

<https://doi.org/10.1038/s44454-026-00028-2>

Environmental gradients explain nearshore microplastic distribution patterns: insights from machine learning models



Jiawei Li^{1,2,4}, Wenjun Sun^{3,4}, Yudong Wang³, Yixuan Cai^{1,2}, Zihao Wang^{1,2}, Xiangyun Xiong³, Xu Xu³ & Yuanyuan Tang^{1,2}✉

Microplastic is a global concern recently, yet the factors associated with particle behavior after entering marine environments remain uncertain. Using three years of observations integrated with unsupervised and machine learning, with feature-family ablation, results show that microplastic distributions in a highly urbanized nearshore consistently align with chemical environmental gradients, particularly nitrogen, total phosphorus, and trace metals. Rather than isolating individual transport drivers, our results indicate that microplastic patterns co-vary with biogeochemical regimes, while the contribution of freshwater and hydrodynamic proxies remains limited under the spatial and temporal resolution considered for classifiers. Variables related to river proximity and runoff potential exhibited lower relative classification importance, reflecting proxy limitations rather than the ecological irrelevance of hydrodynamic processes. Overall, these findings support the interpretation of microplastics as quasi-passive tracers embedded within coastal chemical gradients, integrating signals of eutrophication, wastewater inputs, and industrial activities. By leveraging routinely monitored water-quality and nearshore gradients features, this framework provides a transferable approach for interpreting microplastic patterns after their entry into coastal waters.

United Nations Sustainable Development Goal (SDG) 14.1 calls for reducing marine pollution, including plastics, through comprehensive monitoring and management¹. Coastal cities worldwide have adopted this target as a long-term priority, with strategies extending well beyond 2030². Microplastics—plastic particles <5 mm³—are now a significant global concern due to their widespread presence⁴, environmental persistence⁵, and potential ecotoxicity⁶. Studies indicate that >80% of marine microplastics originate from terrestrial sources⁷. Estuaries and nearshore areas, where urban pollution intersects with natural ecosystems⁸, often contain multiple pollution sources⁹, including domestic sewage¹⁰, industrial wastewater¹¹, agricultural runoff¹², and port logistics¹³. As urbanization accelerates and pollution sources multiply, reliance on a single source, such as freshwater runoff, may be insufficient to account for the dynamic distribution and risks of nearshore microplastic pollution¹⁴. This highlights the need for a broader

understanding of how environmental gradients influence the distribution of microplastics along coastal waters.

Microplastics interact with other pollutants in complex ways. While many studies have examined their spatial distribution, most focus on land-based inputs or static transportation patterns⁴. Mechanistic explanations for the heterogeneity of nearshore microplastic pollution under complex environmental gradients remain limited^{8,15,16}. Microplastics often coexist with nutrients, such as total phosphorus (TP) and nitrogen (e.g., total nitrogen (TN), ammonium (NH₄⁺), nitrate (NO₃⁻; NO₂⁻)), organic pollutants (e.g., COD, petroleum oil), and trace metals^{12,17–20}. Changes in water quality, including chemical and physical characteristics, will directly affect the environmental behavior of microplastics, both chemically and physically, including adsorption^{21,22}, agglomeration²³, and sink-and-resuspension^{24,25}. Environmental features such as temperature, salinity,

¹State Key Laboratory of Soil Pollution Control and Safety, School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China. ²Guangdong Provincial Key Laboratory of Soil and Groundwater Pollution Control, School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China. ³Guangdong Shenzhen Ecological Environment Monitoring Center Station, Shenzhen, China. ⁴These authors contributed equally: Jiawei Li, Wenjun Sun. ✉e-mail: tangyy@sustech.edu.cn

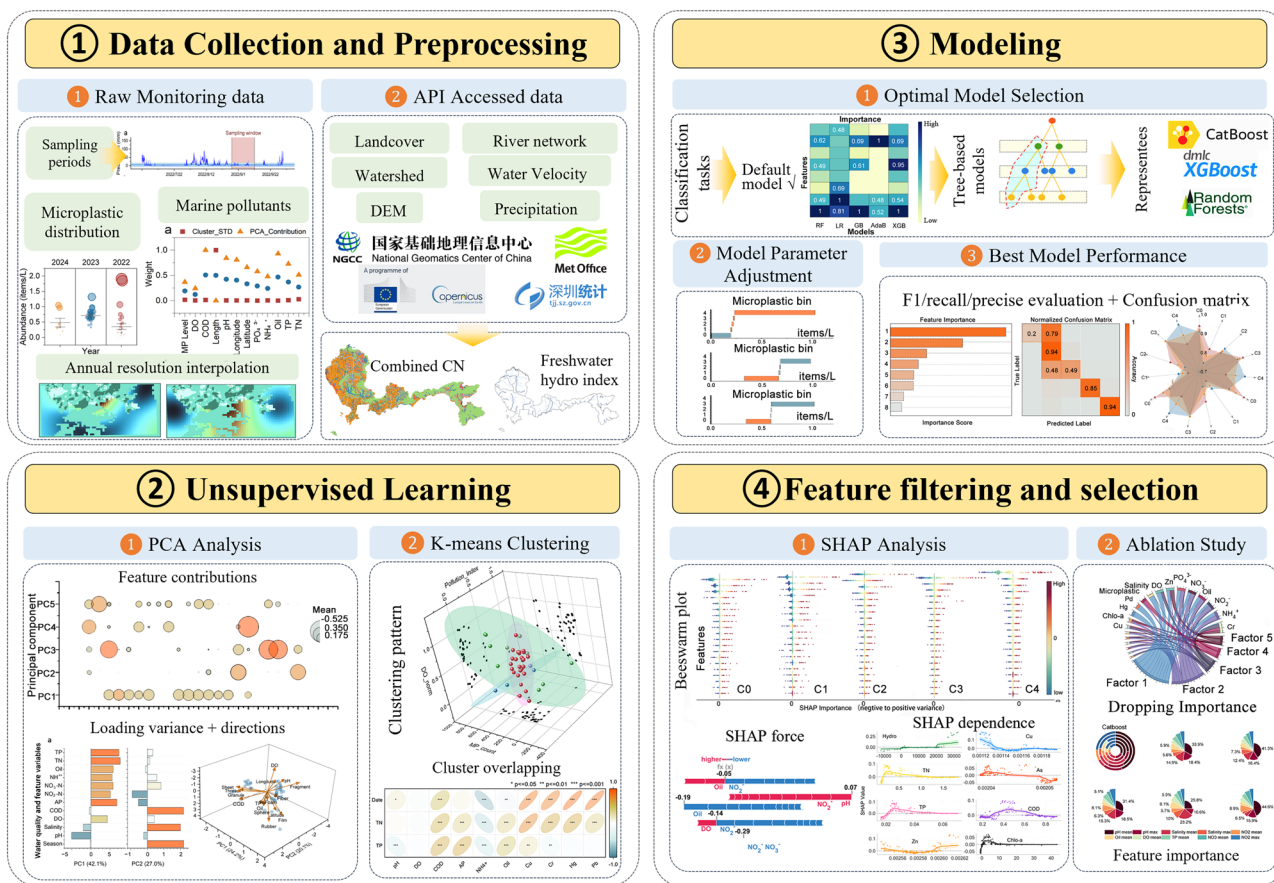


Fig. 1 | Integrated workflow for data-driven assessment of coastal microplastic pollution. The workflow comprises four modules: (1) harmonization of monitoring data with multi-source hydro-environmental predictors; (2) PCA and clustering to reveal dominant variance structures; (3) supervised model selection and

optimization for multi-class pollution classification; and (4) SHAP-based interpretability and feature reduction to isolate the most influential environmental determinants.

dissolved oxygen (DO), and pH can alter microplastic behavior by influencing biofilms and surface properties^{26–29}. Despite recognizing these relationships, studies typically focus on single pollutants^{30,31} or polymer types^{32,33}, thereby overlooking the reciprocal influence between environmental gradients and microplastic distribution^{34,35}.

Moreover, factors such as rainfall-induced changes in pH, temperature, and DO also influence microplastic and environmental behavior, as they often share the same pathways^{24,36}. These feature changes can trigger the redistribution and enrichment of microplastics and associated pollutants, affecting their transport and aggregation^{26,27,31}. Fluctuations in water-quality indicators, such as sudden changes in TP or DO due to hydrodynamic disturbances, can quickly alter the concentration and ecological impact of microplastics, limiting the representativeness of monitoring results^{20,32}. A common assumption in the previous study is that freshwater influx from rainfall and surface runoff is the primary driver of nearshore microplastic distribution^{14,37}. However, this assumption remains untested mainly under high-frequency, multi-parameter monitoring frameworks, a gap this study aims to address.

This research systematically quantifies multiple environmental factors and their relationships with microplastic distribution across various settings^{17,18}. It focuses on two key questions: (i) What are the main factors influencing microplastic pollution intensity and their interactions? (ii) What are the environmental gradients and water-quality characteristics driving the spatial variation of microplastic pollution in nearshore waters^{8,16}? By integrating machine-learning techniques (XGBoost, CatBoost, and Random Forest) for spatial clustering analysis, we aim to develop a multivariate framework for interpreting mechanisms and predicting spatial patterns^{38,39}. This work provides a theoretical and technical foundation for early-warning

systems and differentiated management of emerging pollutant risks in urbanized coastal ecosystems. The integrated workflow is given in Fig. 1²⁴⁰.

Results

Spatial gradients of microplastic and environmental variables

The PCA results reveal notable inter-year variance across the 2022–2024 datasets (Fig. 2)¹⁶. Features available for the 2022 summer and autumn, and 2023 and 2024 summer comprise 67 variables, with over 6000 spatially resampled 2 km grids each year linked to 42 microplastic sampling sites. Pollutants and features were monitored at 85 sites during the 2022 summer and autumn and 2023 summer, and at 106 sites during the 2024 summer. Microplastics showed strong loading contributions in the 2022 summer (PC1 = 39.17%, PC2 = 23.52%) and autumn (PC1 = 68.89%, PC2 = 11.47%) datasets, which are associated with key nearshore water features and pollutants^{8,15}. All-year PCA has the top loadings (>0.9) concentrated on trace metals, nitrogen contents, phosphorus contents, and natural gradient indicators such as salinity and temperature (Fig. 2f)^{12,21,41,42}. The overlapping cluster of the 2022 and 2024 data in PCA space, particularly along PC1 (34.85%) and PC2 (18.75%), suggests that changes in environmental factors such as TP, nitrogen compounds, and heavy metals were similarly influenced by familiar pollution sources or hydrological conditions^{14,43,44}. The observed pollutant concentration gradients reflect consistent water pollution patterns, with similar trends in TP and nitrogen compound concentrations^{12,42}. Hydrological features, influenced by river dynamics, precipitation, and proximity to river mouths, showed lower loadings in PCA variance (below 0.3), indicating their reduced impact on overall environmental variability in the region^{14,43}.

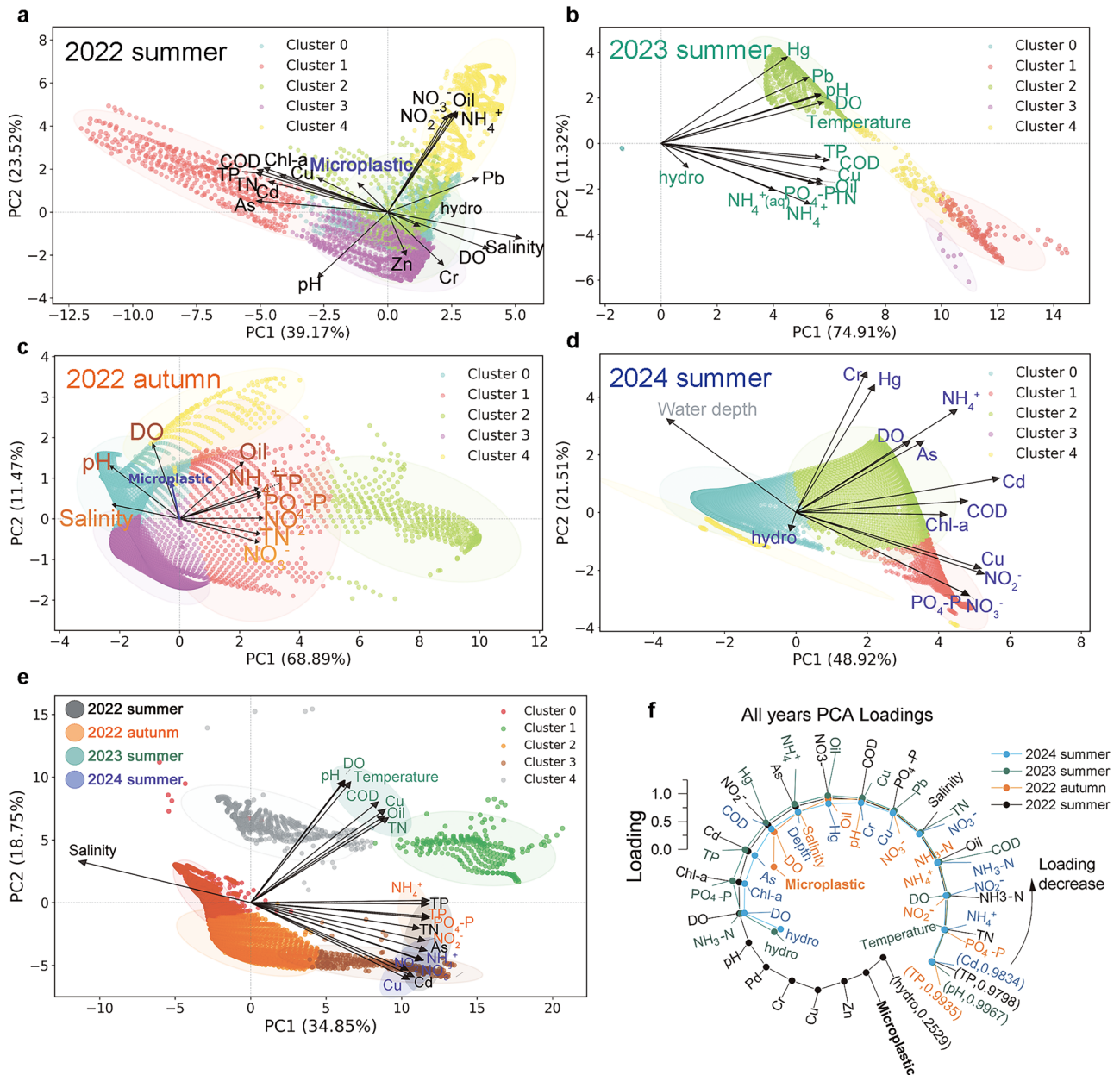


Fig. 2 | PCA biplots and loadings for environmental data across different years and seasons (2022–2024). a PCA result for the 2022 summer dataset, with clusters based on environmental features including microplastics. b PCA for the 2023 summer dataset. c PCA for the 2022 autumn dataset. d PCA for the

2024 summer dataset, showing the relationship between water depth and key chemical factors. e Combined PCA for all years and seasons. f Radial plot of PCA loadings across all years.

Specifically, nitrogen compounds (NO_2^- , NO_3^- , and dissolved NH_4^+), trace metals (Pb, Cd, Hg), and TP are positively correlated with microplastic abundance in the 2022 summer and autumn datasets (Fig. 2a, c). These pollutants may either directly influence microplastic distribution nearshore or overlap with microplastic sources and pathways^{8,11,23}. In contrast, the PCA loadings for the 2023 and 2024 summers (PC1 + PC2 = 86.23% and PC1 + PC2 = 70.43%, respectively) show that nitrogen compounds and trace metals contribute significantly to PCA variance (Fig. 2b, d). This indicates that NH_4^+ , NO_3^- , Pb, and Hg are primarily associated with pollutant gradient patterns nearshore^{20,43,44}.

K-means clustering further reveals distinct spatial variations in the features nearshore^{19,39}. Although microplastic abundance did not contribute to the top 20 PCA loadings (Table S6), clusters containing NH_4^+ , NO_3^- , trace metals (Pb, Hg), TN, and oil demonstrate clear gradient patterns^{12,41}. These variations are especially pronounced in the 2022 summer and autumn

datasets, where the loading vectors for microplastics align with those for pollutants, suggesting that the nearshore distribution of microplastics is influenced by these contaminants^{22,23}. In particular, nutrients and trace metals are concentrated nearshore, where lower salinity and pH are observed^{21,42}. Additionally, Fig. 2d highlights a positive correlation in the 2024 summer between water depth, DO, and COD, indicating that these pollutants exhibit patterns along nearshore feature gradients such as temperature, depth, distance to the river mouth, and water quality^{11,16,43}.

Eutrophication, organic pollution, and hydrological mixing are the primary environmental forces shaping spatial variability across the 2022 summer and autumn, and 2024 summer, with 2023 summer data separated. It may suggest a similar environmental gradient, as shown by the distributions and interpolations of the same feature in Figs. S2 and S5. The clustering and loading separation might be explained by the higher frequency of typhoon events and significant freshwater inputs during the

2023 summer^{45,46}. Meanwhile, PCA results across the entire dataset indicate that PC1 captures a dominant nutrient—organic pollution axis, with high positive loadings for nearly all nearshore features except salinity. In contrast, PC2 represents a secondary hydrographic gradient driven primarily by salinity, pH, COD, and TN, with contributions from trace Cu (Fig. 2e). This structure highlights that TP made significant contributions to both PC1 and PC2, particularly in the 2022 and 2024 summer datasets, suggesting similar water pollution characteristics and potential shared sources or processes^{12,20,42}. Nitrogen compounds (NO₃⁻, NH₄⁺) exhibited similar trends in both years, contributing substantially to both PCs, indicating their role as key markers of familiar pollution sources^{14,16}. Trace heavy metals further exhibit consistent co-loading patterns across years, suggesting a persistent pollution-related gradient possibly linked to industrial emissions or other anthropogenic sources^{11,21}. DO also showed a strong correlation with contaminants such as nitrogen compounds and Cu, further supporting the consistency between these datasets⁴³.

Through the combined PCA, K-means clustering, KDE, and VIF analyses, we identified trace metals (Cu and Pb), nitrogen compounds (NO₃⁻, NO₂⁻, NH₄⁺), TP, and DO as key features for driving machine-learning classification^{12,20,21}. Geographically, these gradients are influenced by varying hydrological conditions, with DO likely reflecting aspects of hydrodynamics, which are crucial for training and validating the classification model^{11,43}. Hydrological factors, including river governance levels, precipitation regimes, flow velocity, and distance to river mouths, were retained in the training set and sequentially removed to assess their impact on model performance. This analysis, along with the strong correlations between microplastic abundance and nutrients, suggests that microplastics act primarily as quasi-passive tracers, responding to complex environmental gradients^{8,16,44}. Overall, the unsupervised analyses establish a preview of the nearshore gradients framework for subsequent machine-learning modeling that microplastic distribution patterns are driven by a stable, interpretative hydro-geochemical regimes^{19,38}.

Model performance and feature analysis

Models were trained on a five-class system derived from microplastic abundance data collected in summer 2022. The classification thresholds were determined using natural-breaks (Jenks) discretization, which reflects different levels of microplastic abundance across nearshore regions and allows samples to be categorized into five distinct groups (low to high abundance)⁴⁷. The training dataset consisted of 807 samples, with the class distribution defined as: Class 0: 0–0.1739; Class 1: 0.1739–0.3470; Class 2: 0.3470–0.5047; Class 3: 0.5047–0.8395; Class 4: >0.8395 (unit: items/kg). The data highlighted regional differences in microplastic abundance during the summer of 2022 (Table S2). The classification performance of three machine-learning models—Random Forest, XGBoost, and CatBoost—was assessed on a k-fold test ($k = 5$, Table S7 and S8), an independent validation dataset (2023 summer, Table S9), and the hold-out data (20% of training dataset, Tables S10 and S11). The results indicated a strong predictive capability, with notable differences in how each model handled class imbalances and environmental features^{48–50}.

The Random Forest model achieved an overall accuracy of 98% on the fivefold test, approaching 96% on the 2023 summer independence set and 88% for the 20% held-out validation dataset, indicating strong performance in classifying the majority of the data. The macro-average F1 scores for each set were 0.86, 0.76, and 0.88, respectively, indicating the model's ability to balance precision and recall across classes⁴⁸. Notably, on k-fold, independent, and test set, the class 0 performed exceptionally well with an accuracy, precision, recall, and F1-score of 0.91–1. However, the model struggled to detect minority classes, particularly class 4, with one independent dataset precision dropping to 0.62. Meanwhile, performance on closed classes 0, 1, and 2 also declined, with class 1 showing the most significant drop, with precision falling to 0.49. This suggests that while the Random Forest model excels at identifying dominant patterns, it struggles with less prevalent and similar classes, indicating the need for further tuning or more robust strategies for handling imbalanced data^{38,50,51}. The XGBoost model achieved

96% accuracy, with a macro-average F1-score of 0.76, similar to Random Forest in overall accuracy⁴⁸. However, its performance in identifying minority classes was slightly better, particularly for class 1, which achieved a recall of 0.85 despite a precision of 0.49. This imbalance suggests that while XGBoost can capture the broader patterns in the data, it may be prone to overfitting in certain minority classes⁴⁹. The model showed relatively higher performance in classes 2 and 3, maintaining a balance between precision and recall, indicating its ability to handle intermediate categories more effectively^{19,48}.

The CatBoost model outperformed both Random Forest and XGBoost, achieving the highest accuracy of 98% and a macro-average F1-score of 0.83. It demonstrated balanced performance across all classes, particularly excelling in the minority classes. Class 4 achieved the highest recall of 0.86 and a precision of 0.72, indicating that CatBoost handles imbalanced data more effectively, likely because of its ability to consider interactions between features^{19,38}. This superior performance across all classes indicates that CatBoost is the most suitable model for this problem, as it effectively captures both majority- and minority-class patterns.

Classification performance and global drivers of microplastic pollution

The classification performance of the three machine-learning models is summarized in Fig. 3a–c, which together provide a comprehensive assessment of their ability to predict microplastic abundance classes across the five defined classes using the 2023 summer independent dataset. The radar plots in Fig. 3b compare precision, recall, and F1-score for Random Forest, CatBoost, and XGBoost, showing that all three algorithms achieve consistently high performance across classes⁵⁰. CatBoost exhibits the best balance between precision and recall, particularly for minority classes, demonstrating strong stability and greater capacity to recover the full spectrum of microplastic abundance classes. Class 0, representing the lowest microplastic abundance class, yields near-perfect classification across all models, while class 4, representing the highest concentrations, shows model-dependent variability, with CatBoost again providing the most reliable detection⁵⁰. These patterns reflect each model's ability to handle class imbalance and emphasize CatBoost's robustness in capturing both low- and high-abundance pollution regimes.

The confusion matrices in Fig. 3a for the independent 2023 summer test set further illustrate how the models handle subtle class boundaries, revealing that nearly all misclassifications occur between adjacent classes rather than across extreme class jumps. This behavior indicates that the microplastic concentration gradient exhibits ordinal continuity^{16,19}: samples with similar environmental signatures lie close to one another in feature space, making transitions between neighboring classes more likely than large jumps across multiple categories. For example, samples in class 3 (moderate concentrations) may occasionally be predicted as class 2, but are rarely misclassified as class 0 or class 4. This pattern is consistent with the spatially continuous nature of nearshore microplastic pollution^{8,16} and highlights the importance of defining appropriate class boundaries when constructing multi-class prediction systems^{8,15,16}.

Figure 3i–k presents the SHAP-based interpretation of model predictions, identifying the dominant environmental drivers that shape microplastic distribution patterns once particles enter the coastal system. The global SHAP bar plot in Fig. 3i, k shows that nutrients (NO₂⁻, NO₃⁻, TP) are the most influential predictors of microplastic abundance^{12,20,42}, consistent with the strong associations between eutrophication processes, particulate organic matter, and microplastic retention in nearshore waters. DO also emerges as a significant predictor, with lower DO—typical of organic pollution and eutrophic conditions—being associated with higher microplastic abundance classes^{16,43}. Oil contamination, indicative of industrial and urban runoff, is another major contributor, promoting microplastic aggregation and co-transport²³. The SHAP values in Fig. 3f, k show how these factors co-structure the separation of microplastic classes along chemical gradients, with greater dependence on these factors associated with elevated nutrient, oil, and trace metal concentrations. In

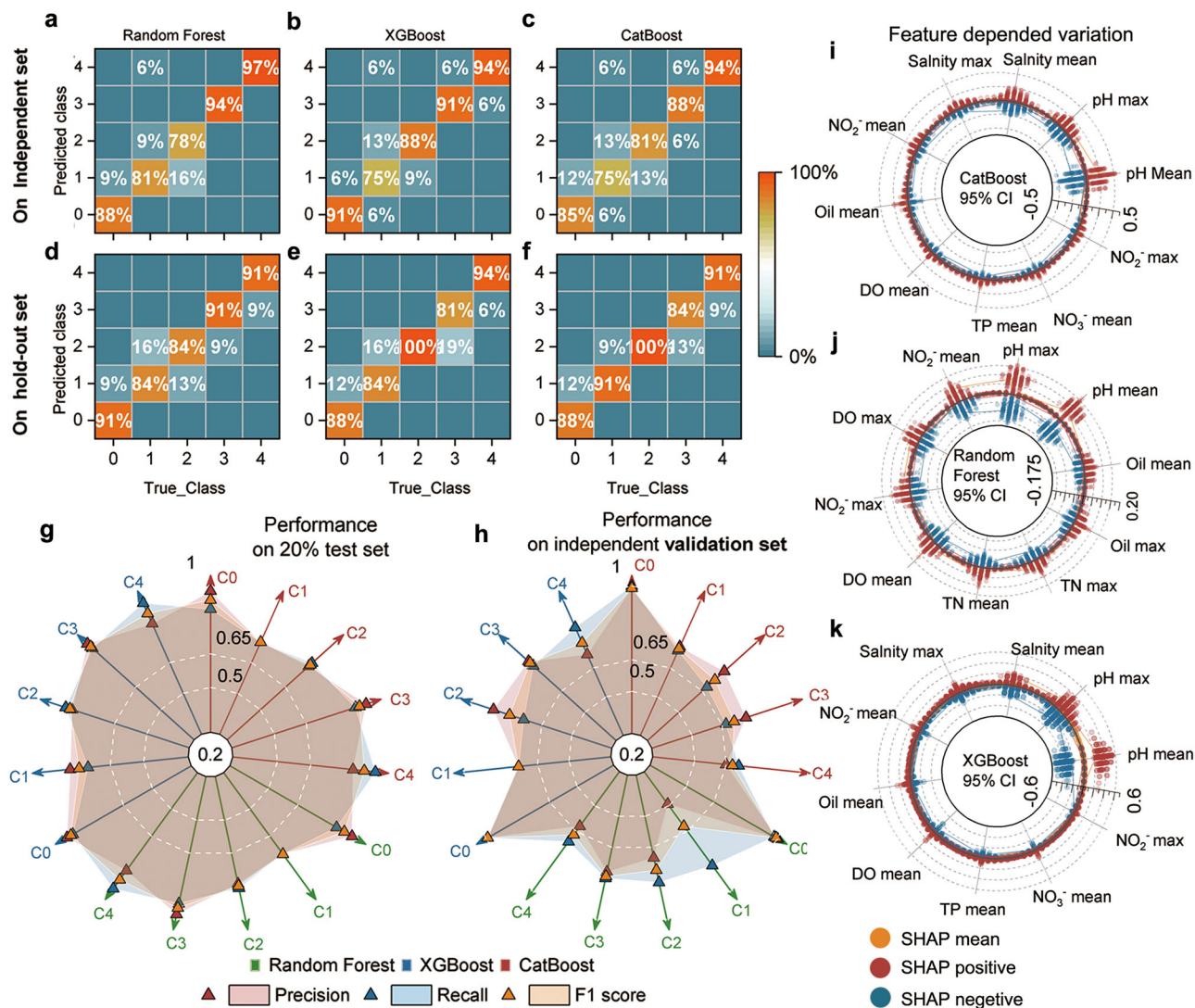


Fig. 3 | Performance and feature-dependence of three classification models. a–c Confusion matrices for Random Forest, XGBoost, and CatBoost were evaluated on the independent validation set. d–f Confusion matrices for the same models evaluated on the hold-out test set. g Class-wise precision, recall, and F1 scores for the three models on the 20% test set. h Class-wise precision, recall, and F1 scores for the

three models on the independent validation set. i–k Feature-dependent variation of model responses for CatBoost, Random Forest, and XGBoost, respectively, showing SHAP mean values and the distribution of positive and negative SHAP contributions (95% confidence intervals (CI), indicated).

contrast, the models’ classification results remain associated with higher salinity and more oxygenated conditions^{11,21}. Together, these SHAP analyses confirm that microplastics act as quasi-passive tracers of overlapping nutrient, organic pollution, and metal gradients^{8,16,20}, reinforcing the conclusion that their spatial distribution is governed primarily by coastal biogeochemical regimes^{14,43,44}.

The combined insights from the model performance metrics, adjacency-based misclassification patterns, and SHAP-derived interpretation provide a unified and robust framework for understanding the environmental drivers of nearshore microplastic pollution. Among the evaluated algorithms, CatBoost remains the most stable and accurate model for predicting microplastic abundance^{19,38}, consistently capturing both low- and high-level classes and aligning closely with the ordinal nature of the observed pollution gradient. SHAP analyses further demonstrate that particularly nitrogen nutrients (TN, NO₂⁻, NO₃⁻, NH₄⁺) and phosphorus, together with DO and oil, constitute the dominant predictors of microplastic variability^{12,16,20}. These variables define the major chemical gradients that organize microplastic distributions across coastal waters, reflecting eutrophication, organic pollution, and

industrial or urban runoff influences^{8,23,43}. Collectively, these patterns underscore that microplastics respond to multifactorial environmental regimes rather than to isolated drivers, reinforcing the necessity of incorporating chemical gradients into predictive modeling and the development of targeted ecological management strategies to mitigate microplastic pollution.

Key drivers of microplastic gradient

Figure 4 provides a comprehensive analysis of the class-specific SHAP fingerprints of microplastic pollution, highlighting the influence of individual environmental predictors and their relative contributions to microplastic class classification¹⁹. The SHAP (Shapley Additive Explanations) analysis allows us to understand how specific environmental features affect microplastic abundance across concentration classes from low to high³⁸. The SHAP beeswarm Plots for Classes 0–4 (Fig. 4a–e) visually represent the directional influence of each predictor, with cooler colors indicating lower feature values and warmer colors indicating higher values. These plots demonstrate how various environmental variables shape the distribution and classification of microplastics in nearshore environments¹⁶.

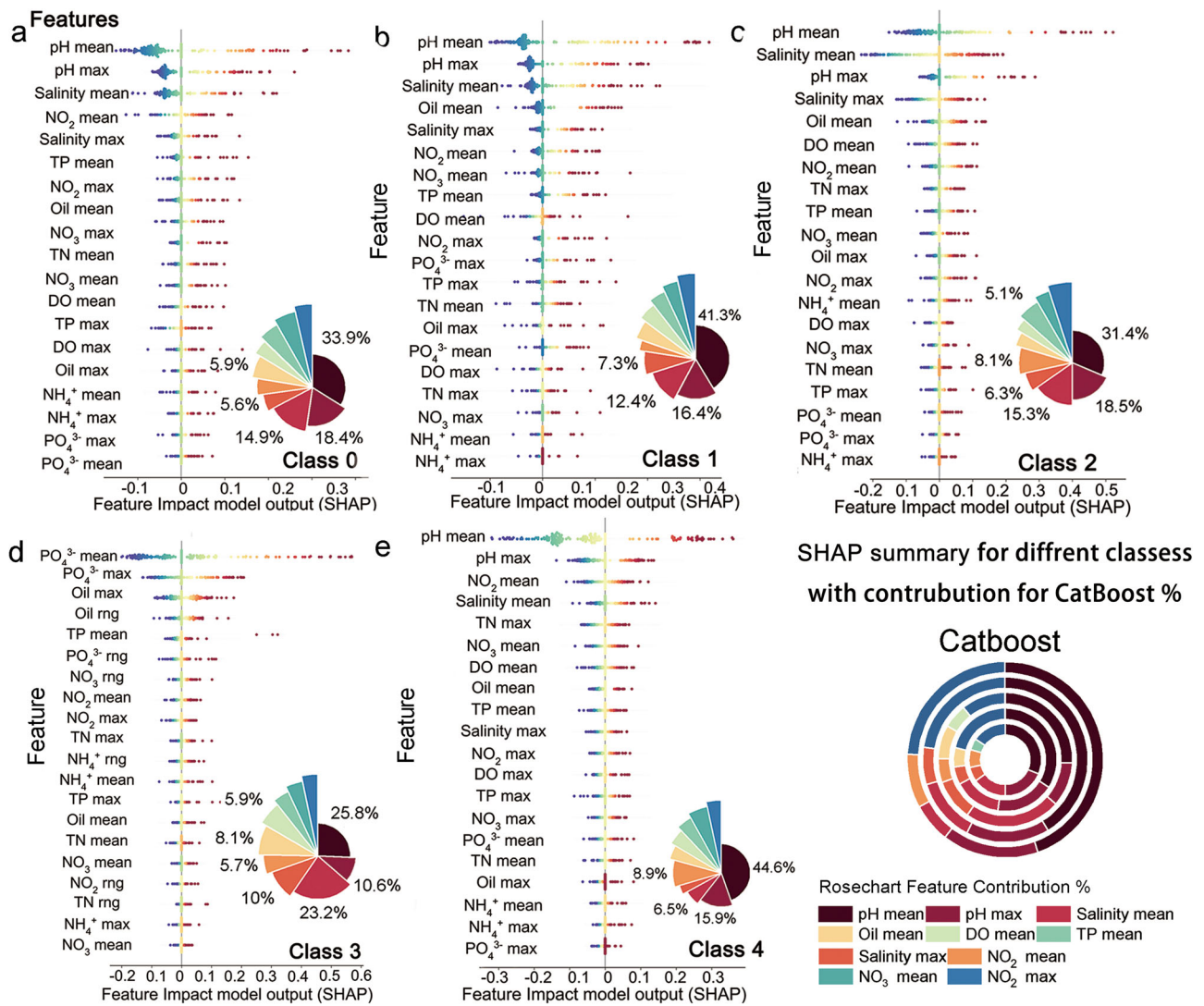


Fig. 4 | Class-specific SHAP fingerprints of microplastic pollution. a–e SHAP beeswarm plots for Classes 0–4 showing the directional influence of individual predictors. Cooler colors represent lower feature values; warmer colors indicate

higher values. Insets show the relative contribution of the main feature groups (pH, salinity, DO, TP, inorganic nitrogen, oil) to each class. f Integrated circular “fingerprnt” summarizing the dominant drivers across all classes.

Our analysis reveals that microplastic abundance is closely linked to nutrient pollution, organic pollutants, and hydrological conditions, with distinct environmental features acting as primary drivers across varying classes¹². Class 0, representing the lowest microplastic abundance, is primarily driven by salinity and pH, with higher values of these parameters correlating with lower microplastic levels. This suggests that areas with more stable, less-polluted waters—characterized by higher salinity and neutral pH—tend to have lower microplastic abundance. Secondary contributors, such as inorganic nitrogen compounds (NO₂⁻, NO₃⁻) and oil, also exert some influence but are less impactful in these less-polluted areas^{21,23}.

As microplastic abundance increases in Class 1, the importance of oil and DO rises, reflecting a shift toward more polluted environments^{16,23}. Elevated oil levels and reduced DO in these regions suggest that organic pollution and oxygen depletion are key factors in microplastic accumulation¹². Additionally, TP and NO₃⁻ become more significant, reinforcing the role of nutrient pollution in driving higher microplastic abundance^{12,14,20}. This trend intensifies in Class 2, where inorganic nitrogen compounds (NO₃⁻, NH₄⁺) and TP emerge as dominant predictors, highlighting the growing link between nutrient pollution and increased microplastic abundance^{19,21}. Oil and DO remain relevant contributors, particularly in areas impacted by organic pollution¹⁹.

In Class 3 and Class 4, where microplastic abundance is highest, the influence of inorganic nitrogen compounds (NH₄⁺ and NO₃⁻), TP, and oil becomes even more pronounced. These classes are strongly associated with nutrient and organic pollution, and the impact of DO, pH, and salinity is reduced. This indicates that nutrient and organic pollutants are the primary drivers of microplastic accumulation in highly polluted coastal areas. In Class 4, the highest microplastic abundance class, the dominance of eutrophication and organic pollution becomes evident, reinforcing the link between high nutrient levels, oil contamination, and elevated microplastic pollution^{8,23}.

The integrated circular “fingerprnt” in Fig. 4f summarizes the dominant drivers across all microplastic classes. This visualization highlights that low microplastic classes (Class 0) are characterized by high salinity and moderate pH, while high-microplastic classes (Class 3 and 4) exhibit elevated levels of inorganic nitrogen, TP, oil, and pH^{12,21}. These findings suggest that microplastic pollution is strongly influenced by nutrient enrichment, organic pollutants, and hydrological conditions (such as pH and salinity)^{16,20,23}. The contrasting environmental features between low- and high-microplastic-abundance classes underscore the complex relationship between microplastic pollution and environmental factors. Areas with higher salinity and more neutral pH typically show lower microplastic abundance class. In contrast, areas with elevated nutrient levels—

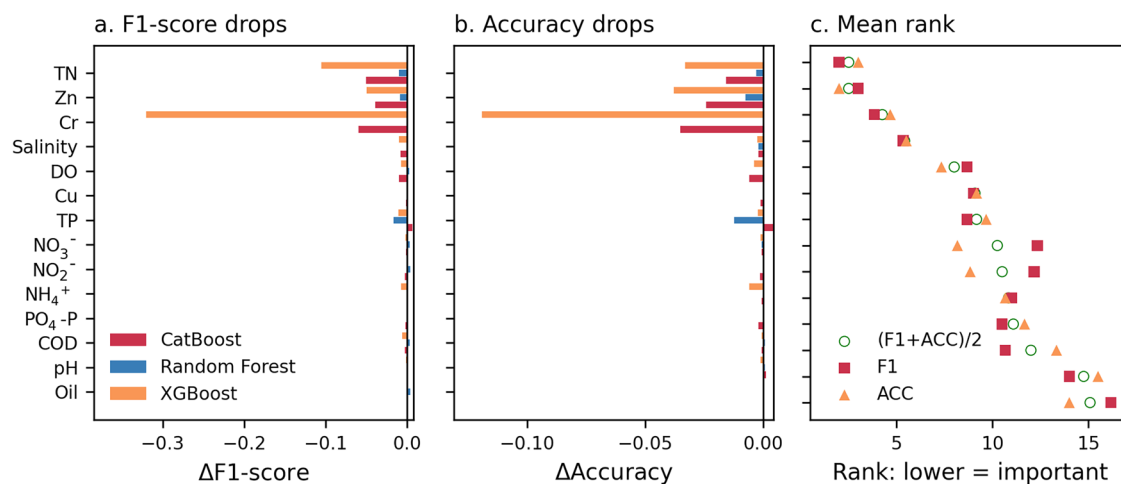


Fig. 5 | Feature-family ablation. **a** Declines in macro F1-score ($\Delta F1$) after removing each feature-family across CatBoost, Random Forest, and XGBoost. **b** Corresponding changes in overall accuracy (ΔACC). **c** Mean importance ranks (lower = more important) summarizing the overall contribution of each family across models.

particularly inorganic nitrogen and phosphorus, along with oil—are strongly associated with higher microplastic abundance classes^{19,27}.

To further validate the environmental gradients identified above, we conducted a feature-family ablation analysis to quantify the unique contribution of each environmental factor to model performance (Table S13). Removing individual feature families revealed that TN, Zn, and Cr form the three most influential gradients shaping nearshore microplastic patterns^{19,38,52}. Ablating TN led to consistent declines in F1 and accuracy across all models, confirming that the nutrients—representing land-derived nutrient loading and eutrophication—act as a stable and universal discriminator of microplastic pollution intensity¹². The Zn family produced the second-largest performance drop, reflecting its role as a proxy for industrial and port-related contamination that co-varies with microplastic sources in urbanized estuaries²¹.

The Cr family exhibited a striking model-specific effect: XGBoost performance deteriorated more sharply than for any other family, indicating that Cr carries strong nonlinear or threshold-dependent signals associated with industrial hotspots¹⁹. In contrast, Random Forest performance remained essentially unchanged, suggesting that other correlated predictors may partially absorb Cr-related information. This divergence highlights the complementary perspectives offered by different algorithms when interpreting anthropogenic pollution gradients²³.

Moderately significant contributors included salinity and DO, whose removal caused smaller yet consistent declines across models, underscoring their role in situating microplastic processes within the estuarine mixing and redox environment¹⁶. In contrast, nutrient species (NH_4^+ , NO_2^- , NO_3^- , PO_4-P), several heavy metals, and hydrographic parameters contributed minimally to the ablation, likely due to high collinearity with TN/Zn or a limited dynamic range in the study region^{11,20}. These results collectively reinforce that microplastic gradients in this coastal system are governed by the interplay of nutrient enrichment, industrial–port activity, and estuarine water-mass structure (Fig. S1), rather than by single-factor drivers^{8,11,20,27}.

Discussion

The combined evidence from PCA, K-means clustering, and machine-learning models supports a coherent picture: chemical environmental gradients are the primary structuring forces organizing microplastic distribution in this nearshore system^{19,38}. Rather than being randomly dispersed, microplastics are consistently aligned with gradients in nutrients, trace metals, and organic pollutants that reflect the intensity and composition of anthropogenic pressures^{12,20,23}.

PCA loadings identify a dominant axis driven by nutrient species (NO_2^- , NO_3^- , NH_4^+), total TP, trace metals (e.g., Pb, Cu), and organic pollutants (e.g., COD, petroleum oil), as well as by salinity, DO, and pH. These

variables explain much of the variance in the environmental data and segregate the study area into contrasting regimes, ranging from eutrophic inner-bay waters to more mixed offshore waters^{16,29,43}. Higher microplastic abundances systematically co-occur with nutrient-enriched and oil-polluted waters, particularly under conditions of lower salinity and altered pH, which are known to promote microbial activity, biofilm development, and particle aggregation^{21,29}. The low salinity of the water indicates that microplastics are concentrated in coastal and estuarine areas, highlighting that human activities, coastal eutrophication, and land-based inputs drive their accumulation in these areas.

K-means clustering independently reinforces this interpretation. Although microplastic abundance itself does not dominate the top PCA loadings, the spatial clusters derived from environmental variables reveal distinct groups in which NO_3^- , NH_4^+ , trace metals (e.g., Pb, Cu), and oil co-accumulate. These shifts indicate reorganization of variable coupling across years, rather than changes in the dominant environmental gradients. Microplastic-rich clusters are preferentially associated with these polluted states, suggesting that microplastics act as integrators of multiple anthropogenic inputs rather than responding to a single source^{11,19,23}. The clustering patterns thus point to a coupled “eutrophication–organic pollution–particle” regime, in which nutrients and organic pollutants accumulate, and microplastics are also retained or re-concentrated.

Importantly, although the dominant environmental gradients underlying these clusters remain consistent, the strongest interannual changes in pairwise correlations primarily involve pH, DO, nutrients (TN, TP, NH_4^+ , PO_4-P), organic matter (COD, oil), chlorophyll-a, and trace metals (Cu, Hg, Cd, Cr, As) (Fig. S7; Table S5). These shifts indicate a reorganization of variable coupling across years rather than changes in the dominant environmental gradients structuring microplastic distributions.

Machine-learning models (Random Forest, XGBoost, CatBoost) provide a third line of evidence. Features related to inorganic nitrogen, TP, oil, and DO consistently achieve high SHAP importance across models, confirming their central role in predicting microplastic classes. CatBoost, in particular, demonstrates strong stability and predictive performance when these gradients are included, underscoring their robustness in modeling nonlinear interactions. Together, these analyses indicate that microplastic patterns in this urbanized coastal system are primarily structured by chemical regimes and biogeochemical state, rather than by the spatial position of near-river input alone.

The feature-family ablation experiment reveals the specific environmental gradients that the models rely on most strongly to predict microplastic pollution (Fig. 5a–c). Systematically removing each family and recalculating performance demonstrates that TN produces the largest and most consistent declines in F1-score and accuracy across all three

algorithms¹⁹. This provides direct evidence that nitrogen enrichment functions as a foundational gradient shaping nearshore microplastic patterns¹². Ecologically, TN reflects processes such as primary production, biofilm growth, and particulate organic matter accumulation, each of which enhances particle adhesion and retention²⁰, explaining its dominant influence.

Heavy-metal families Zn and Cr emerge as the next most influential predictors, but with distinct model-dependent expressions that highlight the complementary strengths of different learning algorithms. Ablation of Zn leads to substantial performance losses—particularly for XGBoost—indicating that this family captures a robust industrial–port pollution axis aligned with known anthropogenic inputs⁵¹. Cr presents an even sharper contrast: its removal induces the largest F1 and accuracy drops in XGBoost, whereas Random Forest performance remains largely unchanged¹⁹. This divergence suggests that Cr encodes nonlinear or threshold-type signals linked to industrial hotspots, which boosting methods detect more sensitively than bagging models. These patterns collectively demonstrate the mechanistic relevance of Zn and Cr in defining chemical regimes associated with microplastic accumulation.

Secondary but consistent effects arise from salinity and DO, which show moderate performance declines and reflect their role as physical and biogeochemical context variables. Although neither acts as a primary pollution source, both variables influence water-mass structure, estuarine mixing, and redox conditions, thereby determining how nutrients, metals, oil, and microplastics are retained or exported²⁹. Their moderate yet uniform influence across models confirms that hydrographic gradients modulate, rather than dominate, microplastic spatial patterns.

In contrast, most remaining families—including NH_4^+ , NO_2^- , NO_3^- , $\text{PO}_4\text{-P}$, COD, Cu, pH, and oil—exhibit only minor performance changes upon removal (Fig. 5a–b), reflecting redundancy rather than lack of ecological relevance. Many of these predictors exhibit high collinearity with TN or Zn, show limited dynamic range, or vary on shorter timescales that are not fully resolved in the dataset^{12,19,23,51}. The integrated ranking (Fig. 5c) therefore clarifies the hierarchy of environmental controls: TN, Zn, and Cr represent the principal chemical axes organizing microplastic distributions, while salinity and DO provide structure related to hydrodynamic and biogeochemical state^{16,43}.

A central assumption in many coastal microplastic studies is that freshwater influx from rainfall and river discharge is the dominant driver of nearshore microplastic distribution. In this study, the performance of proxies for freshwater and hydrodynamics—such as water depth, distance to river mouths, and precipitation-linked flow metrics—consistently exhibits lower explanatory power than chemical gradients in PCA, clustering, SHAP importance, and feature-family ablation^{19,38}. This weak influence reflects the limited information available in hydrodynamic proxies at this spatial and temporal resolution here. In the study region, variations in flow and water depth may be sufficient to disperse and redistribute microplastics, but may not be resolved as strong gradients at the spatial and temporal resolution captured here¹². Chemical variables, in contrast, represent the integrated outcome of sources, transformation, and retention processes. Nutrient and metal concentrations accumulate and decay on timescales more commensurate with microplastic residence time, thereby encoding both the history and intensity of anthropogenic pressures^{16,20,23}. Further DO, salinity, and pH (Fig. S3–S5) might include some of the hydrological information. However, we still find that the chemical variables had a greater impact on the prediction of microplastic abundance classes. As a result, hydrodynamic factors primarily modulate transport and mixing, whereas chemical regimes ultimately determine where microplastics accumulate and persist, as classified using models^{11,19,43}.

These findings carry important implications for monitoring and management. First, they suggest that enhancing high-frequency monitoring of key chemical indicators—such as TN, TP, and trace metals—may yield greater gains in predictive capability than further refining coarse hydrodynamic proxies^{19,38}. Because many of these indicators are already part of routine water-quality programs, incorporating them into microplastic

early-warning frameworks is both feasible and cost-effective¹². Second, the strong dependence of model performance on nutrient and heavy-metal families indicates that reductions in eutrophication and industrial–port emissions are likely to deliver co-benefits in mitigating microplastic risks^{16,23}.

Finally, the combined evidence invites a reframing of microplastics as quasi-passive tracers embedded within an evolving chemical and biogeochemical landscape. Accordingly, our results indicate co-variation between microplastic classes and environmental parameters, rather than causal dominance of freshwater influx. Recognizing this coupling between microplastics and environmental gradients is critical for designing targeted interventions, prioritizing hotspots, and integrating microplastics into broader coastal water-quality management strategies^{14,43}.

In summary, this study advances a mechanistic understanding of nearshore microplastic pollution by demonstrating that chemical environmental gradients—rather than hydrodynamic forcing—serve as the primary structuring drivers of microplastic distribution^{19,38}. Through the integrated application of PCA, K-means clustering, SHAP interpretation, and a feature-family ablation framework, we provide convergent evidence that nitrogen-related pollutants, TP, pH, and salinity consistently outperform spatial in-land input dynamic variables in predicting microplastic abundance once the contaminants enter the sea^{12,23}. This methodological synthesis represents a key scientific contribution, enabling the disentanglement of complex, nonlinear pollutant–microplastic interactions that traditional approaches often overlook.

Our findings further position microplastics as quasi-passive tracers embedded within nutrient-enriched and anthropogenically impacted coastal regimes^{16,20}. The demonstrated dominance of nutrient and contaminant gradients highlights actionable leverage points for management: reducing eutrophication, wastewater inputs, and industrial–port emissions will likely yield measurable benefits for mitigating microplastic risks^{11,27}. Together, these findings establish a new paradigm in which microplastics are not random contaminants but may serve as potential tracers of the chemical regimes that define the modern coastal environment^{19,23,27}.

Methods

Study area and monitoring network

The study region covers the nearshore waters of Shenzhen, China, including semi-enclosed bays, port-associated areas, river-influenced transition zones, and open coastal waters along the eastern Pearl River Estuary (Shenzhen Government Online). The coastline is heavily urbanized and subject to intensive riverine inputs, port activities, coastal engineering, and marine tourism¹¹. A regularly nearshore water-quality monitoring network under the Chinese National Standard GB-17378 (“Marine Monitoring Specification”, GB-China National Standard code) provides routine measurements of physical and biogeochemical indicators at fixed monitoring sites, which are complemented by 2022 to 2024 summer seasonal ship-based surveys of physical parameters (temperature, water depth, pH, wind direction, precipitation and other weather conditions), and pollutants²³. Regionally, the periods correspond to peaks in runoff, biological activity, and anthropogenic influence. Tidal-phase effects were not explicitly resolved because sampling did not consistently capture complete flood–ebb cycles, and the analysis was conducted at a spatial scale that integrates conditions beyond individual tidal periods. Due to data availability, the weathering dataset, such as wind speed, precipitation, and marine water flow speed, was supplied by the CMEMS (Copernicus Marine Service) Global Wind and Stress (Monthly Production)⁵⁴, and multi-satellite data merged and reprojected cell data with the root mean square error $<1\text{ m/s}$ ^{35,56}. Marine precipitation and flow-velocity datasets were extracted from the NEMO v3.6 model using area averaging at $\sim 9\text{ km}$ resolution⁵⁷. The wind field dataset has a 27 km resolution within the field unit and was used as the control mask without interpolation²¹. Marine surface temperature above 10 m was obtained from the Met Office (Met Office Climate Data Portal)¹¹. Local daily weather, nearshore salinity, temperature, and wind speed information were used for data verification, as the city’s administrative

level is the district daily record (Shenzhen Ocean Metrological Unified Dataset, Shenzhen Statistics) (Table S4).

Marine water microplastics were sampled in summers between 2022 and 2024 (Fig. S10), consistently resolving full tidal cycles across stations and campaigns, using a manta trawl (mesh size 330 μm) to collect near-surface water (0–0.5 m depth). Each tow covered ~100 m³ of seawater at a constant towing speed, with GPS-based positions and towing time recorded for volume normalization. Samples were sieved, digested and filtered following protocols published in previous studies and China national guidelines^{26,27,38}. Suspected plastic particles were visually pre-selected under a stereomicroscope, and a subset was confirmed using μ-FTIR spectroscopy with a minimum match threshold of ≥70% against a polymer reference library²⁶. Field blanks (*n* ≈ 17), laboratory blanks (*n* ≈ 28), and replicate trawls (*n* ≈ 15) were used to detect and correct for procedural contamination; blank levels were subtracted from sample counts where necessary¹¹. Microplastic abundance was expressed as items per liter (items/L), and particles were classified by polymer type, color, and size for morphological characteristics²³.

In addition to standard hydrographic covariates (e.g., wind speed, wind direction, and precipitation), we derived a freshwater influx index to represent the potential for land-based microplastics to be delivered to the coastal sea via rainfall-driven surface runoff and river discharge²¹. Urban runoff is recognized as a significant pathway for microplastics entering nearshore waters, and we therefore quantified surface runoff using the Soil Conservation Service Curve Number (SCS-CN) method, which integrates land cover, precipitation, and hydrological connectivity¹².

Daily surface runoff *Q* was calculated as ref. 47:

$$Q = \frac{(P - 0.2S)^2}{P - 0.8S} \quad (1)$$

where *P* is daily precipitation, which is generally obtained from the Shenzhen monitoring network, and *S* is the potential maximum retention, default as ref. 27:

$$S = \frac{5400}{CN} - 254 \quad (2)$$

The CN characterizes the runoff potential of each land cover–soil complex and reflects soil properties, land-use type, and hydrologic condition. Land cover was obtained from the GLAD land-cover dataset, and CN was extracted from the Soil and Water Conservation Planning of Shenzhen. Based on these data, land-cover weights were assigned as follows: water surface, 1.0; wetland, 0.85; high-density land use, 0.95; low-density land use, 0.85; green land, 0.35; and barren land, 0.35 (Table S3), which were used to derive spatially explicit CN values. The hydrographic covariate proxy is estimated to quantify the potential enhancement from details of riverine and hydrologic features, thereby enabling an assessment of freshwater input.

To translate runoff into an index of freshwater influx to the sea, we combined surface runoff, river hierarchy, and topographic slope. The freshwater inflow *F* was computed as ref. 27:

$$F = Q \times C_r \times R_{sl} \quad (3)$$

where the *C_r* is the river class obtained from the Shenzhen government and provincial monitoring sources, weighted from 1 to 5 according to stream (water path, drain, canal, stream, river, respectively) and aggregated into a freshwater hydrological index (hydro) (Fig. S13, Table S4)¹¹. *R_{sl}* is the slope factor derived from a digital elevation model (30 m × 30 m), representing the propensity for runoff to reach the coastline⁴³. The resulting field was further modulated by a 500 m decay buffer from each river reach toward the coastal sea, and then resampled to the 2 km × 2 km grid centroids used for microplastic observations²⁷. This freshwater influx index was finally combined with microplastic measurements and other hydro-environmental

features (Fig. S11) as a predictor in the supervised learning models (Fig. S10)²³.

Feature selection and data screening

For classification modeling, microplastic abundance for each period was divided into five ordered abundance classes (Class 0–Class 4) using Jenks natural breaks⁵³ to represent ordinal levels along a relative concentration gradient. Abundance values were first mapped onto a 2-km grid and normalized before modeling; class boundaries were then established to partition the normalized distribution into five ordered ranges. Importantly, these classes are relative within each sampling period rather than fixed absolute pollution thresholds, so identical abundance values may correspond to different class labels across years when the overall concentration range varies. The resulting class definitions, representing progressively increasing microplastic abundance levels, are summarized in Table S2.

The pollution and parameter records of conventional monitoring in coastal waters for three years were selected, including nitrate (NO₃⁻), nitrite (NO₂⁻), inorganic ammonium (NH₄⁺), dissolved phosphate (PO₄-P), total nitrogen (TN), TP, oil (petroleum oil), chlorophyll-a (Chl-a), copper (Cu), zinc (Zn), chromium (Cr), cadmium (Cd), mercury (Hg), arsenic (As), pH, salinity, temperature¹². Variables that did not exceed detection limits across sampling periods and regions, as well as predictors exhibiting near-zero variance across all samples, were excluded prior to statistical analysis. Meanwhile, through Kriging’s interpolation under 30 m searching and 0.3 searching fraction of spatial record from 3 years (Figs. S2–S5), the features were excluded in the 2 km grid with records <50% in the while area were filtered before the training and validation process³⁸.

Variance filtering, correlation pruning, and unsupervised rotation-based analyses were applied to examine the robustness and redundancy of environmental features¹³. Pairwise comparisons between selected variables were performed using kernel density estimation (KDE) to approximate their empirical probability density functions, and the degree of similarity between distributions was quantified as the percentage of overlapping area¹¹. Further variance inflation factor (VIF) values are provided in Table S1, and overlap metrics are shown in Fig. S6. Core predictors were selected based on multicollinearity thresholds <10 and low redundancy (*R*² approaching 1)²³. For a feature, its VIF was derived from the following formula¹⁹:

$$VIF_i = \frac{1}{1 - R_i^2} \quad (4)$$

where *R*² is the coefficient of determination, which is obtained by the regression prediction of *X_i* from all other independent variables. The final selected features were validated for multicollinearity with VIF < 10 to ensure model robustness²³. Features with VIFs between 10 and 15 were further screened by combining simple factor analysis (by principal component analysis) and KDE distribution overlap, and reconsidered after the KDE overlapping result (Fig. S8)²⁷. Annual correlation differences were computed between 2023 and 2022 and between 2024 and 2022, respectively, to validate the yearly feature differences between these years (Table S5, Fig. S9). In addition, the correlation structure was reorganized across years. K-means groups were set to 4, as cluster 1 represents the chemical redox features, cluster 2 is nitrogen nutrients (TN, NO₃⁻, NO₂⁻, and NH₄⁺), cluster 3 is phosphorus-related features (TP, PO₄-P), and cluster 4 represents related parameters, such as river mouth distance, precipitation, and surface temperature. The clusters were identified by the correlation coefficient only for feature selection validation.

To create a common spatial framework for both microplastic and environmental data, the study area was overlaid with a 2 km × 2 km grid. Spatial interpolation and grid generation use the geopandas package in Python⁴³. All site measurements falling within a grid cell were aggregated to derive summary statistics, including the mean, maximum, and within-cell range (rng) for each environmental variable¹⁹. Microplastic abundance and class for each cell were calculated from the trawl data within that cell and output as a CSV file, including the grid ID, to serve as a data record search

index. Grid cells with no valid observations for microplastics or key environmental variables were excluded, set to a null value (N/A), and removed from the training and validation datasets. This procedure produced a spatially unified record dataset suitable for both unsupervised analyses and supervised classification, with a uniform spatial resolution, because the dataset resolutions range from meters to approximately 10 kilometers.

Unsupervised learning workflow

To characterize the dominant multivariate gradients in the environmental dataset and identify internally coherent groups of sampling sites, we implemented an unsupervised learning workflow combining PCA and K-means clustering. Before PCA, all numerical variables were standardized to a mean of 0 and unit variance using the StandardScaler transformation. PCA was then applied to the standardized matrix to summarize co-variation patterns among environmental and pollutant variables. Components were retained based on the cumulative explained variance criterion (ratio > 0.9), ensuring that the first two principal components (PC1 and PC2) captured the dominant structure of the dataset. The loading vectors were computed as the eigenvectors of the covariance matrix scaled by the square root of the corresponding eigenvalues, providing a measure of each variable's contribution to the principal axes (Table S7)⁴³.

To facilitate interpretation, only the variables with the most significant absolute loadings were visualized. The top-loading features were ranked by their Euclidean loading magnitude across PC1 – PC2, and the ten most important contributors were summarized in a loading bar plot⁵. For the PCA biplot, loading vectors were projected in the PC1 – PC2 plane and scaled to 70% of the data range to maintain visual proportionality¹⁹. The Microplastic variable was highlighted by rendering its loading arrow and label in red, thereby reflecting its importance in the ordination while keeping other features unobtrusive¹².

To identify unsupervised groups of samples exhibiting similar multivariate profiles, K-means clustering was performed on the PCA scores rather than the raw variables, thereby reducing noise and collinearity before partitioning. The number of clusters (k) was selected based on data interpretability and cluster separation in ordination space³⁸. Cluster membership was then mapped onto the PCA biplot using distinct point markers for each cluster. Point sizes were further scaled by the magnitude of each sample's PCA score (Euclidean distance from the origin in PC1 – PC2 space), providing a visual cue of how strongly each sample is positioned along the dominant gradients (Fig. S7)²⁷. To summarize the dispersion of each cluster in PC space, 95% or 99% confidence ellipses were drawn using the empirical covariance matrix of the PCA scores within each cluster, with the final 95% confidence interval used for visualization. Ellipses were rendered as filled polygons with semi-transparent colors to allow overlap, aiding comparison of cluster orientation and separation.

The final multivariate visualization integrates sample distribution, cluster structure, and variable loadings into a single PCA biplot. Samples are shown as size-scaled points with cluster-specific markers; confidence ellipses enclose clusters; and only the most influential loading vectors are displayed. This unified ordination framework provides an interpretation, an unsupervised representation of how environmental variables jointly organize samples into characteristic multivariate groups. Correlations between microplastic abundance and environmental variables were assessed using Spearman's rank correlation, and the results were summarized in heatmaps using QGIS 3.44.0 and presented in Fig. S6.

We then implemented a supervised machine-learning framework to predict discretized microplastic abundance from physicochemical and pollutant indicators as model inputs. All available yearly datasets (2022 summer) were merged and harmonized, after which the target variable was converted into a five-class categorical label through a standardized binning procedure to enable multi-class classification. Before model training, we ensured the completeness of essential keys (e.g., fid, latitude, longitude). We derived family-level features according to predefined variable groups, yielding a unified feature matrix for subsequent analysis. All predictors were coerced into numerical form, and class weights were computed

from the empirical class distribution to mitigate potential imbalance effects during training¹⁹.

Supervised learning workflow

Three tree-based ensemble models, Random Forest, CatBoost, and XGBoost¹⁹, were applied to classify microplastic abundance classes. All features were standardized²⁷. To evaluate generalization while reducing spatial autocorrelation, we adopted a spatially blocked cross-validation strategy. Grid cells were grouped into 10 spatial clusters using K-means on latitude–longitude coordinates, and these clusters defined the folds for cross-validation (GroupKFold). In each fold, one spatial cluster was withheld for validation, and the remaining clusters were used for training. The significant performance gap between these validation methods indicates strong spatial autocorrelation in the data. Therefore, each model was trained using the folds-grouped result with the same feature set, class weights, and random seed to ensure comparability across algorithms. XGBoost and CatBoost were fitted using their native gradient-boosting procedures, and Random Forest was trained using bootstrap aggregation with weighted impurity splitting⁴³.

After model selection, the final Random Forest, XGBoost, and CatBoost models were fitted to the training data and evaluated on a held-out 20% test set ($N = 162$ cells) from the 2022 summer dataset. Model performance was assessed on an independent validation set (the 2023 summer dataset). None of the test and validation datasets were used in any training or tuning step. Seasonal changes and the 2024 dataset were excluded from the training and test sets because metal records were missing from the 2022 autumn set. Meanwhile, the 2024 dataset records only a 10% microplastic abundance distribution, reducing the number of features available for cross-validation. Targeted classes were defined as Class 0–Class 4 (Table S2), representing low to high levels of microplastic abundance. The thresholds were automatically adjusted to accurately reflect the microplastic distribution pattern in each period, for both the independent validation dataset and additional datasets from different sampling periods. This ensures that the classification captures characteristics across periods and prevents data leakage during annual processes modeling. For each algorithm, predictions were compared with the actual class labels to compute classification accuracy (ACC) and class-specific precision, recall, and F1 Scores. The validation split provided an unbiased assessment of generalization performance, and confusion matrices were generated to visualize misclassification patterns across the five microplastic abundance classes. The best-performing model was selected based on overall predictive accuracy and macro-averaged F1-score. All trained models were saved for reproducibility, together with metadata including feature lists, class weights, and training parameters.

To interpret the fitted classification models in a consistent, model-agnostic way, we employed SHAP (SHAPley Additive exPlanations). For each model, SHAP values were computed for every grid cell and feature, decomposing the predicted class probabilities into additive contributions of individual variables³⁸. Global feature importance was quantified as the mean absolute SHAP value across all samples and classes, and the results were summarized in bar plots.

To examine how the influence of each variable changes across the predictor range and among microplastic classes, we derived class-wise SHAP summaries using a one-vs-rest formulation and visualized them with beeswarm plots and class-specific “fingerprint” diagrams. Positive SHAP values indicate conditions driving predictions towards higher microplastic classes, while negative values indicate contributions towards lower classes. This allowed us to link specific hydro-biogeochemical regimes (e.g., high nutrients, high pH, low salinity) to each microplastic abundance class in an interpretable manner.

All data processing and statistical analyses were performed in Python (version 3.12.0) using standard scientific libraries, including NumPy, pandas, scikit-learn, XGBoost, CatBoost, and SHAP. Figures and maps were generated with Matplotlib, the Seaborn package, and QGIS 3.40.0. The information on the features, parameters, abbreviations, and terms used is provided in Table S14.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to the Statistical Management Measures of the Ministry of Ecology and Environment of China (Order No. 29 of the Ministry of Ecology and Environment on 18 January 2023), but are available from the corresponding author on reasonable request.

Code availability

All data processing, analyses, and model development were performed using custom Python scripts developed by the authors. The code is not publicly available due to data-sharing and institutional restrictions, but can be obtained from the corresponding author upon reasonable request for academic, non-commercial use. Analyses were conducted in Python (version 3.12.0) using standard scientific computing and machine-learning libraries.

Received: 15 September 2025; Accepted: 22 January 2026;

Published online: 01 April 2026

References

- Kathuria, V. & Jardosh, N. Sustainable development goals and their emphasis on managing plastic pollution. In *Routledge Handbook of the UN Sustainable Development Goals Research and Policy* (Routledge, 2025).
- Graham, R. E. D. Achieving greater policy coherence and harmonisation for marine litter management in the North-East Atlantic and Wider Caribbean Region. *Mar. Pollut. Bull.* **180**, 113818 (2022).
- ISO 24187:2023. *Principles for the analysis of microplastics present in the environment* (ISO, 2023).
- Hale, R. C., Seeley, M. E., La Guardia, M. J., Mai, L. & Zeng, E. Y. A global perspective on microplastics. *J. Geophys. Res. Oceans* **125**, e2018JC014719 (2020).
- Bergmann, M., Allen, S., Krumpen, T. & Allen, D. High levels of microplastics in the Arctic sea ice alga *Melosira arctica*, a vector to ice-associated and benthic food webs. *Environ. Sci. Technol.* **57**, 6799–6807 (2023).
- Liu, F., Lorenz, C. & Zhao, G. From land to sea: hydrological source tracking of microplastics in coastal sediments. *Environ. Res.* **283**, 122132 (2025).
- Boucher, J. & Friot, D. *Primary Microplastics in the Oceans: A Global Evaluation of Sources* (IUCN, 2017).
- Su, L. et al. Superimposed microplastic pollution in a coastal metropolis. *Water Res.* **168**, 115140 (2020).
- Browne, M. A. et al. Accumulation of microplastic on shorelines worldwide: sources and sinks. *Environ. Sci. Technol.* **45**, 9175–9179 (2011).
- Woodward, J., Li, J., Rothwell, J. & Hurley, R. Acute riverine microplastic contamination due to avoidable releases of untreated wastewater. *Nat. Sustain.* **4**, 793–802 (2021).
- Brennecke, D., Duarte, B., Paiva, F., Caçador, I. & Canning-Clode, J. Microplastics as vector for heavy metal contamination from the marine environment. *Estuar. Coast. Shelf Sci.* **178**, 189–195 (2016).
- Mockler, E. M. et al. Sources of nitrogen and phosphorus emissions to Irish rivers and coastal waters: estimates from a nutrient load apportionment framework. *Sci. Total Environ.* **601–602**, 326–339 (2017).
- GESAMP. *Sea-based Sources of Marine Litter* (IMO, 2021). GESAMP Reports & Studies No. 108.
- Bai, M., Lin, Y., Hurley, R., Zhu, L. & Li, D. Controlling factors of microplastic riverine flux and implications for reliable monitoring strategy. *Environ. Sci. Technol.* **56**, 48–61 (2022).
- Jin, X. et al. Quantitative assessment on the distribution patterns of microplastics in global inland waters. *Commun. Earth Environ.* **6**, 331 (2025).
- Sun, X. et al. Factors influencing the occurrence and distribution of microplastics in coastal sediments: from source to sink. *J. Hazard. Mater.* **410**, 124982 (2021).
- Jahnke, A. et al. Reducing uncertainty and confronting ignorance about the possible impacts of weathering plastic in the marine environment. *Environ. Sci. Technol. Lett.* **4**, 85–90 (2017).
- Adam, V., von Wyl, A. & Nowack, B. Probabilistic environmental risk assessment of microplastics in marine habitats. *Aquat. Toxicol.* **230**, 105689 (2021).
- Phan, S. & Luscombe, C. K. Recent trends in marine microplastic modeling and machine learning tools: potential for long-term microplastic monitoring. *J. Appl. Phys.* **133**, 020701 (2023).
- Wang, H. et al. Hindcasting harmful algal bloom risk due to land-based nutrient pollution in the Eastern Chinese coastal seas. *Water Res.* **231**, 119669 (2023).
- Niu, L. et al. Metal pollution in the Pearl River Estuary and implications for estuary management: the influence of hydrological connectivity associated with estuarine mixing. *Ecotoxicol. Environ. Saf.* **225**, 112747 (2021).
- Shi, M. et al. Adsorption of heavy metals on biodegradable and conventional microplastics in the Pearl River Estuary, China. *Environ. Pollut.* **322**, 121158 (2023).
- Cui, W. et al. Sorption of representative organic contaminants on microplastics: effects of chemical environmental properties, particle size, and biofilm presence. *Ecotoxicol. Environ. Saf.* **251**, 114533 (2023).
- Hanun, J. N. et al. Weathering effect triggers the sorption enhancement of microplastics against oxybenzone. *Environ. Technol. Innov.* **30**, 103112 (2023).
- Sooriyakumar, P. et al. Biofilm formation and its implications on the properties and fate of microplastics in aquatic environments: a review. *J. Hazard. Mater. Adv.* **6**, 100077 (2022).
- Hurley, R., Woodward, J. & Rothwell, J. J. Microplastic contamination of river beds significantly reduced by catchment-wide flooding. *Nat. Geosci.* **11**, 251–257 (2018).
- Zhang, L. et al. Dynamic distribution of microplastics in mangrove sediments in Beibu Gulf, South China: implications of tidal current velocity and tidal range. *J. Hazard. Mater.* **399**, 122849 (2020).
- Rummel, C. D., Jahnke, A., Gorokhova, E., Kühnel, D. & Schmitt-Jansen, M. Impacts of biofilm formation on the fate and potential effects of microplastic in the aquatic environment. *Environ. Sci. Technol. Lett.* **4**, 258–267 (2017).
- Zhang, B. et al. Spatial and seasonal variations in biofilm formation on microplastics in coastal waters. *Sci. Total Environ.* **770**, 145303 (2021).
- Delacuvellerie, A. et al. Microbial biofilm composition and polymer degradation of compostable and non-compostable plastics immersed in the marine environment. *J. Hazard. Mater.* **419**, 126526 (2021).
- Pang, G. et al. The distinct plastisphere microbiome in the terrestrial-marine ecotone is a reservoir for putative degraders of petroleum-based polymers. *J. Hazard. Mater.* **453**, 131399 (2023).
- Yu, H. et al. Polyethylene microplastics interfere with the nutrient cycle in water-plant-sediment systems. *Water Res.* **214**, 118191 (2022).
- Liu, S., Huang, Z., Yang, C., Yao, Q. & Dang, Z. Effect of polystyrene microplastics on the degradation of sulfamethazine: the role of persistent free radicals. *Sci. Total Environ.* **833**, 155024 (2022).
- Abinandan, S., Praveen, K., Venkateswarlu, K. & Megharaj, M. Microalgae–microplastics interactions at environmentally relevant concentrations: implications toward ecology, bioeconomy, and UN SDGs. *Water Res.* **247**, 120778 (2023).
- Kang, W., Sun, S. & Hu, X. Microplastics trigger the Matthew effect on nitrogen assimilation in marine diatoms at an environmentally relevant concentration. *Water Res.* **233**, 119762 (2023).

36. Shabarova, T. et al. Recovery of freshwater microbial communities after extreme rain events is mediated by cyclic succession. *Nat. Microbiol.* **6**, 479–488 (2021).
37. Werbowski, L. M. et al. Urban stormwater runoff: a major pathway for anthropogenic particles, black rubbery fragments, and other types of microplastics to urban receiving waters. *ACS EST Water* **1**, 1420–1428 (2021).
38. Stokal, M. et al. River export of macro- and microplastics to seas by sources worldwide. *Nat. Commun.* **14**, 4842 (2023).
39. Cai, L. et al. Global models and predictions of plant diversity based on advanced machine learning techniques. *N. Phytol.* **237**, 1432–1445 (2023).
40. Li, M. et al. Implications of seawater characteristics on dissolved heavy metals in near-shore surface waters of the Yellow Sea. *Mar. Pollut. Bull.* **211**, 117469 (2025).
41. Wu, Y. et al. Seasonal variability of microplastic transport modulated by tides: a mass-based assessment in a turbidity maximum zone. *J. Hazard. Mater.* **501**, 140607 (2025).
42. Rummel, C. D. et al. Plastic ingestion by pelagic and demersal fish from the North Sea and Baltic Sea. *Mar. Pollut. Bull.* **102**, 134–141 (2016).
43. Zhen, Y., Wang, L., Sun, H. & Liu, C. Prediction of microplastic abundance in surface water of the ocean and influencing factors based on ensemble learning. *Environ. Pollut.* **331**, 121834 (2023).
44. Schmidt, C., Krauth, T. & Wagner, S. Export of plastic debris by rivers into the sea. *Environ. Sci. Technol.* **51**, 12246–12253 (2017).
45. Yu, F. & Hu, X. Machine learning may accelerate the recognition and control of microplastic pollution: future prospects. *J. Hazard. Mater.* **432**, 128730 (2022).
46. Li, M. et al. Estuarine sediment dynamics influenced by successive typhoons: turbidity maximum zone response and mechanisms in the Pearl River Estuary. *J. Geophys. Res. Oceans* **130**, e2024JC022301 (2025).
47. Wang, F. et al. Impact of typhoon events on microplastic distribution in offshore sediments in Leizhou Peninsula of the South China Sea. *Environ. Pollut.* **348**, 123817 (2024).
48. Naidu, G., Zuva, T. & Sibanda, E. M. A review of evaluation metrics in machine learning algorithms. In *Artificial Intelligence Application in Networks and Systems* (eds. Silhavy, R. & Silhavy, P.) 15–25 (Springer, 2023).
49. Vandal, T., Kodra, E. & Ganguly, A. R. Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. *Theor. Appl. Climatol.* **137**, 557–570 (2019).
50. Prokhorenkova, L. et al. CatBoost: unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **31**, 6638–6648 (2018).
51. Simon, S. M., Glaum, P. & Valdovinos, F. S. Interpreting random forest analysis of ecological models to move from prediction to explanation. *Sci. Rep.* **13**, 3881 (2023).
52. Garnier, J. et al. Detangling past and modern zinc anthropogenic source contributions in an urbanized coastal river by combining elemental, isotope and speciation approaches. *J. Hazard. Mater.* **480**, 135714 (2024).
53. Chen, J., Yang, S. T., Li, H. W., Zhang, B. & Lv, J. R. Research on geographical environment unit division based on the method of natural breaks (Jenks). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **40**, 47–50 (2013).
54. Copernicus Marine Service (CMEMS). Global ocean monthly mean sea surface wind and stress. <https://doi.org/10.48670/moi-00185> (2024).
55. Copernicus Marine Service (CMEMS). Global Ocean Hourly Reprocessed Wind and Stress. <https://doi.org/10.48670/moi-00185> (2024).
56. Copernicus Marine Service (CMEMS). Ocean physical model outputs derived from the NEMO v3.6 model. <https://marine.copernicus.eu> (2024).
57. Met Office. Climate data portal: UKC3 and observational sea surface temperature and meteorological data. <https://climatedataportal.metoffice.gov.uk/search?tags=Climate%2CProjections> (2024).

Acknowledgements

The authors are grateful to the National Natural Science Foundation of China (42277403), Shenzhen Science and Technology Program (20231116225539001; JCYJ20220818100217037), Projects of International Cooperation and Exchange of the National Natural Science Foundation of China (NSFC-UNEP: 32261143459), Natural Science Foundation of Guangdong Province (2021B1515020041), Guangdong Provincial Key Laboratory of Soil and Groundwater Pollution Control (No. 2023B1212060002), and High-level University Special Fund (G03050K001) for financial support. Also, we want to thank the Center for Computational Science and Engineering at Southern University of Science and Technology (SUSTech) and the core research facilities at SUSTech for providing quality resources and services.

Author contributions

J.L. was the primary contributor in writing the manuscript, analyzing, and GIS interpreting the microplastic and monitoring data related to nearshore gradient changes and weather conditions. W.S. performed the histological examination of the annual dataset and pretreated the frequent monitoring data, and was the other major contributor to proofreading the manuscript. Upper authors contributed equally to this work. Y.W. was mainly responsible for data management and preprocessing of monitoring data, especially in designing the sampling methods for nearshore samples. Y.C. and Z.W. contributed to the microplastic sample analysis process and polymer characteristic analysis steps. Xiangyun Xiong served as one of the project administrators and was involved in writing and reviewing the manuscript. Xu Xu was the lead sampling administrator and contributed most to the sampling efforts. Y.T. handled project and funding administration, as well as the main review and revision of the manuscript and conceptual framework. All authors read and approved the final version manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44454-026-00028-2>.

Correspondence and requests for materials should be addressed to Yuanyuan Tang.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026