ROYAL SOCIETY
OF CHEMISTRY

PAPER

Check for updates

Quantifying uncertainty in predicted chemical partition ratios required for chemical assessments

Trevor N. Brown, [ID] *[a] Alessandro Sangion,[a] Li Li [ID] [b] and Jon A. Arnot [ID] [acd]

Three Quantitative Structure Property Relationship (QSPR) software packages, IFSQSAR, OPERA, and EPI Suite are compared and assessed for prediction accuracy, applicability domain (AD) and uncertainty of the predictions. A database of experimental physical–chemical (PC) properties is compiled, merged, and filtered, and the QSPRs are assessed with datasets of octanol–water (K_{OW}), octanol–air (K_{OA}), and air–water (K_{AW}) partition ratios. Upper and lower limits on PC property predictions are proposed based on theory, data, and applications of the properties in hazard screening and risk assessment. Validations of the uncertainty metrics of the QSPR packages are done for the PC properties using experimental data external to all training datasets. The IFSQSAR 95% prediction interval (PI95) calculated from root mean squared error of prediction (RMSEP) captures 90% of the external data, while OPERA and EPI Suite require a factor increase of at least 4 and 2 respectively for their PI95 to capture a similar 90% of the external experimental data. The assessment of QSPR consensus predictions identified future research and experimental testing to improve the predictive models for data-poor chemicals such as polyfluorinated or per-fluorinated alkyl substances (PFAS), ionizable chemicals, and chemicals with complex and multifunctional structures.

Environmental significance

The findings of this work provide decision-makers with better tools to recognize and evaluate the uncertainty associated with physical–chemical (PC) properties when conducting chemical assessments. Reasonable upper and lower bounds on predicted PC properties have been proposed, and three PC property prediction packages have had their prediction uncertainty evaluated and refined against novel datasets. In addition, three major classes of data-poor chemicals have been confirmed as requiring more experimental, theoretical and modelling research: polyfluorinated or per-fluorinated alkyl substances (PFAS); ionizable organic chemicals (IOCs), especially strong acids and bases; and large complex chemicals with multiple heteroatom functional groups.

1 Introduction

Physical–chemical property data are fundamental to determining chemical emissions, fate and transport, hazard screening, exposure, and risk assessment as well as to pharmaceutical and veterinary sciences. Among the most common physical–chemical (PC) properties required for conducting legislated ecological and human health assessment for new and existing organic chemicals are molecular weight (MW; g mol^{-1}), water solubility (S_W; mol L^{-1}), vapor pressure (VP; Pa), the octanol–water (K_{OW}), octanol–air (K_{OA}), and air–water (K_{AW}) partition ratios,[1] and dissociation constants (*e.g.*, pK_a) for ionizable organic chemicals (IOCs). Partition ratios have

[a]ARC Arnot Research & Consulting, Toronto, Ontario, Canada. E-mail: trevor.n. brown@gmail.com

[b]School of Public Health, University of Nevada, Reno, Nevada, USA

[c]Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada

[d]Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada

volume units for the two different phases, *e.g.*, m^3-water/m^3-octanol for K_{OW}. The K_{AW} is Henry's Law Constant (H; Pa m^3 mol^{-1}) with units m^3-water/m^3-air converted as H/RT, where R is the ideal gas law constant (8.314 m^3 Pa K^{-1} mol^{-1}) and T is the system temperature (K).[2] This form is specifically named the Henry's volatility constant, and the unit conversion is made with the assumption of ideal gas behaviour, which is suitable for environmental temperature and pressure. Physical–chemical property data are critically important for environmental fate models and physiologically-based biokinetic (PBK) models and for designing and interpreting *in vivo* toxicity[3] and bioaccumulation[4] tests and *in vitro* bioassays (*i.e.*, new approach methods[5,6]). Model estimates for outdoor or indoor environmental fate and transport, toxicokinetics, toxicity, bioactivity, bioaccumulation, exposure, and risk can only be as reliable as the required model input parameters, *i.e.*, "garbage in = garbage out".[7,8] Uncertainty in PC data is inherent whether the data are measured or modelled[9,10] and, in many cases, chemical assessment outcomes are sensitive to the selected PC values, *e.g.*,[6,11–14] Mackay and colleagues pioneered efforts for collecting and critically evaluating experimental PC data and

This journal is © The Royal Society of Chemistry 2025

Environ. Sci.: Processes Impacts, 2025, **27**, 3457–3470 | 3457

providing these data in the form of handbooks.[15] Moreover, Mackay and others[1,16–19] developed methods based on thermodynamic theory to evaluate chemical properties like K_{OW}, K_{OA}, K_{AW}, VP, S_W and solubility in octanol (S_O) for reliability and consistency in a holistic manner using thermodynamic cycles (TC) and the three solubility approach.

It is not feasible to measure PC properties for the several thousand chemicals requiring evaluation and predictive methods are necessary.[20] Methods for predicting PC properties include *in silico* models such as Quantitative Structure–(Activity) Property Relationships (QS(A)PRs)[21–30] and quantum chemistry/ *ab initio*[31] methods, and empirical models such as Poly-Parameter Free Linear Energy Relationships (PPLFER).[32,33] QSPRs are specific to predicting chemical properties, whereas QSARs are more general and may include reactivity and toxicology end points. We use the term QSPR here but the guidance from various sources, which refer to QSARs, also applies. Organisation for Economic Co-operation and Development (OECD) guidance for QSAR development and validation for applications in regulatory decision-making[34,35] includes consideration of the applicability domain (AD) for a predicted property.[36] AD has been defined by experts as "*the response and chemical structure space in which the model makes predictions with a given reliability*",[37] and the OECD guidance document on validation of QSARs ("OECD QSAR principles"),[34] and the QSAR Assessment Framework (QAF)[36] have also adopted this definition. The AD and the reliability are intrinsically linked as implied by the quote, but in the five OECD QSAR principles they are listed as two different principles: "(3). A defined AD; and (4). Appropriate measures of goodness-of-fit, robustness, and predictivity". The QAF also acknowledges that AD and reliability are linked stating "*applicability domain informs the reliability of the prediction*" but again assesses them separately for convenience. In our previous work we have used the term uncertainty[23,25] defined as the inverse of reliability, *i.e.*, high reliability means low uncertainty, and low reliability means high uncertainty. In the QAF the term uncertainty has a broader meaning,[36] but it is used here only as the inverse of the term reliability. Previous related work examined the AD for some PC property QSPRs without specifically investigating the uncertainty,[38] and provided guidance on selecting and harmonizing measured or predicted values for chemical properties.[9] In this work we evaluate and compare both AD and the uncertainty of select QSPRs, especially in the context of data-poor chemicals.

QSPRs from different research groups frequently implement AD in different ways and many methods have been explored in the literature.[38–41] Our previous method development work implemented AD in IFSQSAR using chemical similarity, leverage (a distance metric related to the linear regression), a check based on atoms and bonds not found in the training data, and the range of experimental values in the training data,[25,42] and refined the uncertainty for partitioning properties.[25] The current work compares IFSQSAR Ver. 1.1.2[21–26] to two other QSPR software packages that provide predictions for many of the same properties: Estimation Programs Interface (EPI) Suite™ Ver. 4.11,[27,28] and OPEn (Quantitative) Structure–activity/property Relationship App (OPERA) Ver. 2.9.[29] EPI Suite

does not explicitly provide AD or uncertainty metrics in its outputs, but the documentation identifies chemical structures which are more prone to prediction uncertainty, and suggests simple AD checks by comparing the properties of chemicals to those in the training data. OPERA provides AD with its output which are based on similar methods as IFSQSAR,[29] and provides an expected prediction range as an uncertainty metric.

The primary objective of this study is to better understand and communicate the prediction uncertainty and ADs of the selected QSPR software packages for K_{OW}, K_{OA}, and K_{AW} of neutral organic chemicals and the neutral form of IOCs. This review provides guidance for selecting PC property data for chemical assessments and for integrated testing strategies to systematically address uncertainty in measured and predicted properties. There are chemical classes such as quaternary amines, surfactants, and chemicals with strong specific binding which are out of the AD of partitioning-based models, and these are out of the scope of this work. A general overview of the models selected is provided along with some methods for estimating the prediction uncertainty. Predictions from different K_{OW}, K_{OA}, and K_{AW} models are then compared with a large set of chemicals undergoing regulatory evaluation. A method for choosing which model outputs to include in consensus predictions is described and the predicted values are also compared to measured data which are external to the training datasets of the models selected for this study. Chemical classes and structural features for which uncertainty in the property predictions are large are identified and general recommendations are provided to address these uncertainties.

2 Methods

2.1 Theory

Partitioning and solubility experimental data are only measurable within a certain range of values determined by current methods and technologies available. Standardized Testing Guidelines for PC properties have been developed by the OECD, *e.g.*,[43–46] and a summary of the range of values for which testing methods are applicable is available.[9] Even high-quality measurements of PC property values will always have some amount of experimental uncertainty. Any experimental method which does not directly measure the partitioning in an octanol/ water system, such as high-performance liquid chromatography (HPLC) measurements of log K_{OW},[45] is a type of model which is being used to interpolate or extrapolate from direct measurements, such the slow-stirring method.[44] Even more pedantically, detectors such as mass spectrometers do not directly yield concentrations, a signal is measured and models are used to convert the signal to a concentration, so direct measurements are also based on models. PPLFERs are a type of model based on empirical relationships between partitioning and other experimental properties which correlate with molecular interactions. Relationships such as TC and the three solubility rule[16] are simple models applied in this work and some of the underlying data to predict values of K_{OW}, K_{OA}, and K_{AW} which would not be measurable using other methods. Section SI-1 outlines more of the theory and application of TC

3458 | *Environ. Sci.: Processes Impacts*, 2025, **27**, 3457–3470

This journal is © The Royal Society of Chemistry 2025

and the three solubility approach. QSPRs are a type of model and what distinguishes them from the experimentally based models is that the descriptors used are entirely theoretical and derived from representations of the chemical structure. Predictions with a model will always introduce uncertainty into PC properties, with interpolated values typically having less uncertainty than extrapolated values. One of the questions that motivates this work is how far is it reasonable and practical to extrapolate a model from its training data. Some QSPRs, such as EPI Suite and IFSQSAR, will very easily extrapolate far beyond the limits of the experimental data for PC properties. We explore various hypotheses for setting boundaries on partitioning and solubility predictions using different theory- and data-based methods and propose upper and lower boundaries for PC property predictions in terms of applications for chemical assessments. Some of these boundaries have solid theoretical foundations but others are more speculative and arbitrary, and in the interest of brevity they are organized as a "mini-study" in Section SI-2.

2.2 Chemical datasets

2.2.1 Chemical structure dataset. A dataset of about 85 000 discrete organic chemicals has been collected from various regulatory assessment databases on an *ad hoc* basis over the past 15 years. Chemical identities and structures were curated through a semi-automated process involving cross-referencing Chemical Abstract Service (CAS) Registration Number, chemical names, and molecular structures across multiple databases such as PubChem[47] and US EPA's CompTox Chemistry Dashboard[48] to identify and address inconsistencies and errors.[49] Standardized representations for a chemical are stored using canonical Simplified Molecular Input Line Entry System (SMILES) notation,[50,51] while InChIKeys are used for database indexing. We differentiate between isomeric structures which preserve stereochemistry and counterions, and parent structures which are derived by neutralizing the chemical, *i.e.*, removing counterions and stripping isomeric details. We refer to this dataset as the chemical structure dataset and use it in this work to evaluate how the three QSPR software packages perform on a large dataset of relevant chemical structures which are mostly data-poor. Predictions for the properties assessed in this work have been made with the each of the software packages using the parent structures, because all three QSPR software packages use only two-dimensional (atom connectivity) descriptors which neglect stereochemistry. This dataset and the predicted properties can be accessed in the Exposure And Safety Estimation (EAS-E) Suite online platform (**https://www.eas-e-suite.com**, database ver.1.0.1), and can be queried with name, CAS Registration Number, or SMILES.

2.2.2 Experimental property dataset. Experimental data were compiled for a subset of the chemical structure dataset, which is referred to here as the experimental property dataset. The experimental data are linked to the isomeric structures by InChIKey so that data for stereoisomers are kept separate, in comparison to the QSPR predictions which cannot differentiate between stereoisomers. A brief overview of the merging,

standardization, and filtering of experimental databases for K_{OW}, K_{AW}, K_{OA} follows, with extensive details provided in Section SI-3. Experimental databases for S_W, VP, and melting point (MP) are also compiled because they are used for setting limits on the three main partitioning properties (see Section SI-2) due to their relationships with K_{OW}, K_{OA}, and K_{AW} (see Section SI-1).

Most of the experimental K_{OW}, K_{AW}, K_{OA}, VP, S_W, and MP data originate from the PHYSPROP database[52] developed by the EPA Office of Pollution Prevention and Toxics (OPPT) with the Syracuse Research Corporation (SRC). The PHYSPROP data files were downloaded in 2016 and are no longer available on-line and the time stamps on the PHYSPROP files indicate they were last updated in 2008. To address more recent updates to the datasets, the EPI Suite internal experimental databases were searched in batch mode with CAS numbers from the original PHYSPROP datasets. Previous to the current curation efforts, the EPA Office of Research and Development (ORD) updated and curated the SRC datasets, seeking to ensure that chemical identity and chemical structure were correct.[53] The ORD version of the datasets were used to develop OPERA.[29] The OPERA ver.2.6 experimental datasets were downloaded from GitHub and merged with the updated PHYSPROP datasets forming the preliminary experimental property dataset. OPERA ver.2.9 experimental datasets were investigated; however, problems were identified that are difficult to resolve using automated processing. For example, OPERA ver.2.9 and the CompTox dashboard report some PC data as "experimental" when they are actually QSPR predictions, and some values reported as measurements are averages of multiple sources (including some predicted values), and citations to original literature that could be used to resolve these issues are sometimes missing. Further details on the merging of the PHYSPROP and OPERA 2.6 datasets are described in Section SI-3.

Three other high quality datasets were added to the preliminary experimental property dataset and merged with the chemical structure dataset. Any chemicals identified as salts, permanent ions, or inorganics were excluded. The Henry's Law Constant dataset of Sander[2] was incorporated as log K_{AW} after filtering for data that Sander flagged as reliable experimental values. When multiple values were available for a chemical, more recent measurements were selected over older measurements. The log K_{OA} dataset of Baskaran *et al.*[54] was filtered for experimental values measured for dry octanol between 20 and 30 °C, and any data they flagged as unreliable were removed. When more than one value was available the most reliable value, as ranked in the database, and the most recent value was selected. In both datasets when more than one experimental measurement was available for a chemical the arithmetic mean of the log-scale values was used. The Bradley MP dataset[55] was also added to the EAS-E Suite experimental property dataset. The full experimental dataset can be accessed in the EAS-E Suite online platform.

Finally, external validation datasets were defined by filtering the experimental property dataset to remove chemicals in the training datasets of the QSPR software packages assessed in this work. The OPERA QSPR package returns experimental values

This journal is © The Royal Society of Chemistry 2025

Environ. Sci.: Processes Impacts, 2025, **27**, 3457–3470 | **3459**

instead of QSPR predictions if a chemical is in its experimental database, so all chemicals identified in the OPERA 2.6 and 2.9 experimental databases were removed from consideration for external testing. The original SRC PHYSPROP database files typically identify chemicals in the EPI Suite training datasets, and these were also removed from consideration. Chemicals are also matched by CAS number with chemicals in the solute descriptor database used to develop the IFSQSARs,[24] and any chemicals in the training datasets of the QSPRs or PPLFERs were removed from consideration.

2.2.3 Supplementary log K_{OW} experimental property dataset. After removing all training data from the log K_{OW} dataset eight chemicals were left for external validation of the QSPRs, mostly because the OPERA 2.9 experimental database contains virtually all the publicly available data. Two additional log K_{OW} datasets which could be used for external validation were identified: Martel *et al.* 2013[56] and Tshepelevitsh *et al.* 2020;[57] however, there are limitations with both datasets. Both use regressions with HPLC retention time data rather than direct measurements, which is a model of partitioning properties and will introduce more uncertainty into the analysis as discussed in Section 2.1. The current OECD guidance recommends that the HPLC method only be used for log K_{OW} values between 0 and 6, while the Martel dataset reports values up to 7.5 and Tshepelevitsh reports values up to 21. Chemicals in the Martel dataset are large, complex, and frequently ionizable while their calibration chemicals are mostly neutral, and small or monofunctional.[58] Calibration chemicals in the Tshepelevitsh dataset are quite structurally similar to their test chemicals compared to the Martel dataset but only cover a log K_{OW} range from 1 to 8 meaning the large values reported require considerable extrapolation. These limitations should be kept in mind when interpreting the data, but we proceed with using the data for external validation in this work for four reasons. (1) While the data are uncertain both datasets make efforts to confirm the reasonableness of their measurements with model predictions. Martel *et al.* applied four log K_{OW} QSPRs and removed measurements which were inconsistent with any of the QSPRs.[56] Tshepelevitsh *et al.* used quantum chemical calculations to confirm the reasonableness of their measurements, though they noted a tendency for the calculations to over-predict the large values. (2) Both groups considered the pK_a of their test chemicals and adjusted the pH of their system to ensure they were measuring the neutral form. (3) Tshepelevitsh *et al.* used an enhanced calibration method which considered other molecular descriptors related to hydrogen bonding, polarity and size in addition to the retention time, making it more like a PPLFER. (4) Chemicals in both datasets are completely novel; they have not been used for any application and were likely not synthesized or measured anywhere else but in the reported works. This means that they are completely external to the training data of the three QSPR packages compared in this work, and while the data may be more uncertain than direct measurements of log K_{OW} all three packages will be similarly disadvantaged. One caveat to this is that KOWWIN from EPI Suite was one of the four QSPRs used by Martel *et al.* to confirm their measurements, so some data that

were very inconsistent with EPI Suite predictions may have already been removed from the dataset.

2.3 Physical–chemical property predictions with QSPRs

These software packages are briefly described here, and more details are available in the software user guides and original publications, and in Section SI-4. Notably, each software package has a different approach to making PC property predictions and assigning predictions an AD and uncertainty metrics.

2.3.1 Applicability domain and uncertainty. There are various methods for validating QSPR prediction uncertainty, such as having a second external validation dataset, developing multiple QSPRs with cross validation, and Bayesian analysis.[59,60] Uncertainty metrics are frequently too optimistic and under-estimate the deviation between predicted values and experimental values[59,61] and uncertainty metrics need to be validated in addition to validating the QSPRs. For example, in our previous work on partitioning properties <95% of an external dataset of experimental values fell within the 95% prediction interval (PI95),[25] requiring the uncertainty metric to be increased by a factor of at least 1.25. There are several reasons why uncertainty metrics may be too low or difficult to quantify, the most frequent reason is likely to be that the data available for external validation are too few and are not diverse enough to provide a realistic assessment of the uncertainty of the predictions. A QSPR may also be designed in a way that makes extrapolation outside of the range of training data impossible or the data available for external validation may be an "end-point mismatch", *e.g.,* a QSPR trained only on neutral chemicals would likely not be applicable to ionized chemicals because this would represent a different end-point such as with K_{OW} *vs.* octanol–water distribution ratio (D_{OW}).

2.3.2 QSPR software packages. IFSQSAR has been incrementally expanded and updated since 2012,[21] most recently the uncertainty metrics were validated for PC property prediction (IFSQSAR v1.1.0,[25]), and PC property predictions for per- and polyfluoroalkyl substances (PFAS) were improved with new data (IFSQSAR v1.1.1,[26]). Minor updates to the code made for this work will be released as IFSQSAR v1.1.2. IFSQSAR predictions for PC properties are based on the application of PPLFERs and fragment-based QSPRs which are described in detail elsewhere.[25,26,32,62] IFSQSAR quantifies AD using multiple methods and provides an uncertainty metric. Further details of IFSQSAR are provided in Section SI-4.

The EPI Suite software package (v4.11, Nov 2017)[27,28] was used to predict the properties assessed in this work. EPI Suite QSPR predictions lack explicit AD information, and the software only provides general recommendations in the documentation for determining if a chemical is in the AD of the QSPRs. This suggestion is time-consuming for the EPI Suite user and requires some expertise on structural fragments. To address this limitation, we developed an in-house method to explicitly determine the AD of EPI Suite QSPR predictions, provided training set data and model fragment information are available. We apply this method in the EAS-E Suite database and on-line

3460 | *Environ. Sci.: Processes Impacts*, 2025, **27**, 3457–3470

This journal is © The Royal Society of Chemistry 2025

platform providing AD information for EPI Suite predictions discussed in the present study. EPI Suite provides the point estimate for each endpoint, and we additionally used the root mean squared error of prediction (RMSEP) for the validation datasets from the EPI Suite documentation as an estimated uncertainty metric (for details, see Section SI-5). The following values for standard deviation of prediction from external validation datasets shown in the EPI Suite documentation are used as RMSEPs: $\log K_{OW}$: 0.479, $\log K_{AW}$ (bond method): 1.54, $\log K_{OA}$ (root mean squared sum of $\log K_{OW}$ and $\log K_{AW}$): 1.61, $\log S_W$ (WATERNT): 1.045, $\log VP$: 1.057.

The OPERA QSPRs[29] were developed on the same PHYSPROP datasets as the EPI Suite QSPRs, but with further curation of the datasets and chemical structures,[53] a different methodology, and external validation and AD definition adhering to OECD guidance.[34,35] A k-nearest neighbours model was developed for each PC property, where the predicted values are the weighted average of the $k = 5$ nearest neighbours. OPERA applies two complementary approaches for defining the AD for OPERA model predictions and provides an uncertainty metric.

2.4 Model validation, comparison, and consensus predictions

Predicted properties for the chemical structure dataset were compared against each other and to the experimental property dataset to determine the similarities and differences in the predicted properties. AD and uncertainty information for the model predictions were also considered in these analyses. Different chemical classes are identified and their relative abundance with regards to being in or out of the ADs of the models is quantified.

When multiple QSPR predictions are available for a single property the arithmetic mean of logarithmic values, referred to as the "consensus value", is recommended as a reasonable estimate to combine the battery of QSPR predictions for chemical assessments.[63–65] This approach assumes that QSPRs building on different algorithms would contain uncertainties or biases in different directions or aspects and that errant predictions can, therefore, be mitigated to a degree by predictions from other models.[64,66]

Consensus predictions have been calculated using the three PC property packages by taking the arithmetic mean of the partition coefficients or solubilities on the log scale.[67] The IFSQSAR and EPI Suite QSPRs are additive models and can extrapolate outside of their training data, but OPERA QSPR predictions are limited to the range of experimental training set data. Therefore, including the OPERA predictions in every case will bias consensus predictions towards the center of the experimental range which may not be desirable. In all cases the results from IFSQSAR and EPI Suite are included in the consensus value. After testing several approaches, it was decided not to include the OPERA predictions in the consensus values if the OPERA predictions are flagged as out of the AD. See Section SI-5 for more details.

The quantitative uncertainty metric applied in this work is the root mean squared error of prediction (RMSEP) which is an estimate of the prediction uncertainty. The RMSEP can be converted to a prediction interval which is a probabilistic metric. In this work we calculate prediction intervals at the 95% confidence level (PI95), and while this is a common choice the calculations could be made at any other confidence level. Consensus predictions are assigned quantitative uncertainty metrics by summing the RMSEP of the QSPRs that go into them according to summation of error rules. Another uncertainty metric associated with consensus predictions is the root mean squared deviation (RMSD) which shows the spread of the predictions in relation to the consensus. Equations and more details of these metrics are found in Section SI-5.

3 Results and discussion

3.1 Experimental property dataset

Table 1 summarizes the experimental values for K_{OW}, K_{AW}, K_{OA}, VP, and S_W and proposed lower and upper limits PC property predictions for applications in chemical assessments. The limits and percentiles in Table 1 are shown for subsets of the data based on the physical state of chemicals and the data source in Table S1. Fig. S2–S7 show measured and predicted K_{OW}, K_{AW}, K_{OA}, VP, and S_W (plus S_O) as a function of molecular weight (MW). In those figures, the QSPR results have been separated into predictions that are within the AD, and those that are out of the AD. More details and discussion of the upper and lower boundaries can be found in Section SI-2, but in brief upper limits on the solubilities are set based on dimensional limitations and lower limits are set based on extrapolation of what could be quantified with future experimental improvements and checked for reasonableness with available experimental and empirical data. The three solubility approach is

Table 1 Experimental property dataset and range and percentiles, and proposed QSPR prediction limits

Property	Experiment n	QSPR lower limit	Experiment minimum	2.5%	Median	97.5%	Experiment maximum	QSPR upper limit
Log K_{OW}	14 005	−6	−5.08	−1.3	2.03	6.36	11.29	19.3
Log K_{OA}	855	−3.2	−0.95	1.76	5.56	11.47	12.59	22.3
Log K_{AW}	2184	−22.4	−11.38	−6.7	−2.07	1.83	3.52	16.6
Log VP	2982	−14.6	−11.55	−7.4	0.7	5.28	7.79[a]	5.0[a]
Log S_W	5791	−18	−13.17	−8.18	−2.49	1	1.58[b]	1.4[b]

[a] The upper limit is set to atmospheric pressure, the experimental values that exceed this are for chemicals that are gases at standard conditions. [b] The upper limit is an assumed mole fraction of 0.5 for miscible solutes in water, the few experimental values which exceed this are from a 1990s USEPA database which is no longer accessible, so the reason could not be verified but might involve a different way of treating miscible solutes.

This journal is © The Royal Society of Chemistry 2025

Environ. Sci.: Processes Impacts, 2025, **27**, 3457–3470 | **3461**

then used to set upper and lower limits for the partition ratios. For chemicals within the upper and lower limits for all three of the partition ratios K_{OW}, K_{OA}, K_{AW}, it is acceptable to apply multi-media chemical fate and transport models directly using these values. However, for chemicals with predicted properties that are so extreme they fall outside of these limits we propose that the value of the limit should be used instead as input for environmental fate and exposure models. This is because, at such extremes, further increases or decreases in these partition ratios have minimal impact on the predicted environmental

fate and exposure. Instead, physiological and environmental factors, usage, and mode of emission are more important.

3.2 Model validation *vs.* external experimental data

After removing overlap with the training datasets of IFSQSAR and EPI Suite and removing chemicals in the OPERA experimental database there were 166 log K_{OA} and 128 log K_{AW} values available for external validation. The supplemental log K_{OW} dataset contains 754 log K_{OW} values. The very large log K_{OW} values in Fig. 1 (log K_{OW} > 10) should be treated with
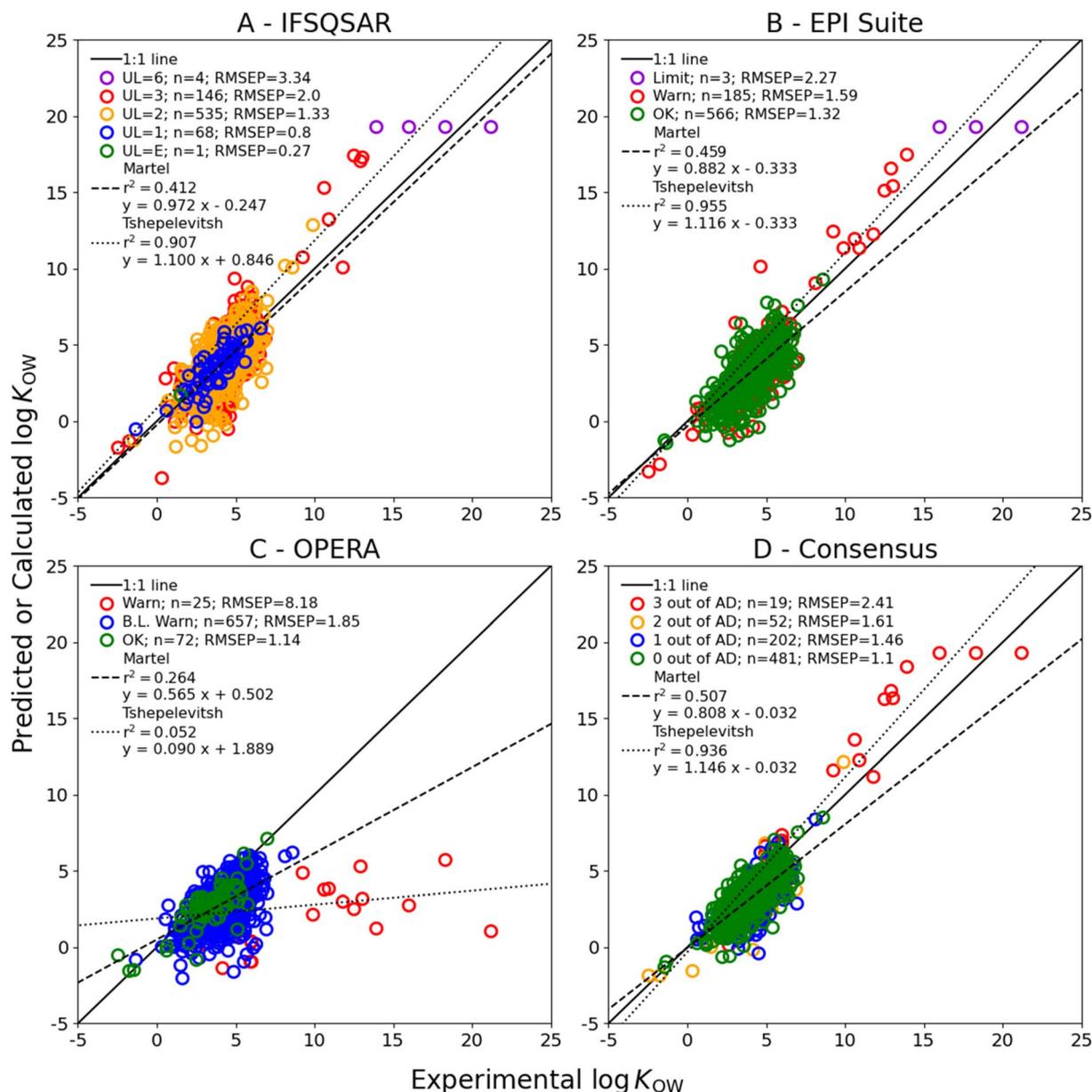

Fig. 1 Predicted or calculated *vs.* experimental values of log K_{OW} for the external validation dataset for (A) IFSQSAR, (B) EPI Suite, (C) OPERA, and (D) consensus values. Martel data[56] $n = 700$ span log K_{OW} 1 to 7.5 and Tshepelevitsh data[57] $n = 45$ span log K_{OW} −1 to 21. Root mean squared error of prediction (RMSEP) are shown for all data based on applicability domain (AD) and regression lines are shown separately for Martel (dashed) and Tshepelevitsh (dotted) data. Uncertainty Level (UL) corresponds to the AD checking of IFSQSAR with E, 0, 1, 2 considered in AD with increasing uncertainty, 3 is out of AD and 6 is a prediction limit violation. EPI Suite and OPERA AD groups are OK and Borderline in AD, "Warn" is out of AD, and Limit is a prediction limit violation.

3462 | *Environ. Sci.: Processes Impacts*, 2025, **27**, 3457–3470

This journal is © The Royal Society of Chemistry 2025

caution, these are the largest values ever reported. The uncertainty assigned to these by the authors[57] is large, with the largest values estimated to be ±3 log units. However, these are the only data available for testing what happens when the models are extrapolated far beyond the training data. IFSQSAR and EPI both have r^2 greater than 0.9 with these data as shown in Fig. 1A and B. Both models over-predict the values with slopes of 1.1, though IFSQSAR overpredicts more than EPI Suite with an intercept 0.5 log units greater, but the chemicals are all from one chemical class so the results may be different for other chemical classes. Fig. 1C shows that OPERA cannot make accurate predictions for these chemicals, because its "five nearest neighbors" algorithm predicts a chemical property as the weighted average of the experimentally determined data of the five most structurally similar chemicals in the training set, which prevents any predictions from exceeding the upper bound of the training set. The consensus predictions shown in Fig. 1D follow the method of removing any OPERA predictions that are flagged as out of AD ("Warning"), more discussion and comparison of the different methods for deciding which predictions to include in the consensus can be found in Section SI-5. Within the experimental range of log K_{OW} values where the OPERA predictions are included the consensus predictions are more accurate, *i.e.,* have lower RMSEP, than any of the individual QSPR packages, as shown in Table S3. Consensus predictions for the large log K_{OW} values from Tshepelevitsh *et al.* have an RMSEP that is between the predictions for IFSQSAR and EPI Suite, likely because a consensus of only two predictions is too few. Including predictions from more QSPRs that can extrapolate beyond the experimental range may improve the accuracy of these consensus predictions.

The external log K_{OA} data from Baskaran *et al.*[54] are mostly organo-halogens that are frequently within the AD of all three QSPR packages (Table S3). The RMSEP of predicted *vs.* experimental values is lowest for OPERA (0.533) and the RMSEP of the consensus predictions is comparable (0.547). The good correlation with the external experimental data ($R^2 = 0.966$) as shown in Fig. S11 is likely due to log K_{OA} being an easily predicted property, and that the data are for well-studied chemical classes within the AD of the QSPRs. There is a tendency for larger scatter above log K_{OA} of 6 because more of these are out of AD or only borderline within AD for IFSQSAR and EPI Suite. The external log K_{AW} dataset comes from the review of Sander 2023 (ref. 2) and includes more diverse chemical classes, which cover a much larger range of values (Fig. S12). The data are still mostly within the AD of the three QSPR packages, *e.g.*, IFSQSAR flagged only one chemical out of the AD. The consensus predictions for all chemicals in the external dataset are more accurate, with a lower RMSEP (1.403), than any of the individual QSPR package predictions, showing the benefit of using consensus predictions.

Table S3 also shows the external validation statistics broken down be chemical state, *i.e.*, gases or liquids, and solids for each of the three main partitioning properties. For log K_{OW} and log K_{OA} most of the chemicals are solids, but log K_{AW} has a nearly equal split between solids and non-solids. The accuracy of predictions for solids is poorer, with higher RMSEP than for

non-solids for all models and all PC properties. A likely explanation for this is that solids are more frequently out of the AD; for all three properties and for all three QSPR packages the solids always have a greater proportion of chemicals that are out of the AD than the non-solids. The AD information of IFSQSAR for the solids that are out of AD indicates egregious extrapolation from the training dataset. Solids tend to be larger and more complex than liquids and gases, they have more functional groups and more combinations of functional groups which pushes them out of the AD of group contribution QSPRs such as those in IFSQSAR and EPI Suite. OPERA predictions are based on a nearest-neighbours approach, so in this case solids are more frequently out of AD because of a lack of similar chemicals in the training data. Because solids are larger and more complex, they will cover a larger chemical space than gases and liquids, and so proportionally more data for solids is needed in the training data to fill in the chemical space and provide adequate nearest neighbours for solid chemicals.

For the EPI Suite and OPERA QSPR packages the accuracy of the uncertainty metrics was validated as was done for IFSQSAR.[25] In brief, the uncertainty metric is estimated as the RMSEP calculated on an external validation dataset, but this tends to underestimate the actual uncertainty, *i.e.*, more chemicals than expected are outside the bounds of the PI95 calculated from the RMSEP. A second external validation dataset is used to fit a scaling factor applied to the calculated RMSEP so that closer to 95% of chemicals are within the PI95. The RMSEP uncertainty metrics were estimated from the original EPI Suite and OPERA validation data as described in Section 2.3.2, and the fraction of chemicals in the external validation datasets from this work within the PI95 are shown as percentages in Table 2. This was done separately for chemicals flagged as in AD and out of AD, because by definition the uncertainty metrics cannot be assumed to be accurate for chemicals that are out of AD. The percentages are first calculated with the uncertainty metrics "as given" as shown in Table 2. The RMSEPs of all IFSQSAR partitioning and solubility QSPRs were scaled by a factor of 1.25 in previous work to make the PI95s capture 95% of the experimental data used in that work, and in this work the fraction is 90% of chemicals in AD and 96% of chemicals out of AD, no further adjustments were made. Less than 95% of chemicals are captured in the IFSQSAR PI95 for log K_{OW} (92%), but the fraction of log K_{OA} and log K_{AW} values captured are even lower. The fraction of log K_{OW} values within the EPI Suite PI95 is much lower than for the other two partition ratios, but the RMSEP is only 0.479 compared to 1.54 and 1.61 for the other partition ratios. The fraction of chemicals in the OPERA PI95 varies from 0.36 to 0.49 and the log K_{OW} PI95 does capture the lowest fraction. These results do not give strong evidence that the experimental log K_{OW} values are more uncertain than the log K_{OA} and log K_{AW} values. However, for log K_{OA} and log K_{AW} there are fewer data and the chemicals are not as diverse as the chemicals with measured log K_{OW}, so the statistics should be treated with some caution. The RMSEP of consensus predictions are calculated using propagation of uncertainty, using the simple assumption of no collinearity. Consensus predictions are only considered to be in AD if all

This journal is © The Royal Society of Chemistry 2025

Environ. Sci.: Processes Impacts, 2025, **27**, 3457–3470 | **3463**

Table 2 Fraction of predictions (%) that are within the prediction interval at the 95% confidence level (PI95) of each QSPR package for each PC property with different scaling factors

Model	Property	% In PI in AD as given	% In PI out AD as given	% In PI in AD adjusted	% In PI out AD adjusted[c]	% In PI out AD readjusted[d]
IFSQSAR	Log K_{OW}	92	96	92	96	96
	Log K_{OA}	88	0	88	0	0
	Log K_{AW}	81	100	81	100	100
	All	90[a]	96	**90**	96	96
EPI Suite	Log K_{OW}	57	47	85	78	90
	Log K_{OA}	100	100	100	100	100
	Log K_{AW}	96	78	100	100	100
	All	69	57	**90**	83	93
OPERA	Log K_{OW}	36	4	91	68	88
	Log K_{OA}	49	100	93	100	100
	Log K_{AW}	44	55	84	95	100
	All	40	29	**90**	81	94
Consensus	Log K_{OW}	59	53	85	78	82
	Log K_{OA}	97	82	100	100	100
	Log K_{AW}	78	70	95	84	90
	All	69	58	89	81	84
IFSQSAR	Uncertainty factor increase[b]	1.25	1.25	1.25	1.25	1.25
EPI suite	Uncertainty factor increase	1	1	2	2	3
OPERA	Uncertainty factor increase	1	1	4	4	6

[a] EPI Suite and OPERA RMSEP are scaled so that their % in PI matches this value, see bolded values. [b] Scaling factor from previous work,[25] no further adjustments in this work. [c] % In PI for out of AD predictions when applying the same scaling factor as in AD predictions. [d] % In PI for out of AD predictions when applying an additional 1.5 scaling factor.

three QSPR packages flag a chemical as in their AD. For EPI Suite and OPERA scaling factors were fitted to make the respective PI95s capture the same percentage of all the experimental values where the predictions were flagged as in AD by IFSQSAR, *i.e.*, 90%. Scaling to reach 90% instead of 95% ensures that the scaling factors are not unduly influenced by any uncertainty in the experimental data. The required scaling factors are shown at the bottom of Table 2. An additional scaling factor of 1.5 was applied only to the out of AD predictions to bring about 95% of chemicals with experimental data within the PI95s of EPI Suite and OPERA, though as previously stated uncertainty for out of AD predictions cannot be assumed to be accurate. The factor 4 increase in RMSEP for OPERA is quite large, it may be the PI of OPERA should be interpreted as $\pm$ RMSEP rather than a PI95, in which case the factor increases for both EPI Suite and OPERA would be about 2.

3.3 Model predictions for data-poor chemicals

Fig. 2 shows log K_{OW} predictions for data-poor chemicals, *i.e.*, 85 000 chemicals in the chemical structure dataset, after removing all chemicals with log K_{OW} values in the experimental property dataset. The three QSPR packages are plotted *vs.* each other, with the range of experimental values and the prediction limits set in Section SI-2, presented in Table 1, shown in black and red boxes. The chemicals plotted are all the neutralized, de-salted chemicals in the chemical structure dataset. The same types of plots for log K_{OA} and log K_{AW} are shown in Fig. S13 and S14 in the SI. The limits of the models are clear in these plots; OPERA predictions go outside of the range of experimental values in very few cases likely because of

specific data points in the OPERA internal database that were excluded from the current work for various reasons, *e.g.*, suspect data points from the expanded OPERA 2.9 database. All the EPI Suite predictions outside of the range of experimental values are flagged as out of AD. IFSQSAR and EPI Suite are correlated over the whole range of values for each property, with the largest scatter in the range of experimental values where most predicted values lay, generally clustered around the 1 : 1 line. At the very upper ranges there is a bias, especially obvious for log K_{AW}, as shown in Fig. S14 where the correlation deviates from the 1 : 1 line. When IFSQSAR and EPI Suite predictions are outside of the range of experimental values, the OPERA predictions tend to stay away from the upper limit of the experimental range. OPERA predictions may be in or out of the AD when the IFSQSAR and EPI Suite predictions are outside of the range of experimental values.

The OPERA and IFSQSAR predictions are both within their respective AD for 60% or more of the chemicals for all three major PC properties, but EPI Suite and IFSQSAR and EPI Suite and OPERA are only both in their AD for less than half the chemicals for every property except for log K_{OW} where the agreement is close to 60%. Agreement between models is best for log K_{OW}, with an average 55.2% and 75.8% of pairs of model predictions agreeing within 1 and 2 log units, respectively. The agreement for log K_{OA} is an average 40.5% and 62.2%, and for log K_{AW} is an average 34.6% and 55.4% within 1 and 2 log units. However, the deviations between model predictions are commonly very large. For example, comparing log K_{AW} predictions between IFSQSAR and OPERA 25.4% of predictions differ by greater than 5 log units. Much of this can
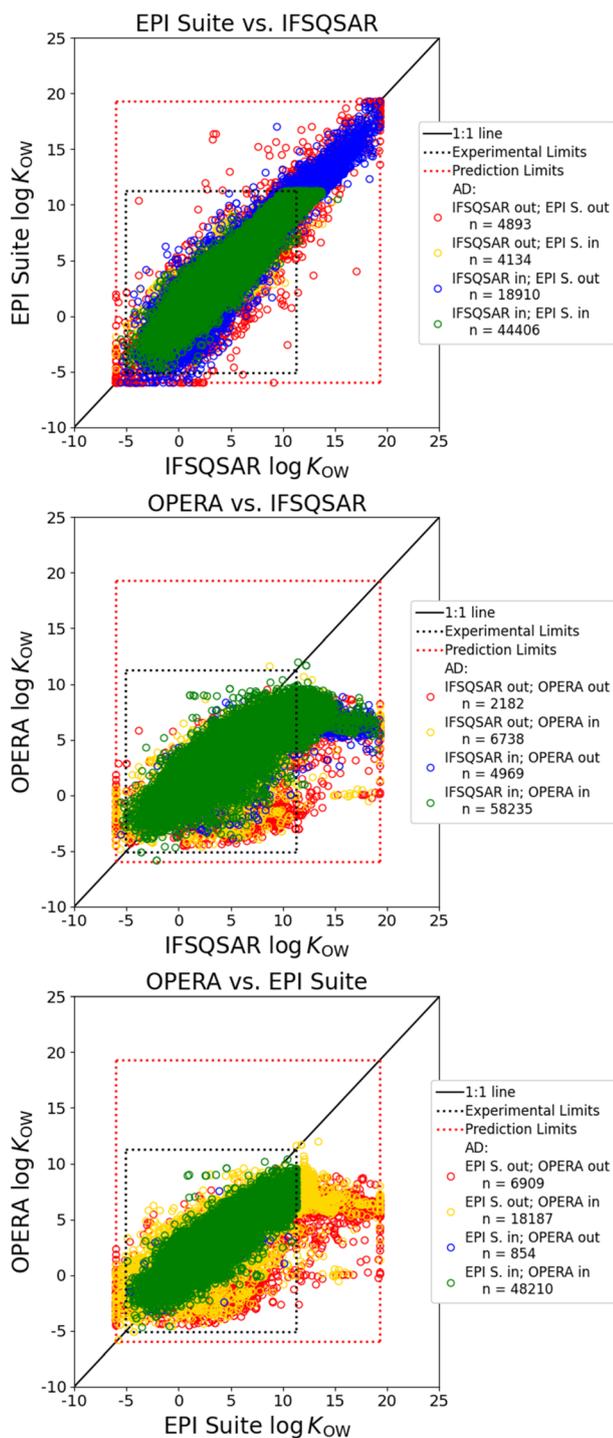
3464 | *Environ. Sci.: Processes Impacts*, 2025, **27**, 3457–3470

This journal is © The Royal Society of Chemistry 2025

Fig. 2 Binary model comparison of log K_{OW} predictions from IFSQSAR, EPI Suite, and OPERA for the chemical structure dataset.

be attributed to chemicals with log K_{AW} values that are out of range of experimental data, but even when the comparison is restricted to cases when both IFSQSAR and OPERA are in their AD (12.6% of predictions), more than 7000 chemicals have predictions that differ by greater than 5 log units. Across all model and partitioning property comparisons 14.2% have a deviation greater than 5 log units, and 3.7% have a deviation

greater than 5 log units when only considering cases where both models are in their respective AD.

Few of the chemicals have been capped at the prediction limits set in Section SI-2, only about 0.3% of the log K_{OW} predictions from IFSQSAR and EPI Suite were capped at the upper or lower prediction limits. More chemicals were capped at the upper or lower limit for log K_{OA} (6.5%) and log K_{AW} (4%), but the fractions were still small. In these cases, the higher predictions of the QSPRs have been replaced with the value of the prediction limit. The chemicals that have been capped typically fall into one of the classes identified in Section SI-2 from the minimum and maximum values of log VP, log S_W, and log S_O such as PFAS, waxy alkanes or fatty acid esters, and complex chemicals with multiple heteroatom functional groups.

Two methods were used to investigate chemicals to identify poorly represented chemical classes and help prioritize future experimental work. First, chemicals that are out of the AD of all three QSPR packages were compared to the chemicals within all three AD. Second, chemicals with consensus RMSD values in the highest 75th percentile were compared to chemicals with RMSD in the lowest 25th percentile. Most of the chemicals identified as being out of all three ADs or as having RMSD above the 75th percentile have consensus values outside of the experimental limits. These typically belong to one of the chemical classes identified in Section 3.1 as having the lowest VP, S_W, and S_O, and are considered experimentally inaccessible using current methods. Instead, this analysis was restricted to chemicals within the experimental limits to investigate which chemicals are poorly represented and have the greatest uncertainty, but which should still be experimentally accessible. The RMSD shows bias towards larger values near the experimental limits, so only chemicals where the IFSQSAR and EPI Suite predictions were at least ±0.674 times the RMSEP (corresponding to a 75% PI) from the upper or lower limit were included.

Next, we seek to better understand which types of chemicals are more likely to fall within or outside the ADs. For this we use solute descriptors (which correlate with molecular interactions) and molar mass to characterize the chemicals. The solute descriptors for chemicals that are in the AD of all three QSPR packages or out of the AD of all three QSPR packages are plotted *vs.* MW for each of the PC properties in Fig. S18–S20. The same plots for chemicals in the lowest and highest 25th percentiles are shown in Fig. S21–S23. An obvious feature in these plots is a group of chemicals that are out of AD or have high RMSD and have L and V, and to a lesser extent S, solute descriptors that follow a distinctly lower trend extending outside the space covered by chemicals that are in the AD or have low RMSD. These chemicals are PFAS which have unique molecular interactions compared to other chemical classes. Recent work has improved the AD of IFSQSAR with regards to this chemical class,[26] but the amount of data available is still small compared to data for other chemical classes meaning many of these chemicals are still out of the AD and have higher uncertainty, especially those with MW greater than 600. This MW range also corresponds to the PFAS with anomalously low S_O shown in

This journal is © The Royal Society of Chemistry 2025

Environ. Sci.: Processes Impacts, 2025, **27**, 3457–3470 | **3465**

Table 3 Fraction (%) of chemicals containing each chemical class in the subsets of the chemical structure dataset that are out of the AD of none (0) or all (3) of the QSPR packages or are in the lowest 25th or highest 75th percentile of consensus root mean squared deviation (RMSD)

Property	Class	% 0 out of AD	% 3 out of AD	% <25 perc. RMSD	% >75 perc. RMSD
Log K_{OW}	Fluorine	11	15.7	13.4	21.7
	Other halogens	20.1	23.5	20.6	22.2
	Heteroatoms	93.1	99.2	88.3	97.0
	Acids	9.2	17.7	8.8	13.9
Log K_{OA}	Fluorine	13.1	0^a	12.0	20.7
	Other halogens	20	0^a	25.7	14.7
	Heteroatoms	89.1	100^a	82.1	94.7
	Acids	6.4	16.7^a	3.1	11.1
Log K_{AW}	Fluorine	5.4	60.8	7.9	15.6
	Other halogens	21.7	7.8	21.6	15.5
	Heteroatoms	90	98.7	85.3	99.0
	Acids	9.9	10.5	7.7	8.0

^{a} There are 6 chemicals out of all 3 AD for log K_{OA} so these numbers are likely not meaningful.

Fig. S10, so that result may also be due to problems with the AD. The AD plot for log K_{OA} in Fig. S19 shows that scarcely any chemicals in the chemical structure dataset are out of the AD of all three QSPR packages, and Fig. S22 shows that, other than PFAS, chemicals with higher consensus RMSD do not have very different molecular interactions than those with lower RMSD. Overall, despite its smaller training dataset, log K_{OA} predictions are more within the AD of the QSPRs, and the QSPRs make more consistent predictions than for log K_{OW} or log K_{AW}.

All the solute descriptors for the chemicals that are out of AD or have high RMSD tend to be larger, with a much higher MW range, and for non PFAS also higher L and V solute descriptors, than chemicals that are in AD or have low RMSD. The S and B solute descriptors correlating with polar interactions and hydrogen bond acceptor strength show extrapolation to higher values meaning that the chemicals are more complex likely containing more heteroatom functional groups. The solute descriptor that correlates with hydrogen bond donor strength (A) shows a different trend than the other solute descriptors, the chemicals that are out of AD or have high RMSD do not tend to have higher A values than those that are in AD or have low RMSD. This may mean that the A solute descriptor is consistently being under-estimated by IFSQSAR for these chemicals. There are some hydrogen bond-donor functional groups that are not represented in the training data, namely the neutral forms of strong acids, because the hydrogen bond donor strength of these chemicals in their neutral form is experimentally inaccessible. As stated in Section 2.2 this comparison for data-poor chemicals only uses the neutral form of chemicals, and the chemical structures in the chemical structure dataset were de-salted and neutralized.

The results from inspecting the solute descriptors were confirmed by inspecting the atoms and functional groups present in the chemicals that are out of AD of the QSPR packages or have RMSD in the highest 75th percentile. First, all atoms in the typical organic subset (C, N, O, Si, P, S, F, Cl, Br, I) were counted in all chemicals in the chemical structure dataset, and then the number of chemicals containing at least one of

each atom type were counted for subsets defined by AD and RMSD groupings. This was done for the three partitioning properties, and trends in the occurrence of each atom type were inspected. Atom types with comparable trends were combined, and some more specific functional groups were also inspected to see if they could better explain the observed trends, the results of this are shown in Table 3. Chemicals containing fluorine are enriched in the subsets of chemicals that are out of AD or have high RMSD, whereas the other halogen atoms either show the opposite trend or no trend. Note that chemicals containing a fluorine are not synonymous with PFAS, but most of the chemicals containing fluorine in the chemical structure dataset are PFAS. Chlorinated and brominated chemicals are well-studied and are well represented in the training data of the different QSPR packages. Iodinated chemicals are less well-represented but in general their PC properties follow similar mechanisms as the chlorinated and brominated chemicals. The heteroatoms N, O, P and S are also enriched in chemicals that are out of AD or have high RMSD. Likewise, Zhang et al.[38] also found chemicals with N, S, and P are more likely to fall outside of the ADs of QSPRs investigated in their work. For log K_{OW} and log K_{OA} more than half of the enrichment of heteroatoms can be explained by the presence of just three strong acid groups: carboxylic, sulfuric, and phosphoric acids.

4 Conclusions and recommendations

The analysis provides some general guidance for the application of QSPR predictions for K_{OW}, K_{OA} and K_{AW} and highlights that in some cases differences in predictions from commonly used software packages can be very large. These results provide important considerations when using QSPR predictions of these properties to inform chemical evaluations. The upper and lower prediction limits, and the classes of chemicals that are predicted to be outside of these limits, inform the limitations of modeling chemical fate and exposure using partitioning-based models. For example, triglycerides are too water insoluble to be distributed in the human body by partitioning so applying

3466 | Environ. Sci.: Processes Impacts, 2025, **27**, 3457–3470

This journal is © The Royal Society of Chemistry 2025

partitioning-based HTTK models to describe their internal distribution would not be valid. Despite their insolubility triglycerides are transported throughout the body by specialized protein-lipid aggregates called lipoproteins,[68] but modelling their distribution would require different non-partitioning physiological models.

Each of the three QSPR packages assessed in this work has merits, and the pre-calculated predictions and corresponding AD as well as consensus values with uncertainty estimates can be accessed in the EAS-E Suite online platform. Each of the packages also has limitations that should be kept in mind when interpreting their results. IFSQSAR PC properties are based on PPLFER equations which have a mechanistic basis correlated to fundamental molecular interactions, this has been exploited in this work to identify chemical classes and functional groups related to extreme property values. IFSQSAR has shown good predictive power for data-poor chemicals classes, *e.g.*, PFAS,[26] and has robust AD and uncertainty estimates.[25] The main limitation of IFSQSAR is that the PPLFER basis means that the predictions for PC properties are an aggregate of four different QSPRs for the solute descriptors, and the AD and uncertainty therefore are also an aggregate. Despite this, it was found that the uncertainty metrics still underestimated the prediction uncertainty by a factor of at least 1.25 when applied to external data. In contrast, the uncertainty metrics of EPI Suite and OPERA underestimated the prediction uncertainty by factors of at least 2 and 4 respectively.

The main merit of EPI Suite is that its QSPR for log K_{OW} has the best predictive power for many of the cases investigated here, and in previous work.[26] The EPI Suite QSPRs for other properties have significantly poorer predictive power, and the definition of AD and uncertainty metrics have been added *post hoc* or are absent entirely. The OPERA QSPRs have good predictive power within their AD, their AD is well defined, and uncertainty metrics are also supplied. The main limitation of OPERA is that its predictive power decreases precipitously when applied to chemicals that are out of its AD. This review shows the ADs provided by OPERA QSPRs do a good job of identifying the cases where the predictions can be expected to have egregious errors due to problems with extrapolation outside of the range of experimental values and structures in its training sets. Based on the current analysis, OPERA values were excluded from consensus predictions with IFSQSAR and EPI Suite only when OPERA predictions are out of their AD. The resulting consensus values showed better predictive power than any of the individual models across the whole range of experimental values.

By comparing the AD and uncertainty metrics of the three QSPR packages three broad chemical classes have been identified as requiring more research. PFAS are a major class of chemicals that require more research, as is well known and identified by other work.[26,69–72] IFSQSAR has improved predictive power after including more partitioning data for PFAS, but there are still too few data compared to other chemical classes. Part of the problem is that PFAS as a class are so diverse, for example some are identified as both the most and least soluble chemicals in octanol. Based on the results in this work, heavy (>600 MW) non-polar PFAS may not be well modelled by partitioning-based

models. Acids and bases, and partitioning of ions in general are also an obvious research need. The partitioning properties of strong acids and bases in their neutral form are, and will remain, experimentally inaccessible, so all QSPRs lack data to calibrate predictions for these chemicals. This limitation will likely only be resolved by studying ion partitioning in general and its relation to partitioning of neutral chemicals. The final class of chemicals identified are large complex chemicals with many heteroatom functional groups. The strong acids and bases are a sub-category of these complex chemicals, and many of the heteroatom functional groups are weak acids or bases so many chemicals in this group have the same research needs. Because of the abundance of polar and H-bonding functional groups and their large size, the chemicals in this class are virtually all solids. Predictions for the partitioning and solubility of solids was found to be more uncertain in previous work,[25] but this is likely to be a simple case of interpolation being more accurate than extrapolation. Large complex structures are more likely to be out of the AD due a lack of similar chemicals in the training dataset, and therefore more uncertain. Increasing the accuracy of predictions for this chemical class will be difficult, because the structures are very diverse. Making measurements for even more complex chemicals might pull this chemical class further within the AD, but this strategy is intractable because the measurements would be even more difficult. A systemic, representative sampling of the known chemical space may be the best approach available, similar to what was done by Martel *et al.*.[56] All three of these chemical classes require more experimental data, but theoretical research and model calculations are also required to advance the science, guide testing strategies, and interpret experimental results.

Author contributions

Trevor N. Brown: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft, writing – review & editing Alessandro Sangion: data curation, visualization, writing – review & editing Li: conceptualization, writing – review & editing Jon A. Arnot: conceptualization, funding acquisition, project administration, supervision, writing – review & editing.

Conflicts of interest

Data availability

The chemical structure dataset and the experimental property dataset can be accessed on the EAS-E Suite platform at **https://www.eas-e-suite.com**. IFSQSAR code developed in this work

This journal is © The Royal Society of Chemistry 2025

Environ. Sci.: Processes Impacts, 2025, **27**, 3457–3470 | **3467**

can be run on the EAS-E Suite platform and precalculated values are available for all chemicals in the dataset.

Supplementary information which provides more details on the data and methods is available. See DOI: **https://doi.org/10.1039/d5em00357a**.

Acknowledgements

References

1 D. Mackay, A. K. Celsie and J. M. Parnis, The evolution and future of environmental partition coefficients, *Environ. Rev.*, 2016, **24**, 101–113.

2 R. Sander, Compilation of Henry's law constants (version 5.0.0) for water as solvent, *Atmos. Chem. Phys.*, 2023, **23**, 10901–12440.

3 OECD, *Test No. 203: Fish,* Acute Toxicity Test, 2019.

4 OECD, Test No. 305: Bioaccumulation in Fish: Aqueous and Dietary Exposure, Organisation for Economic Cooperation and Development, *OECD iLibrary*, 2012.

5 OECD, *Guidance Document on Good In Vitro Method Practices (GIVIMP)*, Report 286, Organisation for Economic Co-operation and Development (OECD), 2018.

6 J. M. Armitage, F. Wania and J. A. Arnot, Application of mass balance models and the chemical activity concept to facilitate the use of *in vitro* toxicity data for risk assessment, *Environ. Sci. Technol.*, 2014, **48**, 9770–9779.

7 A. M. Buser, M. MacLeod, M. Scheringer, D. Mackay, M. Bonnell, M. H. Russell, J. V. DePinto and K. Hungerbuhler, Good modeling practice guidelines for applying multimedia models in chemical assessments, *Integr. Environ. Assess. Manage.*, 2012, **8**, 703–708.

8 OECD, *Guidance Document on the Characterisation, Validation and Reporting of PBK Models for Regulatory Purposes*, Organisation for Economic Co-operation and Development, Paris, FR, 2021.

9 L. Li, Z. Zhang, Y. Men, S. Baskaran, A. Sangion, S. Wang, J. A. Arnot and F. Wania, 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.

10 J. Pontolillo and R. P. Eganhouse, *The search for reliable aqueous solubility (S$_W$) and octanol–water partition coefficient (K$_{OW}$) data for hydrophobic organic compounds: DDT and DDE as a case study*, U.S. Geological Survey, 2001, DOI: **10.3133/wri014201**.

11 F. Wegmann, L. Cavin, M. MacLeod, M. Scheringer and K. Hungerbühler, The OECD software tool for screening chemicals for persistence and long-range transport potential, *Environ. Model. Softw.*, 2009, **24**, 228–237.

12 T. Meyer, F. Wania and K. Breivik, Illustrating sensitivity and uncertainty in environmental fate models using partitioning maps, *Environ. Sci. Technol.*, 2005, **39**, 3186–3196.

13 S. Baskaran and F. Wania, Applications of the octanol–air partitioning ratio: a critical review, *Environ. Sci.: Atmos.*, 2023, **3**, 1045–1065.

14 F. Wania, Y. D. Lei, S. Baskaran and A. Sangion, Identifying organic chemicals not subject to bioaccumulation in air-breathing organisms using predicted partitioning and biotransformation properties, *Integr. Environ. Assess. Manage.*, 2022, **18**, 1297–1312.

15 D. Mackay, W. Y. Shiu, K. C. Ma and S. C. Lee, *Handbook of Physical–Chemical Properties and Environmental Fate for Organic Chemicals, Second Edition*, CRC Press, 2006, vol. 1–4.

16 J. G. Cole and D. Mackay, Correlating environmental partitioning properties of organic compounds: The three solubility approach, *Environ. Toxicol. Chem.*, 2000, **19**, 265–270.

17 R. P. Schwarzenbach, P. M. Gschwend and B. M. Imboden, *Environmental Organic Chemistry*, John Wiley & Sons, 2016.

18 U. Schenker, M. MacLeod, M. Scheringer and K. Hungerbühler, Improving data quality for environmental fate models: a least-squares adjustment procedure for harmonizing physicochemical properties of organic compounds, *Environ. Sci. Technol.*, 2005, **39**, 8434–8441.

19 A. Beyer, F. Wania, T. Gouin, D. Mackay and M. Matthies, Selecting internally consistent physicochemical properties of organic compounds, *Environ. Toxicol. Chem.*, 2002, **21**, 941–953.

20 J. A. Arnot, T. N. Brown, F. Wania, K. Breivik and M. S. McLachlan, Prioritizing chemicals and data requirements for screening-level exposure and risk assessment, *Environ. Health Perspect.*, 2012, **120**, 1565–1570.

21 T. N. Brown, J. A. Arnot and F. Wania, Iterative fragment selection: a group contribution approach to predicting fish biotransformation half-lives, *Environ. Sci. Technol.*, 2012, **46**, 8253–8260.

22 J. A. Arnot, T. N. Brown and F. Wania, Estimating screening-level organic chemical half-lives in humans, *Environ. Sci. Technol.*, 2014, **48**, 723–730.

23 T. N. Brown, J. M. Armitage and J. A. Arnot, Application of an Iterative Fragment Selection (IFS) method to estimate entropies of fusion and melting points of organic chemicals, *Mol. Inform.*, 2019, **38**, e1800160.

24 T. N. Brown, QSPRs for predicting equilibrium partitioning in solvent–air systems from the chemical structures of solutes and solvents, *J. Solution Chem.*, 2022, **51**, 1101–1132.

25 T. N. Brown, A. Sangion and J. A. Arnot, Identifying uncertainty in physical-chemical property estimation with IFSQSAR, *J. Cheminf.*, 2024, **16**, 65.

26 T. N. Brown, J. M. Armitage, A. Sangion and J. A. Arnot, Improved prediction of PFAS partitioning with PPLFERs and QSPRs, *Environ. Sci.: Process. Impacts*, 2024, **26**, 1986–1998.

3468 | *Environ. Sci.: Processes Impacts*, 2025, **27**, 3457–3470

This journal is © The Royal Society of Chemistry 2025

27 W. M. Meylan and P. H. Howard, Atom/fragment contribution method for estimating octanol-water partition coefficients, *J. Pharm. Sci.*, 1995, **84**, 83–92.

28 W. M. Meylan, P. H. Howard and R. S. Boethling, Improved method for estimating water solubility from octanol/water partition coefficient, *Environ. Toxicol. Chem.*, 1996, **15**, 100–106.

29 K. Mansouri, C. M. Grulke, R. S. Judson and A. J. Williams, OPERA models for predicting physicochemical properties and environmental fate endpoints, *J. Cheminf.*, 2018, **10**, 1–19.

30 C. Nieto-Draghi, G. Fayet, B. Creton, X. Rozanska, P. Rotureau, J.-C. de Hemptinne, P. Ungerer, B. Rousseau and C. Adamo, A general guidebook for the theoretical prediction of physicochemical properties of chemicals for regulatory purposes, *Chem. Rev.*, 2015, **115**, 13093–13164.

31 F. Eckert and A. Klamt, *COSMOThermX, COSMOlogic GmbH & Co. KG*, Leverkusen, Germany.

32 M. H. Abraham, Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes, *Chem. Soc. Rev.*, 1993, **22**, 73.

33 K.-U. Goss, Predicting the equilibrium partitioning of organic compounds using just one linear solvation energy relationship (LSER), *Fluid Phase Equilib.*, 2005, **233**, 19–22.

34 OECD, *Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships (QSAR) Models*, Organisation for Economic Cooperation and Development, Environment Directorate, Paris, 2007.

35 OECD, *OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure–Activity Relationship Models*, OECD, Paris, 2004.

36 OECD, *(Q)SAR Assessment Framework: Guidance for the Regulatory Assessment of (Quantitative) Structure–Activity Relationship Models, Predictions, and Results Based on Multiple Predictions*, Report ENV/CBC/MONO(2023)32, Organisation for Economic Cooperation and Development, Paris, 2023.

37 T. I. Netzeva, *Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships–The report and recommendations of ECVAM Workshop 52*, 2005.

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

39 F. Sahigara, K. Mansouri, D. Ballabio, A. Mauri, V. Consonni and R. Todeschini, Comparison of different approaches to define the applicability domain of QSAR models, *Molecules*, 2012, **17**, 4791–4810.

40 R. S. Boethling and J. Costanza, Domain of EPI Suite biotransformation models, *SAR QSAR Environ. Res.*, 2010, **21**, 415–443.

41 N. Aniceto, A. A. Freitas, A. Bender and T. Ghafourian, A novel applicability domain technique for mapping predictive reliability across the chemical space of a QSAR: reliability-density neighbourhood, *J. Cheminf.*, 2016, **8**, 69.

42 T. N. Brown, Predicting hexadecane-air equilibrium partition coefficients (L) using a group contribution approach constructed from high quality data, *SAR QSAR Environ. Res.*, 2014, **25**, 51–71.

43 OECD, *Test No. 104: Vapour Pressure*, 2006.

44 OECD, *Test No. 123: Partition Coefficient (1-Octanol/Water): Slow-Stirring Method*, 2022.

45 OECD, *Test No. 117: Partition Coefficient (n-octanol/water)*, HPLC Method, 2022.

46 OECD, *Test No. 105: Water Solubility*, 1995.

47 S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton, PubChem 2023 update, *Nucleic Acids Res.*, 2022, **51**, D1373–D1380.

48 A. J. Williams, C. M. Grulke, J. Edwards, A. D. McEachran, K. Mansouri, N. C. Baker, G. Patlewicz, I. Shah, J. F. Wambaugh, R. S. Judson and A. M. Richard, The CompTox Chemistry Dashboard: a community data resource for environmental chemistry, *J. Cheminf.*, 2017, **9**, 61.

49 J. Glüge, K. McNeill and M. Scheringer, Getting the SMILES right: identifying inconsistent chemical identities in the ECHA database, PubChem and the CompTox Chemicals Dashboard, *Env. Sci.: Adv.*, 2023, **2**, 612–621.

50 D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Comp. Sci.*, 1988, **28**, 31–36.

51 D. Weininger, A. Weininger and J. L. Weininger, SMILES. 2. Algorithm for generation of unique SMILES notation, *J. Chem. Inf. Comp. Sci.*, 1989, **29**, 97–101.

52 USEPA, *PhysProp EPI Suite Database*, accessed 2016.

53 K. Mansouri, C. M. Grulke, A. M. Richard, R. S. Judson and A. J. Williams, An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling, *SAR QSAR Environ. Res.*, 2016, **27**, 911–937.

54 S. Baskaran, Y. D. Lei and F. Wania, A database of experimentally derived and estimated octanol–air partition ratios (K_{OA}), *J. Phys. Chem. Ref. Data*, 2021, **50**.

55 J.-C. Bradley, A. Lang, A. Williams and E. Curtin, ONS open melting point collection, *Nat. Preced.*, 2011, DOI: **10.1038/npre.2011.6229.1**.

56 S. Martel, F. Gillerat, E. Carosati, D. Maiarelli, I. V. Tetko, R. Mannhold and P.-A. Carrupt, Large, chemically diverse dataset of logP measurements for benchmarking studies, *Eur. J. Pharm. Sci.*, 2013, **48**, 21–29.

57 S. Tshepelevitsh, S. A. Kadam, A. Darnell, J. Bobacka, A. Rüütel, T. Haljasorg and I. Leito, LogP determination for highly lipophilic hydrogen-bonding anion receptor molecules, *Anal. Chim. Acta*, 2020, **1132**, 123–133.

58 A. Guillot, Y. Henchoz, C. Moccand, D. Guillarme, J. L. Veuthey, P. A. Carrupt and S. Martel, Lipophilicity determination of highly lipophilic compounds by liquid chromatography, *Chem. Biodivers.*, 2009, **6**, 1828–1836.

59 D. Wang, J. Yu, L. Chen, X. Li, H. Jiang, K. Chen, M. Zheng and X. Luo, A hybrid framework for improving uncertainty

This journal is © The Royal Society of Chemistry 2025

Environ. Sci.: Processes Impacts, 2025, **27**, 3457–3470 | **3469**

quantification in deep learning-based QSAR regression modeling, *J. Cheminf.*, 2021, **13**, 69.

60 U. Sahlin, Uncertainty in QSAR Predictions, *Altern. Lab. Anim.*, 2013, **41**, 111–125.

61 J. R. Mora, E. A. Marquez, N. Pérez-Pérez, E. Contreras-Torres, Y. Perez-Castillo, G. Agüero-Chapin, F. Martinez-Rios, Y. Marrero-Ponce and S. J. Barigye, Rethinking the applicability domain analysis in QSAR models, *J. Comput.-Aided Mol. Des.*, 2024, **38**, 9.

62 S. Endo and K.-U. Goss, Applications of polyparameter linear free energy relationships in environmental chemistry, *Environ. Sci. Technol.*, 2014, **48**, 12477–12491.

63 A. Tropsha, Best practices for QSAR model development, validation, and exploitation, *Mol. Inform.*, 2010, **29**, 476–488.

64 H. Zhu, A. Tropsha, D. Fourches, A. Varnek, E. Papa, P. Gramatica, T. Oberg, P. Dao, A. Cherkasov and I. V. Tetko, Combinatorial QSAR modeling of chemical toxicants tested against Tetrahymena pyriformis, *J. Chem. Inf. Model.*, 2008, **48**, 766–784.

65 C. Tebes-Stevens, J. M. Patel, M. Koopmans, J. Olmstead, S. H. Hilal, N. Pope, E. J. Weber and K. Wolfe, Demonstration of a consensus approach for the calculation of physicochemical properties required for environmental fate assessments, *Chemosphere*, 2018, **194**, 94–106.

66 Organisation for Economic Co-operation and Development (OECD), *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*, Organisation for Economic Co-operation and Development, Paris, 2007.

67 J. J. Irwin, K. G. Tang, J. Young, C. Dandarchuluun, B. R. Wong, M. Khurelbaatar, Y. S. Moroz, J. Mayfield and R. A. Sayle, ZINC20—A free ultralarge-scale chemical database for ligand discovery, *J. Chem. Inf. Model.*, 2020, **60**, 6065–6073.

68 A. C. Guyton and J. E. Hall, *Guyton and Hall Textbook of Medical Physiology*, Elsevier, 13th edn, 2015.

69 A. O. De Silva, J. M. Armitage, T. A. Bruton, C. Dassuncao, W. Heiger-Bernays, X. C. Hu, A. Kärrman, B. Kelly, C. Ng, A. Robuck, M. Sun, T. F. Webster and E. M. Sunderland, PFAS exposure pathways for humans and wildlife: a synthesis of current knowledge and key gaps in understanding, *Environ. Toxicol. Chem.*, 2021, **40**, 631–657.

70 E. M. Sunderland, X. C. Hu, C. Dassuncao, A. K. Tokranov, C. C. Wagner and J. G. Allen, A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects, *J. Expo. Sci. Environ. Epidemiol.*, 2019, **29**, 131–147.

71 I. S. Gkika, G. Xie, C. A. M. van Gestel, T. L. Ter Laak, J. A. Vonk, A. P. van Wezel and M. H. S. Kraak, Research priorities for the environmental risk assessment of per- and polyfluorinated substances, *Environ. Toxicol. Chem.*, 2023, **42**, 2302–2316.

72 C. Ng, I. T. Cousins, J. C. DeWitt, J. Glüge, G. Goldenman, D. Herzke, R. Lohmann, M. Miller, S. Patton, M. Scheringer, X. Trier and Z. Wang, Addressing urgent questions for PFAS in the 21st century, *Environ. Sci. Technol.*, 2021, **55**, 12755–12765.

3470 | *Environ. Sci.: Processes Impacts*, 2025, **27**, 3457–3470

This journal is © The Royal Society of Chemistry 2025