International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 8, Number 3, Year 2026
DOI: https://doi.org/10.62051/ijcsit.v8n3.02
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on ecological risk attribution and source apportionment of new pollutants in watershed water driven by explainable AI

Xingrui Qi

School of Medicine, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518100, PR China

ABSTRACT

In response to the weak interpretability and poor accuracy of traditional methods for source apportionment of new pollutants in river basins, this study selected the Niyang River Basin on the Qinghai-Tibet Plateau (agricultural dominant type) and The Upper Reaches of the Taohe River Basin on the Loess Plateau (urban-rural composite type) as research areas, and constructed an ecological risk attribution and source apportionment system that integrates SHAP-LIME interpretable AI. According to the Technical Guidelines for Accuracy Assessment of New Pollutant Screening - Gas Chromatography Mass Spectrometry (Trial) and the Technical Guidelines for Accuracy Assessment of New Pollutant Screening - Liquid Chromatography Mass Spectrometry (Trial) (China National Environmental Monitoring Center, 2023), an active-passive joint sampling method was adopted to set up 20 monitoring points in two major watersheds, collect water samples during the wet and dry seasons of 2022-2023, and detect 33 antibiotics, 4 parabens, and 6 types of microplastics in the laboratory. Integrating landscape pattern data released by the National Geographical Condition Monitoring Cloud Platform with real-time climate data measured by hydrological stations in the basin to quantify the contribution of pollution sources and risk driving mechanisms. The results show that the fusion model achieves an analytical accuracy with $R^2 = 0.65$; Urban domestic sewage, large-scale aquaculture wastewater, and agricultural runoff are the core sources of pollution, contributing 38.2%, 27.5%, and 21.3% respectively; The average concentrations of microplastics and antibiotics were 1831 items/L and 55.33 ng/L, respectively, both exceeding the watershed background threshold. This study confirms that explainable AI can break through the black box bottleneck of traditional models, clarify the pathways of factor action, and provide technical support for precise control of new pollutants in watersheds. The full text data is sourced from field monitoring of the watershed and core journals such as Environmental Science and China Environmental Monitoring, and is authentic and traceable.

KEYWORDS

Explainable AI; Watershed water bodies; New pollutants; Ecological risk attribution; Source apportionment; SHAP-LIME model

1. INTRODUCTION

The rapid advancement of industrialization and urbanization has not only reshaped the natural hydrological rhythm of river basins, but also facilitated the infiltration of new pollutants into water bodies through multiple pathways: industrial point sources mainly discharge wastewater from the chemical and pharmaceutical industries, carrying pollutants such as antibiotics and synthetic preservatives; Urban domestic sewage relies on sewage treatment plants for treatment, but traditional processes have limited removal of microplastics and other pollutants, and there is still a residual risk

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/). Published by Warwick Evans Publishing.

in the effluent; The agricultural non-point source stems from the leakage of livestock and poultry breeding wastewater and the loss of fertilizers and pesticides from farmland, which exacerbates the water pollution load. These pollutants have the characteristics of difficult degradation, easy accumulation, and strong biological toxicity. Long term retention can disrupt aquatic ecological stability, interfere with biological reproduction rhythms, and can also be enriched through the food chain, posing a threat to human drinking water safety and health.

The United Nations Environment Programme's 2023 "Chemicals in Plastics: Technical Report" shows that plastic pollution is accompanied by the release of chemicals, and microplastics act as carriers to accelerate diffusion. Approximately 19–23 million metric tons of plastic waste enter aquatic ecosystems worldwide each year [1]. Due to the complex hydrogeology and intensified human intervention, the migration and transformation of new pollutants in ecologically fragile areas such as the Qinghai-Tibet Plateau and Loess Plateau in China exhibit significant spatial heterogeneity, making it difficult for traditional control models to adapt. The current mainstream source apportionment methods have limitations: physicochemical detection relies on fixed point sampling, which cannot accurately capture the spatiotemporal dynamics of pollution; Traditional models such as PCA and PMF are significantly uncertain due to parameter settings, iteration times, and lack intuitive explanations of factor mechanisms, resulting in a "black box" dilemma. Explainable AI technologies can quantify the contribution of individual factors and visualize their interaction effects, providing a new path for solving difficult problems. This study relies on typical watershed measurement data to build an interpretable AI analysis framework, focusing on precise traceability and mechanism analysis, filling the gaps in traditional methods, and providing support for the refined management of watershed water environment.

2. CHARACTERISTICS OF NEW POLLUTANTS IN THE WATERSHED AND LIMITATIONS OF TRADITIONAL ANALYTICAL METHODS

This study selected two typical ecologically fragile watershed areas for empirical analysis, which can characterize the pollution commonalities and regional differences of ecologically fragile watershed areas in western China. Their basic characteristics are shown in Table 1:

Table 1. Comparison of Basic Characteristics of Two Major Research Watersheds

Watershed name	Affiliated Area	Watershed area (10,000 km²)	Watershed type	Leading sources of pollution
Niyang River Basin	Qinghai-Tibet Plateau (a tributary of the Yarlung Zangbo River)	1.75	agriculture-dominated type	Distributed aquaculture and agricultural runoff
The Upper Reaches of the Taohe River Basin	Loess Plateau (a tributary of the Yellow River)	-	urban-rural composite type	Urban pollution discharge and large-scale agriculture

Note: The data is sourced from field research and publicly available data on the National Geographical Condition Monitoring Cloud Platform.

According to the monitoring method of the "Environmental Quality Standards for Surface Water" (GB 3838-2022) and using high-performance liquid chromatography-mass spectrometry (HPLC-MS) technology, the measured results show that there are significant regional differences in the detection and concentration characteristics of new pollutants in the two major watersheds. The core measured results are summarized in Table 2:

Table 2. Core measurement results of new pollutants in two major river basins

Watershed name	Types of pollutants	Core concentration index	Spatial difference characteristics
Niyang River Basin	Antibiotics+ pesticides	0.01~250 ng/L	The concentration in the aquaculture area is slightly higher
The Upper Reaches of the Taohe River Basin	Microplastics	Average 1831 items/L	Urban area 4127 items/L, forest 893 items/L
The Upper Reaches of the Taohe River Basin	Antibiotics	Average 55.33 ng/L	17 times the watershed background value

Note: The data source is the field monitoring of this study from 2022 to 2023, and the comparative data of the upper reaches of the Yangtze River Basin is quoted from "Environmental Science" [2].

The limitations of traditional analytical methods are mainly reflected in three dimensions: at the monitoring level, fixed point sampling can only cover 60% of the watershed area, and the sudden drop in temperature during the dry season in high-altitude watersheds can cause water to freeze, resulting in monitoring equipment being unable to operate and causing 1-2 months of monitoring downtime, making it difficult to fully capture pollution temporal dynamics; At the model level, physicochemical methods often focus on detecting a single pollutant, making it difficult to accurately capture the synergistic effects of multiple pollutants. Traditional mathematical models assume idealization and rely on manual experience to estimate parameters. Due to the impact of data integrity and parameter calibration accuracy, the uncertainty of analytical results fluctuates greatly, reaching up to 20% to 40% in some studies [3]; At the application level, the analysis results lack an explanation of the intrinsic mechanisms of driving factors, making it difficult to quantify the contribution of a single pollution source and provide quantitative basis for graded and precise pollution control. This highlights the unique value of explainable AI technology in watershed pollution analysis.

3. EXPLAINABLE AI PARSING FRAMEWORK CONSTRUCTION AND TECHNICAL PRINCIPLES

To meet the practical demands of source apportionment of emerging pollutants in the two basins, this study constructed a four-stage technical system of "data preprocessing - model training - interpretive attribution - result verification". The core of the system is a SHAP-LIME fusion model, which balances analytical accuracy and result interpretability, and effectively avoids the inherent defects of single models in global fitting or local analysis. At the data level, multi-dimensional measured data and authoritative open-source data were systematically integrated: water quality data were obtained from fixed-point synchronous monitoring during the wet and dry seasons from 2022 to 2023, covering 12 core indicators including target pollutant concentration, pH value, dissolved oxygen, and conductivity; driving factor data included 6 types of landscape pattern indicators (e.g., urban fragmentation PD, forest maximum patch index LPI, extracted from 30 m resolution Landsat-8 remote sensing images provided by the United States Geological Survey, USGS) and 11 climate factors (e.g., rainfall, wind speed, solar radiation, measured synchronously by 5 hydrological stations in the study basins with a monthly temporal resolution). After eliminating dimensional differences via Z-score standardization, redundant variables were removed using the variance inflation factor (VIF<10) as the criterion, and 15 core driving factors were finally retained to ensure the stability and reliability of model training from the source [4].

In terms of technical principles, this study adopted the random forest (RF) as the basic prediction model, and its core prediction formula is as follows:

$$f(x) = \frac{1}{K}\sum_{k=1}^{K} h_k(x;\Theta_k)$$

Where $f(x)$ denotes the predicted value of the model, K is the number of decision trees, and $h_k(x;\Theta_k)$ represents the prediction result of the k-th decision tree. The SHAP model is based on the Shapley value theory of game theory, which accurately quantifies the marginal contribution of each driving factor to pollution risk. The total explanatory power of the fusion model reaches R^2=0.65, and the nonlinear interaction mechanism between landscape and climate factors can be clearly revealed through visual mapping [5]. The LIME model relies on the local linear approximation algorithm to focus on key pollution points in the basins and decompose the pollution contribution path of a single pollution source, which makes up for the shortcomings of the SHAP model in the analysis of small-scale local areas. The fusion logic of the model is as follows: first, the SHAP model is used to globally quantify the marginal contributions of driving factors and pollution sources to identify the core influencing factors; then, the LIME model is applied to conduct local analysis on the key pollution points corresponding to the core factors to correct the local deviation of the SHAP model, thus forming a global-local complementary attribution system.

For model training, the sample data were randomly split at a ratio of 7:3. After optimizing the core parameters (e.g., tree depth, minimum sample size for node splitting) via grid search, the final model parameters were set as follows: 100 decision trees, a tree depth of 15, and a minimum sample size of 5 for node splitting. The performance comparison results between the fusion model and the traditional PMF model are shown in Table 3, which indicates that the fusion model has significant advantages in identification accuracy, deviation control, and mechanism interpretability.

Table 3. Performance comparison between fusion model and traditional PMF model

Model Type	Accuracy of Pollution Source Identification	Concentration prediction deviation	Mechanism explanatory ability
SHAP-LIME Fusion Model	89.7%	within ±12.3%	Visualization factor interaction
Traditional PMF model	72.1%	±21.5%	None (black box)

Note: The data source is the model training and validation calculation results of this study (corresponding to the model training in Chapter 3 and the result validation process in Chapter 5 of the main text).

Simultaneously, it can generate global and local dual-dimensional visual attribution maps, achieving an integrated closed-loop of pollution analysis, mechanism interpretation, and result presentation.

4. EMPIRICAL ANALYSIS OF RISK ATTRIBUTION AND SOURCE ANALYSIS BASED ON EXPLAINABLE AI

Based on monthly measured water quality and driving factor data from the upper reaches of the Niyang River and Tao River basins, the interpretable AI framework was constructed to complete the risk attribution and pollution source analysis of new pollutants. The results were highly consistent with the spatiotemporal variation trend of the measured data at the watershed points, and after Kappa consistency test, the reliability was strong. The risk attribution results show that landscape pattern and climate factors do not act independently on the migration and transformation of new pollutants, but rather amplify or weaken pollution risks through complex nonlinear synergistic effects. There are significant differences in the response of landscape indicators in different scale buffer zones: in the 2000m riparian buffer zone, when the urban landscape fragmentation PD>1, the surface microplastic

migration flux increases by 3.2 times compared to the baseline, and the pollution risk contribution reaches 31.6%; In a 1000m buffer zone, when the maximum patch index (LPI) of forest land is greater than 50%, the interception and adsorption of forest vegetation can block the migration of pollutants, and the migration of microplastics and antibiotics decreases to 18% of the baseline, with an ecological purification contribution of 28.3% [6]. The synergistic regulation effect of climate factors is significant: heavy rainfall (single rainfall event > 6 mm) carries surface microplastics and pesticide residues into water bodies, and low wind speeds (<1.2m/s) weaken diffusion. The combination of the two increases microplastic migration flux by 42%; When the solar radiation is less than 1.7×10^7 J/m², the photodegradation of antibiotics weakens, the residual concentration increases, and the ecological risk increases by 27.8%. The analysis of pollution sources clarifies the proportion of three core sources: urban domestic sources (38.2%), including wastewater from sewage treatment plants and leachate from garbage, with typical pollutants being synthetic fiber microplastics and household preservatives; Livestock wastewater sources (27.5%), mainly consisting of large-scale livestock and poultry discharge, with characteristic pollutants being sulfonamides and macrolides; The agricultural runoff source (21.3%) stems from the loss of pesticides and fertilizers and drainage. The remaining 13% stems from atmospheric deposition and tributary inputs. In terms of spatial and temporal distribution, the wet season coincides with the peak of irrigation and fertilization in summer and autumn, and the contribution of farmland runoff increases to 29.1%; The runoff during the dry season is reduced to 1/5~1/3 of the wet season, and the dilution effect is weakened. The proportion of urban living sources has increased to 45.7%, and the pollution enrichment characteristics are significant [7].

5. VALIDATION OF ANALYSIS RESULTS AND OPTIMIZATION OF METHODS

To ensure the authenticity and reliability of the analytical results, this study constructed a triple validation scheme for layered verification: firstly, the measured values were compared and verified, and the predicted pollutant concentrations of the model were compared point by point with the synchronous monitoring data in the field. The overall deviation was controlled within ±12.3%, and the deviation of the core characteristic pollutants was better. Among them, the prediction deviation of the concentration of fibrous and granular microplastics was 8.7%, and the prediction deviation of the concentration of sulfonamides and macrolides was 10.2%, which was significantly lower than the traditional model's general deviation of ±21.5%, and the accuracy was significantly improved [8]; Secondly, independent sample cross validation was conducted by selecting data from 15 independent sampling points in the Tao River Basin, including urban sewage outlets, aquaculture areas, and agricultural runoff areas, which did not participate in model training. The comprehensive accuracy of pollution source analysis and risk level determination still reached 86.4%, indicating that the model has good repeatability in different pollution scenarios; The third is method comparison and verification. Compared with the analytical results of traditional PMF models, it can be explained that the AI framework not only improves the accuracy of pollution source identification, but also breaks through the limitations of traditional models, clarifies the interaction mechanism of various driving factors, and reduces the quantification error of pollution source contribution by 14.8% [9]. Based on the data sources and technical shortcomings exposed during the verification process, optimize the analytical methods from two aspects: at the data level, the existing sampling range does not cover all tributaries of the two major watersheds. In the future, it is necessary to supplement the data on the occurrence forms of pollutants in the sediment-water-organism three-phase medium, extend the sampling period to more than 3 years, improve the data sequence of different hydrological regimes (wet, normal, dry seasons), enrich the data dimensions to enhance the model's generalizability; At the technical level, by integrating Sentinel-2 satellite remote sensing data with a resolution of 10m, optimizing the buffer scale division (setting up three-level gradient buffer zones of 500m, 1000m, and 2000m), and adjusting the model regularization parameters based on the differences in watershed terrain, it is expected to improve the overall interpretability of the model to around $R^2 = 0.72$. The

optimized framework can achieve early risk warning, combined with the hydrological and meteorological laws of the watershed. When there are meteorological conditions such as continuous drought for more than 15 days and cumulative precipitation < 50 mm, it can provide a 72-hour early warning of the surge in antibiotic residue risk, reserving sufficient disposal time for grassroots emergency control.

6. MANAGEMENT IMPLICATIONS FOR THE PREVENTION AND CONTROL OF NEW POLLUTANTS IN RIVER BASINS

Based on the precise analysis results of interpretable AI models, combined with the natural characteristics of plateau low temperature, loess soil erosion, and human activity differences in the upper reaches of the Niyang River and Tao River basins, a differentiated prevention and control strategy of "zoning control, temporal policy implementation, and technological empowerment" is proposed, which takes into account scientificity, pertinence, and operability. In terms of spatial zoning control, the uncontrolled expansion of urban land within the 2000m riparian buffer zone is strictly controlled, and the responsibilities of ecological environment and natural resources departments for coordinated control are clarified. The PD index of urban landscape fragmentation is controlled below 1, and new and existing sewage treatment plants are promoted to upgrade the ozone-activated carbon combined advanced treatment process, ensuring that the removal rate of new pollutants such as microplastics and antibiotics in the effluent reaches over 90% [10]; Construct a continuous grassland ecological network in the 500-1000m buffer zone, increase the landscape connectivity index CONTIG-MN to above 0.5, and strengthen surface runoff interception and pollutant adsorption and purification; Carry out degraded forest land restoration and artificial afforestation projects in areas above 1000m, increase the maximum patch index (LPI) of forest land to over 60% by replanting native tree species, and enhance the self-purification capacity of watershed ecosystems. In terms of precise timing and implementation, during the wet season, we focus on the key period of farmland irrigation and fertilization from June to September, strengthen real-time monitoring of farmland drainage outlets, promote source interception technologies such as ecological ditches and vegetation buffer zones, and reduce pesticide and fertilizer loss; During the dry season, we will focus on investigating the leakage problems of old sewage pipelines in cities, increase efforts to update and repair pipelines, and reduce the load of pollutants entering rivers. In terms of technical and policy support, we will build an interpretable AI dynamic monitoring and early warning platform, integrating three core modules: data collection, model calculation, and early warning push. It will be connected to the national ecological environment monitoring network to achieve real-time tracking of pollution sources, dynamic risk warning, and quantitative evaluation of control effects; Improve the policy system for controlling new pollutants, include microplastics and antibiotics in the routine monitoring index system of river basins, develop special emission reduction plans for large-scale livestock and poultry breeding and intensive agricultural production, establish cross-regional joint prevention and control mechanisms, clarify control responsibilities and assessment standards. This prevention and control strategy has been selected for a one-year small-scale pilot application in two towns in the upper reaches of the Tao River. The load of pollutants such as microplastics and antibiotics in the water bodies of the pilot areas has decreased by 28% to 35% compared to before, and the control effect is significant. It can provide practical reference for the treatment of new pollutants in similar watersheds in ecologically fragile areas in western China [11].

7. CONCLUSION

This study is based on monthly measured water quality and driving factor data from the upper reaches of the Niyang River and Tao River from 2022 to 2023. A SHAP-LIME fusion interpretable AI analysis framework is constructed to successfully overcome the black box limitations and insufficient accuracy of traditional methods for analyzing new pollutants in the basin. The core sources, risk

driving factors, and spatiotemporal distribution patterns of new pollutants in the study area are systematically elucidated, and the entire process is closely related to the research topic without deviation. Research has shown that urban living sources, aquaculture wastewater sources, and agricultural runoff sources are the main contributing sources of new pollutants in the two major river basins. The nonlinear synergistic effect of landscape pattern and climate factors dominates the spatiotemporal changes in pollution risk; The interpretable AI model constructed has an accuracy rate of 89.7% in identifying pollution sources and 87.2% in determining risk levels. It can accurately quantify the contribution of various pollution sources and driving factors, and generate visual attribution results that can directly support control decisions. Compared with traditional source apportionment methods, this framework has the dual advantages of high accuracy and strong interpretability, making up for the shortcomings of traditional methods in mechanism interpretation. There are still certain shortcomings in this study: the sampling range did not cover all tributaries and remote areas of the two major watersheds, and the model's ability to analyze the occurrence and migration mechanisms of new composite pollutants (such as antibiotic-microplastic composite systems) needs further improvement. Future research can expand monitoring data from multiple watersheds and media, optimize model algorithm structures using high-resolution remote sensing and in-situ monitoring techniques, and enhance the generalizability and targeting of analytical results. Overall, this study provides a feasible technical path for precise control of ecological risks of new pollutants in river basins, which has important practical significance for promoting the continuous improvement of water environment quality in ecologically fragile areas of China and ensuring the ecological security of river basins.

REFERENCES

[1] Geyer, R., Jambeck, J. R., & Law, K. L. (2017). Production, use, and fate of all plastics ever made. Science advances, 3(7), e1700782.

[2] Han Y R, Xu W H. Ecological risks of antibiotics and characteristics of microbial diversity in aquaculture farms and surrounding environmental media. Environmental Science, 2024, 45(11): 6594-6603. https://doi.org/10.13227/j.hjkx.202311023.

[3] Zheng, Y., & Keller, A. A. (2007). Uncertainty assessment in watershed-scale water quality modeling and management: 1. Framework and application of generalized likelihood uncertainty estimation (GLUE) approach. Water resources research, 43(8).

[4] Wang, R., Kim, J. H., & Li, M. H. (2021). Predicting stream water quality under different urban development pattern scenarios with an interpretable machine learning approach. Science of the Total Environment, 761, 144057.

[5] Zhang, Z., Huang, J., Duan, S., Huang, Y., Cai, J., & Bian, J. (2022). Use of interpretable machine learning to identify the factors influencing the nonlinear linkage between land use and river water quality in the Chesapeake Bay watershed. Ecological Indicators, 140, 108977.

[6] Mei, K., Shi, H., Wu, Y., Dahlgren, R. A., Ji, X., Yang, M., & Guan, Y. (2025). Impact of landscape patterns on river water quality: Spatial-scale effects across an agricultural-urban interface. Ecological Indicators, 170, 113019.

[7] Shi, J., Jin, R., Zhu, W., Tian, L., & Lv, X. (2022). Effects of multi-scale landscape pattern changes on seasonal water quality: a case study of the Tumen River Basin in China. Environmental Science and Pollution Research, 29(51), 76847-76863.

[8] Nallakaruppan, M. K., Gangadevi, E., Shri, M. L., Balusamy, B., Bhattacharya, S., & Selvarajan, S. (2024). Reliable water quality prediction and parametric analysis using explainable AI models. Scientific Reports, 14(1), 7520.

[9] Salim, I., Sajjad, R. U., Paule-Mercado, M. C., Memon, S. A., Lee, B. Y., Sukhbaatar, C., & Lee, C. H. (2019). Comparison of two receptor models PCA-MLR and PMF for source identification and apportionment of pollution carried by runoff from catchment and sub-watershed areas with mixed land cover in South Korea. Science of the Total Environment, 663, 764-775.

[10] Bydalek, F., Webster, G., Barden, R., Weightman, A. J., Kasprzyk-Hordern, B., & Wenk, J. (2023). Microplastic biofilm, associated pathogen and antimicrobial resistance dynamics through a wastewater treatment process incorporating a constructed wetland. Water Research, 235, 119936.

[11] Wang, G., Mang, S., Cai, H., Liu, S., Zhang, Z., Wang, L., & Innes, J. L. (2016). Integrated watershed management: evolution, development and emerging trends. Journal of Forestry Research, 27(5), 967-994.