Explainable AI for Predicting Micropollutant Fate and Transport in Coupled Surface Water–Groundwater Systems

Rajkishor, Ajit Das, Aman Kumar
Ajeet Kumar, Abhijeet Raj

Abstract: *Micropollutants move from rivers into aquifers through hyporheic exchange and riverbank filtration, where sorption, redox-dependent biodegradation, photolysis, travel time and dilution jointly control their fate. This study builds an explainable AI framework that predicts the fraction of a micropollutant surviving the river–aquifer–well pathway and audits its explanations against exact Shapley attributions of a process-based reference model. A coupled model of river transport, hyporheic exchange and oxic–anoxic bank filtration generated 12,000 cases for 80 virtual compounds. Under leave-compounds-out validation a multilayer perceptron reached R^2 = 0.96, against 0.81 for XGBoost and 0.60 for a travel-time screening model. SHAP explanations of the perceptron reproduced the true feature ranking almost exactly (Spearman ρ = 0.997), whereas TreeSHAP of the less accurate XGBoost model under-weighted compound charge three-fold (ρ = 0.94). Oxic half-life, aquifer conductivity and sorption dominated attenuation; compound and aquifer properties carried about 85% of the attribution and the river and hyporheic zone less than 10%. Accuracy improved most with more training compounds. Explanations inherit the errors of the model they explain, so their fidelity should be audited before XAI supports riverbank-filtration management.*

Keywords: *Micropollutants; Surface Water–Groundwater Interaction; Riverbank Filtration; Hyporheic Zone; Explainable Ai; Shap; Machine Learning; Contaminant Transport*

Rajkishor (rk.bce22@gmail.com) Assistant Professor, Department of Civil Engineering, Bhagalpur College of Engineering, Bhagalpur, Bihar, India.
Ajit Das (ajit.gecv@gmail.com) Assistant Professor, Department of Civil Engineering, Government Engineering College, Vaishali, Bihar, India.
Aman Kumar (amangecv@gmail.com) Assistant Professor, Department of Civil Engineering, Government Engineering College, Vaishali, Bihar, India.
Ajeet Kumar (ajeet.gecv@gmail.com) Assistant Professor, Department of Civil Engineering, Government Engineering College, Vaishali, Bihar, India.
Abhijeet Raj (abhijeetraj571@gmail.com) Assistant Professor, Geology, Government Engineering College, Vaishali, Bihar, India.

Introduction

1.1 Background

Micropollutants—pharmaceuticals, personal-care products, pesticides and industrial chemicals present at ng L^{-1} to µg L^{-1} levels—are now ubiquitous in surface waters and increasingly detected in groundwater [1]–[3]. Rivers and aquifers are not separate compartments: streams gain from and lose to groundwater along their courses [4], and water repeatedly passes through the hyporheic zone beneath the riverbed, where steep redox gradients and active biofilms transform many compounds [5], [6]. Riverbank filtration exploits this connection deliberately; pumping wells near rivers draw river water through the bank, and sorption, biodegradation and dilution attenuate contaminants on the way [7], [8]. Whether a given compound reaches a well therefore depends on interacting compound properties, river hydraulics, redox conditions and aquifer travel time.

Process-based models represent these interactions but need uncertain compound-specific rate constants; reported half-lives for one substance often differ by more than an order of magnitude [9]. Machine learning (ML) offers fast data-driven prediction [10], [11], but its use in drinking-water decisions depends on whether its reasoning can be trusted [12]. Shapley additive explanations (SHAP) attribute each prediction to its inputs [13]–[15] and are increasingly used to infer which processes control water quality.

1.2 Research Gap and Objectives

A critical question is rarely asked: are the explanations correct? With field data the true drivers are unknown, so an explanation can look plausible while reflecting artefacts of the model, correlated inputs or extrapolation [16], [17]. This study addresses that gap by pairing an explainable ML framework with a process-based reference model whose exact Shapley attributions are computable. Its objectives are to (i) predict the fraction of a micropollutant that survives the coupled river–hyporheic–aquifer pathway to a well under leave-compounds-out validation; (ii) identify global, compartment-level and local drivers with SHAP; (iii) audit the fidelity and stability of explanations against the true attributions; and (iv) test whether models reproduce the response to warming, low flow and increased pumping.

Related Work

2.1 Micropollutant Fate at the Surface Water–Groundwater Interface

Hyporheic passage removes readily biodegradable pharmaceuticals within hours, while persistent compounds such as carbamazepine pass almost unchanged [6], [18]. Bank-filtration removal depends on redox conditions, temperature and residence time [7], [8], compound-specific rates span orders of magnitude [9], and aquifer transport is governed by advection, scale-dependent dispersion [19] and sorption.

2.2 Machine Learning and Explainable AI in Water Science

Random forest [20], gradient boosting [21] and neural networks dominate ML for water quality [10], [11]. Explanation methods include local surrogates [22] and Shapley attributions [13], [15], with exact TreeSHAP for tree ensembles [14]. Known pitfalls are correlated features [17], extrapolation in permutation importance [16] and optimistic accuracy under random splits of grouped data [23]. No study has yet quantified SHAP fidelity for coupled surface water–groundwater transport against a known ground truth.

Materials and Methods

3.1 Coupled Process-Based Reference Model

The reference model follows a micropollutant from an upstream river section to a riverbank-filtration well (Figure 1). The surviving fraction is the product of a surface-water term, dilution and a groundwater term:

$$F = C_{well}/C_0 = f_{SW} \cdot \beta_{bf} \cdot f_{GW} \quad (1)$$

where β_{bf} is the bank-filtrate share of pumped water. In the river, first-order photolysis (depth-averaged, with light attenuation increasing with DOC), slow water-column biodegradation and hyporheic removal act over the travel time $\tau = L/u$:

$$f_{SW} = \exp[-(\lambda_{ph} + \lambda_w + q_h/d \cdot \{1 - \exp(-\lambda_{ox} R^{1/2} t_h)\})\, \tau] \quad (2)$$

where q_h is the hyporheic exchange rate, d the depth and t_h the hyporheic residence time. Biodegradation rates follow $\lambda = \ln 2/t_{1/2}$ with a temperature correction $1.07^{(T-20)}$. Sorption uses $K_d = K_{oc}\, f_{oc}$, reduced ten-fold for anions and doubled for cations, and the retardation factor $R = 1 + \rho_b K_d/n$. In the aquifer the flow path is split into an oxic segment, whose length increases with dissolved oxygen and decreases with DOC, and an anoxic segment with its own half-life. Each segment uses the steady one-dimensional advection–dispersion solution with decay:

$$C/C_{in} = \exp\{(x/2\alpha_L)\,[1 - (1 + 4\lambda R\alpha_L/v)^{1/2}]\} \quad (3)$$

with seepage velocity $v = Ki/n$ and longitudinal dispersivity $\alpha_L = 0.1x$ [19]. The target is $\log_{10}F$, capped at −5 (100,000-fold removal) to represent a practical quantification limit.

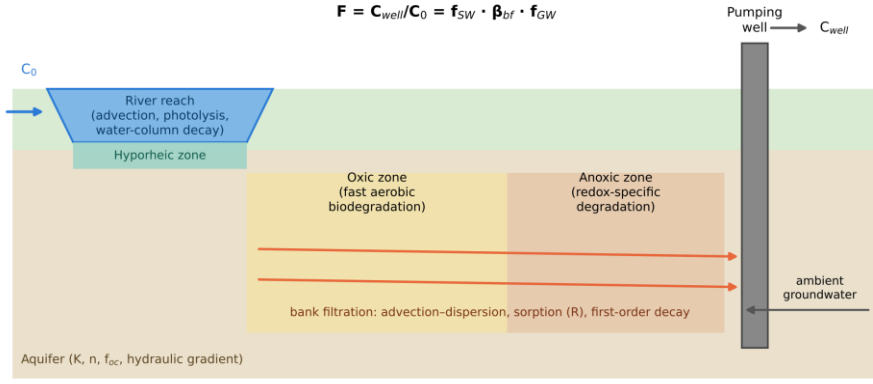


Figure 1. Conceptual model of the coupled river–hyporheic–aquifer pathway from the river inflow to the pumping well.

3.2 Virtual Compounds, Sites and Dataset

Eighty virtual compounds were sampled across property ranges typical of polar to moderately hydrophobic micropollutants, and each was combined with 150 randomly sampled site settings, giving 12,000 cases (Table 1). Water temperature and dissolved oxygen were made physically correlated through oxygen solubility ($r = -0.31$), and two null features, pH and electrical conductivity, were included but had no effect in the model, to test whether explanations wrongly assign them importance. Gaussian noise of 0.06 log units was added to the training targets. The median surviving fraction was $10^{-0.85}$ and 15% of cases reached the quantification cap; median river and aquifer travel times were 1.1 and 67 days.

Table 1. Input variables and sampled ranges

Variable (unit)	Range	Process role
log K_{oc} (L kg^{-1}); charge	0.5–4.0; −1, 0, +1	Sorption and retardation
$t_{1/2,ox}$; $t_{1/2,anox}$ (d)	2–1000; 0.3–1000	Redox-specific biodegradation
$t_{1/2,photo}$ at surface (h)	1–1000	Direct photolysis in river
Reach length L (km); velocity u (m s^{-1}); depth d (m)	5–100; 0.1–1.0; 0.5–5	River travel time, light
Hyporheic exchange q_h (d^{-1}); residence t_h (h)	0.03–3; 1–48	Hyporheic removal
Temperature T (°C); DO (mg L^{-1}); DOC (mg L^{-1})	8–32; correlated; 1–15	Kinetics, redox, light
K (m d^{-1}); gradient i; well distance x (m)	3–316; 0.01–0.08; 20–200	Aquifer travel time
Porosity n; f_{oc}; bank-filtrate share β_{bf}	0.2–0.4; 0.0002–0.005; 0.3–0.95	Sorption, dilution
pH; EC (µS cm^{-1})	6.5–8.5; 200–1500	Null features (no effect)

3.3 Machine Learning Models and Validation

Four ML models were compared with a conventional screening model that multiplies plug-flow aquifer travel time by the oxic decay rate: ridge regression, random forest (300 trees) [20], XGBoost (800 trees, depth 6) [21] and a multilayer perceptron (MLP; two hidden layers of 128 and 64 units, early stopping). Because compounds share all their property values across sites, a random split would leak compound information into the test set [23]. All models were therefore evaluated with leave-compounds-out five-fold cross-validation, and explanations were computed on a fixed hold-out set of 16 unseen compounds (2,400 cases).

3.4 Explainable AI and Fidelity Audit

The SHAP value of feature j for prediction $f(x)$ is its Shapley value [13], [15]:

$$\varphi_j = \Sigma_{S \subseteq N \setminus \{j\}} \, [|S|!(|N| - |S| - 1)!/|N|!] \, [f_{S \cup \{j\}}(x) - f_S(x)] \quad (4)$$

computed with an interventional background of training cases (200 for TreeSHAP, a 100-case subsample for the permutation estimator), using exact TreeSHAP for

XGBoost [14] and the permutation estimator for the MLP. Because the reference model is cheap to evaluate, the same estimator was applied to it, giving the true attributions φ_j^{true} for 600 hold-out cases. Fidelity was measured by the Spearman and Kendall correlation between global importances (mean $|\varphi_j|$), overlap of the top-eight features, the correlation between estimated and true attributions for each feature and case, and the share of importance given to the null features. Stability was the rank correlation of importances across the five cross-validation models.

Results

4.1 Predictive Performance

Table 2 and Figure 2(a) compare the models. The MLP was clearly the most accurate, with $R^2 = 0.964$ and 89% of predictions within ±0.5 log units, followed by XGBoost ($R^2 = 0.809$) and random forest (0.725); ridge regression and the screening model explained only about 60% of the variance. The screening model, which ignores the river, redox zonation, dispersion and dilution, cannot represent compounds whose removal is controlled by anoxic or surface-water processes. Tree models generalised poorly to unseen compounds: XGBoost reached $R^2 = 0.91$ with a random split but only 0.81 with compounds held out, showing how random splitting overstates skill. On the 16 hold-out compounds the MLP gave $R^2 = 0.968$ and RMSE = 0.31 (Figure 2(b)).

Table 2. Leave-compounds-out cross-validation (mean ± standard deviation over five folds)

Model	R^2	RMSE (log units)	MAE (log units)	Within ±0.5 log (%)
Screening model	0.597 ± 0.088	1.05 ± 0.11	0.63 ± 0.05	64.3
Ridge regression	0.618 ± 0.073	1.02 ± 0.05	0.84 ± 0.04	31.7
Random forest	0.725 ± 0.050	0.87 ± 0.08	0.57 ± 0.07	63.1
XGBoost	0.809 ± 0.040	0.73 ± 0.10	0.49 ± 0.08	65.0
MLP	**0.964 ± 0.004**	**0.32 ± 0.05**	**0.21 ± 0.03**	**89.3**

Figure 2(c) shows that accuracy depends strongly on the number of distinct compounds in the training data. With 20 compounds the MLP reached $R^2 = 0.76$ and XGBoost 0.61; with 160 compounds these rose to 0.988 and 0.87. For monitoring design, sampling more compounds is therefore more valuable than sampling more sites for the same compounds.

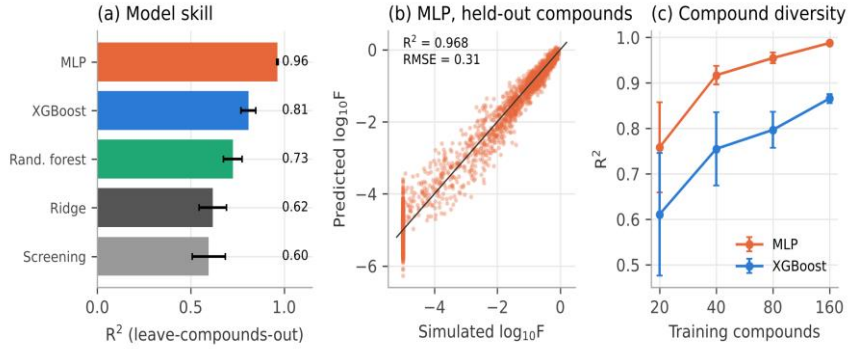


Figure 2. (a) Leave-compounds-out R^2 of the five models; (b) MLP predictions for 16 unseen compounds; (c) effect of the number of training compounds on R^2 (mean ± s.d. over folds).

4.2 Global Drivers of Attenuation

Figure 3 shows the SHAP summary for the MLP. Oxic half-life, aquifer conductivity and log K_{oc} dominated, each with mean |SHAP| of about 0.6 log units: long half-lives and short travel times raised the surviving fraction, while strong sorption lowered it by lengthening the retarded residence time. Hydraulic gradient, well distance, charge (anions passing more readily), f_{oc} and temperature formed a second group; river and hyporheic variables ranked lower.

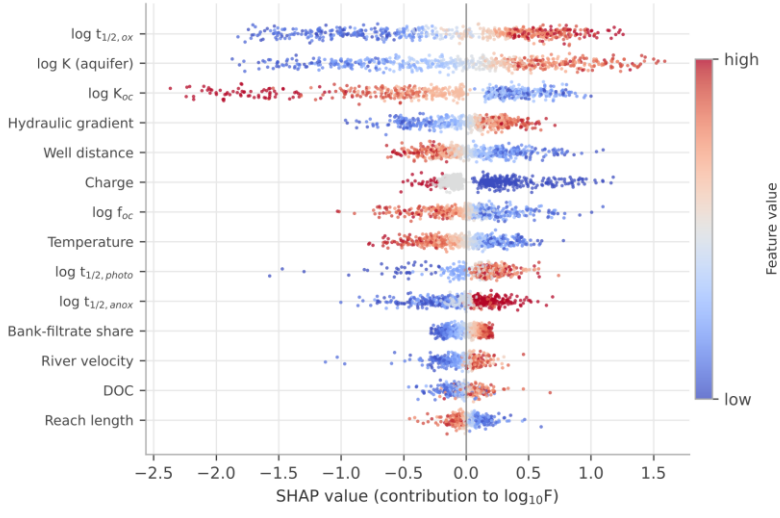


Figure 3. SHAP summary (beeswarm) plot of the MLP for 600 hold-out cases; colour shows the feature value.

4.3 Fidelity and Stability of Explanations

Table 3 and Figure 4 compare the ML explanations with the true attributions. The MLP explanations were almost exact: the global ranking matched the truth with $\rho = 0.997$, all eight leading features were identified, importance errors for these features averaged 1.7%, and the median correlation between estimated and true local attribution vectors was 0.99. XGBoost explanations were good but distorted ($\rho = 0.94$); most notably, compound charge received a mean |SHAP| of 0.07 against a true 0.23, and the anoxic half-life was over-weighted (0.34 versus 0.19), mirroring the tree model's poorer generalisation to unseen compounds: an explanation can be no more faithful than the model it explains. Both models gave less than 1% of importance to the null features, and XGBoost permutation, gain and random-forest impurity importances ($\rho = 0.94$, 0.94 and 0.90) were all less faithful than SHAP of the best model.

Table 3. Fidelity and stability of explanations against simulator ground truth

Metric	MLP (permutation SHAP)	XGBoost (TreeSHAP)
Spearman ρ of global ranking	0.997	0.940
Kendall τ of global ranking	0.981	0.819
Top-8 features recovered	8 of 8	7 of 8
Mean importance error, top 8 (%)	1.7	17.9
Instance-level r, top-8 features	0.980	0.960
Median local r (per case)	0.985	0.948
Importance on null features (%)	0.67	0.61
Cross-fold rank stability (mean ρ)	0.989	0.993
Cross-fold CV of top-8 importances (%)	1.8	5.6

Grouping the attributions by compartment (Figure 4(b)) shows that compound properties carried 48% of the true attribution and the aquifer and bank-filtration variables 38%, while the river (6%), water quality and redox (8%) and hyporheic exchange (0.4%) contributed much less. The MLP reproduced these shares within one percentage point. Explanations were also stable: rankings from the five cross-validation models correlated with $\rho \geq 0.98$.

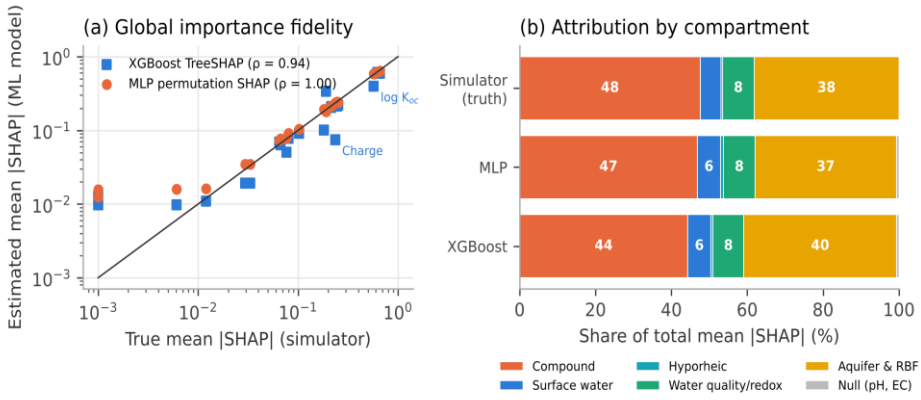


Figure 4. (a) Estimated versus true global SHAP importance for the 21 features (null features plotted at 10^{-3}); (b) attribution shares by compartment for the simulator, MLP and XGBoost.

4.4 Process-Level Dependence and Interactions

Figure 5 compares SHAP dependence for four key features. The MLP reproduced the true shapes closely ($r \geq 0.98$): the contribution of oxic half-life rises monotonically with persistence, the contribution of sorption falls sharply above log $K_{oc} \approx 2.5$, aquifer conductivity acts almost log-linearly, and warmer water lowers the surviving fraction by accelerating biodegradation. SHAP interaction values of XGBoost [14] showed that pairwise interactions carried 47% of the attribution magnitude, led by oxic half-life × conductivity and log K_{oc} × f_{oc}—exactly the couplings between degradation rate and residence time, and between the two sorption terms, that Eqs. (2)–(3) imply.

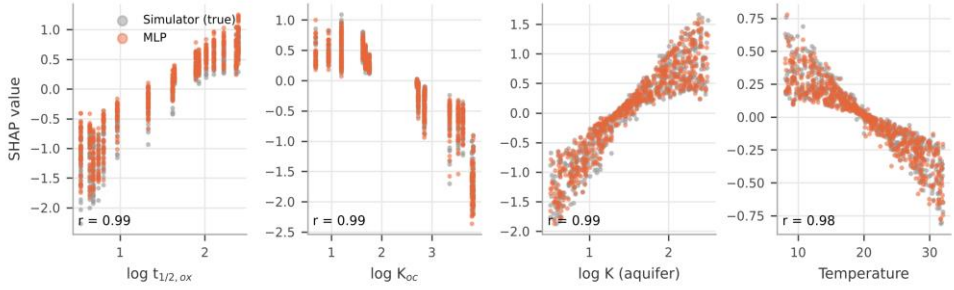


Figure 5. SHAP dependence of the MLP (orange) compared with the true simulator attributions (grey) for oxic half-life, log K_{oc}, aquifer conductivity and temperature.

4.5 Local Explanations

Figure 6 explains two individual hold-out cases. In the weakly attenuated case (true $\log_{10}F$ = −0.05; MLP −0.11), a long oxic half-life, high conductivity and weak sorption all pushed the prediction towards complete breakthrough, and a high bank-filtrate share limited dilution. In the strongly attenuated case (true −3.21; MLP −3.08), a short oxic half-life lowered the prediction by about 1.3–1.6 log units, but cold water (8 °C) partly offset this by slowing degradation. In both cases the MLP identified the same leading features, with the same signs and similar magnitudes as the true attributions.

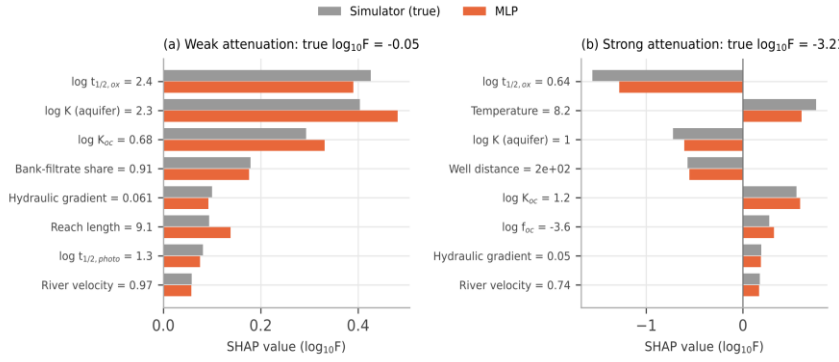


Figure 6. Local explanations for (a) a weakly and (b) a strongly attenuated hold-out case: true simulator attributions (grey) and MLP SHAP values (orange) for the eight most influential features.

4.6 Scenario Response

Table 4 tests whether the models reproduce the response of the system to three what-if scenarios applied to the hold-out cases. All three scenarios changed the median surviving fraction by less than 0.2 log units, but the effects were consistent in sign. Warming by 3 °C and low flow increased removal, and doubling the pumping gradient decreased it by shortening travel times. The MLP tracked the case-by-case changes better than XGBoost (r = 0.77–0.80 versus 0.26–0.75), although both overestimated the small warming and low-flow effects, and the MLP had the sign right in only 80% of pumping cases. Explaining a model is therefore not the same as validating its response to change.

Table 4. Median change in $\log_{10}F$ under three scenarios and agreement with the simulator

Scenario	Simulator	MLP (r; sign %)	XGBoost (r; sign %)	Cases affected (%)
Warming +3 °C	-0.041	-0.065 (0.77; 99)	-0.089 (0.26; 96)	47
Low flow (u × 0.5, d × 0.6)	-0.034	-0.077 (0.80; 89)	-0.099 (0.75; 96)	43
Pumping (i × 2, β_{bf} × 1.15)	+0.185	+0.153 (0.79; 80)	+0.223 (0.36; 98)	87

r is the correlation between predicted and simulated changes; sign % is the share of cases with $|\Delta| > 0.05$ whose sign was predicted correctly.

Discussion

5.1 Implications for Management and Monitoring

For riverbank filtration the decisive controls are compound persistence, sorption and aquifer residence time, the last of which operators can influence through well placement and pumping. The explanations of a model are only as good as its predictions: XGBoost, which looked adequate under random validation, misrepresented compound charge. Leave-compounds-out validation and, where possible, a fidelity check against a process model should precede interpretation, and training data should prioritise chemical diversity (R^2 rose from 0.76 to 0.99 between 20 and 160 compounds).

5.2 Limitations

The data come from a process-based model rather than field observations; this enables the fidelity audit, but real systems with transient flow, seasonal redox shifts, non-linear sorption and censored measurements will be harder to predict and explain. Interventional SHAP with correlated inputs such as temperature and oxygen can create unrealistic feature combinations [16], [17], and compound properties, treated here as known, are themselves uncertain [9].

Conclusion

This study developed and audited an explainable AI framework for micropollutant fate in coupled surface water–groundwater systems. A multilayer perceptron predicted the surviving fraction at a bank-filtration well for unseen compounds with

R^2 = 0.96, outperforming tree ensembles and a travel-time screening model. Its SHAP explanations reproduced the true drivers almost exactly (ρ = 0.997), identifying oxic half-life, aquifer conductivity and sorption as dominant and the river and hyporheic zone as secondary, whereas the explanations of a less accurate tree model distorted the role of compound charge. Model accuracy appeared to matter more for explanation fidelity than the choice of explanation method, and the chemical diversity of the training data mattered most for accuracy. Future work will apply the framework to monitoring data from riverbank-filtration sites, add transient flow and redox dynamics, and combine field-trained models with process models to audit their explanations.

References

R. P. Schwarzenbach *et al.*, "The challenge of micropollutants in aquatic systems," *Science*, vol. 313, no. 5790, pp. 1072–1077, 2006, doi: 10.1126/science.1127291.

Y. Luo *et al.*, "A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment," *Sci. Total Environ.*, vol. 473–474, pp. 619–641, 2014, doi: 10.1016/j.scitotenv.2013.12.065.

D. J. Lapworth, N. Baran, M. E. Stuart, and R. S. Ward, "Emerging organic contaminants in groundwater: A review of sources, fate and occurrence," *Environ. Pollut.*, vol. 163, pp. 287–303, 2012, doi: 10.1016/j.envpol.2011.12.034.

T. C. Winter, "Relation of streams, lakes, and wetlands to groundwater flow systems," *Hydrogeol. J.*, vol. 7, no. 1, pp. 28–45, 1999, doi: 10.1007/s100400050178.

F. Boano *et al.*, "Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications," *Rev. Geophys.*, vol. 52, no. 4, pp. 603–679, 2014, doi: 10.1002/2012RG000417.

J. Lewandowski, A. Putschew, D. Schwesig, C. Neumann, and M. Radke, "Fate of organic micropollutants in the hyporheic zone of a eutrophic lowland stream: Results of a preliminary field study," *Sci. Total Environ.*, vol. 409, no. 10, pp. 1824–1835, 2011, doi: 10.1016/j.scitotenv.2011.01.028.

K. M. Hiscock and T. Grischek, "Attenuation of groundwater pollution by bank filtration," *J. Hydrol.*, vol. 266, no. 3–4, pp. 139–144, 2002, doi: 10.1016/S0022-1694(02)00158-0.

S. K. Maeng, S. K. Sharma, K. Lekkerkerker-Teunissen, and G. L. Amy, "Occurrence and fate of bulk organic matter and pharmaceutically active compounds in managed aquifer recharge: A review," *Water Res.*, vol. 45, no. 10, pp. 3015–3033, 2011, doi: 10.1016/j.watres.2011.02.017.

J. Greskowiak, E. Hamann, V. Burke, and G. Massmann, "The uncertainty of biodegradation rate constants of emerging organic compounds in soil and groundwater – A compilation of literature values for 82 substances," *Water Res.*, vol. 126, pp. 122–133, 2017, doi: 10.1016/j.watres.2017.09.017.

V. Rodriguez-Galiano, M. P. Mendes, M. J. Garcia-Soldado, M. Chica-Olmo, and L. Ribeiro, "Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain)," *Sci. Total Environ.*, vol. 476–477, pp. 189–206, 2014, doi: 10.1016/j.scitotenv.2014.01.001.

M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019, doi: 10.1038/s41586-019-0912-1.

C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, arXiv:1705.07874, doi: 10.48550/arXiv.1705.07874.

S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.

E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowl. Inf. Syst.*, vol. 41, no. 3, pp. 647–665, 2014, doi: 10.1007/s10115-013-0679-x.

G. Hooker, L. Mentch, and S. Zhou, "Unrestricted permutation forces extrapolation: Variable importance requires at least one more model, or there is no free variable importance," *Stat. Comput.*, vol. 31, no. 6, Art. no. 82, 2021, doi: 10.1007/s11222-021-10057-z.

K. Aas, M. Jullum, and A. Løland, "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values," *Artif. Intell.*, vol. 298, Art. no. 103502, 2021, doi: 10.1016/j.artint.2021.103502.

U. Kunkel and M. Radke, "Fate of pharmaceuticals in rivers: Deriving a benchmark dataset at favorable attenuation conditions," *Water Res.*, vol. 46, no. 17, pp. 5551–5565, 2012, doi: 10.1016/j.watres.2012.07.033.

L. W. Gelhar, C. Welty, and K. R. Rehfeldt, "A critical review of data on field-scale dispersion in aquifers," *Water Resour. Res.*, vol. 28, no. 7, pp. 1955–1974, 1992, doi: 10.1029/92WR00607.

L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

M. T. Ribeiro, S. Singh, and C. Guestrin, "“Why should I trust you?”: Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

D. R. Roberts *et al.*, "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," *Ecography*, vol. 40, no. 8, pp. 913–929, 2017, doi: 10.1111/ecog.02881.