

## The First Attempt at Non-Linear in Silico Prediction of Sampling Rates for Polar Organic Chemical Integrative Samplers (POCIS)

Thomas H. Miller,<sup>†</sup> Jose A. Baz-Lomba,<sup>‡</sup> Christopher Harman,<sup>§</sup> Malcolm J. Reid,<sup>‡</sup> Stewart F. Owen,<sup>||</sup> Nicolas R. Bury,<sup>⊥</sup> Kevin V. Thomas,<sup>‡</sup> and Leon P. Barron<sup>\*,†</sup>

<sup>†</sup>Analytical & Environmental Sciences Division, Faculty of Life Sciences and Medicine, King's College London, 150 Stamford Street, London, SE1 9NH United Kingdom

<sup>‡</sup>Norwegian Institute for Water Research (NIVA), Oslo, NO-0349, Norway

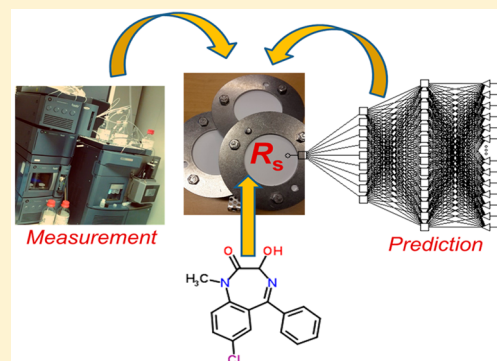
<sup>§</sup>Norwegian Institute for Water Research (NIVA), Grimstad, NO-4879, Norway

<sup>||</sup>AstraZeneca, Global Environment, Alderley Park, Macclesfield, Cheshire SK10 4TF, United Kingdom

<sup>⊥</sup>Division of Diabetes and Nutritional Sciences, Faculty of Life Sciences and Medicine, King's College London, Franklin Wilkins Building, 150 Stamford Street, London, SE1 9NH, United Kingdom

### S Supporting Information

**ABSTRACT:** Modeling and prediction of polar organic chemical integrative sampler (POCIS) sampling rates ( $R_s$ ) for 73 compounds using artificial neural networks (ANNs) is presented for the first time. Two models were constructed: the first was developed ab initio using a genetic algorithm (GSD-model) to shortlist 24 descriptors covering constitutional, topological, geometrical and physicochemical properties and the second model was adapted for  $R_s$  prediction from a previous chromatographic retention model (RTD-model). Mechanistic evaluation of descriptors showed that models did not require comprehensive a priori information to predict  $R_s$ . Average predicted errors for the verification and blind test sets were  $0.03 \pm 0.02 \text{ L d}^{-1}$  (RTD-model) and  $0.03 \pm 0.03 \text{ L d}^{-1}$  (GSD-model) relative to experimentally determined  $R_s$ . Prediction variability in replicated models was the same or less than for measured  $R_s$ . Networks were externally validated using a measured  $R_s$  data set of six benzodiazepines. The RTD-model performed best in comparison to the GSD-model for these compounds (average absolute errors of  $0.0145 \pm 0.008 \text{ L d}^{-1}$  and  $0.0437 \pm 0.02 \text{ L d}^{-1}$ , respectively). Improvements to generalizability of modeling approaches will be reliant on the need for standardized guidelines for  $R_s$  measurement. The use of in silico tools for  $R_s$  determination represents a more economical approach than laboratory calibrations.



### INTRODUCTION

Contamination of the aquatic environment with herbicides, pesticides, pharmaceuticals, and personal care products (PPCPs), among other contaminants, has been the focus of environmental monitoring campaigns over the last two decades. Reported concentrations and associated adverse effects of these contaminants has led to the introduction of legislative procedures to monitor and assess risk associated with pollutants, such as the EU water framework directive and the EU registration, evaluation, authorization, and restriction of chemicals (REACH).<sup>1,2</sup>

High frequency sampling campaigns often involve the use of grab or composite sampling, but are practically difficult and costly to manage for monitoring longer-term fluctuations in contaminant concentrations in the aquatic environment. These methods are also often labor intensive with respect to sampling and can lead to considerable cost during instrumental analysis. More recently, however, the development and use of passive sampling devices (PSDs) is increasing due to their capability for a time-integrated approach to averaging contaminant concen-

trations in surface waters as well as influent and effluent wastewater over extended periods.<sup>3</sup> PSDs minimize sample preparation and allow in situ enrichment of analytes which may potentially reduce limits of quantification in comparison to those achieved by point sampling.<sup>4</sup> Passive sampling devices in some fields are well-established, such as use of semipermeable membrane devices (SPMD) for organochlorines<sup>5</sup> and other similarly hydrophobic compounds.<sup>6–8</sup> However, one type of PSD which is emerging currently is the polar organic chemical integrative sampler (POCIS). These samplers have been used to determine the occurrence of a range of chemically diverse, and comparatively polar to moderately nonpolar compounds.<sup>9–13</sup> However, for quantitative studies, POCIS suffer from some limitations, mainly relating to the reliability of derived estimations of the sampling rates ( $R_s$ ) from

Received: March 21, 2016

Revised: June 30, 2016

Accepted: July 1, 2016

Published: July 1, 2016

experimental measurements, as well the lack of a well-developed performance reference compound (PRC) exposure correction method.<sup>14–16</sup> One further hindrance is that reported sampling rate data are few and methods for their estimation vary which leads to limited transferability across other locations or studies.<sup>17</sup>

Given the time-intensive nature of determining  $R_s$  experimentally, it is possible that computational modeling approaches could offer a solution that would enable prediction of sampling rate data for compounds without the need for experimental determination. A previous investigation by Stephens et al.<sup>18</sup> evaluated the use of an empirical method (Sherwoods correlation) to determine PSD kinetic parameters for a limited number of compounds showing maximum errors of +40 and –20% for the estimation of the aqueous boundary layer mass transfer coefficient ( $k_f$ ). In contrast to empirical methods for estimating specific parameters, quantitative structure–property relationship (QSPR) models are becoming more frequently used in ecotoxicology where a set of  $x$  variables are used to predict a response,  $y$ .<sup>19</sup> The variables are often molecular descriptors that cover constitutional, topological, geometrical, and physicochemical properties which can then be used to model a desired output. Models can vary from simple linear regression approaches to complex nonlinear functions where such models are often designed by machine learning methods. Two well-known machine learning methods are support vector machines (SVMs) and artificial neural networks (ANNs) and have been used successfully in related areas such as the prediction of bioconcentration factors (BCFs), octanol–water partition coefficients ( $\log P$ ) and biosolid/water partition coefficients ( $K_d$ ), as well as for suspect compound screening via prediction of chromatographic retention time.<sup>20–27</sup> The use of SVMs for environmental applications is still in its infancy and substantial programming capability is required for routine application. On the other hand, ANNs are well-known and more user-friendly software has been available for many years. ANNs comprise a layered structure (normally three), each with a different purpose. The input layer contains the molecular descriptor data for each compound for training, verification and blind testing and the output layer is the response. The hidden layer sits in between and contains several nodes, and often multiple sublayers of such nodes, where linear or nonlinear functions are used to relate the descriptors to the output layer. The residual errors are monitored and reduced by using iterative algorithms which adjust weights associated with the nodes in the hidden layer. Thus, such modeling approaches could greatly increase the applicability of POCIS in environmental monitoring studies through bypassing the need for laboratory and in situ calibrations.

The aim of this work was to investigate the potential of ANNs to model and predict  $R_s$  for POCIS devices for a range of pharmaceuticals, endocrine disrupting chemicals, pesticides, herbicides and drugs of abuse. The objectives were to identify suitable analyte molecular descriptors to build, train and test a range of suitable model types and architectures and then finally to externally validate the approach for predicting  $R_s$  for several compounds which were, for comparison, determined in parallel by laboratory calibration. To the authors' knowledge, this represents the first study to draw together, harmonize and predict the published  $R_s$  data for ionizable pharmaceutical compounds on POCIS. Ultimately, where such tools can provide adequate predictions using new data generated in the

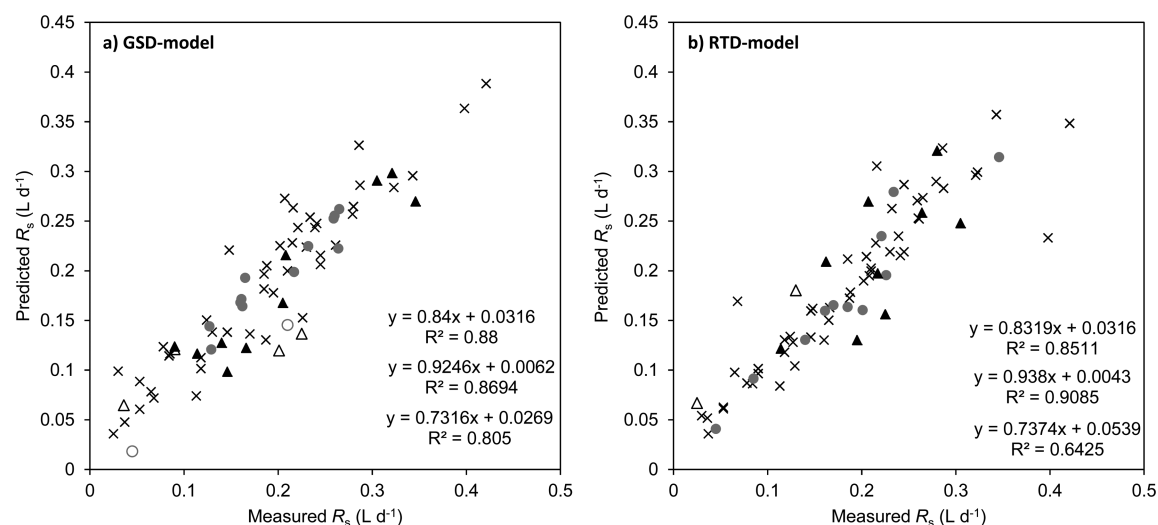
future, this approach could reduce the analytical burden of laboratory estimations of  $R_s$ .

## MATERIALS AND METHODS

**Selection of Data Sets, Molecular Descriptors and ANN Models.** A working data set derived from the literature (2007–present) was used to build, train and optimize models for  $R_s$  prediction on POCIS. A total of  $n = 73$  compound  $R_s$  data were derived from Fauvelle et al.<sup>28</sup> and Morin et al.,<sup>24</sup> which were generated using similar experimental conditions to give the largest combined data set of all studies. Compounds included herbicides, pesticides, endocrine disrupting compounds and pharmaceuticals. Where duplicate compound  $R_s$  data existed, both values were removed entirely from the data set (six compounds). Generally, in these cases  $R_s$  differed and it was uncertain which value was correct or whether an average was appropriate for modeling. Simplified molecular input line entry system (SMILES) strings were generated from Chempidder (Royal Society of Chemistry, UK). Using these,  $n = 185$  molecular descriptors were generated from Parameter Client freeware (Virtual Computational Chemistry Laboratory, Munich, Germany) and an additional  $n = 16$  descriptors were from ACD laboratories Percepta software (Advanced Chemistry Development Laboratories, ON, Canada).

Two models were generated using two separate sets of descriptors covering constitutional, topological, geometrical and physicochemical properties that were investigated for their comparative prediction performance. The first subset of 24 descriptors (see [Supporting Information \(SI\), Table S1](#)) was generated using a genetic feature selection algorithm to produce the genetically selected descriptor model (GSD-model). Genetic feature selection algorithms follow evolutionary concepts to convert input descriptors into binary strings, in this case to prioritise descriptors for  $R_s$  prediction. Using a process similar to natural selection, prioritised strings are crossed to form a new population of strings. The generational “breeding” of strings produced an optimized selection of input variables for application to prediction. The parameters for the GA were as follows; population = 100, generation = 100, mutation rate = 0.1 and crossover rate = 1. In an alternative approach, a much simpler descriptor data set previously used to model elution from reversed-phase liquid chromatography (RPLC) stationary phases was investigated to assess any improvement (see [SI Table S2](#)).<sup>23,25</sup> This model is referred to as the retention time descriptor model (RTD-model). POCIS devices contain a divinylbenzene and *N*-vinylpyrrolidone copolymer, which enabled dual polar and nonpolar interactions for retention. As retention on reversed-phase chromatographic columns is governed predominantly by hydrophobic interactions too, it is possible that these same descriptors will also be important in passive sampling. No retention data was available for the studies by Fauvelle et al.<sup>28</sup> and Morin et al.<sup>24,29</sup> However, correlation between  $R_s$  and 21 corresponding retention times ( $t_R$ ) gathered on a  $C_{18}$  stationary phase in a study by Bade et al.<sup>30</sup> showed a weak relationship ( $R = 0.472$ ).

For both descriptor subsets, several network types were tested for predictive ability using Trajan 6.0 neural network software (Trajan Software Ltd., Lincolnshire, UK) and these included radial basis function (RBF), generalized regression neural networks (GRNNs) and multilayer perceptrons (MLPs). Following training and optimization using both data sets, the GSD- and RTD-models were produced. The GSD-model



**Figure 1.** Measured  $R_s$  against predicted  $R_s$  for (a) the GSD-model and (b) the RTD-model. Crosses, circles and triangles are the training, verification and test subsets, respectively. Open circles and triangles indicate predicted inaccuracies of >30% of the measured value.

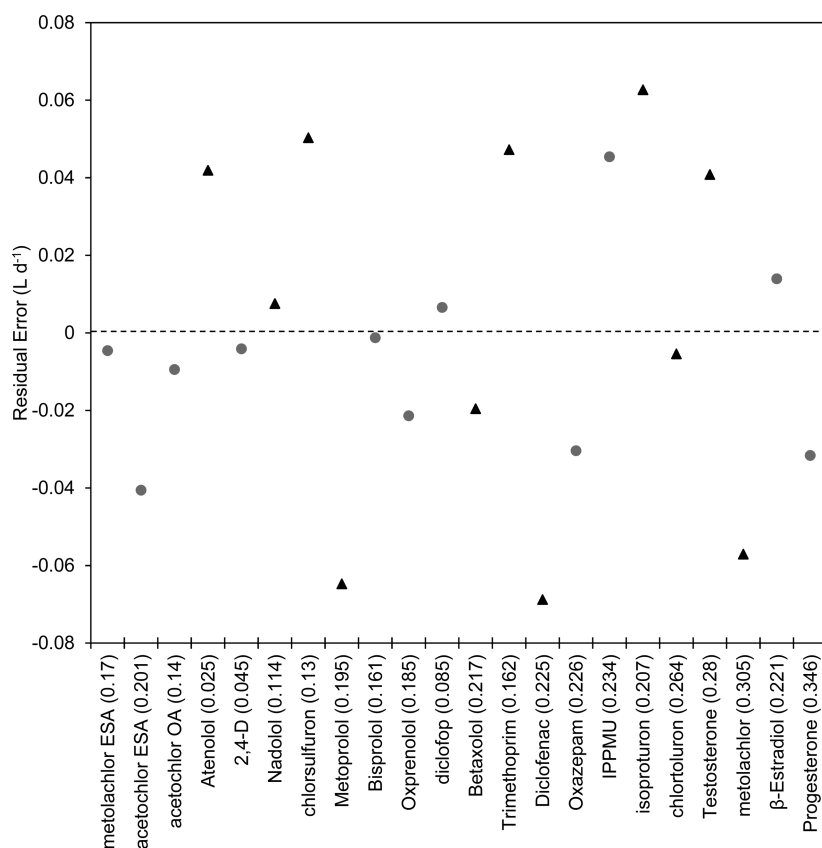
architecture was a four-layer MLP with 24 descriptors in the input layer (independent variables); two hidden layers containing 17 and 14 nodes and the dependent variable output layer ( $R_s$ ). Training involved two types of algorithms, the first was back-propagation (BP) and the second was conjugate gradient descent (CGD). The data set was split into 45:14:14 cases for the training, verification and test subsets (optimized). The RTD-model architecture was also a four-layer MLP using both BP and CGD. The first and fourth layers were the inputs (using the set of descriptors previously used for chromatographic retention modeling) and outputs ( $R_s$ ), respectively, and the second and third layers (hidden layers) contained 14 and 9 nodes, respectively. The division of cases included 51 compounds for training, 11 compounds for verification and 11 compounds for blind testing (optimized).<sup>27</sup> All cases were randomly selected to avoid bias. The verification data set was used to characterize network predictive performance during training and also to allow regularisation to prevent overfitting. The test set was then used to validate the model to ensure that the model generalized well to new cases. The optimized models were selected based on the lowest errors and consistency across the training, verification and test subsets.

**Laboratory Calibration of Sampling Rates to Test Model Generalizability.** Sampling rates ( $L d^{-1}$ ) were determined using a static renewal method over a 14 day exposure period and in a similar manner to data in the literature which were used for modeling here.<sup>28</sup> Briefly, 3 L of high-density polyethylene vessels were filled with ultrapure water, the pH adjusted to 7.6 with 20 mg  $L^{-1}$   $NaHCO_3$  and spiked with the mixture of respective compounds to expose the POCIS. Each vessel contained three POCIS devices for exposure to an aqueous-based standard mixture of 200 ng  $L^{-1}$  of each target compound (solvent <0.001%). This standard solution was prepared and replaced daily in 3 L volumetric flasks to maintain the nominal concentration. Following this, all three POCIS were removed from each vessel at day 4, 7, and 14, rinsed with ultrapure water and frozen at  $-20$  °C. Extraction of POCIS sorbents was performed using a wash phase of 5 mL of ultrapure water and then elution using 5 mL of MeOH. Eluate was dried under nitrogen at 35 °C for 40 min. The dried residue was then reconstituted in 0.5 mL of starting mobile phase. The analysis of the benzodiazepines was

performed on an Acquity UPLC system coupled to a Xevo G2 S QTOF mass analyzer (Milford, MA) with an online Oasis HLB Direct Connect HP loading column. Analyte separation was performed on an Acquity UPLC BEH  $C_{18}$  column (1.7  $\mu m$ , 50  $\times$  2.1 mm) from Waters (Milford, MA) at 50 °C. Gradient elution (0.6 mL  $min^{-1}$ ) for analyte separation was with 0.1% (v/v) formic acid in water (phase A) and 0.1% formic acid in methanol (phase B). Full method details for the laboratory calibration experiments and analysis are given in the SI.

## RESULTS AND DISCUSSION

**$R_s$  Prediction Using a GSD-Model.** Following genetic feature selection, a 24–17–14–1 MLP yielded the best performance using 24 input descriptors with  $R^2 = 0.8800$ , 0.8694, and 0.8050 for training, verification and blind test sets respectively (sum of squared residual errors were 0.084, 0.062, and 0.116, respectively). Therefore, this model initially seemed quite promising for application to prediction of  $R_s$  for new compounds (Figure 1a). Many shortlisted descriptors were derived from topological indices, but some others were expected to have more importance for this application, such as those that describe molecular hydrophobicity. These include the octanol–water partition coefficient ( $\log P$ ) and the distribution ratio between octanol and water ( $\log D_{ow}$ ). The latter takes into account the ionised proportion of a compound at a particular pH and is dependent on the  $\log P$  and the  $pK_a$  of all ionizable functional groups in a molecule. An investigation by Booi et al.,<sup>31</sup> demonstrated that uptake rates in SPMDs correlated well with  $\log P$  where  $R_s \approx P^{-0.044}$ . Correlations between  $\log P$  and  $R_s$  has also been observed for POCIS devices.<sup>32–34</sup> Assessment of the collinearity with  $R_s$  (SI Table S7) showed rather unsurprisingly for so many ionizable compounds that  $\log D_{ow}$  had, by far, the highest correlation ( $R = 0.59$ ), but was insufficient by itself to describe sorption to POCIS sorbents. To the authors knowledge, no previous investigations have used  $\log D_{ow}$  to model  $R_s$ , although it has been weakly correlated with  $R_s$ .<sup>35</sup> Furthermore, interinput descriptor collinearity also existed and especially for constitutional descriptors such as the number of non-H bonds (nBO) and the sum of conventional bond orders (SCBO); as well as topological descriptors such as log Narumi simple topological



**Figure 2.** RTD-model residual plot of predicted  $R_s$  values for the verification and test subset only, ordered in parentheses by their ascending distribution ratio values between octanol and water ( $\log D_{ow}$ ). Circles and triangles represent the verification and test subset, respectively. The measured  $R_s$  values are displayed in parentheses on the  $x$ -axis. 2,4-D (2,4-dichlorophenoxyacetic acid), ESA (ethanesulfonic acid), OA (oxanilic acid), and IPPMU (isoproturon-monodemethyl).

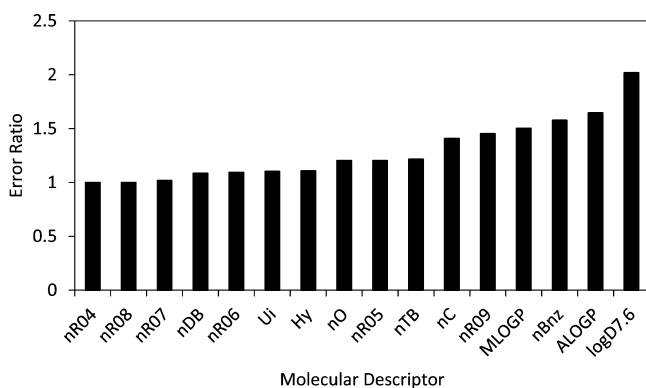
index (S<sub>nar</sub>), second Zagreb index (ZM2). Pearson's coefficients were  $\geq 0.8$  for these descriptors with at least eight other descriptors. Therefore, though genetic algorithms short-listed useful descriptors for potential  $R_s$  modeling here, back-interpretation of model sensitivity to descriptor data for derivation of mechanistic understanding of physicochemical POCIS uptake mechanisms would be limited. However, as a tool to predict  $R_s$ , the training set overall displayed good accuracy within 22% of the measured value on average. In comparison, the verification and test subsets were predicted on average within 19% of their measured values showing consistency across all subsets (SI, Figure S1). For particular blind test cases, however, some notably large inaccuracies were observed such as for sotalol (80% inaccuracy) where the lower hydrophobicity of this molecule may explain poorer correlation with  $R_s$ .<sup>31</sup> Larger errors were also recorded for acetochlor ethanesulfonic acid (40%), diclofenac (39%) and sulcotrione (38%). The verification subset contained two largely inaccurate predictions (2,4-dichlorophenoxyacetic acid at 59% and timolol at 31%), but all remaining compounds were within 20% of measured  $R_s$ . Larger inaccuracies may be related to poor learning from selected training data. For example, mesotrine had a 59% inaccuracy to the measured value in the training set which may explain the poor prediction of another structurally similar compound, sulcotrione, in the test set. Overall, inaccuracy was most prevalent for sulfonate-containing compounds where genetic selection did not sufficiently prioritise descriptors for this portion of cases for reliable  $R_s$  prediction. As the number of available cases expands, genetic

selection of descriptors may improve for such compounds in the future. It is also unclear whether sulfonate bearing molecules are subject to steric and/or repulsive forces arising from the PES membrane. Furthermore, larger inaccuracies (>30%) in the full data set generally corresponded to compounds with  $R_s < 0.1$  such as 2,4-dichlorophenoxyacetic acid, sotalol, sulcotrione and nicosulfuron. However, when predictive accuracy was plotted against  $R_s$  for all compounds, no correlation was observed for other compounds with  $R_s < 0.1$  (SI, Figures S2 and S3).

**$R_s$  Prediction Using a RTD-Model.** The correlation of predicted versus measured  $R_s$  for the RTD-model is shown in Figure 1b. The error (sum squared) for the subsets were 0.092, 0.062, and 0.121 for the training, verification and test sets, respectively. The model was, again, a four-layered MLP with a 16:14:9:1 architecture. Generally, acceptable correlations were achieved for the training, verification and blind test sets ( $R^2 = 0.8511, 0.9085,$  and  $0.6425,$  respectively) though this model performed slightly worse (training and test) than the GSD-model. The training subset showed several larger errors which corresponded to the compounds *t*-butylphenol (149%), 2,4-dichlorophenol (41%), and simazine (41%). The compound sulfamethoxazole showed an 81% overestimation of its experimentally determined  $R_s$ . As discussed earlier, this large inaccuracy was also reflected in the GSD-model which showed an overestimation of 230% for sulfamethoxazole which also bears a sulfonate group. Overall, however, the model showed relatively good predictions of  $R_s$  (mean absolute error for training set = 15%; and for both verification and test subsets =

22%). The average error  $\pm$  standard deviation across the verification and blind test subsets was  $0.03 \pm 0.02$  L day<sup>-1</sup> showing acceptable overall predictive accuracy for  $R_s$ . Atenolol, the compound with the lowest  $R_s$ , yielded poor prediction accuracy (predicted  $R_s = 0.067$ , measured  $R_s = 0.025$ ) which was initially thought to be due its higher polarity in comparison to others selected for this study. However, no correlation was observed between predictive accuracy and  $\log D_{ow}$  (Figure 2). Average predictive mean error of the verification and blind test sets both reduced to  $\sim 15\%$  upon removal of the atenolol data-point. Importantly, as very polar compounds are generally not retained well by *n*-vinylpyrrolidone-*co*-divinylbenzene-based polymer sorbents, inaccuracy in measured  $R_s$  may be compound specific as a result, which in turn may contribute to RTD-model prediction errors. This highlights the lack of consistent measurements available for training of such models for predictive purposes. Nonetheless, considering this performance alongside the potential for inaccuracy in  $R_s$  data from different laboratory calibrations, predictions using these models were considered reasonable.

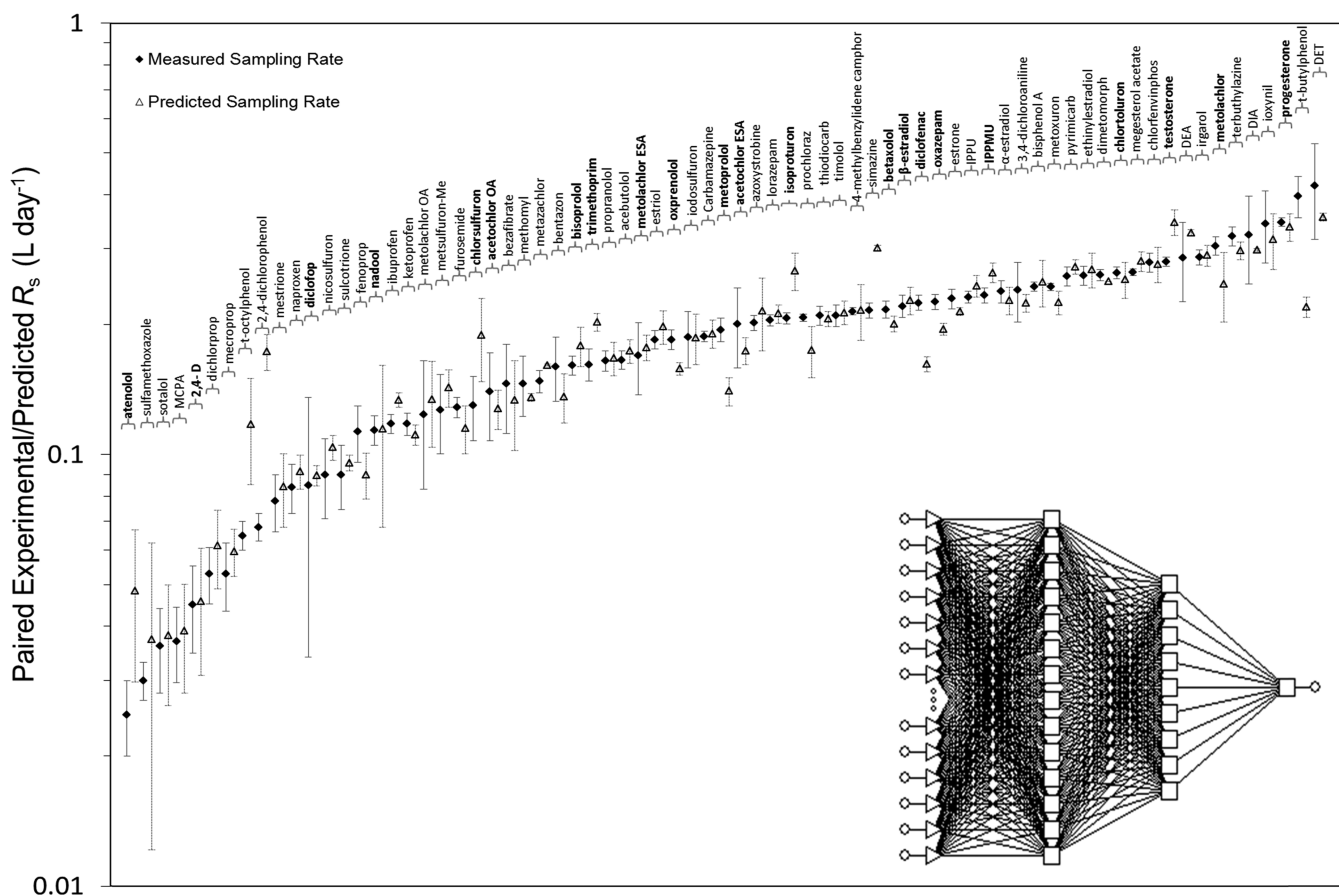
**Model Interpretation and Descriptor Contribution to  $R_s$  Prediction.** Given the level of multicollinearity observed for GSD-model descriptors, a sensitivity analysis could only be performed to identify the relative contribution of each descriptor to predictions in the RTD-model. This was represented as the error ratio, i.e. the ratio between the model error using all descriptors and the model error when one descriptor was removed. However, like in the GSD-model, the use of sensitivity analysis to further mechanistic understanding of sorption processes should be approached with caution if some individual descriptors display multicollinearity (please refer to SI Tables S1–S3 for full descriptor details and data). The  $\log D_{ow}$ , the Moriguchi octanol–water partition coefficient (MlogP), the Ghose-Crippen octanol–water partition coefficient (AlogP) and the number of Benzene rings (nBnz) were the top four descriptors used by the RTD-model (Figure 3). This is in agreement with Bäuerlein et al., who showed that hydrophobicity and *pi*-*pi* interactions (e.g., via benzene rings) were important for adsorption to HLB sorbents in batch experiments<sup>36</sup> and which can also affect diffusion. Other important descriptors were the number of triple bonds (nTB;



**Figure 3.** Sensitivity analysis of the optimized RTD-model. Acronyms: nDB/nTB = number of double/triple bonds; nC/nO = number of carbon/oxygen atoms; nR04–nR09 = number of 4–9 membered rings; Ui = unsaturation index; Hy = hydrophilic factor; nBnz = number of benzene-like rings; MlogP/AlogP = Moriguchi/Ghose-Crippen logarithm of octanol–water partition coefficient;  $\log D_{7.6}$  = logarithm of distribution ratio between octanol and water at pH 7.6.

error ratio = 1.2165), number of five-membered rings (nR05; error ratio = 1.2041) and number of nine-membered rings (nR09; error ratio = 1.4544). The importance of the *n*-membered ring descriptors could be attributed to molecular size and flexibility thus affecting the diffusivity of molecules through the water boundary layer (WBL), PES membrane (pore = 0.1  $\mu$ m) or pores of the HLB copolymer (80 Å).<sup>37–39</sup> A previous investigation showed that size descriptors were also important for predicting soil sorption coefficients for pesticides.<sup>39</sup> We also previously showed these descriptors were important for ANN-based predictions of pharmaceutical sorption to soils and sludge.<sup>40</sup> In addition to those mentioned above, the number of carbons (nC), number of oxygens (nO) and hydrophilic factor (Hy) also showed that they were important to the RTD-model. Hy relates to the number of hydrophilic groups in the molecule such as hydroxyls, thiols and sulfonates. As polar surface area has been previously shown to influence interactions with HLB sorbents, it is logical that hydrophilicity/polarity related descriptors would have some importance.<sup>36</sup> Several authors have suggested that diffusion is the main factor governing uptake rates in PSDs.<sup>41</sup> We have attributed the importance of the descriptors mainly to sorbent interactions so far, but it is also possible that these same descriptors could relate to diffusion processes due to the number of molecular properties that will affect it including dipole moments, polarizability, molecular size (including hydration radius) and electrostatic charge.<sup>42</sup> The genetic feature selection algorithm did not select some recognized diffusion-related descriptors, such as molecular weight as a simple example. However, it did select other descriptors that showed interdependencies on factors affecting diffusion such as number of atoms, number of rotatable bonds, and electrotopological states.  $R_s$  has been attributed mainly in the past to diffusion processes in partition samplers such as silicone rubbers.<sup>41</sup> The portion of  $R_s$  governed by diffusion in adsorption samplers using HLB-type sorbents in POCIS remains unclear especially whether sorption of analytes via hydrogen bonding, dipole–dipole, dipole–induced dipole, van der Waals and *pi*-*pi* interactions plays a more significant role. It is also possible that the models presented here for  $R_s$  prediction could be developed and improved further with additional or alternative descriptors, such as diffusion coefficients. However, adding such descriptors may introduce a greater uncertainty into the model as estimates can be based on several different approaches.<sup>43–45</sup> In addition, diffusion coefficients will be affected by numerous environmental factors and hydrodynamic conditions that would be difficult to replicate or control in situ. Inclusion of larger numbers of descriptors to cover all the processes involved will likely inhibit model generalizability. Indeed, ANNs learn more holistically, making predictions possible without the need for such comprehensive a priori information. However, such a holistic approach obviously limits deeper understanding of the precise contribution of individual mechanisms involved in POCIS.

By comparison, the GSD-model featured many more topological and geometrical descriptors than in the RTD-model. These descriptors showed multicollinearity and therefore the sensitivity analysis could not be performed reliably (Table S7). Simply adding noncollinear descriptors to the RTD-model is also disadvantageous at this point. As the number of descriptors increases, overfitting of data is more likely to occur and would require significantly more case examples for valid application.<sup>46</sup> Model complexity will also



**Figure 4.** Comparison of the measured and predicted  $R_s$  values and their respective variances against the variance in predicted  $R_s$  ( $n = 73$ ) from replicate RTD-models ( $n = 3$ ). Inset: Optimized 16–14–9–1 model architecture. Compounds in bold represent the verification and blind test cases. All others were used for model training.

limit the ability of the network to generalize when predicting unknown compounds therefore a smaller number of descriptors (and nodes in the hidden layer(s)) is ultimately more beneficial.

**Reproducibility of Predicted and Experimentally Determined  $R_s$ .** Model performance and generalizability is limited by the quality of input data. Measured  $R_s$  can differ considerably even within calibration studies performed in the same laboratory. The largest variance in measured  $R_s$  used corresponded to diclofop which had a 60% relative standard deviation (RSD),  $n = 15$ .<sup>28</sup> Many of the reported  $R_s$  values vary by more than 2-fold depending on the methodology used for their estimation.<sup>29</sup> Although pH and temperature during collection of both sets of data used herein were similar, the type of calibration experiment applied was slightly different (flow-through and static renewal) therefore the resulting differences in the  $R_s$  estimates from each investigation could have affected the performance of a model. For the six compounds common to both calibration methods that were removed from the original data set used for model optimization, the absolute difference in  $R_s$  was  $0.088 \pm 0.072$  L day<sup>-1</sup> between measurements.

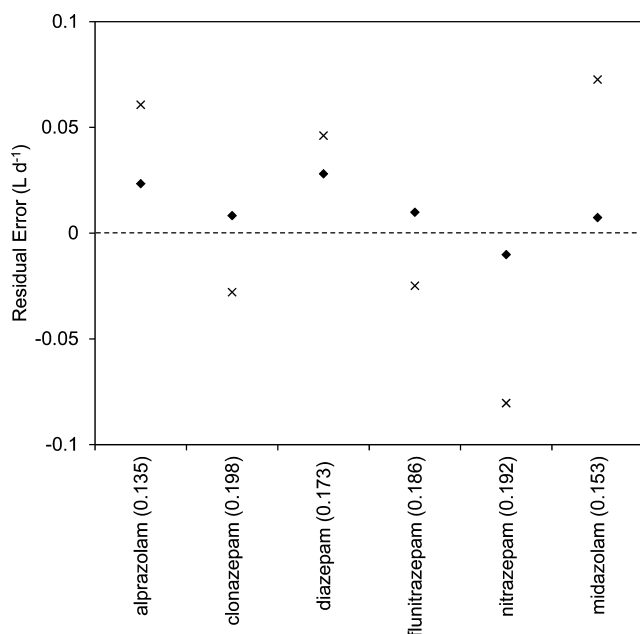
The average % RSD of measured  $R_s$  data used herein was 11% (mean deviation was  $\pm 0.017$  L day<sup>-1</sup>) (Figure 4). In 45% of all cases, the % RSDs of predicted  $R_s$  across triplicate network trained ab initio were better than the % RSDs of the measured data. For several specific cases, such as DET, DIA, diclofop and ioxynil, the experimental variation was relatively

large when compared to the variation in predicted  $R_s$ . Such deviation in experimentally derived sampling rates can be attributed to several similar factors to those already discussed above (e.g., temperature, pH, flow rate etc.). Figure 4 shows that for cases which had poor predictive accuracy with respect to the mean true value, such as for acetochlor ESA, the standard deviation of the predicted  $R_s$  overlapped with the reported experimental variance. A review by Harman et al., suggests that literature reported  $R_s$  data should only be considered as an approximation.<sup>47</sup> However, in the absence of a standardized method for POCIS calibration, either in the laboratory or in the field, it would seem that  $R_s$  modeling in this way offers similar accuracy and precision without being labor or resource intensive. Calibration experiments for each compound can take several weeks, requiring a large mass of reference material for static renewal and flow through experiments, or very frequent and accurate water sampling for in situ experiments. Furthermore, given that models developed herein are derived from a very limited number of training cases, any new reported  $R_s$  data generated by similar methods to those used herein will likely enable better generalizability in the future, as was observed with retention time predictions in reversed-phase liquid chromatography.<sup>27</sup>

**External Application to  $R_s$  Prediction.** To further support the application of the optimized modeling approach,  $R_s$  data for several additional benzodiazepines were experimentally determined in our laboratory using a similar approach. In the previous sections, blind test compounds

were structurally diverse which is logical for testing model accuracy.<sup>48</sup> However, for this experiment, structural similarity was deliberately chosen to externally test its discriminative power. Despite this similarity, it was expected that measured  $R_s$  could be different on POCIS given their slight differences in chromatographic retention on  $C_{18}$  phases. The retention order of the benzodiazepines was as follows: oxazepam (3.26 min) nitrazepam (3.26 min), clonazepam (3.29 min), lorazepam (3.29 min), alprazolam (3.31 min), midazolam (3.32 min), flunitrazepam (3.37 min) and diazepam (3.58 min). As discussed previously, measurement of  $R_s$  often suffers from some imprecision. The calibration experiment performed here was not exempt from this either. Two compounds, lorazepam ( $R_s$ : 0.205 L d<sup>-1</sup>) and oxazepam ( $R_s$ : 0.226 L d<sup>-1</sup>), were originally present in the training set and verification set respectively during model development. The  $R_s$  values for these compounds were experimentally determined again here to characterize the variance between the selected calibration method used here and the method by Morin et al.<sup>24</sup> The  $R_s$  determined here varied by approximately 0.1 L d<sup>-1</sup> for both compounds (lorazepam: 0.302 L d<sup>-1</sup> and oxazepam: 0.327 L d<sup>-1</sup>). This observation showed again that the difference in calibrations between flow-through and static renewals is not negligible and was an unavoidable limitation of the calibration experiment used here. Standard deviations for the six compounds ranged from  $\pm 0.024$  to  $\pm 0.055$  L day<sup>-1</sup> ( $n = 9$ ). Overall, the average RSD for all compounds was  $20 \pm 6\%$  (flunitrazepam: 19%; clonazepam: 13%; nitrazepam: 13%; midazolam: 23%; diazepam: 23%; and alprazolam: 29%) and this variance was consistent with other studies.<sup>29</sup>

As shown in Figure 5, both the GSD- and RTD-models predicted  $R_s$  well to within the measured value for all six compounds. The two largest errors in the RTD-model corresponded to those substances with the highest  $R_s$  variance (diazepam and alprazolam at 16 and 17%, respectively), but the



**Figure 5.** Residual plot of the predicted  $R_s$  values for the GSD-model (cross) and RTD-model (diamond) for external prediction validation (as a blind test application) using six additional benzodiazepines. Measured  $R_s$  values are displayed in parentheses.

four remaining compounds showed little inaccuracy ( $\leq 5\%$ ). In terms of absolute inaccuracy of the measured  $R_s$  however, examination of the RTD-model residual errors showed that for all compounds except nitrazepam, that predictions were slightly overestimated. The GSD-model performed worse by comparison (Figure 5). The two largest errors corresponded to nitrazepam and midazolam that were 37% and 43% inaccurate, respectively. The remaining compound inaccuracies were alprazolam (28%), clonazepam (18%), diazepam (10%) and flunitrazepam (19%). The average absolute error for the GSD-model predictions was 0.0437 L d<sup>-1</sup> and all compounds were predicted within  $\pm 0.075$  L d<sup>-1</sup>. By contrast, the RTD-model had an average absolute error of 0.0145 L d<sup>-1</sup> for these benzodiazepines (and  $R_s$  for all compounds were predicted within 0.03 L day<sup>-1</sup>). These predictions again demonstrated that predicted  $R_s$  were similar enough to those determined by experimental determination to be practical.

Passive sampling for nonhydrophobic compounds is mainly used for screening purposes and as a semiquantitative technique. Furthermore, in situ exposures are difficult to quantify accurately as laboratory calibrations may not translate well into field  $R_s$  due to several factors such as biofouling and other matrix- or environmentally related effects on diffusion, for example. In addition, for reliable quantification the performance reference compound approach has limited availability and application for polar passive sampling due to the strong retention of analytes on HLB sorbents.<sup>49</sup> However, modeling approaches could potentially overcome these limitations if models were built from in situ calibration data. It is also possible that estimation of  $R_s$  by in silico approaches may offer a viable alternative for compounds where  $R_s$  data cannot be estimated by field studies due to poor correlation of concentrations in water to sample mass on the PSD.<sup>35</sup> Lastly, the two different approaches to the molecular descriptor selection presented show acceptable predictive accuracy for polar compound passive sampling. However, the use of descriptors derived for  $t_R$  prediction in a model for  $R_s$  prediction holds significant potential for application to new compounds based solely on their SMILES strings by simultaneously allowing preliminary identification (by  $t_R$  and high resolution  $m/z$ , for example) and estimation of  $R_s$  using the same descriptors.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.6b01407.

Additional POCIS extraction and measurement conditions; details on molecular descriptors and data; descriptor collinearity results; and predicted versus measured data for individual compounds(PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Phone: +44 20 7848 3842; fax: +44 20 7848 4980; e-mail: leon.barron@kcl.ac.uk.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was conducted under funding from the Biotechnology and Biological Sciences Research Council (BBSRC) CASE

industrial scholarship scheme (Reference BB/K501177/1) and AstraZeneca Global SHE research programme. AstraZeneca is a biopharmaceutical company specialising in the discovery, development, manufacturing and marketing of prescription medicines, including some products measured here. Financial support: EU International Training Network SEWPROF (Marie Curie-FP7-PEOPLE, grant number 317205). Funding bodies played no role in the design of the study or decision to publish. The authors declare no financial conflict of interest.

## REFERENCES

- (1) European Commission, Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *OJ*. **2006**, L327, 1 - 849.
- (2) European Commission, Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy *OJ*. **2000**, L327, 1 - 73.
- (3) Vrana, B.; Allan, I. J.; Greenwood, R.; Mills, G. A.; Dominiak, E.; Svensson, K.; Knutsson, J.; Morrison, G. Passive sampling techniques for monitoring pollutants in water. *TrAC, Trends Anal. Chem.* **2005**, *24* (10), 845–868.
- (4) Miège, C.; Budzinski, H.; Jacquet, R.; Soulier, C.; Pelte, T.; Coquery, M. Polar organic chemical integrative sampler (POCIS): application for monitoring organic micropollutants in wastewater effluent and surface water. *J. Environ. Monit.* **2012**, *14* (2), 626–635.
- (5) Prest, H. F.; Richardson, B. J.; Jacobson, L. A.; Vedder, J.; Martin, M. Monitoring organochlorines with semi-permeable membrane devices (SPMDs) and mussels (*Mytilus edulis*) in Corio Bay, Victoria, Australia. *Mar. Pollut. Bull.* **1995**, *30* (8), 543–554.
- (6) Gale, R. W. Three-compartment model for contaminant accumulation by semipermeable membrane devices. *Environ. Sci. Technol.* **1998**, *32* (15), 2292–2300.
- (7) Bartkow, M. E.; Booiij, K.; Kennedy, K. E.; Müller, J. P.; Hawker, D. W. Passive air sampling theory for semivolatile organic compounds. *Chemosphere* **2005**, *60* (2), 170–176.
- (8) Huckins, J. N.; Petty, J. D.; Orazio, C. E.; Lebo, J. A.; Clark, R. C.; Gibson, V. L.; Gala, W. R.; Echols, K. R. Determination of uptake kinetics (sampling rates) by lipid-containing semipermeable membrane devices (spmds) for polycyclic aromatic hydrocarbons (pahs) in water. *Environ. Sci. Technol.* **1999**, *33* (21), 3918–3923.
- (9) Fedorova, G.; Randak, T.; Golovko, O.; Kodes, V.; Grabicova, K.; Grabic, R. A passive sampling method for detecting analgesics, psycholeptics, antidepressants and illicit drugs in aquatic environments in the Czech Republic. *Sci. Total Environ.* **2014**, *487*, 681–687.
- (10) Bailly, E.; Levi, Y.; Karolak, S. Calibration and field evaluation of polar organic chemical integrative sampler (POCIS) for monitoring pharmaceuticals in hospital wastewater. *Environ. Pollut.* **2013**, *174*, 100–105.
- (11) Harman, C.; Reid, M.; Thomas, K. V. In situ calibration of a passive sampling device for selected illicit drugs and their metabolites in wastewater, and subsequent year-long assessment of community drug usage. *Environ. Sci. Technol.* **2011**, *45* (13), 5676–5682.
- (12) Martínez Bueno, M. J.; Herrera, S.; Munaron, D.; Boillot, C.; Fenet, H.; Chiron, S.; Gómez, E. POCIS passive samplers as a monitoring tool for pharmaceutical residues and their transformation products in marine environment. *Environ. Sci. Pollut. Res.* **2016**, *23*, 1–11.
- (13) Baz-Lomba, J. A.; Reid, M. J.; Thomas, K. V. Target and suspect screening of psychoactive substances in sewage-based samples by UHPLC-QTOF. *Anal. Chim. Acta* **2016**, *914*, 81–90.
- (14) Mazzella, N.; Lissalde, S.; Moreira, S.; Delmas, F.; Mazellier, P.; Huckins, J. N. Evaluation of the use of performance reference compounds in an oasis-hlb adsorbent based passive sampler for improving water concentration estimates of polar herbicides in freshwater. *Environ. Sci. Technol.* **2010**, *44* (5), 1713–1719.
- (15) Harman, C.; Allan, I. J.; Bäuerlein, P. S. The challenge of exposure correction for polar passive samplers the PRC and the POCIS. *Environ. Sci. Technol.* **2011**, *45* (21), 9120–9121.
- (16) Harman, C.; Booiij, K. Letter to the editor concerning the viewpoint; “Recognizing the limitations of performance reference compound (prc)-calibration technique in passive water sampling. *Environ. Sci. Technol.* **2013**, *48* (1), 3–3.
- (17) Bartelt-Hunt, S. L.; Snow, D. D.; Damon-Powell, T.; Brown, D. L.; Prasai, G.; Schwarz, M.; Kolok, A. S. Quantitative evaluation of laboratory uptake rates for pesticides, pharmaceuticals, and steroid hormones using POCIS. *Environ. Toxicol. Chem.* **2011**, *30* (6), 1412–1420.
- (18) Stephens, B. S.; Kapernick, A.; Eaglesham, G.; Mueller, J. Aquatic passive sampling of herbicides on naked particle loaded membranes: accelerated measurement and empirical estimation of kinetic parameters. *Environ. Sci. Technol.* **2005**, *39* (22), 8891–8897.
- (19) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: the correlation and quantitative prediction of chemical and physical properties from structure. *Chem. Soc. Rev.* **1995**, *24* (4), 279–287.
- (20) Toropov, A. A.; Toropova, A. P.; Lombardo, A.; Roncaglioni, A.; Benfenati, E.; Gini, G. CORAL: Building up the model for bioconcentration factor and defining its applicability domain. *Eur. J. Med. Chem.* **2011**, *46* (4), 1400–1403.
- (21) Padmanabhan, J.; Parthasarathi, R.; Subramanian, V.; Chattaraj, P. K. QSPR models for polychlorinated biphenyls: n-Octanol/water partition coefficient. *Bioorg. Med. Chem.* **2006**, *14* (4), 1021–1028.
- (22) Bade, R.; Bijlsma, L.; Sancho, J. V.; Hernández, F. Critical evaluation of a simple retention time predictor based on LogKow as a complementary tool in the identification of emerging contaminants in water. *Talanta* **2015**, *139*, 143–149.
- (23) Miller, T. H.; Musenga, A.; Cowan, D. A.; Barron, L. P. Prediction of chromatographic retention time in high-resolution anti-doping screening data using artificial neural networks. *Anal. Chem.* **2013**, *85* (21), 10330–10337.
- (24) Morin, N.; Camilleri, J.; Cren-Olivé, C.; Coquery, M.; Miège, C. Determination of uptake kinetics and sampling rates for 56 organic micropollutants using “pharmaceutical” POCIS. *Talanta* **2013**, *109*, 61–73.
- (25) Munro, K.; Miller, T. H.; Martins, C. P. B.; Edge, A. M.; Cowan, D. A.; Barron, L. P. Artificial neural network modelling of pharmaceutical residue retention times in wastewater extracts using gradient liquid chromatography-high resolution mass spectrometry data. *J. Chromatogr. A* **2015**, *1396*, 34–44.
- (26) Luan, F.; Xue, C.; Zhang, R.; Zhao, C.; Liu, M.; Hu, Z.; Fan, B. Prediction of retention time of a variety of volatile organic compounds based on the heuristic method and support vector machine. *Anal. Chim. Acta* **2005**, *537* (1–2), 101–110.
- (27) Barron, L.; McEneff, G. Gradient liquid chromatographic retention time prediction for suspect screening applications: A critical assessment of a generalised artificial neural network-based approach across ten multi-residue reversed-phase analytical methods. *Talanta* **2015**, *147* (147), 261–270.
- (28) Fauvelle, V.; Mazzella, N.; Delmas, F.; Madarassou, K.; Eon, M.; Budzinski, H. Use of mixed-mode ion exchange sorbent for the passive sampling of organic acids by polar organic chemical integrative sampler (POCIS). *Environ. Sci. Technol.* **2012**, *46* (24), 13344–13353.
- (29) Morin, N.; Miège, C.; Coquery, M.; Randon, J. Chemical calibration, performance, validation and applications of the polar organic chemical integrative sampler (POCIS) in aquatic environments. *TrAC, Trends Anal. Chem.* **2012**, *36*, 144–175.
- (30) Bade, R.; Bijlsma, L.; Miller, T. H.; Barron, L. P.; Sancho, J. V.; Hernández, F. Suspect screening of large numbers of emerging contaminants in environmental waters using artificial neural networks for chromatographic retention time prediction and high resolution



mass spectrometry data analysis. *Sci. Total Environ.* **2015**, 538, 934–941.

(31) Booij, K.; Hofmans, H. E.; Fischer, C. V.; Van Weerlee, E. M. Temperature-dependent uptake rates of nonpolar organic compounds by semipermeable membrane devices and low-density polyethylene membranes. *Environ. Sci. Technol.* **2003**, 37 (2), 361–366.

(32) Mazzella, N.; Dubernet, J.-F.; Delmas, F. Determination of kinetic and equilibrium regimes in the operation of polar organic chemical integrative samplers: Application to the passive sampling of the polar herbicides in aquatic environments. *J. Chromatogr. A* **2007**, 1154 (1), 42–51.

(33) Thomatou, A.-A.; Zacharias, I.; Hela, D.; Konstantinou, I. Passive sampling of selected pesticides in aquatic environment using polar organic chemical integrative samplers. *Environ. Sci. Pollut. Res.* **2011**, 18 (7), 1222–1233.

(34) Ibrahim, I.; Togola, A.; Gonzalez, C. Polar organic chemical integrative sampler (POCIS) uptake rates for 17 polar pesticides and degradation products: laboratory calibration. *Environ. Sci. Pollut. Res.* **2013**, 20 (6), 3679–3687.

(35) Moschet, C.; Vermeirssen, E. L. M.; Singer, H.; Stamm, C.; Hollender, J. Evaluation of in-situ calibration of Chemcatcher passive samplers for 322 micropollutants in agricultural and urban affected rivers. *Water Res.* **2015**, 71, 306–317.

(36) Bäuerlein, P. S.; Mansell, J. E.; ter Laak, T. L.; de Voogt, P. Sorption behavior of charged and neutral polar organic compounds on solid phase extraction materials: which functional group governs sorption? *Environ. Sci. Technol.* **2012**, 46 (2), 954–961.

(37) Vrana, B.; Mills, G. A.; Kotterman, M.; Leonards, P.; Booij, K.; Greenwood, R. Modelling and field application of the Chemcatcher passive sampler calibration data for the monitoring of hydrophobic organic pollutants in water. *Environ. Pollut.* **2007**, 145 (3), 895–904.

(38) Rusina, T. P.; Smedes, F.; Klanova, J.; Booij, K.; Holoubek, I. Polymer selection for passive sampling: A comparison of critical properties. *Chemosphere* **2007**, 68 (7), 1344–1351.

(39) Gramatica, P.; Corradi, M.; Consonni, V. Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. *Chemosphere* **2000**, 41 (5), 763–777.

(40) Barron, L.; Havel, J.; Purcell, M.; Szpak, M.; Kelleher, B.; Paull, B. Predicting sorption of pharmaceuticals and personal care products onto soil and digested sludge using artificial neural networks. *Analyst* **2009**, 134 (4), 663–670.

(41) Rusina, T. P.; Smedes, F.; Koblizkova, M.; Klanova, J. Calibration of silicone rubber passive samplers: Experimental and modeled relations between sampling rate and compound properties. *Environ. Sci. Technol.* **2010**, 44 (1), 362–367.

(42) Nightingale, E., Jr. Phenomenological theory of ion solvation. Effective radii of hydrated ions. *J. Phys. Chem.* **1959**, 63 (9), 1381–1387.

(43) Schramke, J. A.; Murphy, S. F.; Doucette, W. J.; Hintze, W. D. Prediction of aqueous diffusion coefficients for organic compounds at 25 C. *Chemosphere* **1999**, 38 (10), 2381–2406.

(44) La-Scalea, M. A.; Menezes, C. M. S.; Ferreira, E. I. Molecular volume calculation using AM1 semi-empirical method toward diffusion coefficients and electrophoretic mobility estimates in aqueous solution. *J. Mol. Struct.: THEOCHEM* **2005**, 730 (1), 111–120.

(45) Gharagheizi, F. Determination of diffusion coefficient of organic compounds in water using a simple molecular-based method. *Ind. Eng. Chem. Res.* **2012**, 51 (6), 2797–2803.

(46) Hawkins, D. M. The problem of overfitting. *J. Chem. Inf. Comp. Sci.* **2004**, 44 (1), 1–12.

(47) Harman, C.; Allan, I. J.; Vermeirssen, E. L. M. Calibration and use of the polar organic chemical integrative sampler—a critical review. *Environ. Toxicol. Chem.* **2012**, 31 (12), 2724–2738.

(48) Bowden, G. J.; Maier, H. R.; Dandy, G. C. Optimal division of data for neural network models in water resources applications. *Water Resour. Res.* **2002**, 38 (2), 2–1–2–11.

(49) Alvarez, D. A.; Petty, J. D.; Huckins, J. N.; Jones-Lepp, T. L.; Getting, D. T.; Goddard, J. P.; Manahan, S. E. Development of a passive, in situ, integrative sampler for hydrophilic organic

contaminants in aquatic environments. *Environ. Toxicol. Chem.* **2004**, 23 (7), 1640–1648.