 94, 107–121.

Kah, M., Sigmund, G., Xiao, F., Hofmann, T., 2017. Sorption of ionizable and ionic organic compounds to biochar, activated carbon and other carbonaceous materials. Water Res. 124, 673–692.

Karickhoff, S.W., 1981. Semi-empirical estimation of sorption of hydrophobic pollutants on natural sediments and soils. Chemosphere 10 (8), 833–846.

Kickham, P., Otton, S.V., Moore, M.M., Ikonomou, M.G., Gobas, F.A.P.C., 2012. Relationship between biodegradation and sorption of phthalate esters and their metabolites in natural sediments. Environ. Toxicol. Chem. 31 (8), 1730–1737.

Li, L., Arnot, J.A., Wania, F., 2018a. Towards a systematic understanding of the dynamic fate of polychlorinated biphenyls in indoor, urban and rural environments. Environ. Int. 117, 57–68.

Li, L., Arnot, J., Wania, F., 2018b. Revisiting the contributions of far- and near-field routes to aggregate human exposure to Polychlorinated Biphenyls (PCBs). Environ. Sci. Technol. 52 (12), 6974–6984.

Li, Y., Oliver, D.P., Kookana, R.S., 2018. A critical analysis of published data to discern the role of soil and sediment properties in determining sorption of per and polyfluoroalkyl substances (PFASs). Sci. Total Environ. 628–629, 110–120.

Lu, Z., MacFarlane, J.K., Gschwend, P.M., 2016. Adsorption of organic compounds to diesel soot: frontal analysis and polyparameter linear free-energy relationship. Environ. Sci. Technol. 50 (1), 285–293.

MacKay, A.A., Vasudevan, D., 2012. Polyfunctional ionogenic compound sorption: challenges and new approaches to advance predictive models. Environ. Sci. Technol. 46 (17), 9209–9223.

Mackay, D., Arnot, J.A., 2011. The application of fugacity and activity to simulating the environmental fate of organic contaminants. J. Chem. Eng. Data 56 (4), 1348–1355.

Mackay, D., Paterson, S., Shiu, W.Y., 1992. Generic models for evaluating the regional fate of chemicals. Chemosphere 24 (6), 695–717.

Mayer, P., Toräng, L., Glæsner, N., Jönsson, J.Å., 2009. Silicone membrane equilibrator: measuring chemical activity of nonpolar chemicals with poly(dimethylsiloxane) microtubes immersed directly in tissue and lipids. Anal. Chem. 81 (4), 1536–1542.

Nelson, D.W., Sommers, L.E., 1996. Total carbon, organic carbon, and organic matter. In: Methods of Soil analysis: Part 3 Chemical Methods, 5. Wiley, pp. 961–1010.

Neumann, M., Schliebner, I., 2019. Protecting the Sources of Our Drinking Water: The Criteria for Identifying Persistent, Mobile and Toxic (PMT) Substances and Very Persistent and Very Mobile (vPvM) Substances Under EU Regulation REACH (EC) No 1907/2006. UBA TEXTE 127/2019. German Environment Agency (Umweltbundesamt; UBA). Dessau-Roßlau, Germany.

OECD, 2000. Test No. 106: Adsorption – Desorption Using a Batch Equilibrium Method. Organisation for Economic Co-operation and Development.

OECD, 2020. Overview of Concepts and Available Guidance related to Integrated Approaches to Testing and Assessment (IATA). Organisation for Economic Cooperation and Development. Series on Testing and Assessment No.

Paya-Perez, A.B., Riaz, M., Larsen, B.R., 1991. Soil sorption of 20 PCB congeners and six chlorobenzenes. Ecotoxicol. Environ. Saf. 21 (1), 1–17.

Pignatello, J.J., Mitch, W.A., Xu, W., 2017. Activity and reactivity of pyrogenic carbonaceous matter toward organic compounds. Environ. Sci. Technol. 51 (16), 8893–8908.

Pignatello, J.J., Xing, B., 1996. Mechanisms of slow sorption of organic chemicals to natural particles. Environ. Sci. Technol. 30 (1), 1–11.

Plata, D.L., Hemingway, J.D., Gschwend, P.M., 2015. Polyparameter linear free energy relationship for wood char–water sorption coefficients of organic sorbates. Environ. Toxicol. Chem. 34 (7), 1464–1471.

Reichenberg, F., Mayer, P., 2006. Two complementary sides of bioavailability: accessibility and chemical activity of organic contaminants in sediments and soils. Environ. Toxicol. Chem. Int. J. 25 (5), 1239–1245.

Reid, B.J., Jones, K.C., Semple, K.T., 2000. Bioavailability of persistent organic pollutants in soils and sediments—A perspective on mechanisms, consequences and assessment. Environ. Pollut. 108 (1), 103–112.

Schwarzenbach, R.P., Gschwend, P.M., Imboden, D.M., 2017. Environmental Organic Chemistry, 3rd edit. Wiley.

Sigmund, G., Gharasoo, M., Hüffer, T., Hofmann, T., 2020. Deep learning neural network approach for predicting the sorption of ionizable and polar organic pollutants to a wide range of carbonaceous materials. Environ. Sci. Technol. 54 (7), 4583–4591.

Spark, K.M., Swift, R.S., 2002. Effect of soil composition and dissolved organic matter on pesticide sorption. Sci. Total Environ. 298 (1), 147–161.

Ulrich, N., Endo, S., Brown, T.N., Watanabe, N., Bronner, G., Abraham, M.H., Goss, K.U. 2017. UFZ-LSER database v 3.2 [Internet]. Helmholtz Centre for Environmental Research-UFZ. Leipzig, Germany. Available from http://www.ufz.de/lserd (Accessed on December 1, 2022).

Webster, E., Mackay, D., Wania, F., 1998. Evaluating environmental persistence. Environ. Toxicol. Chem. 17 (11), 2148–2158.

Zhang, Z., Sangion, A., Wang, S., Gouin, T., Brown, T.N., Arnot, J.A., Li, L., 2024. Chemical space covered by applicability domains of quantitative structure-property relationships and semi-empirical relationships in chemical assessments. Environ. Sci. Technol. 58 (7), 3386–3398.