Proceedings of CONF-MCEE 2027 Symposium: Innovative Environmental Engineering Strategies for a Sustainable Future
DOI: 10.54254/2755-2721/2027.MA38438

Failure-Aware Explainable Decision Support for Circular Sulfate Adsorbent Design under Hydrochemical Domain Shift

Sophia Chen

Great Neck South High School, New York, USA
essaysci1@outlook.com

Abstract. This study develops a failure-aware, explainable environmental decision-support framework for circular sulfate adsorbent design and asks whether predictive transportability and engineering usefulness remain aligned under hydrochemical domain shift. A 58-sample Semarang groundwater dataset served as the development domain. Random forest achieved internal AUROC 0.710 ± 0.165, and a target-free matrix-complexity index (MCI) combined EC, Ca^{2+}, Mg^{2+}, Cl^{-}, and HCO_3^{-} into four treatment regimes. Using the fixed Semarang model, discrimination was retained in the 25-sample Bayat basin (AUROC 0.794; balanced accuracy 0.762) but was lost in a larger cross-country Mekong challenge (n = 118; AUROC 0.491; 95% CI 0.367–0.617). In contrast, the fixed-reference MCI remained strongly associated with sulfate in Mekong (rs = 0.699; 95% CI 0.587–0.782), with 73 of 78 high-sulfate events in the complex-matrix domain. To assess consistency between the engineering interpretation and independent material evidence, open-access sulfate-adsorption studies were synthesized at study level. Zirconium modification increased reported pomelo-peel-biochar qmax from 1.02 to 35.21 mg/g (34.5-fold), while $CaCl_2$-modified kelp carbon retained substantial reported kinetic uptake in lake water (569.37 mg/g versus 624.39 mg/g in simulated water). The combined evidence supports a workflow in which external model validity governs predictive use, whereas transparent hydrochemical regimes and independent experimental evidence guide capacity, selectivity, stability, and functionalization priorities.

Keywords: explainable machine learning, sulfate adsorption, domain shift, external validation, circular materials

1. Introduction

Sulfate is common in groundwater, surface water, and industrial effluents. At elevated concentrations it can affect taste, scaling, corrosion, and downstream water use; the U.S. Environmental Protection Agency lists 250 mg/L as a secondary drinking-water level based on aesthetic effects [1]. Sulfate treatment is nevertheless strongly matrix dependent because dissolved ions alter ionic strength, surface charge, ion pairing, precipitation behavior, and competition for adsorption sites.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2027.MA38438

Waste-derived biochar and biosorbents provide a circular route for water treatment by converting agricultural and food residues into functional sorbent platforms [2]. Vegetable peels are inexpensive precursors, but raw biomass is rarely an optimized sulfate sorbent. Sulfate-specific studies show that anion-affinity modification can be decisive: zirconium-oxide modification increased the reported sulfate capacity of pomelo-peel biochar from 1.02 to 35.21 mg/g [3], while calcium-modified biomass carbon has also been developed for sulfate uptake [4]. The engineering problem is therefore not simply to identify a low-cost peel, but to match surface chemistry, material form, and regeneration strategy to the hydrochemical environment.

Data-driven water-quality modelling offers a complementary opportunity. Machine learning has been applied to sulfate prediction in groundwater [5], and explainable-AI methods increasingly support interpretation of water-quality models [6]. Recent groundwater-ML reviews identify cross-watershed generalization and external validation as persistent gaps [7]. In aqueous pollutant-removal modelling, group-aware and independent external tests can expose large differences between interpolation and genuine transfer [8]. Yet model-transfer diagnostics and adsorbent-design decisions are still usually treated separately. The engineering question is therefore not only whether sulfate can be predicted, but what information remains safe to use when predictive relationships shift across hydrochemical domains.

The study makes five linked contributions. First, it develops an explainable high-sulfate screening workflow in Semarang under repeated cross-validation. Second, model parameters, probability cutoffs, sulfate threshold, and MCI reference rules are fixed before Bayat and Mekong evaluation, establishing a locked external-validation design. Third, a failure-aware gate separates predictive deployment from hydrochemical engineering interpretation when external discrimination or calibration deteriorates. Fourth, the target-free MCI translates the water matrix into four capacity-, selectivity-, and stability-oriented treatment regimes. Fifth, open-access sulfate-adsorption studies are synthesized at study level to assess whether published material behavior is directionally consistent with these design requirements. The workflow is summarized in Figure 1.

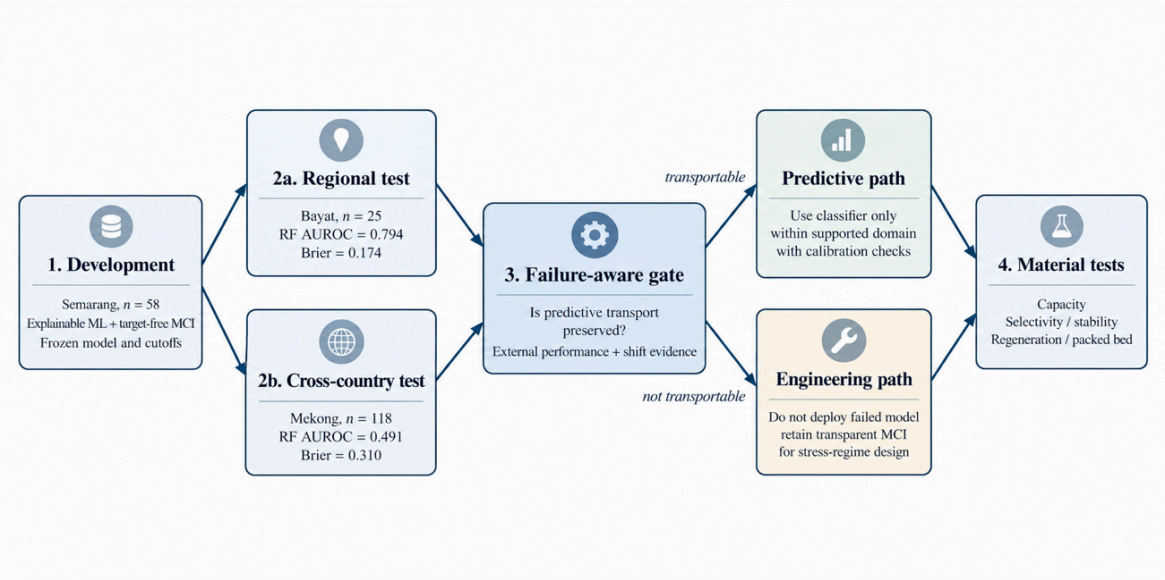


Figure 1. Failure-aware decision-support workflow

DOI: 10.54254/2755-2721/2027.MA38438

2. Data and methodology

2.1. Development and locked external hydrochemical datasets

The development analysis used the PANGAEA dataset of Putranto, Rüde, and Irawan, containing 58 groundwater samples from Semarang, Java, Indonesia [9]. Regional external validation used 25 groundwater samples from the Bayat Melange Complex, Klaten, Indonesia [10]. A stronger cross-country challenge used the 2021 Mekong Delta dataset, which contains 119 unique monitoring-well events from five deep (semi-)confined aquifers and reports the same core predictors together with sulfate [11]. The Mekong dataset comprised 123 analytical records representing 119 unique monitoring-well events. For four events with replicate measurements, chemistry values were averaged at the event level. One event lacked the required major-ion/sulfate measurements and was excluded from locked multivariate evaluation, leaving 118 complete events. Laboratory HCO_3^- was used for the primary analysis, with field HCO_3^- retained for sensitivity analysis.

2.2. Statistical screening and high-sulfate definition

Sulfate was summarized by median, interquartile range, and observed range. Aquifer differences were evaluated with the Kruskal-Wallis test. Spearman correlations were calculated between sulfate and the major hydrochemical variables, with nine simultaneous tests adjusted using the Benjamini-Hochberg false-discovery-rate procedure. A high-sulfate screening class was defined as SO_4^{2-} at or above the dataset upper quartile (64.3 mg/L), yielding 15 high-sulfate and 43 lower-sulfate samples. This threshold is a data-derived engineering screening level and is not a drinking-water standard.

2.3. Explainable machine-learning screening

Eight predictors were used: pH, EC, K^+, Ca^{2+}, Mg^{2+}, Na^+, Cl^-, and HCO_3^-. Sulfate itself was excluded from the input matrix, and TDS was omitted because of near-collinearity with EC. Positive-valued predictors were transformed as log(1+x), while pH remained untransformed. Three classifiers were compared: class-weighted logistic regression, random forest, and Extra Trees. Tree models used 300 estimators, maximum depth 3, minimum leaf size 3, square-root feature sampling, balanced class weights, and a fixed random seed. Internal performance was estimated with repeated stratified five-fold cross-validation (10 repeats; 50 held-out evaluations). Model-specific probability cutoffs were selected from Semarang out-of-fold predictions by maximizing Youden's J and were then fixed for external evaluation. Bayat and Mekong were evaluated using the fixed models and thresholds with AUROC, balanced accuracy, sensitivity, specificity, and Brier score. Random-forest external AUROC uncertainty was estimated by 2,000 event-level bootstrap resamples. All model specifications and decision thresholds were determined from the Semarang development data before external evaluation.

2.4. Matrix-complexity index, external projection, and engineering translation

A unitless, target-free matrix-complexity index (MCI) was defined from EC, Ca^{2+}, Mg^{2+}, Cl^-, and HCO_3^-. In Semarang, each variable was converted to its empirical percentile rank and the five ranks were averaged with equal weights: $MCIi = (1/5)\ \Sigma j\ Pij$. Equal weighting limited additional parameter estimation in the small development sample. Samples were divided at the fixed Semarang median MCI (0.509) and sulfate upper quartile (64.3 mg/L), producing a 2 × 2 treatment-regime map. For Bayat and Mekong, every MCI component was scored against the Semarang empirical

DOI: 10.54254/2755-2721/2027.MA38438

reference distribution rather than re-ranked locally, preserving a common decision space across domains. Regimes were linked to literature-supported material forms: activated winter-melon peel [12], potato-peel biochar [13], cucumber-peel alginate composites [14], and sulfate-oriented Zr- or Ca-functionalized biochars. The MCI was used as a summary of concurrent ion-matrix burden relevant to adsorption selectivity and stability.

2.5. Robustness and integrity checks

Robustness analyses were specified to test whether the engineering interpretation depended on data-handling choices or a single MCI component. For Mekong, event-mean duplicate consolidation was compared with first-replicate retention and with treating all analytical rows separately. Laboratory versus field HCO_3^- was evaluated as a measurement-source sensitivity analysis. For both external datasets, the MCI was recalculated five times while omitting one component at a time, with all remaining components still referenced to Semarang and the 0.509 cutoff left unchanged. Bootstrap intervals were calculated for external MCI–sulfate Spearman correlations, and Wilson intervals were used for the fraction of high-sulfate samples occupying the complex-matrix domain.

2.6. Open experimental evidence benchmark

Open-access sulfate-adsorption studies were synthesized as a study-level experimental benchmark [15, 16]. Reported qmax, kinetic qe, and multicomponent findings were retained in their published study contexts. Analyses were restricted to within-study descriptive comparisons, including a modification enhancement factor (MEF = qmodified/qunmodified) for Zr-modified versus unmodified pomelo-peel biochar and a lake/simulated-water kinetic uptake ratio for $CaCl_2$-modified kelp carbon. Because the latter waters differed in starting chemistry and sulfate concentration, the ratio was interpreted as a descriptive transport indicator rather than a causal estimate of matrix effects. The benchmark was used to assess directional consistency between published material responses and the proposed engineering requirements of functionalization, capacity, and matrix tolerance.

3. Results

3.1. Development-basin hydrochemical structure

Across the 58 samples, sulfate ranged from 6.4 to 166.1 mg/L (mean 44.2 mg/L; median 37.0 mg/L; interquartile range 16.8–64.3 mg/L). Median sulfate concentrations were 15.8 mg/L in the Damar aquifer, 56.5 mg/L in Garang, and 40.5 mg/L in Quaternary marine groundwater; the distributions differed significantly (Kruskal-Wallis $H = 17.92$, $p < 0.001$). After FDR correction, sulfate remained positively associated with Na^+ ($r_s = 0.408$, $q = 0.013$) and negatively associated with Ca^{2+} ($r_s = -0.355$, $q = 0.028$) and Mg^{2+} ($r_s = -0.326$, $q = 0.038$). These results justify treating sulfate burden and the surrounding ion matrix as separate design dimensions rather than using sulfate concentration alone.

3.2. Internal screening and nested external transportability

Table 1 summarizes internal and locked external performance. Random forest achieved the strongest internal discrimination (AUROC 0.710 ± 0.165), while Extra Trees gave the highest internal balanced accuracy (0.660 ± 0.127). Using the fixed Semarang random-forest model and threshold,

DOI: 10.54254/2755-2721/2027.MA38438

Bayat retained useful discrimination (AUROC 0.794; balanced accuracy 0.762; sensitivity 0.857; specificity 0.667; Brier 0.174). Under the cross-country Mekong challenge, all three classifiers lost discrimination: AUROC values were 0.488 for logistic regression, 0.491 for random forest, and 0.443 for Extra Trees. Random-forest balanced accuracy was 0.593 (sensitivity 0.910; specificity 0.275), Brier score worsened to 0.310, and the AUROC 95% bootstrap interval was 0.367–0.617. The simultaneous loss of ranking and calibration indicates that model-based screening should be restricted to domains in which external discrimination and calibration are maintained.

Table 1. Internal validation and external classifier performance

Model	Internal AUROC	Internal BA	Bayat AUC / BA	Mekong AUC / BA
Logistic regression	0.569 ± 0.168	0.578 ± 0.134	0.762 / 0.746	0.488 / 0.505
Random forest	0.710 ± 0.165	0.626 ± 0.149	0.794 / 0.762	0.491 / 0.593
Extra Trees	0.697 ± 0.171	0.660 ± 0.127	0.770 / 0.516	0.443 / 0.378

Cross-validated permutation analysis ranked Mg^{2+}, Na^{+}, Ca^{2+}, and EC as the four most informative random-forest predictors in Figure 2, with mean held-out AUROC decreases of approximately 0.102, 0.078, 0.023, and 0.018. Fold-to-fold uncertainty was wide, so the ordering is interpreted as predictive relevance rather than a causal hierarchy. Directional interpretation is supported by the independent correlation analysis: Na^{+} increased with sulfate in Semarang, whereas Ca^{2+} and Mg^{2+} decreased. The predictor analysis therefore identifies hydrochemical dimensions for subsequent adsorbent validation.

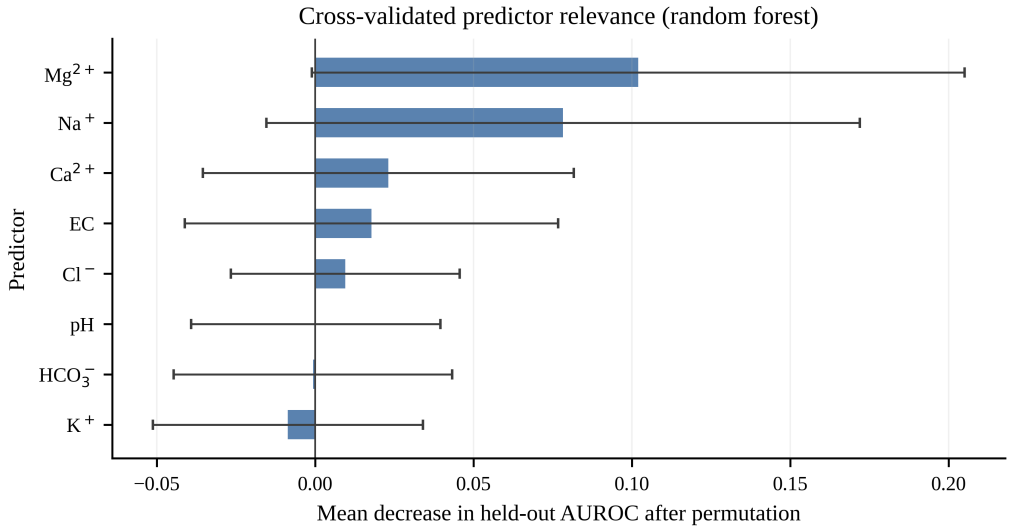


Figure 2. Cross-validated random-forest permutation importance

3.3. Hydrochemical domain shift and engineering-rule stability

In Figure 3, the fixed treatment map reveals a different pattern from the classifiers. Semarang contained 19 lower-sulfate/simpler-matrix, 24 lower-sulfate/complex-matrix, 10 high-sulfate/simpler-matrix, and 5 high-sulfate/complex-matrix samples. Bayat placed 6 of 7 high-sulfate samples in the complex-matrix domain. In Mekong, sulfate ranged from 3 to 2,097 mg/L (median 120.5 mg/L); 78 of 118 complete Events exceeded the fixed 64.3 mg/L screen, and 73 of those 78

DOI: 10.54254/2755-2721/2027.MA38438

(93.6%; Wilson 95% CI 85.9–97.2%) occupied the complex-matrix domain. Sulfate remained strongly associated with the Semarang-referenced MCI (rs = 0.699; 95% bootstrap CI 0.587–0.782; p = 1.39 × 10−18). At the same time, the conditional hydrochemical profile changed sharply: across the eight model predictors, the mean absolute change in feature–sulfate Spearman correlation relative to Semarang was 0.462 in Bayat and 0.714 in Mekong. In Mekong, sulfate correlated positively with EC (rs = 0.852), Na^+ (0.834), Ca^{2+} (0.718), and Mg^{2+} (0.756), whereas Semarang showed negative Ca^{2+} and Mg^{2+} associations. This relationship shift provides a direct explanation for the failed transported classifier. Table 2 summarizes the four hydrochemical regimes and material design requirements.

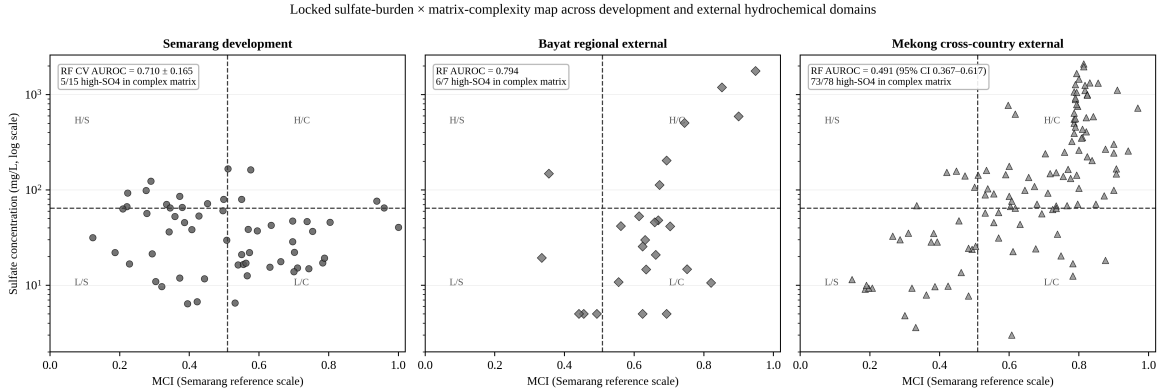


Figure 3. Sulfate burden and matrix complexity across three groundwater domains

Table 2. Four hydrochemical regimes and corresponding adsorbent design priorities

Hydrochemical regime	n (Semarang / Bayat / Mekong)	Primary design requirement	Proposed material route
Lower SO_4 / simpler matrix	19 / 4 / 23	Low-cost capacity screening	Activated/carbonized peel baseline
Lower SO_4 / complex matrix	24 / 14 / 17	Selectivity under ion competition	Functionalized carbon or immobilized composite
High SO_4 / simpler matrix	10 / 1 / 5	Higher anion capacity	Ca- or Zr-functionalized peel biochar
High SO_4 / complex matrix	5 / 6 / 73	Capacity + selectivity + stability	Oxide-loaded immobilized composite / packed-bed polishing

3.4. Robustness and failure-mode diagnostics

The failure-aware interpretation was stable to several perturbations (Table 3). In Mekong, replacing event means with first technical replicates left random-forest AUROC unchanged at 0.491, while treating all 122 complete analytical rows separately gave AUROC 0.488. Replacing laboratory HCO_3^- with field HCO_3^- gave RF AUROC 0.496 and MCI–sulfate rs = 0.699. Leave-one-component-out MCI analysis also preserved the engineering conclusion: Bayat retained 6/7 high-sulfate samples in the complex domain for every omitted component, and Mekong retained 65–75 of 78; its MCI–sulfate rs remained 0.616–0.802. These sensitivity analyses show that external model

DOI: 10.54254/2755-2721/2027.MA38438

performance was stable across duplicate-handling and bicarbonate-source choices, while the MCI regime pattern remained robust to omission of individual components.

Table 3. External robustness and domain-shift diagnostics

Diagnostic	Bayat	Mekong	Interpretation
RF Brier score	0.174	0.310	Calibration deteriorates in cross-country transfer
MCI–SO_4 rs (95% bootstrap CI)	0.480 (0.045–0.797)	0.699 (0.587–0.782)	Target-free matrix summary remains informative
High-SO_4 in complex matrix	6/7 (85.7%)	73/78 (93.6%)	Complex treatment burden persists externally
Leave-one-component-out MCI	rs 0.345–0.599; 6/7 retained	rs 0.616–0.802; 65–75/78 retained	Stable across component omission

3.5. Open experimental adsorption benchmark

Table 4 summarizes independent material-level evidence. In the pomelo-peel study, Zr oxide modification raised qmax from 1.02 to 35.21 mg/g, giving MEF = 34.5 and supporting the role of anion-affinity functionalization for peel-derived carbon scaffolds [3]. For $CaCl_2$-modified kelp carbon, reported pseudo-second-order qe values were 624.39 mg/g in simulated water and 569.37 mg/g in lake water (descriptive ratio = 0.912); because the starting waters were not identical, the comparison indicates that substantial uptake persisted in a natural matrix [4]. An independent Zea mays study reported 16.4 and 7.4 mg/g at 25 °C for two modified adsorbent routes, a 2.22-fold route difference [15]. A separate ternary nitrate-sulfate-phosphate study using acid-modified Zea mays biochar found that sulfate remained competitively adsorbed in the multicomponent system, reinforcing the importance of co-ion selectivity alongside adsorption capacity [16]. Together, these studies provide external material evidence consistent with the proposed engineering design priorities.

Table 4. Literature-based sulfate adsorption performance benchmark

Open experimental evidence	Published sulfate response	Derived descriptor	What it supports
Pomelo peel PBC → ZrBC [3]	qmax 1.02 → 35.21 mg/g	MEF = 34.5×	Strong same-study support for anion-affinity functionalization
$CaCl_2$-modified kelp carbon [4]	kinetic qe: SW 624.39; lake 569.37 mg/g	lake/SW = 0.912	Substantial uptake persists in natural water; descriptive cross-matrix comparison
Modified Zea mays routes [15]	25 °C capacity: 16.4 vs 7.4 mg/g	2.22× route difference	Material formulation changes sulfate capacity
Competitive Zea mays biochar [16]	ternary NO_3^-/SO_4^{2-}/PO_4^{3-} test	qualitative multicomponent evidence	Supports explicit co-ion selectivity testing

DOI: 10.54254/2755-2721/2027.MA38438

4. Discussion

4.1. From prediction to intelligent environmental decision support

The central contribution is a failure-aware environmental decision-support workflow that separates predictive confidence from engineering usefulness. This distinction is consistent with recent evidence that apparently strong water-treatment ML performance can collapse when evaluation moves from interpolation to genuinely unseen groups or independent external data [8], and with groundwater-ML reviews that identify transferability across hydrogeological settings as a major unresolved need [7]. In the present analysis, the Mekong result activated the failure-aware gate: model-based screening was withheld, while transparent target-free chemistry summaries were retained for material-test design.

4.2. Predictive versus engineering transportability under domain shift

Bayat and Mekong answer different transfer questions. Bayat preserved random-forest discrimination (AUROC 0.794; 95% bootstrap CI 0.492–0.993), although uncertainty is wide because n = 25. Mekong reduced random-forest AUROC to 0.491 (95% CI 0.367–0.617), with Brier score worsening from 0.174 to 0.310; logistic regression and Extra Trees deteriorated similarly. The larger change in the feature–sulfate correlation profile in Mekong further indicates conditional relationship shift rather than a single-algorithm failure. The MCI serves instead as a fixed, target-free description of co-ion burden used to specify experimental stress conditions. Its leave-one-component-out stability supports this engineering role. Accordingly, transported classifier scores are used only when external discrimination and calibration are preserved; otherwise, the locked hydrochemical regime defines the conditions required for subsequent laboratory validation.

4.3. Implications for circular adsorbent design

The regime map changes how waste-derived carbon is framed: the precursor is a renewable scaffold, while sulfate performance depends on surface chemistry and the surrounding ion matrix. The open experimental benchmark adds independent triangulation. Zr modification increased pomelo-peel-biochar qmax 34.5-fold [3], providing direct same-study evidence that anion-affinity functionalization can dominate raw-carbon performance. $CaCl_2$-modified kelp carbon maintained high reported kinetic uptake in both simulated and lake water [4], while the Zea mays studies showed substantial formulation dependence and explicit multicomponent behavior [15, 16]. These observations are consistent with using low-complexity regimes for baseline capacity screening and reserving complex/high-sulfate regimes for functionalization, selectivity, regeneration, particle containment, and hydraulic-stability tests. They support the direction of the design rules but do not experimentally validate the exact MCI boundaries or rank candidate materials for the present waters.

4.4. Limitations and validation priorities

The study has four important limitations. First, Semarang contains only 58 development samples, so model uncertainty is substantial and the 64.3 mg/L cutoff is an engineering screen rather than a health standard. Second, Bayat is small, whereas Mekong differs strongly in geography, salinity structure, sampling period, and analytical workflow; its failed classifier transfer indicates domain shift rather than intrinsic algorithm failure. Third, the MCI is a transparent rank-based engineering summary rather than a mechanistic speciation model. Fourth, the adsorption benchmark integrates

DOI: 10.54254/2755-2721/2027.MA38438

heterogeneous published experiments; differences in precursor, modification, pH, dose, and initial sulfate preclude pooled effect estimates and direct factorial comparison. A prospective controlled experiment crossing material functionalization with the four locked hydrochemical regimes would provide the most direct next-stage validation.

5. Conclusion

This study develops a failure-aware bridge between environmental AI and circular sulfate-adsorbent engineering. The Semarang model showed moderate internal discrimination and regional Bayat transfer, but lost discrimination in the cross-country Mekong domain, where sulfate-ion relationships changed substantially. The locked external evaluation therefore withheld model-based screening in Mekong while retaining a target-free hydrochemical regime map. MCI remained associated with sulfate in Mekong and placed 73 of 78 high-sulfate events in the complex-matrix domain. Open experimental benchmarking added material-level evidence: Zr modification increased pomelo-peel-biochar qmax 34.5-fold, $CaCl_2$-modified kelp carbon retained substantial sulfate uptake in natural lake water, and modified Zea mays studies demonstrated formulation- and multicomponent-dependent behavior. The principal contribution is a structured separation between prediction, hydrochemical engineering interpretation, and external experimental evidence, with each evidence layer applied within the scope supported by its data.

Data availability

The Semarang development dataset is openly available from PANGAEA under CC BY 3.0 (https://doi.org/10.1594/PANGAEA.862987) [9]. The Bayat regional external dataset is available under CC BY 3.0 (https://doi.org/10.1594/PANGAEA.875165) [10], and the Mekong Delta challenge dataset under CC BY 4.0 (https://doi.org/10.1594/PANGAEA.974221) [11]. Mekong replicate analytical records were consolidated at the event level, and 118 complete events were used for locked external modelling. The adsorption benchmark is based on study-level values reported in the cited open-access publications [3, 4, 15, 16].

References

[1] U.S. Environmental Protection Agency. (2026). Secondary drinking water standards: Guidance for nuisance chemicals. https://www.epa.gov/sdwa/secondary-drinking-water-standards-guidance-nuisance-chemicals

[2] Olunusi, S. O., Adeleye, P. G., Oladipupo, S. B., Adeleye, A. T., & John, K. I. (2024). Biomass-derived biochar in wastewater treatment—A circular economy approach. *Waste Management Bulletin, 1*(4), 1–14. https://doi.org/10.1016/j.wmb.2023.07.007

[3] Ao, H., Cao, W., Hong, Y., Wu, J., & Wei, L. (2020). Adsorption of sulfate ion from water by zirconium oxide-modified biochar derived from pomelo peel. *Science of the Total Environment, 708*, 135092. https://doi.org/10.1016/j.scitotenv.2019.135092

[4] Tian, B., Song, Y., Wang, R., Wang, Y., Wang, T., Chu, J., Qiao, Z., Li, M., Lu, J., & Tong, Y. (2023). Adsorption of sulfate ions from water by $CaCl_2$-modified biochar derived from kelp. *RSC Sustainability, 1*, 898–913. https://doi.org/10.1039/D2SU00136E

[5] Tian, Y., Liu, Q., Ji, Y., Dang, Q., Sun, Y., He, X., Liu, Y., & Su, J. (2024). Prediction of sulfate concentrations in groundwater in areas with complex hydrogeological conditions based on machine learning. *Science of the Total Environment, 923*, 171312. https://doi.org/10.1016/j.scitotenv.2024.171312

[6] Makumbura, R. K., Mampitiya, L., Rathnayake, N., Meddage, D. P. P., Henna, S., Dang, T. L., Hoshino, Y., & Rathnayake, U. (2024). Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence techniques. *Results in Engineering, 23*, 102831. https://doi.org/10.1016/j.rineng.2024.102831

DOI: 10.54254/2755-2721/2027.MA38438

[7] Elsayed, A., Levison, J., Binns, A., Larocque, M., & Goel, P. (2026). A review of machine learning applications in the prediction of selected groundwater quality parameters: Key lessons, knowledge gaps, and future directions. *Science of the Total Environment, 1027*, 181693. https://doi.org/10.1016/j.scitotenv.2026.181693

[8] Jiang, S., Yang, Y., Mao, Y., Yan, H., Zhou, C., & Cheng, X. (2026). Group-aware and external validation distinguish interpolation from transfer in machine learning for aqueous pollutant removal. *Journal of Hazardous Materials, 517*, 143594. https://doi.org/10.1016/j.jhazmat.2026.143594

[9] Putranto, T. T., Rüde, T., & Irawan, D. E. (2016). Hydrochemical properties of groundwater samples in Semarang area, Java Island, Indonesia (1992, 1993, 2003, 2006, and 2007) [Data set]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.862987

[10] Putranto, T. T. (2017). Hydrochemical properties of groundwater samples in the Bayat Melange Complex, Klaten, Indonesia [Data set]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.875165

[11] Osenbrück, K., Steinel, A., Montcoudiol, N., Manh, L. V., Degtjarev, A., Fastert, T., Königer, P., Viol, M., Dee, M. W., & Bäumle, R. (2025). Hydrochemistry and isotope data from groundwater monitoring wells in the Mekong Delta, Vietnam [Data set]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.974221

[12] Chen, Y., Zhang, R., Wang, H., & Li, Y. (2024). Alkali activated winter melon peel as absorbent: Characterization and its adsorption performance for heavy metal ions in aqueous solutions. *Desalination and Water Treatment, 317*, 100055. https://doi.org/10.1016/j.dwt.2024.100055

[13] Singh, M., Rano, S., Roy, S., Mukherjee, P., Dalui, S., Gupta, G. K., Kumar, S., & Mondal, M. K. (2022). Characterization of organophosphate pesticide sorption of potato peel biochar as low cost adsorbent for chlorpyrifos removal. *Chemosphere, 297*, 134112. https://doi.org/10.1016/j.chemosphere.2022.134112

[14] Singh, R. J., Martin, C. E., Barr, D., & Rosengren, R. J. (2021). Cucumber peel bead biosorbent for multi-ion decontamination of drinking water collected from a mine region in New Zealand. *Environmental Technology, 42*(16), 2461–2477. https://doi.org/10.1080/09593330.2019.1703824

[15] Tejada-Tovar, C., Villabona-Ortíz, Á., Gonzalez-Delgado, A. D., Herrera, A., & Viera De la Voz, A. (2021). Efficient sulfate adsorption on modified adsorbents prepared from Zea mays stems. *Applied Sciences, 11*(4), 1596. https://doi.org/10.3390/app11041596

[16] Villabona-Ortíz, A., Ortega-Toro, R., & Tejada-Tovar, C. (2022). Selective and competitive adsorption of anions in solution on porous adsorbent from Zea mays steams: Kinetic and equilibrium study. *Water, 14*(18), 2906. https://doi.org/10.3390/w14182906