Article

A Framework for Developing Tools to Predict PFAS Physical–Chemical Properties and Mass-Partitioning Parameters

Mark L. Brusseau

Environmental Science Department, The University of Arizona, Tucson, AZ 85721, USA; brusseau@arizona.edu

Abstract: A framework for developing predictive models for PFAS physical–chemical properties and mass-partitioning parameters is presented. The framework is based on the objective of developing tools that are of sufficient simplicity to be used rapidly and routinely for initial site investigations and risk assessments. This is accomplished by the use of bespoke PFAS-specific QSPR models. The development of these models entails aggregation and curation of measured data sets for a target property or parameter, supplemented by estimates produced with quantum–chemical ab initio predictions. The application of bespoke QSPR models for PFAS is illustrated with several examples, including partitioning to different interfaces, uptake by several fish species, and partitioning to four different biological materials. Reasonable correlations to molar volume were observed for all systems. One notable observation is that the slopes of all of the regression functions are similar. This suggests that the partitioning processes in all of these systems are to some degree mediated by the same mechanism, namely hydrophobic interaction. Special factors and elements requiring consideration in the development of predictive models are discussed, including differences in bulk-phase versus interface partitioning processes.

Keywords: PFOS; PFOA; adsorption; uptake; QSPR models

1. Introduction

Studies have identified that PFAS are present in essentially all environmental and ecological compartments. This ubiquitous distribution poses significant potential risks to human and ecosystem health. Accurate characterizations of exposure risks and the development of effective mitigation measures are both essential to reducing the impacts of PFAS contamination. Implementing accurate risk assessments and developing effective mitigation measures requires an in-depth understanding of the distribution of PFAS in the environment and their associated transport and fate behavior. Research conducted over the past two decades has shown that PFAS mobility and distribution is influenced by several mass-transfer (i.e., partitioning) processes [1–4]. These partitioning processes are governed by the physical–chemical properties of PFAS, which are a function of their molecular structure.

The fundamental properties of aqueous solubility and vapor pressure determine the propensity for PFAS to reside in the aqueous or vapor phases, which has a significant impact on mobility given that these phases are the standard conduits for contaminant transport in the environment. The acid dissociation constant (i.e., pKa) is another important property for PFAS given that many PFAS are ionizable. PFAS mobility and distribution is also influenced by mass partitioning to a number of other phases and interfaces. For example, PFAS may partition to bulk organic immiscible liquids (NAPL) such as hydrocarbon fuels and chlorinated solvents, as well as to the interfaces between NAPL and water [3,5–7]. PFAS have been demonstrated to adsorb at (or partition to) air–water interfaces [3,6,8–14]. This process has been demonstrated to be critical for PFAS in multiple systems, including retention and leaching in vadose zones, accumulation at aerosol surfaces in the atmosphere, accumulation at atmosphere–surface water interfaces, and controlling removal for foam

fractionation and other treatment methods. A multitude of studies have examined the partitioning of PFAS to solid phases, i.e., sorption by soils, sediments, and engineered media [15–20]. The degree to which PFAS undergo solid-phase sorption is a key factor mediating retention and transport in the subsurface and for the performance of several treatment methods. Finally, many studies have investigated the uptake of PFAS by flora and fauna, i.e., mass partitioning to biological media, as discussed in recent reviews [21–23]. This uptake controls the bioaccumulation of PFAS within organisms, the bioconcentration within the food web, and the distribution within the ecosystem.

Clearly, a detailed, quantitative understanding of PFAS distribution and transport in the environment requires knowledge of the mass-partitioning behavior of PFAS between the many relevant phases present, and how their physical–chemical properties influence such partitioning. However, physical–chemical properties and mass-partitioning parameters have been measured for only a relatively few PFAS. Furthermore, measuring them for the thousands of existing PFAS is not practical considering the costs and time required. Hence, methods are needed to accurately predict PFAS physical–chemical properties and mass-partitioning parameters to support site investigations, risk assessments, and transport and fate modeling. While several valuable efforts have been made to develop predictive methods to date, there is a continued need for further research to resolve remaining uncertainties and limitations and to develop an array of approaches that address different application needs.

The objective of this work is to present a framework for developing cost-effective tools for predicting PFAS properties and parameters. The approaches available for the prediction of physical–chemical properties will first be discussed, followed by an examination of prior applications focused specifically on PFAS. A framework for developing a cost-effective approach will then be presented, along with illustrative examples and discussion of special considerations, uncertainties, and limitations.

2. Methods for Predicting Physical–Chemical Properties and Parameters

Several methods exist for predicting the properties and mass-partitioning parameters of chemical compounds. These various methods have been used for many years and successfully applied to a wide range of compounds. The methods can be divided into two general categories—empirical methods and physical-modeling methods.

A widely used empirical prediction method is the quantitative-structure/property-relationship (QSPR) approach. QSPR models are based on developing correlations between measured properties or parameters that describe a specific behavior of the compounds of interest and molecular features of those compounds. The molecular features are represented by so-called descriptors, such as molar volume, that can be readily determined independently. Data sets comprising the measured property or parameter are used to develop and test the models. Linear free-energy relationships (LFERs) are another empirical-based approach wherein correlations are established between two properties for a set of compounds, one property for which predictions are desired and another property for which measured data are more readily available. The octanol–water partition coefficient is a commonly used reference property for the latter. An advantage of these methods is their relative ease of use and ability to be applied rapidly, which supports routine application for site investigations and risk assessments. A primary disadvantage of these methods is the need for training data sets that are representative of the target compounds for which predictions are being made.

Ab initio physical-modeling methods employ quantum–chemical or molecular–mechanical calculations that are based on atomistic models of the molecule in question and its disposition in the relevant solvent. The defined model is used to generate information about the system, such as energy states, electron densities, and solvation free energies, that are then used to predict the target property or parameter. These methods are essentially independent of the need for measured-data training sets. This is a distinct advantage compared to the empirical methods. A primary disadvantage of these methods is their

greater complexity and required computational resources. As a result, they are not as easy and rapid to use as the empirical methods, which generally limits their routine application.

A formal program titled the Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) Challenges, funded by the National Institutes of Health, has been in progress for several years to support the blind testing of multiple prediction methods [http://www.samplchallenges.org/ (accessed 12 February 2024)]. These efforts have provided several valuable insights into the best methods and practices for predicting physical–chemical properties. The results of the two most recent challenges, SAMPL6 and SAMPL7, which tested the ability to predict octanol–water partition coefficients (log Kow) for several series of organic compounds, demonstrated that the best performing prediction methods were those based on quantum–chemistry and QSPR approaches [24,25]. Notably, it was shown that empirical-based methods such as QSPR models can perform equally as well as the ab initio physical-modeling methods. However, the accuracy of the QSPR methods is dependent upon the quality of the training data sets used to develop the QSPR relationships. In particular, the size and representativeness of the training data sets were more important than the specific approaches and descriptors employed. Of the two general physical-modeling methods, the quantum-chemical based methods were superior to those based on molecular-mechanical approaches. Quantum-chemical based methods employing COSMOtherm [26], a commercial property-prediction platform, achieved the overall highest scores of the physical-modeling methods.

Both categories of methods have been applied to predicting PFAS properties and mass-partitioning parameters. Several studies have developed and applied bespoke QSPR and LFER models for predicting specific PFAS properties. These bespoke empirical models are developed through the use of measured data sets obtained for specific PFAS for a specific property or parameter. For example, LFER models for several partitioning parameters, including air–water, octanol–water, and octanol–air, have been developed for fluorotelomer alcohols (FTOHs) and olefins [27]. QSPR models for several properties including aqueous solubility, vapor pressure, and octanol–water partitioning were developed for data sets comprising a number of PFAS, including perfluorocarboxylic acids (PFCAs), perfluorosulfonic acids (PFSAs), and FTOHs [28,29]. QSPR models for air–water interfacial adsorption have been developed for a wide range of PFAS, including all headgroup types [8–10,14,30]. QSPR and LFER models have also been developed for solid-phase sorption of PFAS [31–35]. Finally, several studies have investigated the uptake of PFAS by plants or animals and observed that bioconcentration factors exhibit correlations to PFAS chain length [21–23,36–42], which has led to development of QSPR models.

The PFAS-specific QSPR models that have been developed to date have generally been shown to produce reliable estimates of properties and parameters for PFAS that are represented by the training data sets used to develop the models. They provide valuable confirmation that PFAS-specific QSPR models can provide a robust method to rapidly predict relevant parameters. However, the models developed so far have typically focused on a small subset of PFAS. Further development and testing is required to incorporate a wider range of PFAS and environmental conditions.

Large-scale general purpose (i.e., non-PFAS specific) software platforms have also been used to predict PFAS properties. Two of the most widely used software platforms employed in prior studies of PFAS are EPI Suite and COSMOtherm. EPI Suite is a web-based platform operated by the U.S. EPA that comprises an extensive database of measured data and a suite of individual prediction modules that employ QSPR-based estimation methods. It therefore serves as a prime representative of an estimation-based method, and it is freely available for use. As previously noted, COSMOtherm is a commercial platform that is based on quantum–chemical calculations. Studies comparing the performance of multiple methods, including COSMOtherm and EPI Suite, for predicting measured PFAS properties have demonstrated that COSMOtherm generally produced reliable estimates of PFAS properties while EPI Suite performed comparatively poorly, with the poor performance attributed to a lack of PFAS-specific training data [34,43,44]. The effectiveness of

COSMOtherm for predicting PFAS properties was further demonstrated in studies that compared COSMOtherm predictions alone to measured data [45,46]. Finally, a recent study demonstrated that a QSPR-based method and COSMOtherm both produced accurate predictions of hexadecane–air partition coefficients for a wide range of PFAS [47].

The results of the PFAS-specific studies compiled in the preceding paragraph are consistent with the results obtained from the SAMPL testing program discussed above. First, COSMOtherm appears to produce generally reliable predictions and therefore serves as an effective independent source of predictions for cases where measured data are not available. Second, empirical-based methods such as QSPR models can perform equally as well, and better in some cases, compared to COSMOtherm and other physical-modeling methods. However, the performance of the empirical methods is highly dependent upon the quality and representativeness of the training data.

3. Framework Development

The objective of the framework is to provide an integrated approach for developing cost-effective tools for predicting PFAS physical–chemical properties and mass-partitioning parameters. The framework is focused on developing prediction tools that are readily accessible and easy to use so that they can be routinely and rapidly employed for site-characterization and risk-assessment applications by consultants, regulators, and other interested parties. This has been identified as a critical need [48]. The effort is accomplished by using PFAS-specific QSPR models as the basis of the prediction methods. It is recognized that prediction tools may be developed for other objectives, and that the associated frameworks and resultant tools may differ from the ones presented herein.

The framework is based on the following principles and workflow logic:

- There is a critical need for rapid, easy-to-use methods to predict physical–chemical properties and mass-partitioning parameters for the thousands of existing PFAS of all structure types
- Empirical methods such as QSPR models have been demonstrated to produce accurate predictions of properties and parameters for a wide range of compounds, including PFAS, as long as representatives of the target compounds are present in the QSPR database
- PFAS can be grouped by molecular structure (e.g., headgroup type) and representative members of each structure type included in QSPR models
- Large-scale empirical-based software platforms such as EPI Suite currently do not produce reliable predictions of PFAS properties for most cases, due to a lack of PFAS-specific training data
- Physical-modeling methods such as COSMOtherm have been demonstrated to produce reliable predictions of properties and parameters for a wide range of compounds, including PFAS
- Due to their commercial nature, relative complexity, and higher computation demands, platforms such as COSMOtherm are not currently suitable for the rapid, day-to-day routine use required for environmental site investigations and screening assessments
- COSMOtherm can be employed to develop validated predictions of properties to fill in data gaps for measured PFAS properties and parameters.
- Curated measured and COSMOtherm-predicted data can be used to develop bespoke PFAS-specific QSPR models that can provide rapid and accurate predictions
- The PFAS-specific QSPR models will provide a critical tool that is accessible and easy to use for consultants, regulators, and other interested parties.

The basic framework comprises the following elements. (1) Aggregate relevant measured data sets for the selected property or parameter. (2) Curate the compiled database. Prior studies that have compiled measured PFAS property data have shown that significant differences can occur between measured values from different studies, due to a number of factors. Therefore, the aggregated data sets should be curated following best practices to ensure data representativeness. Li et al. provide a recent comprehensive overview of

this issue [49]. (3) Identify data gaps present in the curated database. To be more broadly applicable, the prediction tools should be inclusive of all major PFAS structure types. Data gaps may often involve missing data for certain structure types. (4) Fill in the data gaps with validated predicted values. COSMOtherm or a similar platform can be used to generate independent predictions of properties and partitioning parameters for representative PFAS structure types. (5) Develop bespoke PFAS-specific QSPR models. The measured and independently-predicted data sets are combined to develop a PFAS-specific QSPR model for predicting the selected property or mass-partitioning parameter. The two key elements comprising steps 4 and 5 will be briefly discussed.

As previously noted, it is not practical to measure properties for the many PFAS that exist. In addition, it is not possible to measure some properties for certain PFAS due to a number of factors, such as constraints in experimental methods or analysis, or commercial availability of the PFAS. Given its prior successful applications to PFAS systems, COSMOtherm serves as an example physical-modeling method that can be used to produce predictions to resolve critical data gaps.

COSMOtherm is based on the approach termed "Conductor-like Screening Model for Real Solvents" that was developed by Klamt [26]. The basis of this model is an assumption that solute molecules are embedded within a cavity that is formed in the solvent (i.e., the basis for the hydrophobic interaction mechanism). The solvent, typically aqueous solution for many environmental applications, is treated as a virtual conductor environment. Both the polarization charge density on the molecule surface and the total energy of the molecule is calculated with a quantum chemical field algorithm. The strength of electrostatic, hydrogen bonding, and dispersion interactions between the molecule and the solvent are also calculated. The strength of these interactions is then used to predict the chemical potential of the solute molecules in the solvent using statistical thermodynamics. The magnitude of each selected physical–chemical property is then determined based on the calculated chemical potentials. A key element in the application of COSMOtherm is the identification of the most representative 3-D conformation of the molecule in the solvent under the relevant conditions. This can involve significant effort and time. The application of COSMOtherm to PFAS has been demonstrated in prior studies, with the methods developed and tested [34,43–47].

The final prediction tools will comprise bespoke PFAS-specific QSPR models. There are many publications available providing guidance on best practices for the development of QSPR models [50,51]. The guiding principle in developing the models for this application is that of ease-of-use for routine implementation. One of the key elements of the QSPR method is the selection of the molecular descriptors. Several molecular-descriptor types are available, including size-based descriptors such as molecular mass and molar volume, constitutional descriptors based on the numbers of a specific type of atom or bond (e.g., carbon number), descriptors characterizing molecular structure (such as the molecular connectivity index), and complex 3-D geometrical descriptors [52]. To support ease-of-use, it is important to focus on descriptors that are readily available in reference materials or are easily determined.

Another important element is the number of descriptors to employ. Many QSPR models employ a single descriptor, while others use multiple. In the interest of simplicity, a parsimonious approach would focus initially on developing single-descriptor models. This approach is anticipated to produce reasonable results for some properties and parameters. However, it is also anticipated that the single-descriptor approach may fail for other properties or parameters. In these cases, a polyparameter approach can be used wherein multiple descriptors are employed. However, it is important to keep in mind that as the number of descriptors increases, the more complicated the model becomes, wherein at some point diminishing returns limit the ease-of-use.

The issues of which descriptor to employ and the number of descriptors needed is illustrated with QSPR models developed for air–water interfacial adsorption. Brusseau and colleagues have developed bespoke QSPR models for predicting PFAS air–water interfacial

adsorption coefficients [8–10,14]. An example is presented in Figure 1. It is observed that the model is successful for all major structure types. In addition, the single-descriptor approach is sufficient in this case. The descriptor employed, molar volume, is a readily available parameter that can be determined in different ways, including from the quotient of molar mass and mass density. Therefore, the model meets both key criteria of being widely applicable as well as simple to use.
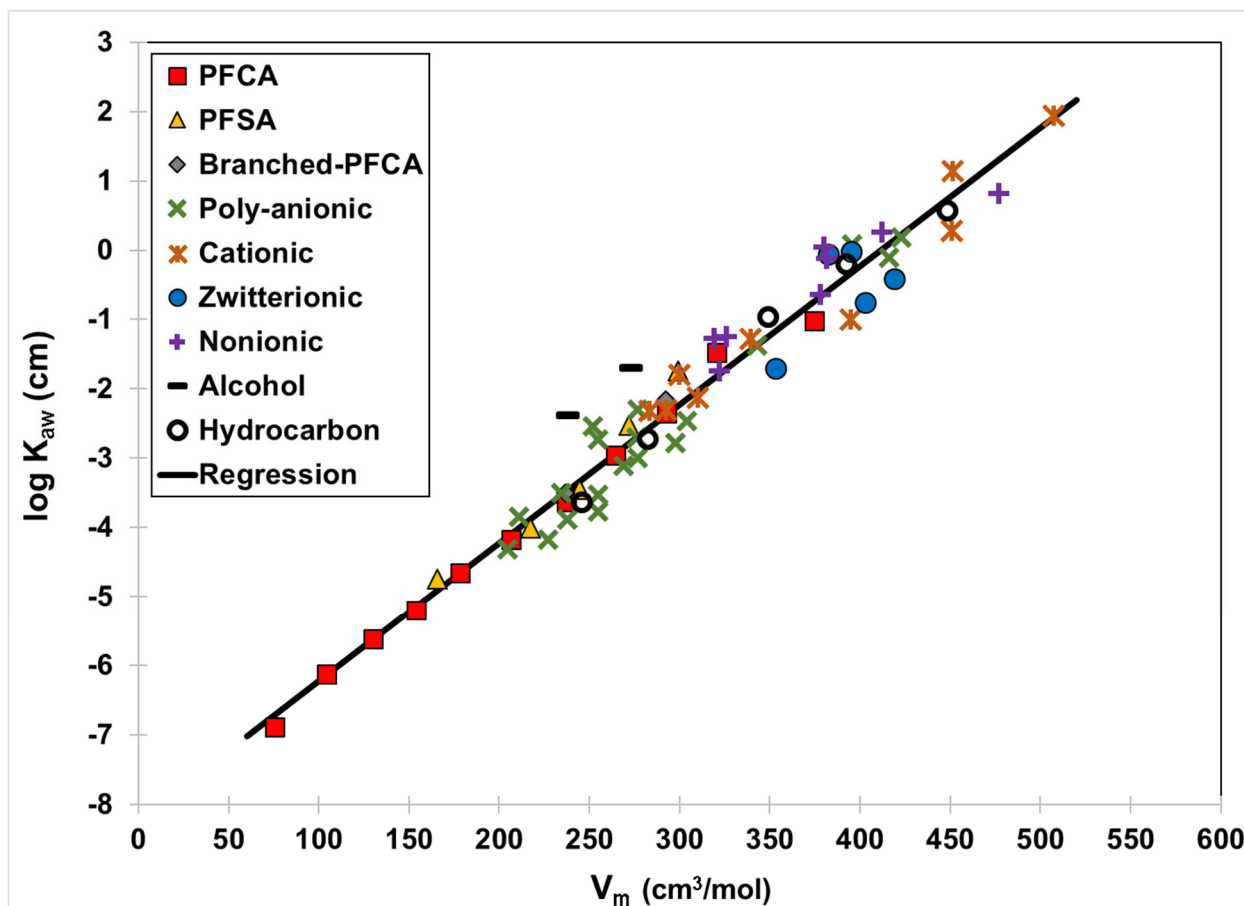

Figure 1. QSPR model to predict air–water interfacial adsorption coefficients (K_{aw}) for PFAS as a function of molar volume (V_m). The data include measurements for 61 different individual PFAS representing all types of molecular structures. From Brusseau and Van Glubt [14].

Other descriptors are commonly employed in PFAS studies, including fluorinated carbon number, carbon number, and molecular weight. These descriptors have been used with some success for applications involving primarily PFCAs and PFSAs. However, molar volume was demonstrated to be a superior descriptor in representing a wider range of structures for air–water interfacial adsorption [9]. For example, the poor performance of fluorinated carbon number is shown in Figure 2 for a subset of the data presented in Figure 1. It is representative for the PFCAs, PFSAs, and branched PFCAs, consistent with prior studies wherein fluorinated carbon number has been used for QSPR models for other partitioning parameters. However, fluorinated carbon number is clearly inadequate for the more complex PFAS structures.
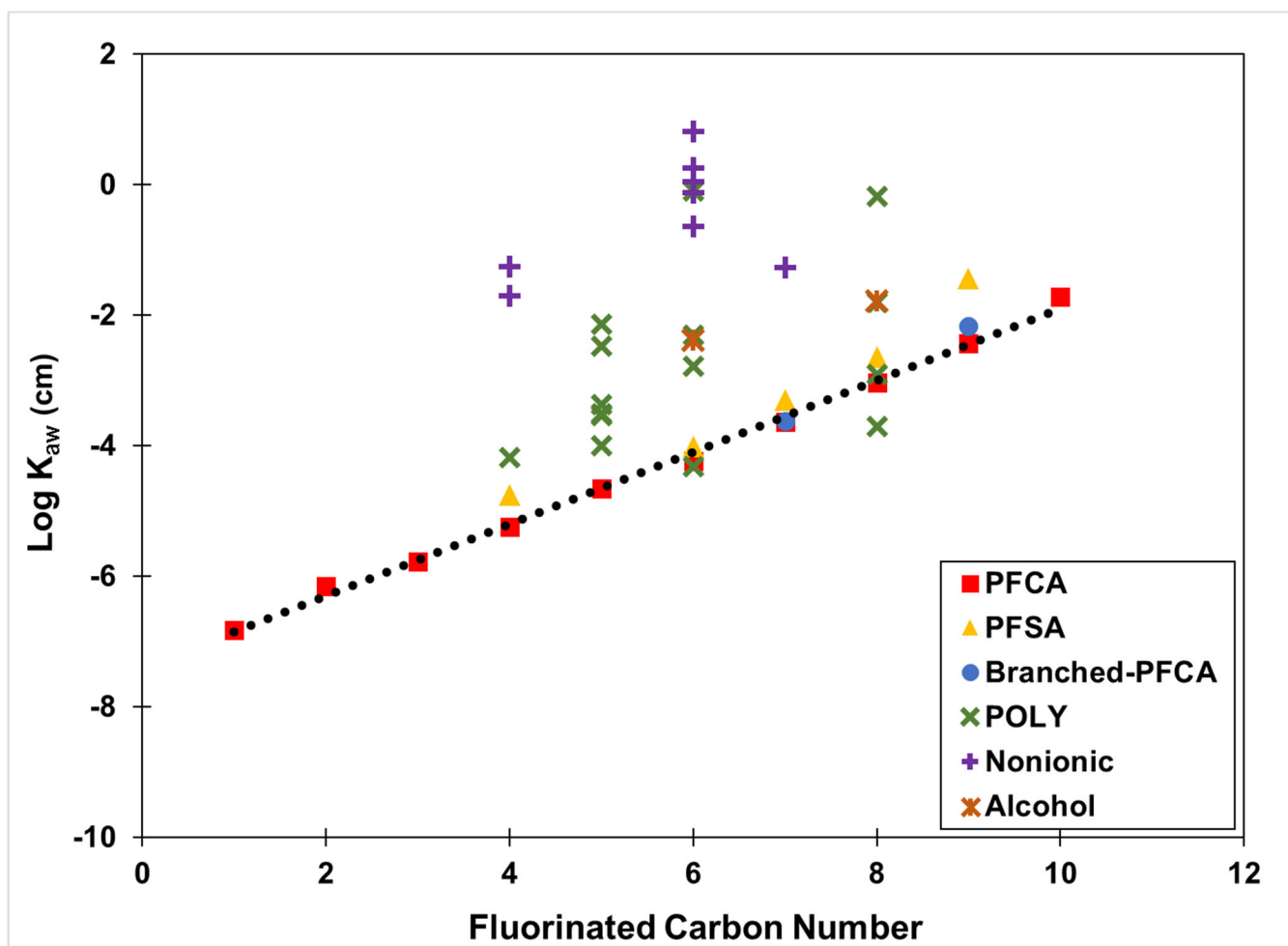
Figure 2. QSPR model to predict air–water interfacial adsorption coefficients (K_{aw}) for PFAS as a function of fluorinated carbon number. The data comprise a subset of data presented in Figure 1. Revised from Brusseau [9].

As noted, the QSPR model presented in Figure 1 is successful for all major structure types. However, the single-descriptor model failed for PFAS with very large headgroups. In this case, a two-descriptor model was developed to successfully predict coefficients for these PFAS [30].

The issue of descriptor selection can be further illustrated by examining recent models used to predict the log of the organic-carbon normalized sorption coefficient (K_{oc}). Coppola et al. [33] used EPI Suite to predict log K_{oc} values for several PFAS based on log K_{ow}. A four-descriptor QSPR model was recently developed for predicting log K_{oc} values for several PFAS [31]. The two models produce similar predictions of log K_{oc}. For example, the first model predicts log K_{oc} values of 1.34, 2.82, and 3.35 for perfluorobutanoic acid, perfluorooctanoic acid, and perfluorooctane sulfonic acid, respectively, whereas the second model predicts values of 1.36, 2.65, and 3.21. The observation that the four-parameter model produces similar results compared to the single-parameter model illustrates that the inclusion of additional parameters, which complicates the model, does not always lead to improved accuracy. In addition, obtaining values for some of the descriptors used in the four-parameter model (e.g., "lowest unoccupied molecular orbital energy") requires the use of chemical modeling programs, which significantly reduces ease of use.

4. Example Application and Special Considerations

An example of combining measured data with COSMOtherm-predicted data sets to create a bespoke PFAS-specific QSPR model is illustrated in Figure 3 for vapor pressure. COSMOtherm was successful in producing predicted values for PFAS for which measured data were not available. A single-descriptor model is observed to be adequate for this case, at least for the limited range of molecular structures represented. In addition, fluorinated carbon number serves as a representative descriptor, in contrast to the prior example of air–water interfacial adsorption. Further testing would be required to determine if fluorinated carbon number would be effective for a broader range of structure types.
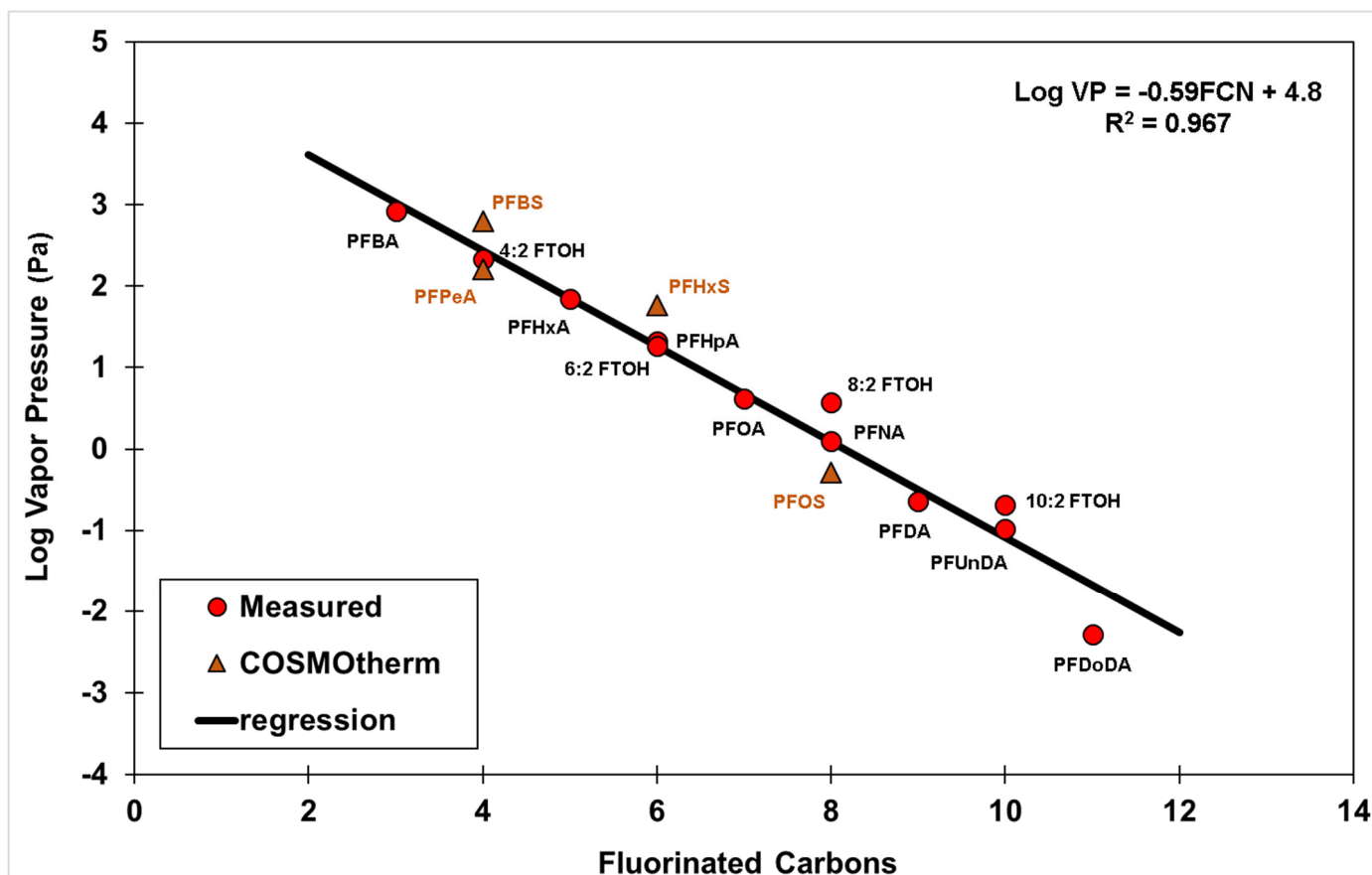

Figure 3. QSPR model for PFAS vapor pressures developed from an integrated set of measured values and values predicted with COSMOtherm. Data combined from [34,45,46] and citations included therein.

A critical consideration in the development of property-prediction tools is the dependency of the particular property or parameter on system conditions. For example, bulk-phase partitioning processes such as air–water, octanol–water, and octanol–air can generally be treated as linear, which means that the associated partition coefficients (H, K_{ow}, and K_{oa}) are constants independent of concentration. Conversely, interfacial partitioning processes such as air–water interfacial adsorption, NAPL–water interfacial adsorption, and solid-phase adsorption are typically nonlinear, and can be treated as linear only under certain conditions. Hence, the associated interfacial adsorption coefficients may be functions of concentration. This complicates the development of predictive models.

Another difference between bulk and interfacial partitioning is the relative impacts of solution chemistry, wherein it is typically of greater import for the latter processes. For example, Abusallout et al. [53] conducted an important test by measuring the impact of solution chemistry on the Henry's coefficients of two PFAS. Measurements conducted

with a deionized water matrix were compared to matrices comprised of two tertiary wastewater effluents, tap water produced by a conventional treatment plant treating surface water, and groundwater from a monitoring well near an active U.S. Air Force base. Various water-quality parameters such as dissolved organic carbon, sulfate, nitrate, chloride, calcium carbonate, and turbidity varied greatly among the waters. However, the Henry's coefficients were statistically similar across all matrices, including the deionized water, for both PFAS. These results indicate that mass transfer between water and air was not significantly influenced by the composition of the solution. In contrast, numerous studies have shown that air–water interfacial adsorption and in particular solid-phase adsorption is sensitive to solution composition.

Another factor contributing to the relative complexity of interfacial partitioning/adsorption processes is the need to consider the nature of the interface, which for many environmental systems is physically and/or geochemically heterogeneous. In contrast, bulk phases are typically treated as comparatively simple homogeneous media. As a result of the preceding and other factors, the development of prediction tools for parameters associated with interfacial partitioning may often be more difficult than for bulk-phase partitioning.

The range in complexity of environmental interfaces is illustrated in Figure 4, wherein three common interfaces are presented. The air–water interface is the most physically and geochemically homogeneous of the three, particularly in the absence of constituents accumulated at the interface. The surfaces of granular activated carbon (GAC) particles are physically heterogeneous while comparatively geochemically homogeneous. Finally, soils are comprised of many different components and, as a result, their surfaces are both physically and geochemically heterogeneous.
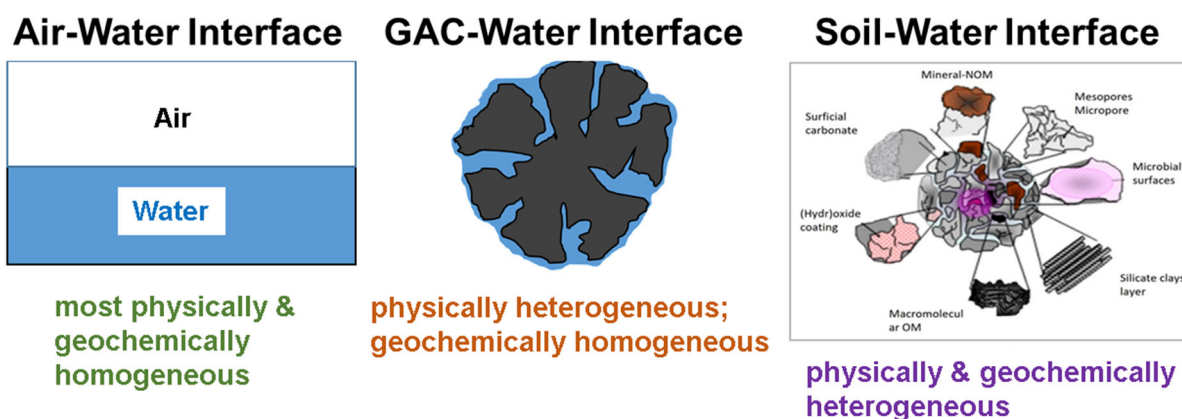

Figure 4. Comparison of three different environmentally relevant interfaces.

The disparity in interface complexity can have a significant impact on the development of predictive tools. This is illustrated in Figure 5, wherein QSPR models are presented for air–water interfacial adsorption, GAC adsorption, and soil adsorption of PFAS. Note that the regression for the soil data is based only on the long-chain PFAS. The linear functions satisfactorily represent the first two processes. Conversely, while the linear function is suitable for the long-chain PFAS for soil adsorption, the values for the shorter-chain PFAS deviate from the linear regression. Hence, the one-descriptor model does not capture the enhanced adsorption measured for the shorter-chain PFAS. Prior studies have hypothesized that the enhanced adsorption of the shorter-chain PFAS is related to the impact of additional mechanisms such as electrostatic interactions with inorganic constituents of the soil [15,16,20].
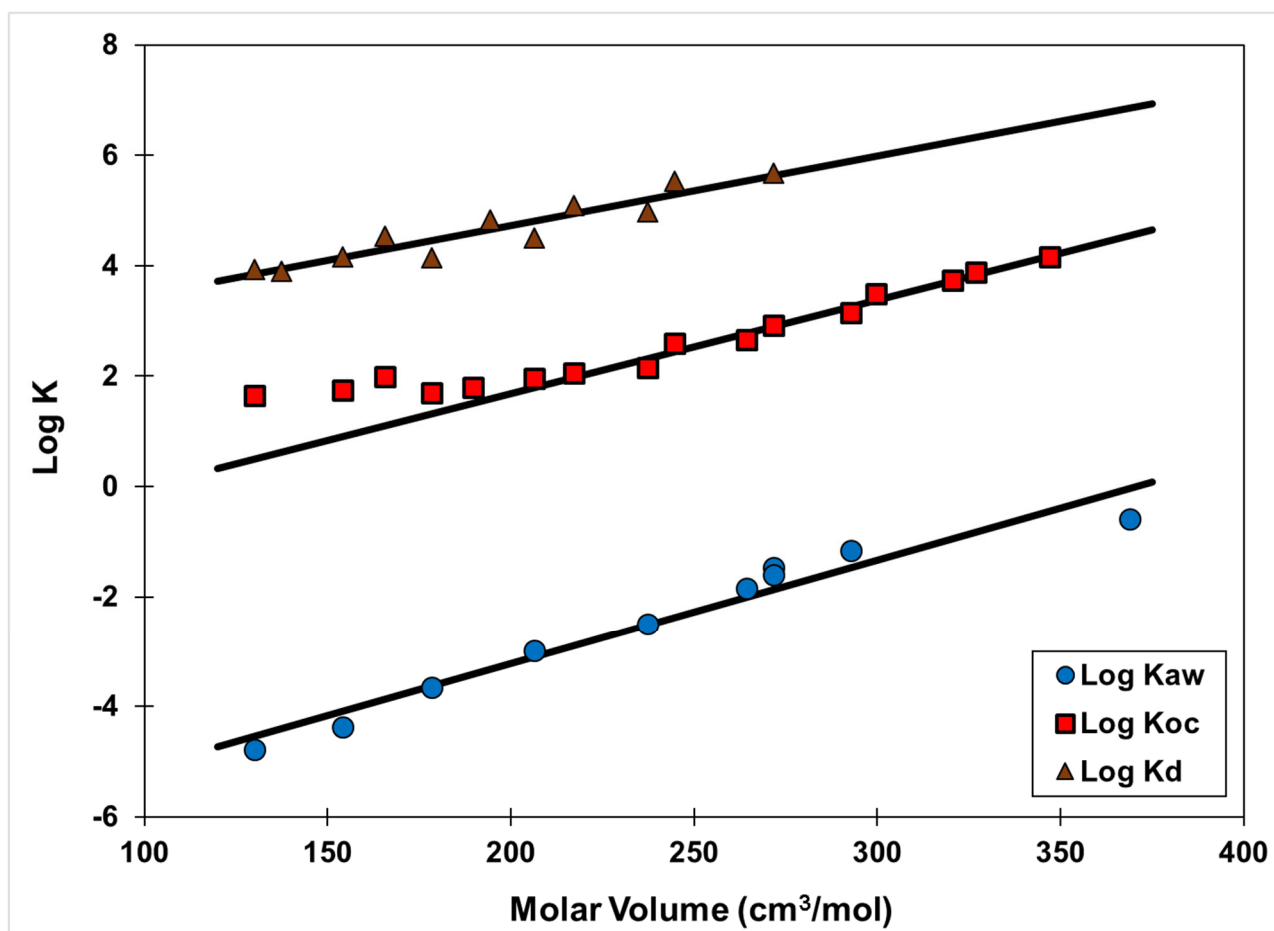
Figure 5. QSPR models for PFAS adsorption: (bottom data) at the air–water interface (K_{aw}), (middle data) by soils (K_{oc}), and (top data) by granular activated carbon (K_d). Revised from Brusseau [20]. Regression equation for adsorption by soils and sediments: $\log K_{oc} = 0.016\ V_m - 1.6$, $r^2 = 0.98$; Regression equation for adsorption by GAC: $\log K_d = 0.013\ V_m - 2.2$, $r^2 = 0.89$; Regression equation for adsorption at the air–water interface: $\log K_{aw} = 0.019\ V_m - 7.1$, $r^2 = 0.95$.

In summary, PFAS adsorption by soils and other geomedia is often nonlinear, may be influenced by solution chemistry, and is impacted by soil properties. As a result, it is likely that the development of predictive tools for solid-phase adsorption by soils and other geologic media will be one of the most difficult propositions of all properties and parameters. This difficulty is supported by the results of studies that have compared predictions obtained with general-purpose platforms such as EPI Suite and COSMOtherm to measured data sets [33,34]. These studies have reported that predicted values of the log K_{oc} compared poorly to measured values.

Numerous studies have investigated the uptake of PFAS by plants or animals. Several of these studies have shown that bioconcentration factors exhibit correlations to PFAS chain length [21–23,36–40]. To illustrate the application of QSPR analysis to such data, a model is presented in Figure 6 for uptake of several PFCAs and PFSAs by several freshwater fish species. The model provides a reasonably good representation of the data. Notably, the slope of the QSPR function for the data set is essentially identical to the slope of the QSPR model for air–water interfacial adsorption (Figure 5). Given that air–water interfacial adsorption is governed by the hydrophobic-interaction mechanism, the similarity in slopes suggests that uptake of PFAS by fish is also mediated to some degree by hydrophobic interaction.
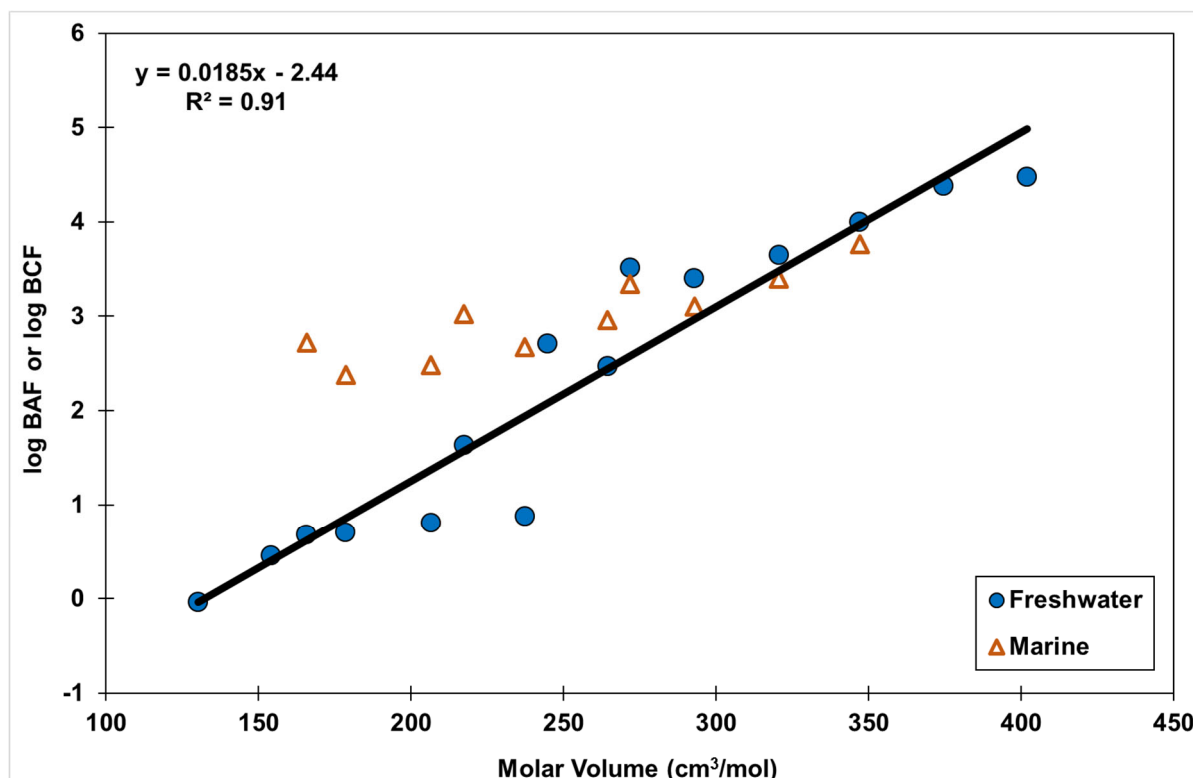
Figure 6. QSPR model for PFAS uptake by fish. The freshwater data are compiled from 10 studies, and include measurements for 13 species (Lewis et al. [23]). The marine data are compiled from 2 studies (Cheng et al. [40]; Jeon et al. [41]), and include measurements for 20 species. The regression function is fit only to the freshwater data.

Data sets for marine fish species are compared to the freshwater data in Figure 6. It is observed that the bioconcentration factors for the longer-chain PFAS determined for the marine fish are similar to those for the freshwater species. Conversely, the values for the shorter-chain PFAS are significantly greater for the marine species. Recall that similar behavior was observed for adsorption of shorter-chain PFAS by soils (Figure 5). One possible reason for the observed disparity is that uptake of the shorter-chain PFAS is influenced to a greater extent by specific electrostatic-mediated interactions with tissue constituents, and that these interactions are enhanced under the higher salinity conditions present in marine systems. Detailed investigations would be required to test this and other possible factors.

There are several factors that can complicate the assessment and quantification of PFAS uptake by plants and animals. These complications can lead to uncertainty in the magnitudes and robustness of bioconcentration factors, and concomitantly complicate the development of predictive tools. One factor is that investigations of uptake measure concentrations in different tissues and components in different studies. This is particularly relevant for animal studies wherein a wide range of tissues may be measured. For example, studies on PFAS uptake by fish have characterized concentrations in muscle, blood, liver, kidney, other organs, as well as whole body measurements.

Another critical factor is that uptake in most cases is likely mediated by interactions with specific constituents of the plant or animal [37,39]. Many studies have quantified PFAS uptake or adsorption by specific constituents such as phospholipid membranes, serum albumin, and other proteins. This research has demonstrated that PFAS appear to be associated primarily with proteins and cell membranes. They also can interact with lipids, but to a much lesser extent than standard hydrophobic contaminants. Some studies have measured the magnitude to which PFAS partition to proteins, membranes, or lipids. The magnitude of the uptake or adsorption is often quantified through the determination

of distribution coefficients (K_d). This work has shown that the measured distribution coefficients are functions of chain length. The development of predictive models will need to consider and account for these interactions with specific constituents. Detailed discussions of the development of predictive models for PFAS interactions with biological components are presented in prior works [37,39].

An example set of QSPR analyses for partitioning of several PFCAs and PFSAs to different biological constituents is presented in Figure 7. Data aggregated from multiple studies are reported for four specific constituents, bovine serum albumin, structural muscle protein, phospholipid membrane, and lipid. The QSPR regressions provide reasonable representations of the data, although the PFSAs values exhibit deviations in some cases. Notably, PFAS are observed to have significantly greater association with serum albumin and phospholipid-based membrane compared to muscle protein and in particular lipid. In addition, the slopes of the regression functions, particularly for the muscle protein and phospholipid membrane data, are very similar to the slopes determined for the fish-uptake data as well as the air–water interfacial adsorption data.
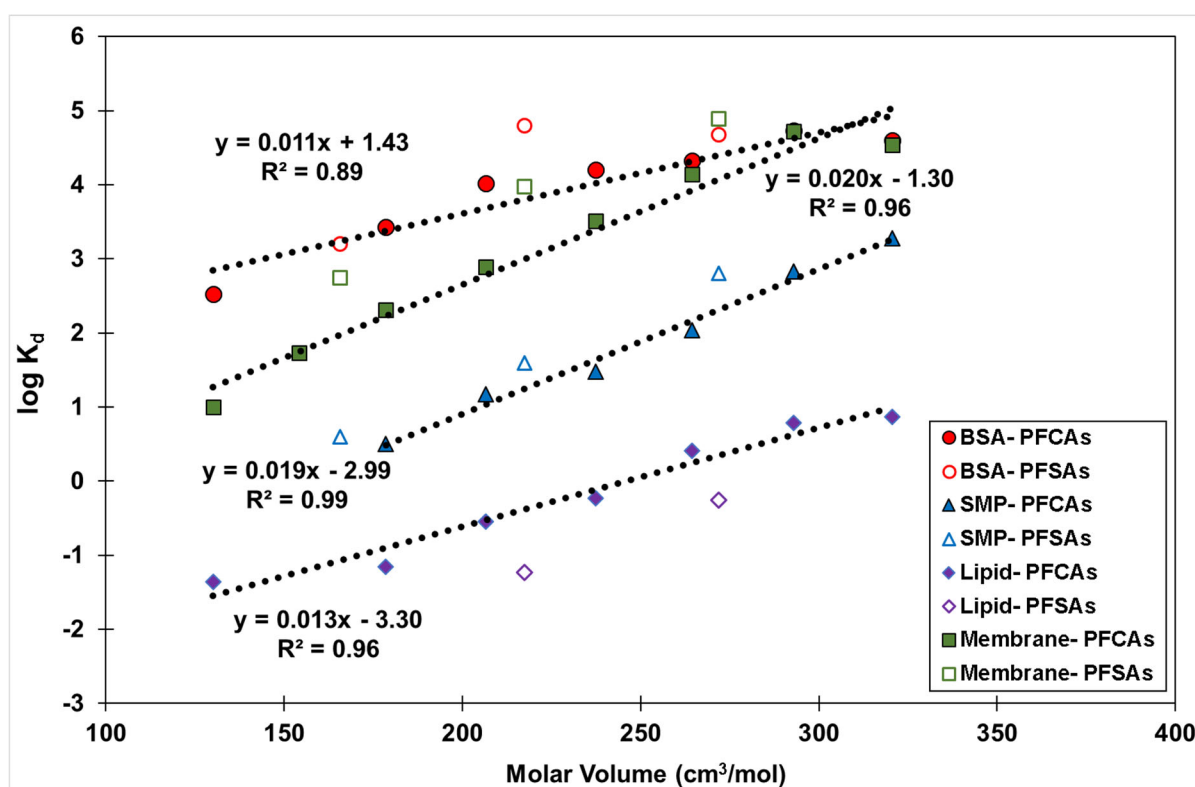

Figure 7. Measured distribution coefficients for interaction of PFAS with different biological constituents. BSA = bovine serum albumin, SMP = structural muscle protein, membrane = phospholipid membrane. Data compiled from the following sources: [53–58]. The regression lines are developed for the PFCA data only.

5. Conclusions

This work presents a framework for developing tools to predict PFAS physical–chemical properties and mass-partitioning parameters. The framework is based on the objective of developing tools that are of sufficient simplicity to be used rapidly and routinely for initial site investigations and risk assessments. This is accomplished with the use of bespoke PFAS-specific QSPR models. The development of these models entails aggregation and curation of measured data sets for a target property or parameter, supplemented by estimates produced with quantum–chemical ab initio predictions.

The effort and associated costs to develop a particular PFAS-specific QSPR model within this framework will likely be similar to those required to develop other types

of predictive tools focused on PFAS. These involve data collection, data curation, and data-gap filling, all of which are key to essentially all predictive-model development efforts. However, the effort and cost to use the developed models based on the presented framework are anticipated to be substantially less than those that may be associated with other tools. For example, COSMOtherm or a similar platform is employed in the presented framework to fill in data gaps. However, these platforms are not used for generating the actual predictions; this is done with the bespoke QSPR models. In contrast, the use of COSMOtherm or similar platforms may serve as the basis for other predictive tools, where they are used as the primary source of predicted parameters. In this case, there would be a greater level of effort and cost for the application.

The application of bespoke QSPR models for PFAS properties was illustrated with several examples. This included adsorption of PFAS by soils and sediments, by GAC, and at the air–water interface, as well as uptake by several fish species and partitioning to four different biological constituents. Reasonable correlations to molar volume were observed for all systems. One notable observation is that the slopes of all of the regression functions are similar. This suggests that the partitioning processes in all of these systems are to some degree mediated by the same mechanism, namely hydrophobic interaction. Of course, additional mechanisms of interaction are certainly in effect in many cases, with their significance a function of the specific medium, PFAS involved, and other factors.

As discussed herein and illustrated with examples, bespoke QSPR models are able to produce reasonably accurate predictions under the conditions for which they were developed. The development of PFAS-specific models for a specific target property or parameter is an advantage of the approach presented herein, wherein it is likely that the developed model will be highly representative for that specific system. In contrast, the accuracy of the large-scale general purpose (i.e., non-PFAS specific) software platforms such as EPI Suite for predicting PFAS properties has been demonstrated to be highly variable. PFAS-specific QSPR models have been demonstrated in some cases to perform better than general-purpose platforms, and they will likely continue to do so until the general-purpose platforms include a wider range of PFAS in the database.

There are several questions to be addressed in the development and testing of a given predictive model. A primary one is whether or not the model is broadly applicable to the full range of PFAS molecular structures, or if it is limited to certain subsets. Data reported to date comprise primarily PFCAs and PFSAs, with few measurements for precursor PFAS and other non-anionic PFAS. Models limited to specific PFAS types are useful however, as long as the limitations are clearly noted. The measurements and predictions have to date focused essentially exclusively on single-component PFAS systems, with minimal investigation of PFAS mixtures. Finally, a key question to investigate is the applicability of laboratory measurements and ab initio-based predictions to field-scale systems.

Data Availability Statement: The original contributions presented in the study are included in the article; the sources of the data sets aggregated from the literature are cited in the relevant figures.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Ahrens, L. Polyfluoroalkyl compounds in the aquatic environment: A review of their occurrence and fate. *J. Environ. Monit.* **2011**, *13*, 20–31. [CrossRef] [PubMed]
2. Krafft, M.P.; Riess, J.G. Per- and polyfluorinated substances (PFASs): Environmental challenges. *Curr. Opin. Colloid Interface Sci.* **2015**, *20*, 192–212. [CrossRef]
3. Brusseau, M.L. Assessing the potential contributions of additional retention processes to PFAS retardation in the subsurface. *Sci. Total Environ.* **2018**, *613–614*, 176–185. [CrossRef]

4.   Bryant, J.D.; Anderson, R.; Bolyard, S.C.; Bradburne, J.T.; Brusseau, M.L.; Carey, G.; Chiang, D.; Gwinn, R.; Hoye, B.R.; Maher, T.L.; et al. PFAS Experts Symposium 2: Key advances in poly- and perfluoroalkyl characterization, fate, and transport. *Remediat. J.* **2022**, *32*, 19–28. [CrossRef]

5.   McKenzie, E.R.; Siegrist, R.L.; McCray, J.E.; Higgins, C.P. The influence of non-aqueous phase liquids (NAPL) and chemical oxidant application on perfluoroalkyl acid (PFAA) fate and transport. *Water Res.* **2016**, *92*, 199–207. [CrossRef]

6.   Brusseau, M.L.; Yan, N.; Van Glubt, S.; Wang, Y.; Chen, W.; Lyu, Y.; Dungan, B.; Carroll, K.C.; Holguin, F.O. Comprehensive retention model for PFAS transport in subsurface systems. *Water Res.* **2019**, *148*, 41–50. [CrossRef] [PubMed]

7.   Van Glubt, S.; Brusseau, M.L. Contribution of nonaqueous phase liquids the retention and transport of per and polyfluoroalkyl substances (PFAS) in porous media. *Environ. Sci. Technol.* **2021**, *55*, 3706–3715. [CrossRef]

8.   Lyu, Y.; Brusseau, M.L.; Chen, W.; Yan, N.; Fu, X.; Lin, X. Adsorption of PFOA at the air-water interface during transport in unsaturated porous media. *Environ. Sci. Technol.* **2018**, *52*, 7745–7753. [CrossRef]

9.   Brusseau, M.L. The influence of molecular structure on the adsorption of PFAS to fluid-fluid interfaces: Using QSPR to predict interfacial adsorption coefficients. *Water Res.* **2019**, *152*, 148–158. [CrossRef]

10.   Brusseau, M.L.; Van Glubt, S. The influence of surfactant and solution composition on PFAS adsorption at fluid-fluid interfaces. *Water Res.* **2019**, *161*, 17–26. [CrossRef]

11.   Silva, J.A.; Martin, W.A.; Johnson, J.L.; McCray, J.E. Evaluating air-water and NAPL-water interfacial adsorption and retention of perfluorocarboxylic acids within the vadose zone. *J. Contam. Hydrol.* **2019**, *223*, 103472. [CrossRef] [PubMed]

12.   Costanza, J.; Arshadi, M.; Abriola, L.M.; Pennell, K.D. Accumulation of PFOA and PFOS at the air–water interface. *Environ. Sci. Technol. Lett.* **2019**, *6*, 487–491. [CrossRef]

13.   Schaefer, C.E.; Culina, V.; Nguyen, D.; Field, J. Uptake of poly and perfluoroalkyl substances at the air–water interface. *Environ. Sci. Technol.* **2019**, *53*, 12442–12448. [CrossRef] [PubMed]

14.   Brusseau, M.L.; Van Glubt, S. The influence of molecular structure on PFAS adsorption at air-water interfaces in electrolyte solutions. *Chemosphere* **2021**, *281*, 130829. [CrossRef] [PubMed]

15.   Nguyen, T.M.H.; Braunig, J.; Thompson, K.; Thompson, J.; Kabiri, S.; Navarro, D.A.; Kookana, R.S.; Grimison, C.; Barnes, C.M.; Higgins, C.P.; et al. Influences of chemical properties, soil properties, and solution pH on soil-water partitioning coefficients of per- and polyfluoroalkyl substances (PFASs). *Environ. Sci. Technol.* **2020**, *54*, 15883–15892. [CrossRef]

16.   Guelfo, J.L.; Higgins, C.P. Subsurface transport potential of perfluoroalkyl acids at aqueous film-forming foam (AFFF)-impacted sites. *Environ. Sci. Technol.* **2013**, *47*, 4164–4171. [CrossRef]

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

18.   Fabregat-Palau, J.; Vidal, M.; Rigol, A. Modelling the sorption behaviour of perfluoroalkyl carboxylates and perfluoroalkane sulfonates in soils. *Sci. Total Environ.* **2021**, *801*, 149343. [CrossRef]

19.   Du, Z.; Deng, S.; Bei, Y.; Huang, Q.; Wang, B.; Huang, J.; Yu, G. Adsorption behavior and mechanism of perfluorinated compounds on various adsorbents—A review. *J. Hazard. Mater.* **2014**, *274*, 443–454. [CrossRef]

20.   Brusseau, M.L. Differential sorption of short-chain versus long-chain anionic per- and poly-fluoroalkyl substances by soils. *Environments* **2023**, *10*, 175. [CrossRef]

21.   Mei, W.; Sun, H.; Song, M.; Jiang, L.; Li, Y.; Lu, W.; Ying, G.-G.; Luo, C.; Zhang, G. Per- and polyfluoroalkyl substances (PFASs) in the soil–plant system: Sorption, root uptake, and translocation. *Environ. Inter.* **2021**, *156*, 106642. [CrossRef] [PubMed]

22.   Lesmeister, L.; Lange, F.T.; Breuer, J.; Biegel-Engler, A.; Giese, E.; Scheurer, M. Extending the knowledge about PFAS bioaccumulation factors for agricultural plants—A Review. *Sci. Total Environ.* **2021**, *766*, 142640. [CrossRef] [PubMed]

23.   Lewis, A.J.; Yun, X.; Spooner, D.E.; Kurz, M.J.; McKenzie, E.R.; Sales, C.M. Exposure pathways and bioaccumulation of per- and polyfluoroalkyl substances in freshwater aquatic ecosystems: Key considerations. *Sci. Total Environ.* **2022**, *822*, 153561. [CrossRef] [PubMed]

24.   Isik, M.; Bergazin, T.D.; Fox, T.; Rizzi, A.; Chodera, J.D.; Mobley, D.L. Assessing the accuracy of octanol–water partition coefficient predictions in the SAMPL6 Part II log P Challenge. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 335–370. [CrossRef] [PubMed]

25.   Bergazin, T.D.; Teilker, N.; Zhang, Y.; Mao, J.; Gunner, M.R.; Francisco, K.; Ballatore, C.; Kast, S.M.; Mobley, D.L. Evaluation of log P, pKa, and log D predictions from the SAMPL7 blind challenge. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 771–802. [CrossRef] [PubMed]

26.   Klamt, A. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235. [CrossRef]

27.   Goss, K.-U.; Bronner, G.; Harner, T.; Hertel, M.; Schmidt, T.C. The partition behavior of fluorotelomer alcohols and olefins. *Environ. Sci. Technol.* **2006**, *40*, 3572–3577. [CrossRef] [PubMed]

28.   Bhhatarai, B.; Gramatica, P. Prediction of aqueous solubility, vapor pressure and critical micelle concentration for aquatic partitioning of perfluorinated chemicals. *Environ. Sci. Technol.* **2011**, *45*, 8120–8128. [CrossRef]

29.   Kim, M.; Li, L.Y.; Grace, J.R.; Yue, C. Selecting reliable physicochemical properties of perfluoroalkyl and polyfluoroalkyl substances (PFASs) based on molecular descriptors. *Environ. Pollut.* **2015**, *196*, 462–472. [CrossRef]

30.   Brusseau, M.L. QSPR-based prediction of air-water interfacial adsorption coefficients for nonionic PFAS with large headgroups. *Chemosphere* **2023**, *340*, 139960. [CrossRef]

31. Jiang, L.; Xu, Y.; Zhang, X.; Xu, B.; Xu, X.; Ma, Y. Developing a QSPR model of organic carbon normalized sorption coefficients of perfluorinated and polyfluoroalkyl substances. *Molecules* **2022**, *27*, 5610. [CrossRef] [PubMed]

32. Brusseau, M.L. Estimating the relative magnitudes of PFAS adsorption to solid-water and air/oil-water interfaces. *Environ. Pollut.* **2019**, *254*, 113102. [CrossRef] [PubMed]

33. Coppola, G.; Polesello, S.; Ferrario, C.; Peruzzo, M.; Lava, R.; Mazzola, M. Does the partitioning of perfluoroalkyl acids to soil depend on their chain length? *Integ. Environ. Assess. Manag.* **2022**, *18*, 1747–1752. [CrossRef] [PubMed]

34. Lampic, A.; Parnis, J.M. Property estimation of per-and polyfluoroalkyl substances: A comparative assessment of estimation methods. *Environ. Toxicol. Chem.* **2020**, *39*, 775–786. [CrossRef] [PubMed]

35. Rayne, S.; Forest, K. Congener-specific organic carbon-normalized soil and sediment-water partitioning coefficients for the C1 through C8 perfluoroalkyl carboxylic and sulfonic acids. *J. Environ. Sci. Health A* **2009**, *44*, 1374–1387. [CrossRef]

36. Martin, J.W.; Marbury, S.A.; Solomon, K.R.; Muir, D.C.G. Bioconcentration and tissue distribution of perfluorinated acids in Rainbow trout (oncorhynchus mykiss). *Environ. Toxic. Chem.* **2003**, *22*, 196–204. [CrossRef]

37. Conder, J.M.; Hoke, R.A.; De Wolf, W.; Russell, M.H.; Buck, R.C. Are PFCAs bioaccumulative? A critical review and comparison with regulatory criteria and persistent lipophilic compounds. *Environ. Sci. Technol.* **2008**, *42*, 995–1003. [CrossRef] [PubMed]

38. Inoue, Y.; Hashizume, N.; Yakata, N.; Murakami, H.; Suzuki, Y.; Kikushima, E.; Otsuka, M. Unique physicochemical properties of perfluorinated compounds and their bioconcentration in common carp *Cyprinus carpio* L. *Arch. Environ. Contam. Toxicol.* **2012**, *62*, 672–680. [CrossRef]

39. Ng, C.A.; Hungerbuhler, K. Bioaccumulation of perfluorinated alkyl acids: Observations and models. *Environ. Sci. Technol.* **2014**, *48*, 4637–4648. [CrossRef]

40. Cheng, H.; Lv, C.; Li, J.; Wu, D.; Zhan, X.; Song YZhao, N.; Jin, H. Bioaccumulation and biomagnification of emerging poly- and perfluoroalkyl substances in marine organisms. *Sci. Total. Environ.* **2022**, *851*, 158117. [CrossRef]

41. Jeon, J.; Kannan, K.; Lim, H.K.; Moon, H.B.; Kim, S.D. Bioconcentration of perfluorinated compounds in blackrock fish, *Sebastes schlegeli*, at different salinity levels. *Environ. Toxic. Chem.* **2010**, *29*, 2529–2535. [CrossRef] [PubMed]

42. Blaine, A.C.; Rich, C.D.; Hundal, L.S.; Lau, C.; Mills, M.A.; Harris, K.M.; Higgins, C.P. Uptake of perfluoroalkyl acids into edible crops via land applied biosolids: Field and greenhouse studies. *Environ. Sci. Technol.* **2014**, *47*, 14062–14069. [CrossRef] [PubMed]

43. Arp, H.P.H.; Niederer, C.; Goss, K.-U. Predicting the partitioning behavior of various highly fluorinated compounds. *Environ. Sci. Technol.* **2006**, *40*, 7298–7304. [CrossRef] [PubMed]

44. Endo, S.; Hammer, J.; Matsuzawa, S. Experimental determination of air/water partition coefficients for 21 per- and polyfluoroalkyl substances reveals variable performance of property prediction models. *Environ. Sci. Technol.* **2023**, *57*, 8406–8413. [CrossRef] [PubMed]

45. Wang, Z.; MacLeod, M.; Cousins, I.T.; Scheringer, M.; Hungerbühler, K. Using COSMOtherm to predict physicochemical properties of poly- and perfluorinated alkyl substances (PFASs). *Environ. Chem.* **2011**, *8*, 389–398. [CrossRef]

46. Zhang, M.; Suuberg, E.M. Estimation of vapor pressures of perfluoroalkyl substances (PFAS) using COSMOtherm. *J. Hazard. Mat.* **2023**, *443*, 130185. [CrossRef] [PubMed]

47. Hammer, J.; Endo, S. Volatility and nonspecific van der Waals interaction properties of per- and polyfluoroalkyl substances (PFAS): Evaluation using hexadecane/air partition coefficients. *Environ. Sci. Technol.* **2022**, *56*, 15737–15745. [CrossRef] [PubMed]

48. SERDP. Summary Report: Strategic Workshop on Management of PFAS in the Environment. SERDP-ESTCP, November. 2022. Available online: https://www.nist.gov/publications/summary-report-strategic-workshop-management-pfas-environment (accessed on 12 February 2024).

49. Li, L.; Zhang, Z.; Men, Y.; Baskaran, S.; Sangion, A.; Wang, S.; Arnot, J.A.; Wania, F. Retrieval, selection, and evaluation of chemical property data for assessments of chemical emissions, fate, hazard, exposure, and risks. *ACS Environ. Au* **2022**, *2*, 376–395. [CrossRef] [PubMed]

50. Dearden, J.C.; Cronin, M.T.D.; Kaiser, K.L.E. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR QSAR Environ. Res.* **2009**, *20*, 241–266. [CrossRef]

51. Roy, K.; Kar, S.; Das, R.D. *A Primer on QSAR/QSPR Modeling Fundamental Concepts*; Springer: Berlin/Heidelberg, Germany, 2015.

52. Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2009.

53. Abusallout, I.; Holton, C.; Wang, J.; Hanigan, D. Henry's Law constants of 15 per- and polyfluoroalkyl substances determined by static headspace analysis. *J. Hazard. Mater. Lett.* **2022**, *3*, 100070. [CrossRef]

54. Alesio, J.L.; Slitt, A.; Bothun, G.D. Critical new insights into the binding of poly- and perfluoroalkyl substances (PFAS) to albumin protein. *Chemosphere* **2022**, *287*, 131979. [CrossRef] [PubMed]

55. Allendorf, F.; Berger, U.; Goss, K.-U.; Ulrich, N. Partition coefficients of four perfluoroalkyl acid alternatives between bovine serum albumin (BSA) and water in comparison to ten classical perfluoroalkyl acids. *Environ. Sci. Process. Impacts* **2019**, *21*, 1852–1863. [CrossRef]

56. Allendorf, F.; Goss, K.-U.; Ulrich, N. Estimating the Equilibrium Distribution of Perfluoroalkyl Acids and 4 of Their Alternatives in Mammals. *Environ. Toxic. Chem.* **2021**, *40*, 910–920. [CrossRef] [PubMed]

57. Droge, S.T.J. Membrane-water partition coefficients to aid risk assessment of perfluoroalkyl anions and alkyl sulfates. *Environ. Sci. Technol.* **2019**, *53*, 760–770. [CrossRef] [PubMed]

58. Ebert, A.; Allendorf, F.; Berger, U.; Goss, K.U.; Ulrich, N. Membrane/water partitioning and permeabilities of perfluoroalkyl acids and four of their alternatives and the effects on toxicokinetic behavior. *Environ. Sci. Technol.* **2020**, *54*, 5051–5061. [CrossRef] [PubMed]