Beyond activated carbon properties and hydrophobicity

Data-driven assessment of organic micro-pollutant treatability and mechanistic insights

Wang, Zichu; Wang, Qi; Kyritsakas, Greg; Yang, Min; Yu, Jianwei; Rietveld, Luuk C.

DOI
10.1016/j.watres.2025.124079

Publication date
2025

Document Version
Final published version

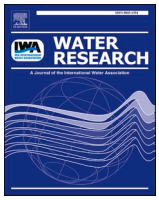
Published in
Water Research

Water Research 285 (2025) 124079

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres



Beyond activated carbon properties and hydrophobicity: Data-driven assessment of organic micro-pollutant treatability and mechanistic insights

Zichu Wang [a,b,d,1], Qi Wang [a,b,1,*], Greg Kyritsakas [c], Min Yang [a,b,d], Jianwei Yu [a,b,d,*], Luuk C. Rietveld [c]

[a] National Engineering Research Center of Industrial Wastewater Detoxication and Resource Recovery, Research Center for Eco-Environmental Sciences, Beijing 100085, China
[b] State Key Laboratory of Environmental Aquatic Chemistry, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, 100085, Beijing, China
[c] Delft University of Technology, Department of Water Management, PO Box 5048, 2600 GA Delft, the Netherlands
[d] University of Chinese Academy of Sciences, 100049, Beijing, China

ARTICLE INFO

Keywords:
Adsorption potential assessment
Interpretable machine learning
Adsorbability
Hydrophobicity
Alignment
Activated carbon adsorption

ABSTRACT

Activated carbon (AC) is widely used for organic micro-pollutants (OMPs) removal, yet adsorbability evaluation remains challenging due to molecular diversity and adsorbent heterogeneity, especially given the limitations of traditional assessment metrics such as hydrophobicity (logD). This study proposed a machine learning (ML)-driven assessment strategy by aligning the adsorbability of various AC adsorbents with a hypothetical "Standard AC" to evaluate the adsorbabilities across 56 OMPs. XGBoost, RF, and ET models achieved high prediction accuracy on the test set (R^2 = 0.88–0.98, RMSE = 0.17–0.38, MAE = 0.13–0.27), and were further validated against a published experimental dataset. Interpretable ML analysis identified a logD threshold of ≈ 2, at which the dominant adsorption mechanisms transitioned from hydrophobic interactions for OMPs with higher hydrophobicity to π-π interactions, hydrogen bonding, and pore-filling for those with lower hydrophobicity. Adsorbability increased with molecular weight, as flexible molecules (rotatable bond ratio > 0.012) overcame steric hindrance in micropores, enhancing pore-filling efficiency through improved accessibility. By introducing a standardized, data-driven adsorbability reference and elucidating the intrinsic interplay between molecular properties and adsorption mechanisms, this study offers a robust framework for knowledge-informed treatability evaluation and a practical benchmark to guide adsorption process design.

1. Introduction

Activated carbon (AC) has been widely used as an effective method for removing a broad range of organic micro-pollutants (OMPs) (Bhatnagar et al., 2013; Shannon et al., 2008), owing to high surface area (typically 800–1500 m^2/g) and well-developed microporous structure (pore size <2 nm), which enhance the adsorption of diverse OMPs. Although materials such as biochar and carbon nanotubes can also be engineered for high adsorption performance, AC offers superior practicality due to greater production capacity, stable supply, lower leaching risk compared to biochar, and cost advantage over carbon nanotubes, making it a more practical choice in water treatment applications.

The assessment of adsorbability and the screening of high-performance AC candidates have been active research areas since the early days of bulk industrial chemical production (e.g., toluene) through to the more recent focus on trace-level algae-producing odorants (e.g., 2-methylisoborneol) and mobile pharmaceuticals (e.g., ciprofloxacin) (Schumann et al., 2023; Schwarzenbach et al., 2006; Sobek et al., 2023; Zhang, 2023). However, the evaluation of OMP adsorbability is usually conducted on adsorbents with varying pore structure and surface chemistry, which results in the lack of a reliable reference list for AC adsorption of OMPs for the design and operation of adsorption facilities (Saeidi et al., 2025). While hydrophobicity, often quantified by the log transformed distribution coefficient (logD)—a pH-dependent measure of differential solubility between octanol and water—remains a widely

* Corresponding authors at: National Engineering Research Center of Industrial Wastewater Detoxication and Resource Recovery, Research Center for Eco-Environmental Sciences, Beijing 100085, China.
E-mail addresses: qiwang@rcees.ac.cn (Q. Wang), jwyu@rcees.ac.cn (J. Yu).
[1] The authors contributed equally to this work.

https://doi.org/10.1016/j.watres.2025.124079
Received 11 April 2025; Received in revised form 30 May 2025; Accepted 21 June 2025
Available online 27 June 2025

used and convenient predictor of adsorbability (Crini et al., 2018), its validity is largely limited to specific OMP classes (e.g., antibiotics, pharmaceuticals, algae-derived odorants), where hydrophobicity dominates adsorption behavior (Gagliano et al., 2020; Nam et al., 2014a; Pivokonsky et al., 2021). As structural complexity increases—such as in compounds with non-aromatic moieties or diverse polar functional groups—other mechanisms, including π-π interactions, hydrogen bonding, and pore-filling and surface functionalities, become more influential (Gayathiri et al., 2022; Grimm et al., 2025, 2023; Islam et al., 2024a, 2024b; Tong et al., 2019), while the relative role of hydrophobicity diminishes. This fundamental limitation stems from inability of logD to account for the full spectrum of adsorbate-adsorbent interactions governing AC adsorption systems (Aminul Islam et al., 2024).

To objectively determine the adsorption of various types of OMPs, a modeling technique capable of integrating multiple adsorption-driven forces, adsorbent properties, and background water quality factors is required. Machine learning (ML), with its ability to extract complex relationships from high-dimensional data, has been widely applied across environmental science and engineering fields, including removal prediction, material optimization, ecological analysis, and system interaction modeling (Han et al., 2023; Jia et al., 2023; Richards et al., 2023; Wang et al., 2022; Zhi et al., 2024). These techniques could potentially extend adsorption prediction beyond studies limited to a narrow range of datasets, such as those involving only one or two types of AC adsorbents and a few OMPs of a particular category with varying functional groups. Traditional adsorption models developed with the multiple linear regression (MLR) method are inadequate to describe the adsorbability of different categories of OMPs (Ersan et al., 2019; Giraudet et al., 2006). Existing ML adsorption models—mostly developed by pioneering environmental engineering data scientists—have exhibited strong predictive performance using large, heterogeneous adsorption datasets (Haider Jaffari et al., 2023; Zhang et al., 2020; Zhu et al., 2022). Advanced interpretative techniques, such as Shapley Additive Explanations (SHAP) and Partial Dependence Plots (PDPs), provide adsorption researchers with a more aligned perspective, enabling a deeper understanding of the interactions among key adsorption driving forces (e.g., hydrophobicity, π-π interactions, hydrogen bonding, and pore-filling) by quantifying feature contributions and interaction effects.

Herein, this study aims to establish a reliable reference of OMP adsorbability and provide mechanistic insights by applying an interpretable ML model to high-dimensional adsorption data. The main advances are specified as follows:

(i). Developed an ML-based framework that homogenizes adsorbability assessment across OMPs through a conceptual "Standard AC", which minimized the heterogenetic influence of 52 AC adsorbents (cf. Fig. 1).
(ii). Identified category-specific adsorption mechanisms by analyzing the differences in hydrophobic effects against π-π and hydrogen bonding interactions using interpretable ML approaches.
(iii). Investigated molecular flexibility as a key factor influencing micropore adsorption, emphasizing the role of structural adaptability in OMP adsorption.

By establishing a standardized, data-driven reference for OMP adsorbability, this study proposed a unified framework that enables consistent comparison across diverse OMPs and AC. Through interpretable ML analysis, we further revealed how key molecular features—including hydrophobicity, π-π, hydrogen bonding interactions, and conformational flexibility—jointly govern adsorption behavior. Together, these contributions support a knowledge-informed approach to evaluating OMP treatability and offer a practical basis for guiding adsorption process optimization.

2. Materials and methods

2.1. Data collection and feature selection

This study used single-component isothermal adsorption data, incorporating OMP properties, adsorbent characteristics, and equilibrium adsorption concentrations ($\log(C_e)$). The dataset covered 56 OMPs including PPCPs, odorants, pesticides, and industrial chemicals, and the adsorbents covered 52 ACs, which together comprised the adsorption isotherm dataset of 1557 data points to ensure robust model training and validation. In the dataset, the odorant adsorption data were obtained from powdered activated carbon (PAC) adsorption experiments in ultrapure water, which are local laboratory data (State Key Laboratory of Environmental Aquatic Chemistry, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences), while the data of other OMPs belonging to granular activated carbon is a collection

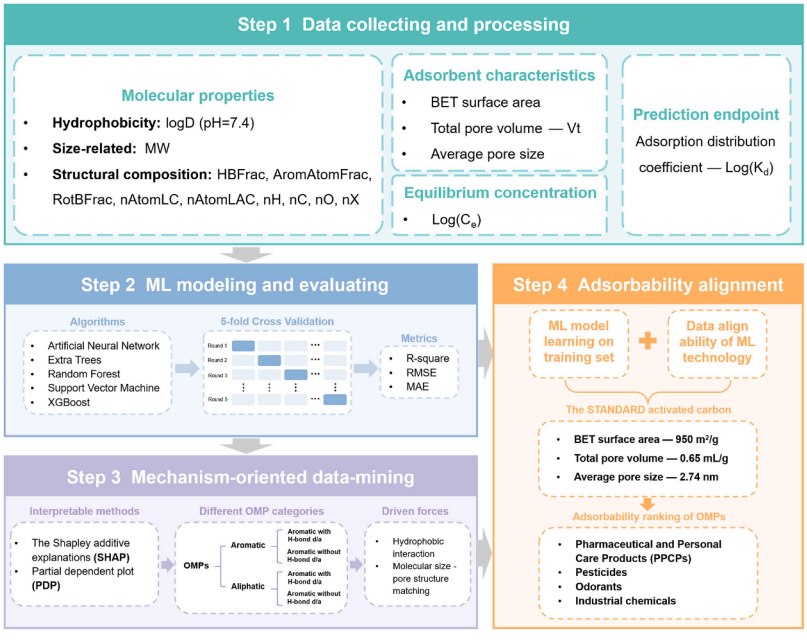


Fig. 1. Diagram of the ML prediction and interpretation framework for adsorbability assessment of organic micro-pollutants.

from existing literature (de Ridder et al., 2010). All equilibrium adsorption data were rigorously screened to include only systems where target OMPs were the only adsorbates in organic-free aqueous background (e.g., deionized water or ultrapure water). This selection protocol eliminates potential biases from competitive adsorption of high-concentration background organic matter in authentic water samples, ensuring direct comparability across all experimental and literature-derived data. The original dataset is available in Support Information.

Given the mechanistic interpretability is one of the core objectives, the feature screening process followed a strategy combining statistical redundancy reduction and interpretability-oriented selection. Descriptor collinearity was assessed using Pearson correlation analysis, and those with correlation coefficients above 0.8 were considered

Table 1
The evaluation of adsorbability of OMPs in pure water on AC predicted by XGBoost. "Evaluated log(K_d)" is the predicted value of log(K_d) under a certain C_e condition. Note that the predicted dosage referred to the required dose of F400 AC for 1 log removal (i.e., 90 %) of OMPs.

Category	Pollutant	logD (pH = 7.4)	Evaluated log(K_d) (C_e = 1 μmol/L)	Predicted dosage (mg/L) (C_e = 1 μmol/L)	Evaluated log(K_d) (C_e = 0.1 μmol/L)	Predicted dosage (mg/L) (C_e=0.1 μmol/L)
PPCPs	Phenyl salicylate	3.98	2.81	14.04	3.48	2.99
	Ciprofloxacin	−0.76	2.74	16.55	3.24	5.19
	Carbamazepine	2.77	2.70	18.13	3.44	3.29
	Sulfamethoxazole	−0.07	2.47	30.45	3.20	5.72
	Tetracycline	−0.66	2.43	33.06	2.99	9.29
	Tryptophan	−1.09	2.38	37.10	3.22	5.48
	Vanillin	1.08	2.20	57.41	3.05	8.02
	Salicylic acid	−1.52	1.85	127.62	2.99	9.13
	3, 5-Dimethoxybenzoic acid	−2.03	1.84	129.99	3.10	7.15
	D-phenylalanine	−1.19	1.83	131.69	3.13	6.60
	Acetylsalicylic acid	−2.16	1.64	205.20	3.10	7.21
Pesticides	Lindane	4.35	2.82	13.77	3.74	1.65
	Fluorobenzene	2.12	2.76	15.68	3.00	8.95
	2-Phenylphenol	3.32	2.71	17.48	3.42	3.44
	2, 4-Dichlorobenzoic acid	−0.63	2.52	27.34	3.34	4.11
	2, 3-Dichloroaniline	2.35	2.49	29.02	3.46	3.13
	1, 2-Dibromo-3-chloropropane	2.60	2.39	36.86	3.37	3.88
	Tetrachloroethylene	2.62	2.28	47.68	3.39	3.70
	4, 6-Dinitro-o-cresol	0.20	2.18	59.49	2.88	11.86
	1, 2-Dibromoethane	1.87	1.99	92.98	2.75	15.86
	1,1,1-Trichloroethane	2.08	1.79	144.43	2.56	24.93
Odorants	Geosmin	3.17	2.52	27.22	3.36	3.94
	2-Methylisoborneol	2.27	2.50	28.67	3.07	7.70
	Catechol	1.36	2.49	29.19	2.57	24.31
	2-Isopropyl-5,5-dimethyl-1,3-dioxane	2.39	2.44	32.69	2.88	11.76
	2, 5-Dimethylphenol	2.70	2.44	32.76	3.12	6.81
	2,3,5-Trimethylphenol	3.21	2.42	33.84	3.46	3.11
	2-Butyl-4-methyl-1,3-dioxolane	2.19	2.34	41.14	3.03	8.37
	2-Butyl-5,5-dimethyl-1,3-dioxane	2.80	2.33	41.82	3.29	4.67
	Methylphenol	2.18	2.28	46.81	2.96	9.95
	2-Ethyl-5,5-dimethyl-1,3-dioxane	1.91	2.16	62.31	2.62	21.52
	2-Isobutyl-4-methyl-1,3-dioxolane	2.03	1.99	91.57	3.02	8.69
Industrial Chemicals	1, 2, 4-Trichlorobenzene	3.79	2.92	10.75	3.65	2.03
	2, 4, 6-Trimethyl aniline	2.69	2.91	10.99	3.33	4.20
	4-Chlorophenol	2.26	2.91	11.18	3.43	3.34
	2-Chlorobiphenyl	4.23	2.88	11.86	3.52	2.74
	2-Nitrobiphenyl	3.56	2.80	14.26	3.44	3.25
	Biphenyl	3.62	2.79	14.61	3.51	2.80
	Nitrobenzene	1.91	2.68	18.67	3.12	6.83
	Toluene	2.49	2.52	26.88	3.43	3.36
	Ethylbenzene	2.93	2.46	31.40	3.46	3.15
	4-Tert-butylphenol	3.21	2.39	36.68	3.36	3.92
	4-Nitrotoluene	2.43	2.39	36.95	3.43	3.37
	Phenol	1.67	2.31	44.38	2.80	14.30
	Bis(2-chloro-1-methylethyl) ether	2.29	2.30	44.61	3.03	8.41
	Benzene	1.97	2.29	45.86	2.93	10.53
	Aniline	1.14	2.25	51.13	2.42	34.36
	2-Chloro-5-nitrobenzoic acid	−1.34	2.23	53.58	3.01	8.87
	1, 1, 1, 2-Tetrachloroethane	2.41	2.22	54.02	2.99	9.27
	Trichloroethene	2.18	2.20	57.33	2.84	13.08
	p-Anisidine	0.98	2.10	71.55	2.85	12.62
	1, 1, 2-Trichloroethane	2.17	2.07	76.45	2.79	14.71
	2-Carboxybenzaldehyde	−1.72	2.07	77.23	3.13	6.69
	2-Isopropyl-1,3-dioxolane	1.36	1.95	100.83	2.16	62.35
	3-Aminobenzoic acid	−1.78	1.84	129.33	2.97	9.75
	Methyl t-butyl ether	1.18	1.20	567.79	1.79	146.30

redundant. In parallel, SHAP values were calculated to quantify marginal contribution of each feature to model predictions (Fig. S1), capturing feature influence across varying input conditions. Descriptors exhibiting strong collinearity, low SHAP importance, or overlapping physicochemical interpretations were excluded. For example, topoRadius and TopoPSA were removed due to their high correlation with molecular weight (MW), redundant size-related meaning, and low SHAP rankings, as were ALogP and MLogP due to overlapping representation regarding hydrophobicity. However, descriptors with clear and distinct physical relevance to adsorption mechanisms were retained—even if their SHAP importance was moderate. For instance, the fraction of rotatable bonds (RotBFrac) was preserved for the unique role in describing molecular flexibility, a key factor influencing pore accessibility and adsorption. After screening, the input descriptors covered 11 organic molecular properties, three AC characteristics (BET surface area, total pore volume, and average pore size), and log(C_e), as detailed in Table 1.

Since pH conditions were not consistently available, logD at pH = 7.4 was selected as a standardized descriptor of hydrophobicity. The hydrophobicity (logD at pH = 7.4) was obtained from the cheminformatics website (http://www.chemicalize.com). Size-related molecular metrics (MW), and structural composition (nAtomLC, nAtomLAC, nH, nC, nO, nX) were obtained from PaDEL-Descriptor (Yap, 2010). Additionally, in order to better evaluate the role of hydrogen bonding, π-π interactions, and molecular rotatability, we constructed three new descriptors: HBFrac, the fraction of hydrogen bond acceptors or donors; AromAtomFrac, the fraction of aromatic atoms; and RotBFrac, the fraction of rotatable bond, which are detailed in Table S1. ue to inconsistencies in the reporting of experimental details across data sources, certain adsorbent-specific features—such as the surface chemistry of AC—could not be uniformly obtained and were thus not included in the input features.

Distribution coefficient (log (K_d)), which directly quantifies solute distribution between solid and liquid phases, was selected as the prediction endpoint. Unlike adsorption capacity (q_e), K_d is less influenced by experimental variables such as initial (C_0) and equilibrium (C_e) concentrations, making it more suitable for integrating data from multiple data sources. This choice also minimizes dependency on input C_e reducing potential bias and enhancing the predictive sensitivity to molecular and adsorbent descriptors. Moreover, since OMP concentrations in the dataset span several orders of magnitude (from ng/L to μg/L), K_d serves to mitigate the impact of data imbalance related to concentration variability and allows better utilization of literature data reported at higher concentration levels (e.g., in the mg/L range) with reduced bias. The C_e and q_e data were directly extracted from numerical tables in the reference publication (Zhang et al., 2020) and our laboratory records, and the corresponding K_d values were calculated as follows:

$$K_d = \frac{q_e}{C_e} \tag{1}$$

Here the sample-size to feature-size ratio was 98.63, much higher than the lowest acceptable threshold (10) and close to the best ratio (100), which thus can be considered sufficient data to accomplish valid predictions (Zhu et al., 2023).

2.2. Model development and evaluation metrics

We selected five algorithms for performance evaluation due to their reported efficiency in previous studies (Huang et al., 2023; Li et al., 2022; Mesellem et al., 2021; Pauletto et al., 2021; Zhu et al., 2021), specifically including artificial neural network (ANN), random forest (RF), support vector machine (SVM), extra tree (ET), extreme gradient boosting (XGBoost). The models were implemented using Python with the scikit-learn library. The model structures and complexity were carefully controlled to match the available dataset size (n = 1557). For example, the ANN used in this study is a shallow network with only one hidden layer, while tree-based models such as RF and ET were regularized by limiting tree depth and number of estimators. These construction limitations ensured that all models remained within a reasonable complexity range to avoid overfitting and maintain stable performance.

First, to ensure robust model development and to prevent data leakage from interconnections between adsorption data in the training and testing sets originating from the same isotherm, the dataset was partitioned based on units of isotherm (Yang et al., 2023; Zhang et al., 2020). Specifically, 90 % of the original data was allocated for model development (training and validation), while the remaining 10 % was reserved for testing. Random seeds were fixed during model training to ensure consistent dataset splitting across multiple runs, thereby supporting the reproducibility of this study. Due to the significant impact of separation seeds on data splitting results, we chose to randomly assign separation seeds 20 times for data separation and model training to determine the accuracy between the 5 algorithms. Then, the model was developed with 5-fold cross-validation and grid search to identify the best hyperparameter combinations for each algorithm, as detailed in the Supporting Information.

Three metrics were employed to evaluate model performance: 1) R-square (R^2), measuring the proportion of the variance in the dependent variable that is predictable from the independent variables, a key indicator of the goodness-of-fit of the model; 2) mean-absolute error (MAE), representing the average absolute difference between predicted and actual values, providing a straightforward measure of prediction accuracy; 3) root-mean-square error (RMSE), indicating the square root of the average squared differences between predicted and actual values, offering an objective measure of prediction error that penalizes larger discrepancies more heavily.

2.3. Adsorbability assessment methodology

To ensure a more accurate assessment of OMP adsorbability for water treatment applications, where lab-scale ultra-porous adsorbents may lead to overestimation, we introduced a hypothetical “Standard AC” as a reference for evaluating the adsorbability of 56 OMPs in the dataset. This hypothetical “Standard AC” is proposed based on a national regulation for AC products in water treatment (CJ/T 345–2010) (Ministry of Housing and Urban-Rural Development, 2010) and the widely used commercial AC, F400 (Calgon, USA), to ensure the representativeness and practical applicability of the “Standard AC” for realistic adsorption process evaluation in water treatment applications. Key parameters—BET surface area (950 m^2/g), total pore volume (0.65 mL/g), and average pore size (2.74 nm)—were selected to standardize adsorption predictions. The adsorption distribution coefficient log(K_d) values for the 56 OMPs were then predicted under these standardized conditions, ensuring a consistent and objective comparison of adsorbabilities across different compounds. By introducing a standardized adsorbent, this method ensured a fair and consistent comparison of OMP adsorbabilities.

The adsorbability evaluation was performed based on XGBoost, RF, and ET with fixed configurations that have been trained, validated, and tested by the original isothermal adsorption dataset (all uniformly with “random seed = 42”). Since the concentrations of these 56 OMPs in the actual water differ considerably, from ng/L to mg/L, log(K_d) was calculated under two scenarios to avoid extrapolation at concentration conditions where the data are less distributed: a high concentration level scenario (C_e = 1 μmol/L) and a low concentration level scenario (C_e = 0.1 μmol/L) (de Ridder et al., 2010). Furthermore, to visualize the differences in the ease of adsorption, the dosage of AC required to remove one log unit of each OMP (removed from C_0 = 10 μmol/L to C_e = 1 μmol/L at high concentration level, and removed from C_0 = 1 μmol/L to C_e = 0.1 μmol/L at low concentration level) were also shown by predicted log(K_d).

$$Dosage = \frac{C_0 - C_e}{q_e} = \frac{C_0 - C_e}{K_d \times C_e} \quad (2)$$

where C_0 is the initial concentration of solutes (μmol/L), q_e is the carbon loading of solutes (μmol/g).

In the validation of the actual adsorbent (F400 AC), the log(q_e) data for OMPs from the reference (de Ridder et al., 2010) were extracted for a total of 10 overlapping substances in our dataset. As described in Section "Assessment of organic micro-pollutant adsorbability", the well-trained models after algorithm screen (XGBoost, RF, ET) are used to predict the log(q_e) data of these 10 OMPs on F400 (BET = 986.03 m^2/g, Vt = 0.54 mL/g), to analyze the correlation between the reference data and our log (q_e) prediction results.

2.4. Interpretable methods

The SHAP and PDP approaches were used to interpret the adsorption mechanisms for the most accurately predicted algorithm among the five trained models. SHAP assigns a Shapley value to each feature, offering a detailed insight into feature importance. This approach allows for a comprehensive understanding of how each feature influences the prediction outcome (Cao et al., 2024; Haider Jaffari et al., 2023; Zhu et al., 2021). Additionally, the Shapley interaction index was used to enhance understanding of significant interaction effects among multidimensional features (Grabisch and Roubens, 1999; Tsai et al., 2023). This extension not only quantifies the joint influence of features on model output but also reveals complex dependencies that cannot be captured by SHAP values alone. Besides SHAP analysis, PDPs were used to illustrate how changes in feature values dynamically influence the model predictions. PDPs display the marginal effect of one or two features on the predicted outcome while accounting for the average effects of all other features. Two-factor PDPs, for example, reveal more complex bilateral dependencies by capturing interaction effects between two features, offering detailed numerical interpretations.

3. Results and discussion

3.1. Dataset statistics and model performance

3.1.1. Feature distribution and collinearity

The Pearson correlation coefficient, which measured the strength and direction of monotonic relationships between pairs of features, was calculated for all feature pairs of OMP properties. After screening, the input descriptors to the model covered 11 organic molecular properties, three AC characteristics (BET surface area, total pore volume, and average pore size), and the equilibrium adsorption concentration (log (C_e)), as detailed in Table S1. As shown in Fig. 2a, none of the 11 selected molecular descriptors exhibited a correlation coefficient (ρ) greater than 0.8, which was consistent with the model construction requirements and indicated a low risk of overfitting due to strong correlations. Additionally, the distributions of each feature were visualized using boxplots (Fig. S2), providing an overview of their central tendency, variability, and potential outliers.

3.1.2. Model performance

Fig. 2b shows the comparison of ANN, RF, ET, SVM, and XGBoost predicting log(K_d) from the training and testing set based on 11 molecular predictors, 3 adsorbent characteristics, and log(C_e) for the adsorption of OMPs on AC. Based on the average values of R^2, RMSE, and MAE across 20 random data splits, the predictive performance of the five models followed the order: XGBoost > ET > RF ≈ ANN > SVM. Given the relatively lower stability of ANN and SVM, the ensembled XGBoost, ET, and RF were considered as more robust models, with XGBoost emerging as the top performer (R^2 = 0.88–0.98, RMSE = 0.17–0.38, MAE = 0.13–0.27). This is due to the combination of several single learners, which, on the one hand, enhances robustness to noise and outliers in large datasets. On the other hand, employing voting and averaging among the individual learners reduces reliance on training data and minimizes the risk of overfitting (Ren et al., 2016; Sagi and Rokach, 2018). The RF algorithm shows strong performance due to its ensemble learning approach, which reduces overfitting by averaging multiple decision trees (Talekar, 2020). ET, a variant of RF, further enhances RF by increasing randomness in the split selection process, leading to improved generalization (Geurts et al., 2006; Jaxa-Rozen and

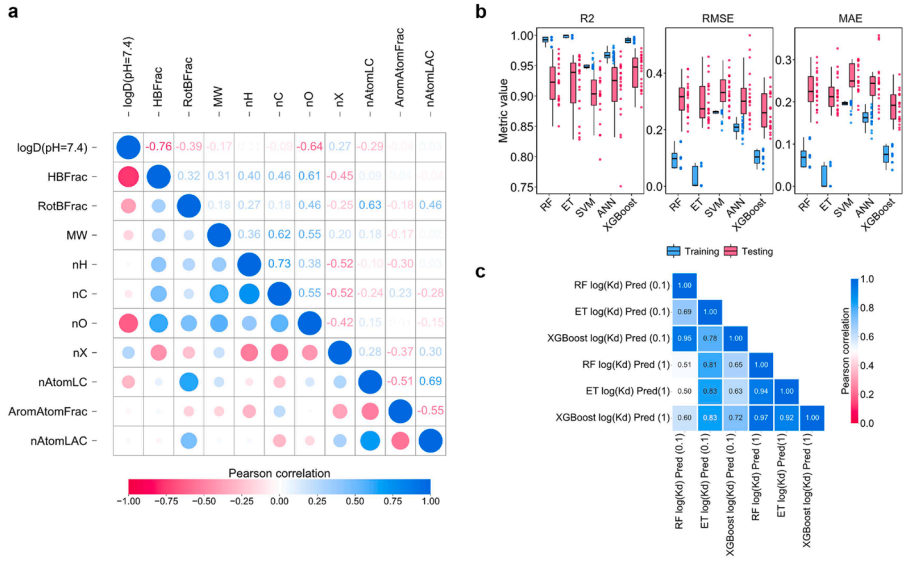


Fig. 2. a) Heatmap of Pearson correlation analysis between 11 selected molecular descriptors for organic micro-pollutants; b) Comparison of adsorption distribution coefficient log(K_d) predictive performance for training and test set by artificial neural network (ANN), extra tree (ET), random forest (RF), support vector machine (SVM), and extreme gradient boosting (XGBoost). The prediction results were obtained based on 20 times random splitting of the dataset; c) Heatmap of Pearson correlation analysis between the predicted values of adsorption distribution coefficient (log(K_d)) and logD (pH = 7.4) by the three algorithms for the adsorption equilibrium concentration (C_e = 0.1 μmol/L) and (C_e = 1 μmol/L).

Kwakkel, 2018). XGBoost, known for its gradient boosting framework, excels by effectively handling both bias and variance, optimizing performance through iterative learning (Bentéjac et al., 2020). Therefore, the subsequent comparison and validation were based on XGBoost, RF, and ET models, and interpretable analyses were conducted based on the XGBoost model (all uniformly with "random seed = 42").

3.2. Assessment of OMP adsorbability

As shown in Fig. 2c, the Pearson correlation analysis indicated that XGBoost, RF, and ET models consistently predicted log(K_d) across both concentration levels (C_e = 0.1 μmol/L and C_e = 1 μmol/L), with correlations reaching 0.97, indicating that the proposed models can provide a reliable assessment of OMP adsorption performance. Table 1 presents a ranked list of OMP adsorbability based on the "Standard AC" assessment, which was broadly categorized into pharmaceuticals and personal care products (PPCPs), pesticides, odorants, and industrial chemicals. This methodological reference list allowed water treatment practitioners to conduct a preliminary assessment of the OMP treatability by AC dosing before considering site-specific competitive effects from natural organic matter (NOM) (Nam et al., 2014b; Newcombe et al., 1997; Pelekani and Snoeyink, 1999) and decide if oxidation is required to enhance the OMP abatement and guarantee the safety of produced water. Moreover, environmental regulators and policymakers can leverage this list and the associated model to identify OMPs with low adsorbability, providing a foundation for targeted control measures.

Further, to validate the accuracy of our adsorbability assessment of OMPs on the real AC, the results were validated with a previous study that has well-documented OMP loading (log(q_e)) data on F400 AC (de Ridder et al., 2010). Fig. 3 shows the validation results of the adsorbability assessment against an actual ranking on F400 AC. The carbon loading data (log(q_e)) predictions (calculated by the predicted log(K_d) on F400 AC) of all three models at high concentration levels exhibited substantial concordance with one another with regard to the reported log(q_e) for OMPs, with the correlation coefficient (ρ) exceeding 0.78 and the highest predictive agreement of 0.91 for ET. At the low concentration level, the correlation coefficient (ρ) between RF and reported log (q_e) was 0.81.

An analysis of the adsorption ranking in Table 1 revealed that the low logD of sulfamethoxazole, indicative of a weak hydrophobic interaction with the surface of AC, is insufficient to explain its high adsorbability at both high and low concentration levels. π-π interactions in a moderately alkaline aqueous environment, a higher pore permeability due to smaller molecular sizes, and dimerization reactions resulting from the formation of oxygen-containing functional groups may be important factors that allow sulfamethoxazole to have a high adsorbability while having a low hydrophobicity (Adeyanju et al., 2022; Nam et al., 2014a; Nielsen et al., 2014; Zhu et al., 2021). Similarly, tryptophan (Belhamdi et al., 2019; Farbun and Trykhlib, 2023) and acetylsalicylic acid (Rakic et al., 2015), despite their low logD values (−1.09 and −2.16, respectively), exhibited high adsorbability due to π-π interactions facilitated by their aromatic structure.

At low concentration levels, a group of emerging odorants, namely 2-butyl-5,5-dimethyl-1,3-dioxane (2-BDD), 2-butyl-4-methyl-1,3-dioxolane (2-BMD), 2-isobutyl-4-methyl-1,3-dioxolane (2-IMD), 2-isopropyl-5,5-dimethyl-1,3-dioxane (2-IDD), 2-ethyl-5,5-dimethyl-1,3-dioxane (2-EDD) exhibited a moderate positive correlation (Pearson correlation = 0.84) between hydrophobicity and removal efficiency, which may be attributed to their similar saturated heterocyclic structures leading to the same adsorption driving force (Liu et al., 2025; Wang et al., 2024), making the strength of hydrophobicity a crucial factor in their adsorbability. In addition, odorants often occur at trace concentrations (ng/L) in natural waters (Lu et al., 2023; Whelton and Dietrich, 2004), limiting the availability of data on these substances at higher concentration levels. Consequently, the adsorption capacities at high concentrations, as shown in Table 1, differed substantially from those at low concentrations. The lack of opportunities to evaluate odorant adsorption at high concentration levels deteriorated the reliability of the extrapolated assessment at C_e = 1 μmol/L (and thus reduced the correlation in Fig. 2c and Fig. 3).

3.3. Hydrophobicity unreliably predicts adsorbability

The analysis indicated that logD is not a reliable predictor for the adsorption capacity of OMPs on AC, although the calculation of logD considered the actual dissociation state of OMPs, it is not a reliable predictor for the evaluation of the adsorption capacity of OMPs on AC. Fig. 4a shows the poor correlation between the actual hydrophobicity (logD) of OMPs and the predicted values of log(K_d), especially for the adsorption at the low concentration level (Pearson R^2 = 0.26 - 0.44), and only reached 0.65 - 0.72 at high pollutant concentration levels. In addition, Fig. 4b-c show the linear relationship between logD and log (K_d) predicted on the "Standard AC" with the XGBoost model for the four classes of OMPs. At high concentrations, the Pearson correlation coefficient was 0.65, while at low concentrations, it was only 0.39. This suggested that the explanatory power of logD is strong for pollutants with limited structural variability but fails to extend effectively across different categories (Chang et al., 2015; Mariangela Grassi, 2012). As previously discussed, the observation of efficient removal of some hydrophilic OMPs reinforced the inadequacy of logD as a standalone predictor (de Ridder et al., 2012; Liu et al., 2017; Nam et al., 2014a).

The local SHAP values at each data point on the logD scale provided a

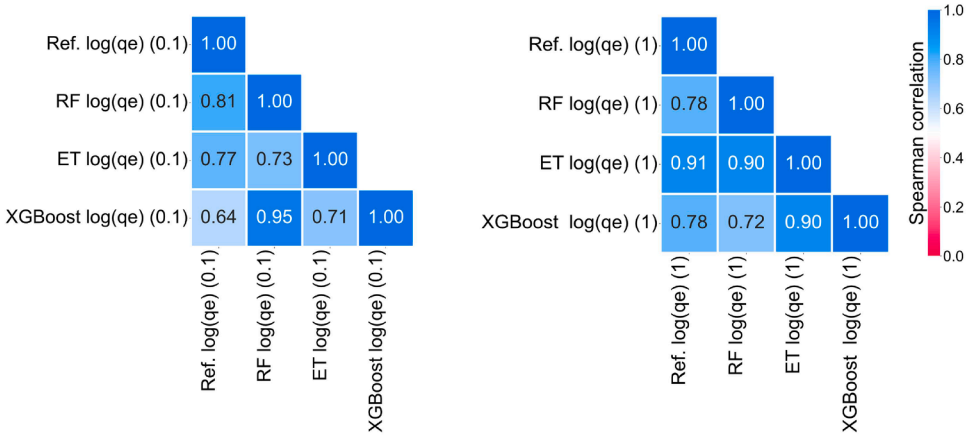


Fig. 3. Spearman correlation analysis of predicted loading (log(q_e)) between three models of this work and reference data for the adsorption equilibrium concentration (C_e = 0.1 μmol/L) and (C_e = 1 μmol/L). The reference data is from the article published by de Ridder et al. (de Ridder et al., 2010).

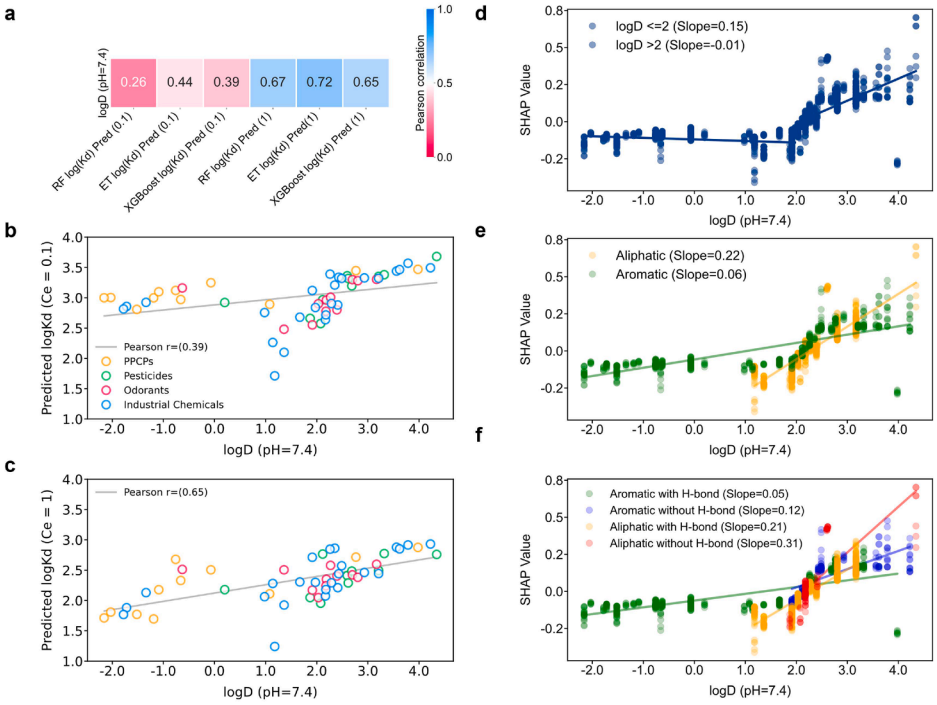


Fig. 4. a) Heat map of Pearson correlation between logD (pH = 7.4) and log(K_d) predicted by RF, ET, and XGBoost at low (C_e = 0.1μmol/L) and high (C_e = 1 μmol/L) concentration level; b-c) Scatterplot between logD and log(K_d) predicted by XGBoost for PPCPs, pesticides, odorants, and industrial chemicals at b) low concentration (C_e = 0.1μmol/L) and c) high concentration (C_e = 1 μmol/L), where "r" in the legend is the Pearson correlation coefficient; (d-f) SHAP Partial dependence plot of logD to reveal the hydrophobicity effect on adsorption: Organic micro-pollutants categorized into d) logD > 2 and logD ≤ 2; e) aromatic and aliphatic solutes; f) four groups according to aromaticity and H-bond donor/accepter.

more detailed description of the dependence on hydrophobicity interaction. Fig. 4d-f demonstrate the dependence of log(K_d) on hydrophobicity (logD) using Shapley values, with positive values representing the promotion of adsorption by hydrophobicity and negative values representing inhibition of adsorption. We further classified the OMPs according to the presence or absence of aromatic rings and H-bond donor/accepter groups to visualize the difference in dependence on adsorption. In Fig. 4e, the slope of the fitted trendline for the SHAP value of logD for aliphatic solutes was 0.22, higher than that for aromatic solutes (only 0.06), likely due to the fact that the aliphatic solutes lack the driving force of π-π interaction and have limited hydrogen bonding interactions, are more dependent on hydrophobicity. In Fig. 4f, aromatic solutes with H-bond donor/accepter groups (green dots) and aliphatic solutes without H-bond donor/accepter groups (red dots) exhibited the lowest and highest SHAP slopes, with values of 0.05 and 0.31, respectively, suggesting that hydrophobicity contributes the least to the adsorption of organic solutes with a variety of driving forces, while in the absence of hydrogen bonding or π-π interactions, the importance of hydrophobicity increased substantially (Park et al., 2020).

In Fig. 4d, the results indicated that for solutes with a logD <2, the SHAP values ranged from −0.4 to 0.0, with a minimal overall variation. This suggests that the hydrophobicity of these organic solutes has a weak negative effect on their adsorption, but the effect is not significant and does not play a major role in adsorption. In contrast, when logD > 2, the SHAP values showed a monotonically increasing trend with increasing logD in a wider range of 0.0–0.8, reflecting the increasing dependence on the hydrophobicity. Although the current boundary value of logD = 2 attained from data mining is not necessarily accurate, it is nonetheless clear that hydrophobicity manifests a facilitating effect on the adsorption process of only the OMPs with high logD values. The adsorption process of hydrophilic OMPs, on the other hand, may be dominated by pore-filling, hydrogen bonding, or π-π interaction (Liu et al., 2017; Moreno-Castilla, 2004). Additionally, the PDPs of SHAP values for each category with high hydrophobicity (logD ≥ 2) were analyzed separately, as shown in Fig. S4. The trends for each category were consistent with the results shown in Fig. 4b and 4c.

3.4. *Size-structure matching in determining adsorption*

Through size-exclusion and pore-filling effects, the interaction between the molecular weight (MW) of OMPs, the average pore size of AC, and the fraction of rotatable bonds (RotBFrac) are also crucial for OMP adsorbability by influencing molecular accessibility to pores, steric hindrance, and adsorption affinity (Gayle et al., 1997; Guo et al., 2008, 2007; Karanfil and Dastgheib, 2004; Quinlivan et al., 2005). According to Fig. 5a, the adsorption performance of OMPs increased with MW, across the average pore size range. Larger OMPs enhance van der Waals interactions and provide a greater surface affinity area with more accessible adsorption sites, thereby increasing their adsorbability within micropores and narrow mesopores through improved pore-filling and steric stabilization (Gayle et al., 1997; Liu et al., 2017). In addition, log (K_d) reached the highest predicted adsorption performance when the average pore size was < 2 nm, which indicated that a well-developed microporous structure (< 2 nm) is most favorable for OMP adsorption due to enhanced pore-filling effects. Furthermore, compared to aromatic OMPs with logD > 2, the high SHAP ranking of average pore size for aromatic OMPs with logD < 2 (cf. Fig. S5) suggests a stronger reliance on pore-filling mechanisms, as the absence of dominant hydrophobic interactions necessitates alternative adsorption pathways.

Another key factor limiting the effectiveness of adsorption of larger solutes is pore accessibility (Chang et al., 2015; Moreno-Castilla, 2004). As shown in Fig. 5b, the fraction of rotatable bond (normalized by MW) greater than 0.012 improved the OMP adsorption, especially in the micropore range (< 4 nm). Larger molecules with high flexibility, indicated by the fraction of rotatable bonds, can navigate and adsorb effectively in both micropores and mesopores, whereas rigid molecules might be restricted to larger pore spaces due to steric hindrance. This finding, on the one hand, aligns with the intuitive understanding that

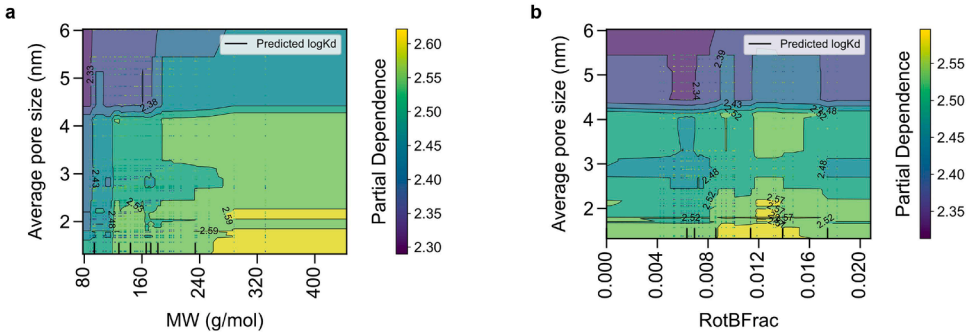


Fig. 5. Two-factor partial dependence plot (PDP) of molecular structure and pore size: a) PDP between molecular weight (MW) and average pore size of AC; b) PDP between the average pore size of AC and the rotatable bond fraction (RotBFrac).

the increased flexibility allows these molecules to better adapt to the pore structures, thereby enhancing the interaction with the adsorbent surface (Apul et al., 2013; Wang et al., 2024). On the other hand, the improvement of pore accessibility by molecular flexibility confirms that the pore-filling effect of OMPs during adsorption is more concentrated in micropores and less in mesopores.

4. Conclusions

- The proposed adsorbability list with hypothetical "Standard AC" elevates the treatability evaluation of AC adsorption to a new level—comparable to the use of kinetics constant lists in oxidation processes. The XGBoost, RF, and ET models demonstrated robust predictive performance for OMP adsorption, with XGBoost achieving the highest accuracy. The integrated molecular descriptors, AC properties, and equilibrium concentrations enable cross-comparison of OMP adsorbabilities beyond traditional hydrophobicity-based assessments.
- Interpretable ML analysis identified a clear transition in adsorption mechanisms at logD $\approx$ 2, below which adsorption mechanisms transitioned from hydrophobic interactions to π-π interactions, hydrogen bonding, and pore-filling. Furthermore, AC with an average pore size below 2 nm exhibited superior adsorption efficiency, particularly for larger, flexible molecules (rotatable bond ratio $>$ 0.012), which demonstrated enhanced pore accessibility due to conformational flexibility.
- ML techniques show potential of extending beyond prediction of OMP removal to serve an earlier strategic role in guiding process configuration and decision-making. However, their direct application to natural waters faces limitations due to site-specific competition from background organic matter and unique solution chemistry conditions, such as varying pH and ionic strength. To overcome these challenges, future work must prioritize the establishment of comprehensive datasets for multi-solute adsorption systems. This will be coupled with the crucial expansion of current models to accommodate a more diverse array of AC adsorbents, including chemically activated and surface-functionalized materials. These advances collectively hold immense promise for substantially improving the understanding and resolution of site-specific adsorption problems in water treatment.

Data and code availability statements

All data supporting the findings of this study are available from Supplementary Materials. The core codes for the model comparison and training, and interpretable analysis are available at https://github.com/MitchPrince/ML_adsorption_RCEES. Further codes for analyzing and producing other results are available from the corresponding authors upon request.

CRediT authorship contribution statement

Zichu Wang: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Qi Wang:** Writing – review & editing, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Greg Kyritsakas:** Writing – review & editing, Validation, Software. **Min Yang:** Writing – review & editing, Supervision, Resources, Project administration. **Jianwei Yu:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition. **Luuk C. Rietveld:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

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

Acknowledgements

This work was supported by funds for the National Natural Science Foundation of China (52400018), the National Key R&D Program of China (2022YFC3203603), and the 2024 Jinan Science and Technology Plan, China (202428075).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2025.124079.

Data availability

Data will be made available on request.

References

Adeyanju, C.A., Ogunniyi, S., Selvasembian, R., Oniye, M.M., Ajala, O.J., Adeniyi, A.G., Igwegbe, C.A., Ighalo, J.O., 2022. Recent advances on the aqueous phase adsorption of Carbamazepine. Chembioeng Rev. 9 (3), 231–247.

Aminul Islam, M., Nazal, M.K., Sajid, M., Altahir Suliman, M., 2024. Adsorptive removal of paracetamol from aqueous media: a review of adsorbent materials, adsorption mechanisms, advancements, and future perspectives. J. Mol. Liq. 396.

Apul, O.G., Wang, Q.L., Zhou, Y., Karanfil, T., 2013. Adsorption of aromatic organic contaminants by graphene nanosheets: comparison with carbon nanotubes and activated carbon. Water Res. 47 (4), 1648–1654.

Belhamdi, B., Zoulikha, M., Hamza, L., Trari, M., 2019. The removal and adsorption mechanisms of free amino acid 1-tryptophan from aqueous solution by biomass-based activated carbon by H_3PO_4 activation: regeneration study. Phys. Chem. Earth 114.

Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2020. A comparative analysis of gradient boosting algorithms. Artif. Intell. Rev. 54 (3), 1937–1967.

Bhatnagar, A., Hogland, W., Marques, M., Sillanpää, M., 2013. An overview of the modification methods of activated carbon for its water treatment applications. Chem. Eng. J. 219, 499–511.
Cao, X., Huang, J., Du, K., Tian, Y., Hu, Z., Luo, Z., Wang, J., Guo, Y., 2024. Machine-learning-assisted descriptors identification for indoor formaldehyde oxidation catalysts. Environ. Sci. Technol. 58 (19), 8372–8379.
Chang, E.E., Wan, J.-C., Kim, H., Liang, C.-H., Dai, Y.-D., Chiang, P.-C., 2015. Adsorption of selected pharmaceutical compounds onto activated carbon in dilute aqueous solutions exemplified by acetaminophen, diclofenac, and sulfamethoxazole. Sci. World J. 2015, 1–11.
Crini, G., Lichtfouse, E., Wilson, L.D., Morin-Crini, N., 2018. Conventional and non-conventional adsorbents for wastewater treatment. Environ. Chem. Lett. 17 (1), 195–213.
de Ridder, D.J., Verliefde, A.R.D., Heijman, B.G.J., Gelin, S., Pereira, M.F.R., Rocha, R.P., Figueiredo, J.L., Amy, G.L., van Dijk, H.C., 2012. A thermodynamic approach to assess organic solute adsorption onto activated carbon in water. Carbon N Y 50 (10), 3774–3781.
de Ridder, D.J., Villacorte, L., Verliefde, A.R.D., Verberk, J.Q.J.C., Heijman, S.G.J., Amy, G.L., van Dijk, J.C., 2010. Modeling equilibrium adsorption of organic micropollutants onto activated carbon. Water Res. 44 (10), 3077–3086.
Ersan, G., Apul, O.G., Karanfil, T., 2019. Predictive models for adsorption of organic compounds by Graphene nanosheets: comparison with carbon nanotubes. Sci. Total Environ. 654, 28–34.
Farbun, I.A., Trykhlib, V.A., 2023. Selectivity of tryptophan and phenylalanine adsorption by activated coconut carbon of medical assignment. Theor. Exp. Chem. 59 (1), 59–65.
Gagliano, E., Sgroi, M., Falciglia, P.P., Vagliasindi, F.G.A., Roccaro, P., 2020. Removal of poly- and perfluoroalkyl substances (PFAS) from water by adsorption: role of PFAS chain length, effect of organic matter and challenges in adsorbent regeneration. Water Res. 171.
Gayathiri, M., Pulingam, T., Lee, K.T., Sudesh, K., 2022. Activated carbon from biomass waste precursors: factors affecting production and adsorption mechanism. Chemosphere 294.
Gayle, N., Mary, D., Rob, H., 1997. Influence of characterised natural organic material on activated carbon adsorption: II. Effect on pore volume distribution and adsorption of 2-methylisoborneol. Water Res. 31, 1065–1073.
Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learning. 63 (1), 3–42.
Giraudet, S., Pré, P., Tezel, H., Le Cloirec, P., 2006. Estimation of adsorption energies using the physical characteristics of activated carbons and the molecular properties of volatile organic compounds. Carbon N Y 44 (12), 2413–2421.
Grabisch, M., Roubens, M., 1999. An axiomatic approach to the concept of interaction among players in cooperative games. Int. J. Game Theory. 28 (4), 547–565.
Grimm, A., Conrad, S., Gentili, F.G., Mikkola, J.-P., Hu, T., Lassi, U., Silva, L.F.O., Lima, E.C., dos Reis, G.S., 2025. Highly efficient boron/sulfur-modified activated biochar for removal of reactive dyes from water: kinetics, isotherms, thermodynamics, and regeneration studies. Colloids Surf. A: Physicochem. Eng. Asp. 713.
Grimm, A., dos Reis, G.S., Khokarale, S.G., Ekman, S., Lima, E.C., Xiong, S., Hultberg, M., 2023. Shiitake spent mushroom substrate as a sustainable feedstock for developing highly efficient nitrogen-doped biochars for treatment of dye-contaminated water. J. Water Process. Eng. 56.
Guo, Y., Kaplan, S., Karanfil, T., 2008. The significance of physical factors on the adsorption of polyaromatic compounds by activated carbons. Carbon N Y 46 (14), 1885–1891.
Guo, Y.P., Yadav, A., Karanfil, T., 2007. Approaches to mitigate the impact of dissolved organic matter on the adsorption of synthetic organic contaminants by porous carbonaceous sorbents. Environ. Sci. Technol. 41 (22), 7888–7894.
Haider Jaffari, Z., Jeong, H., Shin, J., Kwak, J., Son, C., Lee, Y.-G., Kim, S., Chon, K., Hwa Cho, K., 2023. Machine-learning-based prediction and optimization of emerging contaminants' adsorption capacity on biochar materials. Chem. Eng. J. 466.
Han, M., Jin, B., Liang, J., Huang, C., Arp, H.P.H., 2023. Developing machine learning approaches to identify candidate persistent, mobile and toxic (PMT) and very persistent and very mobile (vPvM) substances based on molecular structure. Water Res. 244.
Huang, C., Gao, W., Zheng, Y., Wang, W., Zhang, Y., Liu, K., 2023. Universal machine-learning algorithm for predicting adsorption performance of organic molecules based on limited data set: importance of feature description. Sci. Total Environ. 859.
Islam, M.A., Nazal, M.K., Akinpelu, A.A., Sajid, M., Alhussain, N.A., Billah, R.E.K., Bahsis, L., 2024a. Novel activated carbon derived from a sustainable and low-cost palm leaves biomass waste for tetracycline removal: adsorbent preparation, adsorption mechanisms and real application. Diam. Relat. Mater. 147.
Islam, M.A., Nazal, M.K., Akinpelu, A.A., Sajid, M., Alhussain, N.A., Ilyas, M., 2024b. High performance adsorptive removal of emerging contaminant paracetamol using a sustainable biobased mesoporous activated carbon prepared from palm leaves waste. J. Anal. Appl. Pyrolysis 180.
Jaxa-Rozen, M., Kwakkel, J., 2018. Tree-based ensemble methods for sensitivity analysis of environmental models: a performance comparison with Sobol and Morris techniques. Environ. Model. Softw. 107, 245–266.
Jia, X., Wang, T., Zhu, H., 2023. Advancing computational toxicology by interpretable machine learning. Environ. Sci. Technol. 57 (46), 17690–17706.
Karanfil, T., Dastgheib, S.A., 2004. Trichloroethylene adsorption by fibrous and granular activated carbons: aqueous phase, gas phase, and water vapor adsorption studies. Environ. Sci. Technol. 38 (22), 5834–5841.
Li, H., Nasirin, C., Abed, A.M., Olegovich Bokov, D., Thangavelu, L., Abdulameer Marhoon, H., Lutfor Rahman, M., 2022. Optimization and design of machine learning computational technique for prediction of physical separation process. Arab. J. Chem. 15 (4).
Liu, F., Wang, Q., Zietzschmann, F., Yang, F., Nie, S.Z., Zhang, J.Z., Yang, M., Yu, J.W., 2025. Competition & UV 254 projection in odorants vs natural organic matter adsorption onto activated carbon surfaces: is the chemistry right? Water Res. 268.
Liu, L., Tan, S.J., Horikawa, T., Do, D.D., Nicholson, D., Liu, J., 2017. Water adsorption on carbon - A review. Adv. Colloid Interface. Sci. 250, 64–78.
Lu, J.P., Su, M., Su, Y.L., Fang, J., Burch, M., Cao, T.X., Wu, B., Yu, J.W., Yang, M., 2023. MIB-derived odor management based upon hydraulic regulation in small drinking water reservoirs: principle and application. Water Res. 244.
Mariangela Grassi, G.K., Vincenzo Belgiorno, Giusy Lofrano (2012) Removal of emerging contaminants from water and wastewater by adsorption process.
Mesellem, Y., Hadj, A.A.E., Laidi, M., Hanini, S., Hentabli, M., 2021. Computational intelligence techniques for modeling of dynamic adsorption of organic pollutants on activated carbon. Neural Comput. Appl. 33 (19), 12493–12512.
Ministry of Housing and Urban-Rural Development, P., 2010. Activated carbon from coal for drinking water treatment in municipal water plant. CJ/T 345.
Moreno-Castilla, C., 2004. Adsorption of organic molecules from aqueous solutions on carbon materials. Carbon N Y 42 (1), 83–94.
Nam, S.-W., Choi, D.-J., Kim, S.-K., Her, N., Zoh, K.-D., 2014a. Adsorption characteristics of selected hydrophilic and hydrophobic micropollutants in water using activated carbon. J. Hazard. Mater. 270, 144–152.
Nam, S.W., Choi, D.J., Kim, S.K., Her, N., Zoh, K.D., 2014b. Adsorption characteristics of selected hydrophilic and hydrophobic micropollutants in water using activated carbon. J. Hazard. Mater. 270, 144–152.
Newcombe, G., Drikas, M., Hayes, R., 1997. Influence of characterised natural organic material on activated carbon adsorption .2. Effect on pore volume distribution and adsorption of 2-methylisoborneol. Water Res. 31 (5), 1065–1073.
Nielsen, L., Biggs, M.J., Skinner, W., Bandosz, T.J., 2014. The effects of activated carbon surface features on the reactive adsorption of carbamazepine and sulfamethoxazole. Carbon N Y 80, 419–432.
Park, M., Wu, S.M., Lopez, I.J., Chang, J.Y., Karanfil, T., Snyder, S.A., 2020. Adsorption of perfluoroalkyl substances (PFAS) in groundwater by granular activated carbons: roles of hydrophobicity of PFAS and carbon characteristics. Water Res. 170.
Pauletto, P.S., Lütke, S.F., Dotto, G.L., Salau, N.P.G., 2021. Forecasting the multicomponent adsorption of nimesulide and paracetamol through artificial neural network. Chem. Eng. J. 412.
Pelekani, C., Snoeyink, V.L., 1999. Competitive adsorption in natural water: role of activated carbon pore size. Water Res. 33 (5), 1209–1219.
Pivokonsky, M., Kopecka, I., Cermakova, L., Fialova, K., Novotna, K., Cajthaml, T., Henderson, R.K., Pivokonska, L., 2021. Current knowledge in the field of algal organic matter adsorption onto activated carbon in drinking water treatment. Sci. Total Environ. 799.
Quinlivan, P.A., Li, L., Knappe, D.R., 2005. Effects of activated carbon characteristics on the simultaneous adsorption of aqueous organic micropollutants and natural organic matter. Water Res. 39 (8), 1663–1673.
Rakic, V., Rac, V., Krmar, M., Otman, O., Auroux, A., 2015. The adsorption of pharmaceutically active compounds from aqueous solutions onto activated carbons. J. Hazard. Mater. 282, 141–149.
Ren, Y., Zhang, L., Suganthan, P.N., 2016. Ensemble classification and regression-recent developments, applications and future directions. IEEE Comput. Intell. Mag. 11 (1), 41–53.
Richards, C.E., Tzachor, A., Avin, S., Fenner, R., 2023. Rewards, risks and responsible deployment of artificial intelligence in water systems. Nat. Water 1 (5), 422–432.
Saeidi, N., Lotteraner, L., Sigmund, G., Hofmann, T., Krauss, M., Mackenzie, K., Georgi, A., 2025. Towards a better understanding of sorption of persistent and mobile contaminants to activated carbon: applying data analysis techniques with experimental datasets of limited size. Water Res. 274, 123032.
Sagi, O., Rokach, L., 2018. Ensemble learning: a survey. Wires. Data Min. Knowl. 8 (4).
Schumann, P., Muschket, M., Dittmann, D., Rabe, L., Reemtsma, T., Jekel, M., Ruhl, A.S., 2023. Is adsorption onto activated carbon a feasible drinking water treatment option for persistent and mobile substances? Water Res. 235, 119861.
Schwarzenbach, R.P., Escher, B.I., Fenner, K., Hofstetter, T.B., Johnson, C.A., von Gunten, U., Wehrli, B., 2006. The challenge of micropollutants in aquatic systems. Science 313 (5790), 1072–1077.
Shannon, M.A., Bohn, P.W., Elimelech, M., Georgiadis, J.G., Marinas, B.J., Mayes, A.M., 2008. Science and technology for water purification in the coming decades. Nature 452 (7185), 301–310.
Sobek, A., Abel, S., Sanei, H., Bonaglia, S., Li, Z., Horlitz, G., Rudra, A., Oguri, K., Glud, R.N., 2023. Organic matter degradation causes enrichment of organic pollutants in hadal sediments. Nat. Commun. 14 (1), 2012.
Talekar, B., 2020. A detailed review on decision tree and random forest. Biosci. Biotechnol. Res. Commun. 13 (14), 245–248.
Tong, Y., McNamara, P.J., Mayer, B.K., 2019. Adsorption of organic micropollutants onto biochar: a review of relevant kinetics, mechanisms and equilibrium. Environ. Sci-Wat. Res. 5 (5), 821–838.
Tsai, C.P., Yeh, C.K., Ravikumar, P., 2023. Faith-Shap: the faithful Shapley interaction index. J. Mach. Learn. Res. 24.
Wang, H., Wang, Z., Chen, J., Liu, W., 2022. Graph attention network model with defined applicability domains for screening PBT chemicals. Environ. Sci. Technol. 56 (10), 6774–6785.
Wang, Q., Du, Y.N., Li, W.T., Wang, C.M., Zhang, J.Z., Yang, M., Yu, J.W., 2024. Treatability of odorous dioxanes/dioxolanes in source water: how does molecular flexibility and pre-oxidation affect odorant adsorption. Water Res. 266.
Whelton, A.J., Dietrich, A.M., 2004. Relationship between intensity, concentration, and temperature for drinking water odorants. Water Res. 38 (6), 1604–1614.

Yang, M., Zhu, J.-J., McGaughey, A., Zheng, S., Priestley, R.D., Ren, Z.J., 2023. Predicting extraction selectivity of acetic acid in pervaporation by machine learning models with data leakage management. Environ. Sci. Technol. 57 (14), 5934–5946.
Yap, C.W., 2010. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 32 (7), 1466–1474.
Zhang, K., Zhong, S., Zhang, H., 2020. Predicting aqueous adsorption of organic compounds onto biochars, carbon nanotubes, granular activated carbons, and resins with machine learning. Environ. Sci. Technol. 54 (11), 7008–7018.
Zhang, S., 2023. Adsorption of persistent and mobile substances on activated carbon. Nat. Water 1 (4), 311. -311.
Zhi, W., Appling, A.P., Golden, H.E., Podgorski, J., Li, L., 2024. Deep learning for water quality. Nat Water 2, 228–241.
Zhu, J.-J., Yang, M., Ren, Z.J., 2023. Machine learning in environmental research: common pitfalls and Best practices. Environ. Sci. Technol. 57 (46), 17671–17689.
Zhu, X., He, M., Sun, Y., Xu, Z., Wan, Z., Hou, D., Alessi, D.S., Tsang, D.C.W., 2022. Insights into the adsorption of pharmaceuticals and personal care products (PPCPs) on biochar and activated carbon with the aid of machine learning. J. Hazard. Mater. 423 (Pt B), 127060.
Zhu, X., Wan, Z., Tsang, D.C.W., He, M., Hou, D., Su, Z., Shang, J., 2021. Machine learning for the selection of carbon-based materials for tetracycline and sulfamethoxazole adsorption. Chem. Eng. J. 406.