Journal of Hazardous Materials Advances 18 (2025) 100708

Contents lists available at ScienceDirect

Journal of Hazardous Materials Advances

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

Machine learning predictive modelling for sediment risk indices within an urbanized river channel

María Alejandra Pimiento [a,*], Jose Anta [a], Andres Torres [b]

[a] Universidade da Coruña, Water and Environmental Engineering Group, Center for Technological Innovation in Construction and Civil Engineering (CITEEC), A Coruña, Spain

[b] Instituto Javeriano del Agua, Pontificia Universidad Javeriana, Bogotá, Colombia

ARTICLE INFO

Keywords:
Sediment pollution
Risk index
SVM
Enrichment factor
Rainfall

ABSTRACT

Despite the growing application of machine learning (ML) in water quality assessment and pollution source identification, its potential for predicting environmental risk indices in urban stormwater sediments remains largely unexplored. Conventional models struggle to capture complex interactions among hydrological variables, sediments and pollution parameters. This study uses ML techniques to enhance sediment quality assessment to address this gap. The case study focuses on sediments from the Molinos River in Bogotá, Colombia, characterized by particle size distribution (PSD), heavy metal (HM) concentrations, and environmental risk indices. Cohen's Kappa coefficient was used to evaluate the relationship between the enrichment factor (EF) of Ni and Pb, PSD, and hydrological variables as rainfall data. A support vector machine model using an ANOVA kernel, validated through multiple calibration and validation datasets, demonstrated the feasibility of predicting sediment-related risks in urban drainage systems. The best model successfully predicted Pb EF levels for 7 of 8 samples, achieving a Cohen's Kappa coefficient of 0.71 ($p = 0.037$), indicating substantial agreement. These findings highlight the potential of ML models to predict sediment EF using rainfall data, providing a practical tool for environmental risk assessment. By enabling predictions of contamination levels, this methodology enhances decision-making and promotes more sustainable urban water management strategies.

1. Introduction

Urbanization often disrupts the natural flow of rivers, transforming them into conduits for water from diverse sources, thereby degrading water quality compared to pristine environments (Guimarães et al., 2021). A major contributor to this decline is stormwater runoff, which transports particles and pollutants that deteriorate river ecosystems (Rietveld et al., 2022). Since stormwater is a valuable natural resource, its conservation in urban environments is critical for mitigating pollution (Rentachintala et al., 2022). Continuous monitoring and prediction of sediment quality in urban stormwater drainage systems are essential for assessing pollution trends in waterways, improving environmental risk assessment, and informing management strategies (Nawrot et al., 2021; Pimiento et al., 2023).

Assessing pollutant concentrations in urban stormwater sediments is crucial for understanding their toxicity and potential impact on water quality. Several factors influence sediment quality such as pH, organic matter content, and particle size, which regulate pollutant adsorption and precipitation (Rumuri et al., 2023; Skorbiłowicz et al., 2024; Wang et al., 2024). Land use also plays a critical role, as urbanization increases impervious areas, population density, traffic loads, domestic waste generation, and construction and industrial activities, all of which contribute to pollution (Skorbiłowicz et al., 2024; Yang et al., 2022). Additionally, hydrological parameters such as precipitation, dry weather duration, runoff, and flow discharge significantly influence the transport, erosion, and leaching processes of contaminated sediments, shaping their distribution in receiving water bodies (Fatmi et al., 2024; Nie et al., 2024; Pimiento et al., 2018; Sanabria-Morales et al., 2021). Heavy metals (HMs) in sediments originate from both natural and anthropogenic sources. Natural contributors include geological weathering, lithology, and climatic conditions that enhance erosion (Skorbiłowicz et al., 2024), while anthropogenic sources stem from activities such as agriculture, traffic, industrial processes, and urban waste (Fatmi et al., 2024; Nie et al., 2024; Wang et al., 2024). Consequently, various pollution indices have been developed to assess the HM contamination, emphasizing the need for regular sediment monitoring

* Corresponding author.
E-mail address: m.a.pimiento@udc.es (M.A. Pimiento).

https://doi.org/10.1016/j.hazadv.2025.100708
Received 10 September 2024; Received in revised form 21 March 2025; Accepted 31 March 2025
Available online 4 April 2025

to evaluate potential environment and human health risks (Ahmed et al., 2024; Fatmi et al., 2024; Pimiento et al., 2023).

Machine learning (ML) and statistical models have become essential tools for predicting environmental variables using readily available data. These methods effectively handle complex, nonlinear relationships, making them valuable for applications such as flood risk assessment, HM contamination prediction, and agricultural risk forecasting (Ma et al., 2022; Yin et al., 2022; Zhao et al., 2023). Compared to traditional statistical approaches, ML models enhance predictive accuracy, and reliability, and can integrate large datasets, providing deeper insights into environmental processes. When combined with multivariate statistical techniques, ML enables a more comprehensive analysis of pollution sources and key influencing factors (Fatmi et al., 2024; Bi et al., 2024).

Beyond environmental monitoring, ML is increasingly applied in urban water management to improve decision-making and system efficiency. These models assist in predicting flooding events, optimizing drainage networks, and enhancing maintenance strategies while reducing reliance on expensive simulations and software (Boughandjioua et al., 2023). A key advantage of ML is its ability to continuously learn from new data, enabling adaptive and real-time solutions. By automating complex analyses, ML supports proactive urban infrastructure management, offering scalable and cost-effective strategies for long-term sustainability (Kwon and Kim, 2021).

To the best of the authors' knowledge, ML has not been implemented to predict environmental pollution risk indices in urban river sediments. While previous studies have utilized ML for water quality assessment and pollutant source identification, none have specially addressed its potential for evaluating heavy metal contamination risk in stormwater sediments. Given the limitations of conventional models in capturing complex interactions among hydrological, granulometric, and contamination parameters, ML presents a promising alternative for improving sediment quality assessment. Developing such models could significantly enhance our ability to monitor contamination patterns, facilitate decision-making in sediment management, and improve urban drainage infrastructure strategies.

This study aims to develop a methodology for predicting heavy metal environmental risk indices in urban river sediments influenced by stormwater runoff using ML. To achieve this, the research will identify relationships among hydrological, granulometric, and environmental risk index variables in sediments retained in stormwater drainage systems. An ML model will then be developed to predict heavy metal risk levels, followed by validation using continuously measured variables, such as hydrological parameters, to assess the model's effectiveness for urban stormwater management. Ultimately, this approach seeks to provide a decision-support tool for optimizing sediment management in urban drainage systems, contributing to more efficient and sustainable urban water management.

2. Materials and methods

2.1. Study site

The study site (Fig. 1) was the Molinos River (MR), a canalized river within urban stormwater system of Bogotá, Colombia, characterized by pluvial contributions and the accumulation of runoff sediments. The channelized system consists of an open concrete channel. To determine the hydrological and geomorphological characteristics of the MR urban catchment, a Copernicus DEM with a 30 m resolution was used. The basin spans an area of 927.76 ha, with a main channel slope of 11.98 % and a length of 4.22 km. The land use is predominantly residential, meaning that the sediment source are likely a combination of natural materials and urban runoff from impermeable surfaces.

Sediment grab samples were collected at the river sampling point using handmade sediment traps, which were installed on the riverbed and retrieved approximately every fifteen days. This approach ensured replicability and sufficient particle accumulation. The sampling campaigns were conducted between February 2020 and November 2021, resulting in a total of 24 samples.

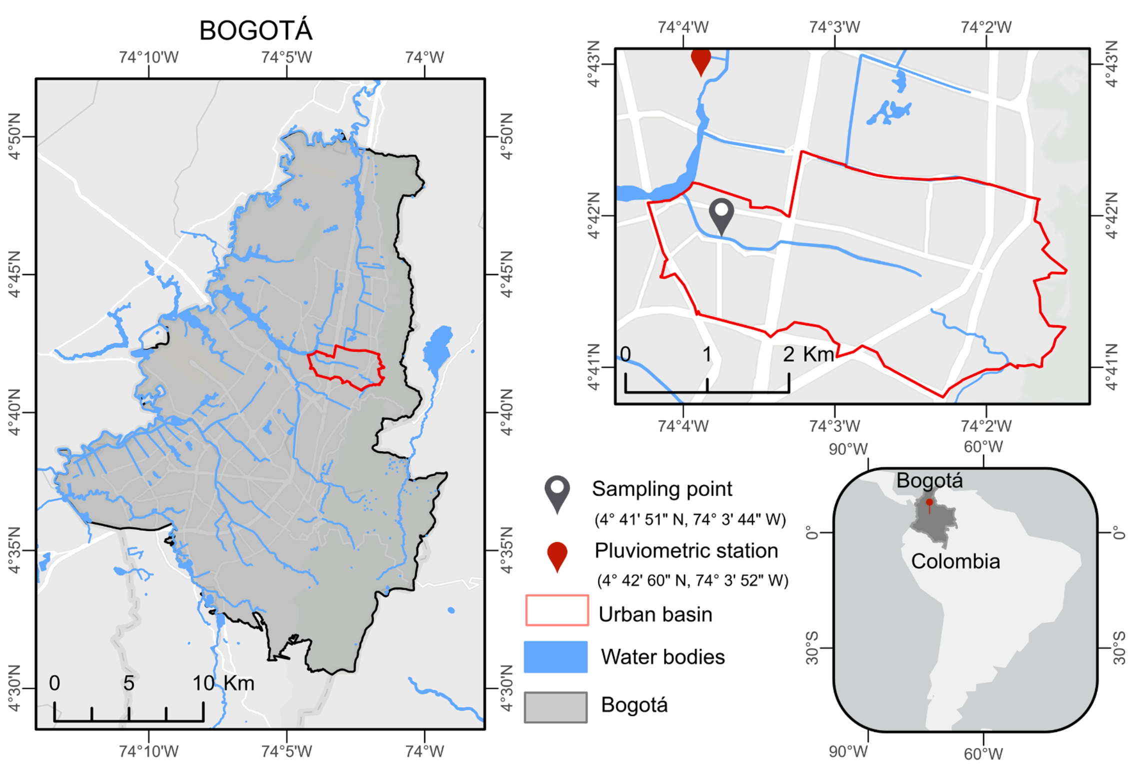

Fig. 1. Study site location: Colombia, Bogotá and the Molinos River Basin (highlighted in red). Locations of the pluviometric station Gustavo Morales and sampling point.

2.2. Rainfall characterization

Rainfall data were obtained from the early warning system by the District Risk Management Institute of Bogotá (IDIGER, for its name in Spanish). The Gustavo Morales pluviometric station, located at latitude 4.716619, and longitude −74.064401 (Fig. 1) provided the data used in this study. The following daily rainfall variables were analyzed: total precipitation height (H_mm), average rainfall intensity (I_{ave}_mm/h), number of consecutive dry days (dw_days), number of days with measurable rainfall (rw_days), and antecedent dry weather period (ADP_days). Fig. 2 shows an example of rainfall variable distribution over time and the sampling period, highlighting the installation and removal of the sediment traps.

2.3. Sediments characterization

The collected sediments samples were transported to the Water Quality Laboratory at Pontificia Universidad Javeriana. A wet portion of each sample was used for PSD analysis, while the remaining portion was stored in a refrigerator for HM total concentration analysis.

PSD analysis was performed using laser diffraction Mastersizer 3000E Particle Size Analyzer within a measurement range of 0.1 µm to 1000 µm (Ltd, 2024). The PSD was determined based on the scattering pattern produced when a laser beam passed through the sample. To ensure accuracy, each sediment sample was homogenizing, and two samples were taken and analyzed in the Coulter device. The analysis was repeated five times, and from the mean PSD, the following characteristic diameters of the sample were determined: D10, D20, D30, D40, D50, D60, D70, D80, D90, and D100.

Based on the literature review, the most common HMs in urban environments were selected for analysis: Cadmium (Cd), Lead (Pb), Cooper (Cu), Zinc (Zn), Chromium (Cr), and Nikel (Ni). From the remaining sediment portion, 0.5 g of material was weighed and digested following EPA Method 3051A method (EPA., 2007). The digestion process involved adding 10.0 ± 0.1 mL of concentrated nitric acid (HNO3) and 3.0 ± 0.1 mL of concentrated hydrochloric acid (HCl). Total HM were measured using a Shimadzu ICP-E 9800 spectrometer, which enables the simultaneous analysis of multiple elements using a multiparametric standard of 19 elements. The instrument was calibrated using 12 standard concentrations ranging from 0.001 mg/L to 10 mg/L. To ensure measurement reliability, all analyses were performed in triplicate.

Several risk indices have been used in the literature to evaluate the impact of HMs in sediments. In Pimiento et al. (2023), various indices were analyzed, including the geo-accumulation index, pollution index, modified pollution index, potential ecological risk index, modified potential ecological risk, pollution load, and enrichment factor were determined for the MR catchment. Since some of these indices require a background concentration (B_n), which has not been reported for the study site, soil B_n concentrations from Rauch & Pacyna (2009) were used as reference. The enrichment factor (EF) was selected for this study because it exhibited the greatest variability across risk classes (Table 1). The EF is commonly used to determine whether contaminants adhered to sediments originate from natural or anthropogenic sources (Gopal et al., 2023; Rumuri et al., 2023).

2.4. Statistical analysis

Statistical analyses were performed to identify relationships between PSD, rainfall variables, and EF. The statistical test were conducted using RStudio open-source software (Team, 2023). The variables were classified into two groups: dependent variables (Y), which included risk indices, and independent variables (X), which included PSD characteristic diameters and rainfall variables defined in Section 2.2.

As EF is a categorical variable, Cohen’s kappa coefficient was employed to assess agreements between two categorical variables (Cohen, 1960). The kappa.test fucntion from the fmsb library (Nakazawa, 2021). was used to perform the test. Additionally, Pearson and Spearman correlation tests were conducted to numerically validate relationships between parametric and non-parametric variables. These tests were executed using the chart.correlation function from the PerformanceAnalytics package (Peterson and Carl, 2020). Additionally,

Table 1
Enrichment factor (EF) classification (adapted from Choi et al., 2020).

Def	deficiency to minimal enrichment	$EF \le 1$
Mod	moderate enrichment	$1 < EF \le 5$
Sig	significant enrichment	$5 < EF \le 20$
High	very high enrichment	$20 < EF \le 40$
Ext	extreme enrichment	$EF > 40$

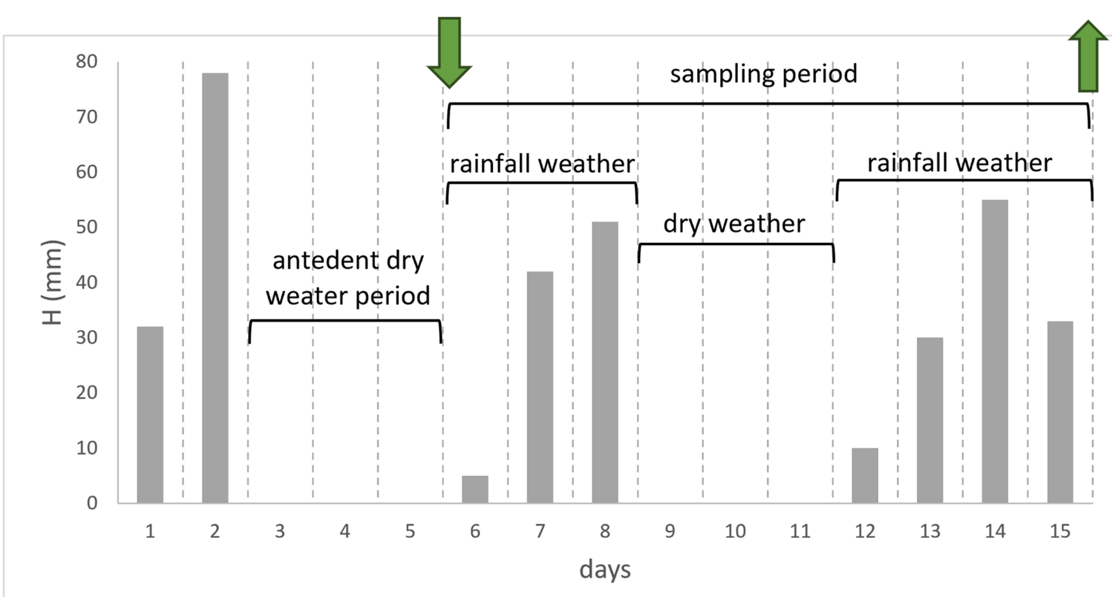

Fig. 2. Example of rainfall variables distribution over time. H: daily precipitation height in mm. Green arrows indicate the installation and removal of the sediment traps (Pimiento, 2023).

support vector machine (SVM) was used as a predictive tool (Galarza-Molina, 2017). The ksmv command from the kernlab library was used (Karatzoglou et al., 2004), which includes the most popular kernel functions for data training and prediction.

2.5. Modelling methodology

Fig. 3 illustrates the proposed methodology for modelling EF risk indices, based on statistical tests that evaluate the relationships between variables. To predict each risk index (Variable Y), it was necessary to identify the corresponding related variables (X variables). The Cohen's Kappa coefficient was used to establish relationships between risk index categories with the PSD and rainfall, which were converted into categories. Variable pairs with a p-value < 0.05 were selected. Since risk index classification is based on a specific threshold, a correlation test was performed to numerically validate the relationships between variables pairs.

A Monte Carlo method was applied to ensure model robustness by randomly selecting thousand calibration and validation datasets. Two-thirds of the observations were randomly selected as calibration data, while the remaining one-third was used for validation. Thus, the model will be evaluated for each validation and calibration data combination, generating a thousand prediction models.

To evaluate model predictability, Cohen's Kappa coefficient was used to compare the categories predicted by the models with the actual measured categories for the validation dataset. All available kernels in the ksvm function were tested to assess their impact on model performance. To select the most effective kernel, models were evaluated based on Cohen's Kappa test results, and those with a p-value <0.05 during the calibration phase were chosen.

Model validation was performed using the designated third of the dataset, incorporating corresponding rainfall variables, which are easily obtainable without direct sampling. Assuming that only rainfall data is available for validation, the characteristic diameters from the PSD data of the first calibration observation were used to supplement all validation observations. This approach ensures method replicability without requiring additional sediment field data collection, thereby reducing analysis costs and improving decision-making in sediment management. The final risk prediction model was selected by assessing the agreement between predicted and observed classes using Cohen's Kappa coefficient and identifying the model with the lowest p-value (<0.05). The model with the highest agreement was then considered the most effective.

3. Results and discussion

3.1. Sediment and rainfall characterization

The PSD of sediments from the MR catchment sediments indicates a median diameter (D50) of 117 μm, with a range of 15 – 307 μm. Approximately 40 % of these sediments are finer than 63 μm, while the remaining 60 % fall within 63 – 1000 μm range. According to the United States Department of Agriculture (USDA) soil classification (García-Gaines and Frankenstein, 2015), these sediments are primarily classified as sand. These findings align with research conducted on urban stormwater in Calgary, which showed that over 75 % of particles are finer than 250 μm, and 43 % are finer than 63 μm, highlighting the prevalence of fine particles (Yan et al., 2024). Similarly, road runoff in Ohio demonstrated a mean PSD comprising 48.7 % sand, 47.4 % silt, and 3.9 % clay, with median D50 values ranging from 24 μm to 72 μm (Winston et al., 2023). These studies collectively suggest that fine particles are a common feature of urban stormwater runoff, indicating high potential for contaminant accumulation due to their large specific surface area.

Analysis of HM in MR sediments revealed undetectable

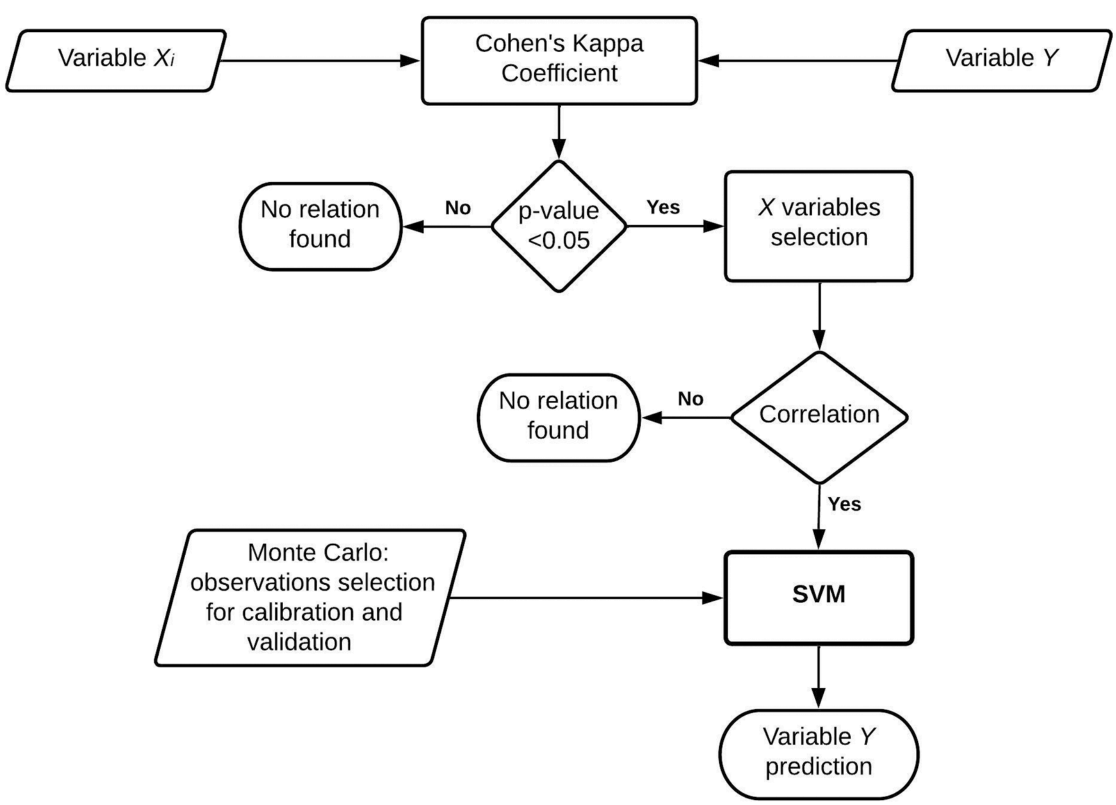

Fig. 3. Modeling methodology. Variable Xi: D1 to D100, precipitation height, rain intensity, antecedent dry weather period, number of dray days. Variable Y: risk indices.

concentrations for Cd and Cr. The median concentrations of Pb, Ni, Zn, and Cu were found to be 8.98, 17.4, 100, and 570.47 mg/kg, respectively. In comparison, the Saïda urban catchment area reported Cd levels ranging from 0.63 to 2.48 mg/kg, Cu levels from 68.2 to 192.95 mg/kg, Cr levels from 65.05 to 167.15 mg/kg, Ni levels from 37.95 to 83.85 mg/kg, Pb levels from 145.75 to 248.25 mg/kg, and Zn levels averaged 178.08 mg/kg (Fatmi et al., 2024). In contrast, Bug River sediments exhibited lower concentrations of Ni, Pb, Cu, and Cr, with average values of 5.31, 5.04, 2.25, and 1.54 mg/kg, respectively (Skorbiłowicz et al., 2024). These findings highlight significant disparities in HM concentrations across various catchment areas, emphasizing the need for localized sediment management strategies to address specific contamination levels, sources, and risk assessment.

As presented in Pimiento et al. (2023), the enrichment factor for each HM was calculated with the measured concentrations, resulting in varied EF classes. Cu and Zn predominantly exhibit extreme enrichment, Ni demonstrates very high enrichment, while Pb indicates significant enrichment. Table 2 illustrates that the EF levels for the Molinos River are considerably higher than those observed in sediments from other urban environments, demonstrating the anthropogenic origins of these HMs. Consequently, the EF is a valuable input for developing models to predict the risk level and inform sediment management alternatives.

Rainfall data were analyzed to derive precipitation variables for each sampling period, as shown in Table 3. No extreme precipitation events or predominant periods of rain or drought were identified at the study area, likely due to its geographic characteristics and the influence of La Niña events during the study period. Consequently, an analysis under distinct hydrometeorological conditions (i.e., dry and wet seasons) was not feasible. However, climatic conditions do not adversely affect the model's performance, as the proposed methodology is adaptable for separating dry and rainy seasons based on the local conditions of the study site.

Table 2
Enrichment Factor in urban sediments. RDS: road deposited sediments, Mod: moderate enrichment, Sig: significant enrichment, High: very high enrichment, Ext: extreme enrichment.

Cu	Zn	Ni	Pb	Sample location	Source
Ext	Ext	High	Sig	Molinos River	Pimiento et al. (2023)
Mod	Mod	-	Mod	RDS	Wang et al. (2019); Zhang et al. (2020)
High	High	Mod	Sig	Stormdrain	Choi et al. (2020)
Sig	Sig	Sig	Sig	Stormwater sedimentation tank	Sałata and Dąbek (2017)

Table 3
Rainfall variables. ADP: antecedent dry weather period in days, H: precipitation height (mm), I_{ave}: average intensity (mm/h).

Sample	ADP	H (mm)	I_{ave} (mm/h)	Sample	ADP	H (mm)	I_{ave} (mm/h)
1	6	19.3	0.47	13	0	117.2	0.33
2	3	84.3	0.23	14	4	17.7	0.05
3	3	11.8	0.03	15	2	27.5	0.08
4	4	30.7	0.09	16	1	158.4	0.44
5	2	46.5	0.13	17	1	6.8	0.02
6	5	24.9	0.07	18	12	16.4	0.04
7	5	75.1	0.21	19	0	14.4	0.05
8	0	2.2	0.01	20	8	59.8	0.17
9	4	3.5	0.01	21	0	77	0.21
10	0	155.93	0.43	22	0	29.9	0.08
11	0	127.65	0.35	23	3	51.5	0.14
12	9	13.4	0.04	24	3	108.4	0.30

Table 4
Molinos river pairwise variables (Var1, Var2) with some degree of agreement determined by Cohen's Kappa coefficient. Pb_EF: enrichmdent factor by Pb, Ni_EF: enrichmdent factor by Ni, ADP: antecedent dry weather period, H: precipitation height (mm), I_{ave}: average intensity (mm/h).

Var1	Var2	p-value	Kappa	Var1	Var2	p-value	Kappa
H	D10	0.014	Moderate (0.44)	D40	Pb_EF	0.015	Fair (0.25)
I_{ave}	D10	0.014	Moderate (0.44)	D50	Pb_EF	0.015	Fair (0.25)
H	D20	0.014	Moderate (0.44)	D60	Pb_EF	0.004	Fair (0.31)
I_{ave}	D20	0.014	Moderate (0.44)	D60	Pb_EF	0.004	Fair (0.31)
H	D30	0.014	Moderate (0.44)	D20	Pb_EF	0.043	Slight (0.20)
I_{ave}	D30	0.014	Moderate (0.44)	D30	Pb_EF	0.043	Sligh (0.20)
H	D40	0.014	Moderate (0.44)	D70	Pb_EF	0.043	Slight (0.20)
I_{ave}	D40	0.014	Moderate (0.44)	D80	Pb_EF	0.043	Slight (0.20)
H	D90	0.014	Moderate (0.44)	D20	Ni_EF	0.023	Slight (0.20)
I_{ave}	D90	0.014	Moderate (0.44)	D30	Ni_EF	0.023	Slight (0.20)
I_{ave}	D90	0.014	Moderate (0.44)	D40	Ni_EF	0.023	Slight (0.20)
ADP	D10	0.039	Fair (0.35)				
ADP	D20	0.039	Fair (0.35)				
ADP	D30	0.039	Fair (0.35)				

3.2. Statistical analysis

The analysis of HM in MR sediments revealed significant relationships for the enrichment factors of niquel (Ni_EF) and lead (Pb_EF), as indicated by Kappa agreements (Table 4). No agreements were found for the other metals analyzed. These EFs are associated with specific PSD and hydrological variables. While Ni_EF exhibited a positive correlation with the antecedent dry weather period (ADP) using the Pearson method, this correlation was not statistically significant (p-value $>$ 0.05). In contrast, Pb_EF demonstrates significant correlations with D20 to D70 (p-value $<$ 0.05) as determined by the Spearman method (Fig. 4). This positions Pb_EF as a crucial target variable for predictive models, with D20 to D80, H, ADP, and I_{ave} serving as the independent variables.

In their study, Fatmi et al. (2024) utilized principal component analysis and cluster analysis to demonstrate the significant impact of hydrological parameters, such as precipitation, rainfall intensity, and the duration of dry weather preceding rainfall events, significantly impact on the dynamics and distribution of heavy metals. During dry periods, contaminated sediments tend to accumulate and function as reservoirs for contaminants, which are then mobilized during rainfall events. Similarly, Nie et al. (2024) empoyed random forest models to identify key factors influencing HM concentrations, including population density, annual average precipitation, and elevation for Cr; annual average precipitation, soil humidity, and soil organic matter for Cd; annual average precipitation for Pb; and population density, annual average precipitation, and slope for Ni.

In comparison, other studies, such as those by Fatmi et al. (2024) and Nie et al. (2024), have demonstrated the crucial role that hydrological parameters play in influencing HM dynamics. These findings reinforce the necessity for localized sediment management strategies. By integrating these insights, future research and environmental management practices can more effectively confront the risks associated with HM contamination. This approach will contribute to enhanced mitigation efforts and improved protection of aquatic ecosystems.

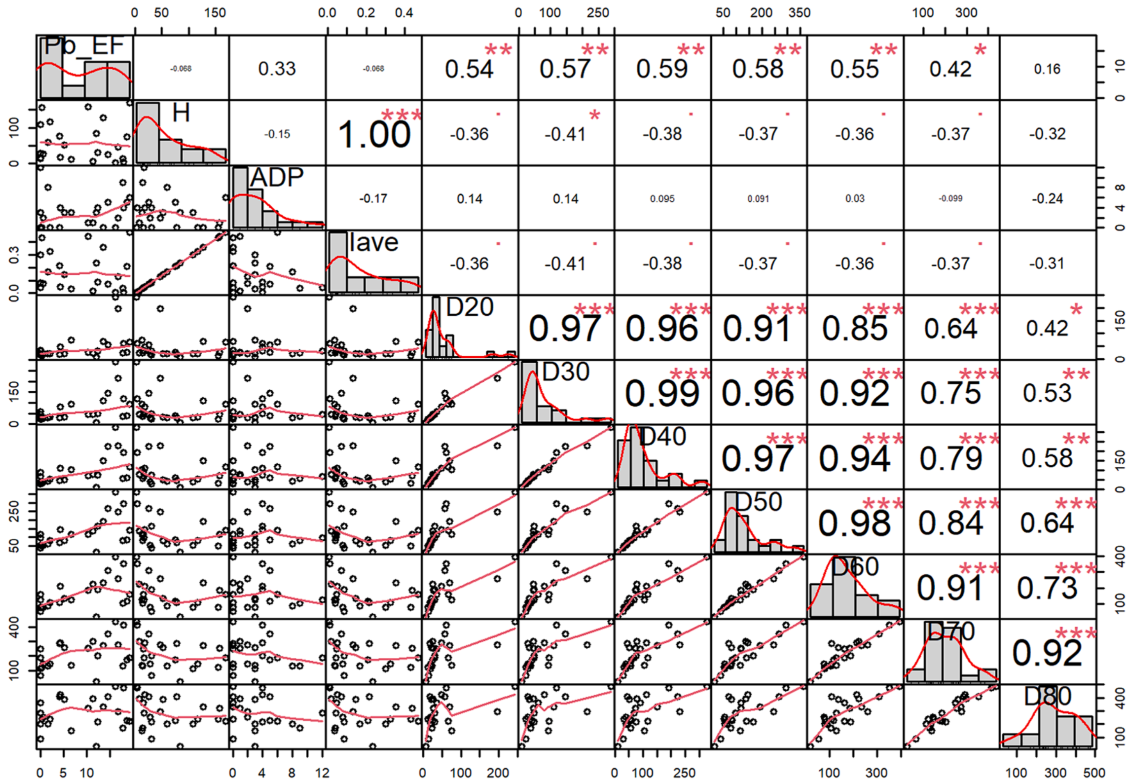

Fig. 4. Spearman correlation matrix. • p-value ≤0.1, * p-value ≤0.05, **p-value ≤0.01, ***p-value ≤0.001. Pb_EF: enrichment factor by Pb, ADP: antecedent dry weather period, H: precipitation height, I_{ave}: average intensity.

3.3. Risk index prediction

During the calibration-validation process of the SVM model, all available kernels were evaluated using the Monte Carlo method. The ANOVA kernel emerged as the optimal choice for calibrating the classification SVM for Pb_EF modelling, incorporating all related variables (PSD and rainfall). This kernel yield a p-value of <0.05 across a thousand calibration iterations. Fig. 5 illustrates the percentage of predicted classes for each observation in the validation models. Among the fourteen observations classified as having significant enrichment, eleven

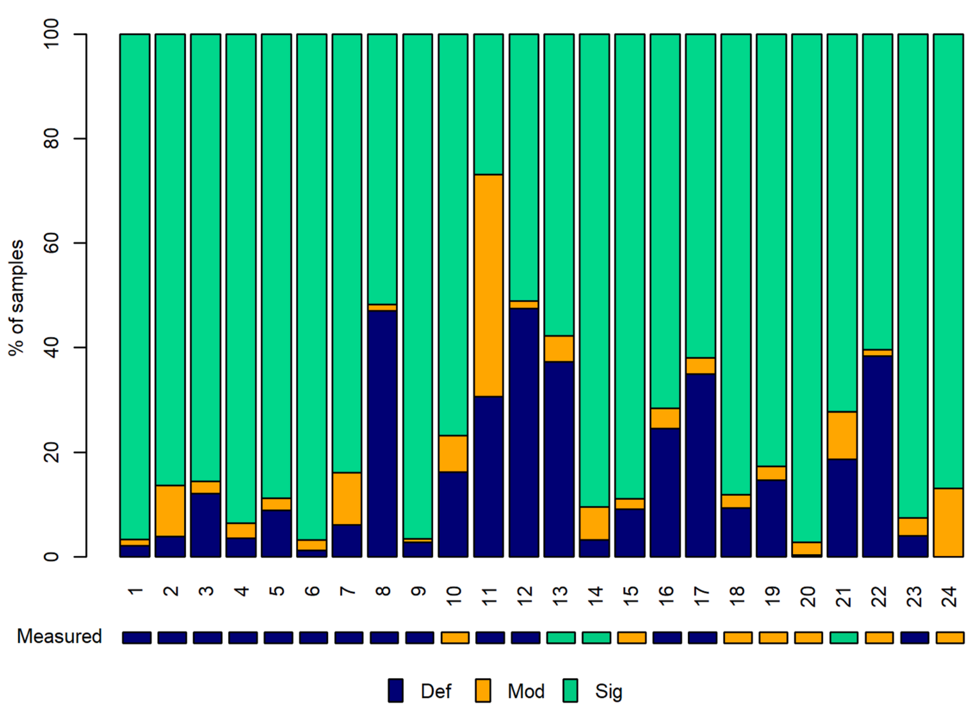

Fig. 5. Comparison of Support Vector Machine (SVM) model predictions with measured outcomes for Pb enrichment factors using validation data. Categories include Def (Deficiency to minimal enrichment), Mod (Moderate enrichment), and Sig (Significant enrichment).

were correctly predicted by >60 % of the models. For samples classified with moderate enrichment, fewer than 10 % of the models predicted the correct level. In the case of deficiency enrichment, one observation achieved up to 38 % prediction accuracy. This variability is largely attributed to the number of analyzed samples and their pronounced distribution of significant enrichment.

Model validation resulted in Cohen's Kappa coefficient of 0.71 and a p-value of 0.037, indicating substantial agreement. Fig. 6 compares the predicted and observed classification for eight validation samples. The model accurately predicted risk levels for seven out of eight observations but exhibited a significant error in the final one. Since all validation observations share the same PSD, corresponding to a calibration observation classified with significant enrichment, variations in risk indices are more likely associated with the rainfall data. Specifically, sample 24 exhibited high precipitation height and average intensity, along with a mean ADP value, compared to all observations. Although the prediction risk level for sample 24 was one degree higher than measured, the results still support effective sediment management decisions, as the risk was not underestimated.

The EF is instrumental in identifying the sources of HMs adhered to sediments. Modeling results assist in defining appropriate disposal strategies to mitigate concentration magnification and secondary environmental effects, particularly when moderate or higher levels of enrichment are detected. Nonetheless, the analyzed sediments could be reused as raw construction materials, provided they comply with local guidelines.

Given the lack of reported research on predicting environmental risk indices in sediments, our findings can be compared to studies that predict HM concentrations in soils using various machine learning techniques. For example, Bi et al. (2024) found that ensemble models based on decision trees, such as Random Forest and Bagging, demonstrated superior predictive accuracy by aggregating multiple trees, thereby reducing model variance and overfitting. Similarly, Yin et al. (2022) employed a genetic algorithm and neural network model to predict the concentrations of eleven soil HMs, achieving high correlation coefficients and emphasizing the value of integrating environmental variables and advanced modeling techniques.

Unlike these approaches, our study focuses on predicting risk levels rather than concentration. Direct risk levels estimation immediate insights into environmental hazard without requiring post-processing and relies on continuously measured data. The SVM model used in this study, which incorporates PSD and rainfall data, highlights the critical role of environmental factors in predicting HM enrichment.

During the development and evaluation of the proposed methodology, it became evident that models cannot be universally applied across different basins due to the unique characteristics of each basin's sediments. Consequently, an initial characterization is recommended, involving the collection of sufficient samples to ensure proper model calibration. Determining the optimal kernel and the most relevant variables is crucial for accurately identifying risk indices. Key contributing factors to this limitation include site-specific PSD distributions, variations in rainfall patterns, and differences in heavy metal sources.

4. Conclusions

This research proposes a methodology to predict the environmental risk index using statistical tools to facilitate sediment management. A case study demonstrated its ability to predict classification levels for risk indices. However, due to variability in climatic conditions and land use, applying this model to other catchments requires prior sediment characterization and model calibration.

Sediment characterization revealed that particles are predominantly sand-sized, indicating a small particle size and a larger specific surface area, which enhances contaminant adsorption. High concentrations of Cu, Pb, Ni, and Zn were detected in sediments, and the enrichment factor was calculated to determine pollutants origins. Results showed varying risk levels among samples, emphasizing the need for sediment management strategies.

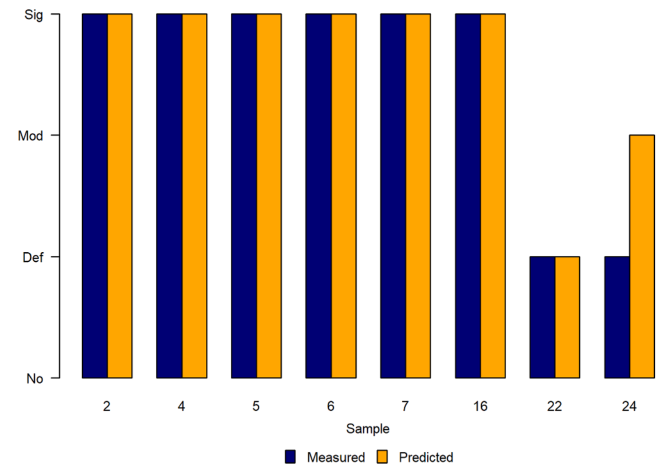

Fig. 6. Support Vector Machine (SVM) model results for Pb enrichment factors using validation data, compared with measured outcomes for eight samples. Labels indicate levels of enrichment: No (deficiency to minimal enrichment), Def (deficiency enrichment), Mod (moderate enrichment), and Sig (significant enrichment).

Statistical tools, including correlation test and Cohen's Kappa, identified relationships between the enrichment factor by Pb and specific characteristic diameters. Additionally, correlations between PSD and rainfall variables were observed. These relationships were used to calibrate and validate a SVM model, demonstrating the potential to predict enrichment factor based on rainfall data. However, validating these relationships with a larger dataset is recommended to improve reliability and robustness in future predictions.

The study contributes to knowledge by demonstrating that environmental risk levels associated with runoff sediments can be predicted using rainfall variables. This predictive tool provides decision-makers with a valuable resource for sediment management. To enhance accuracy, periodic model recalibration is necessary, particularly in response to anthropogenic and climate change influences.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Copilot in order to improve readability. After using this tool, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

CRediT authorship contribution statement

María Alejandra Pimiento: Writing – original draft, Investigation, Data curation. **Jose Anta:** Writing – review & editing, Conceptualization. **Andres Torres:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Maria Alejandra Pimiento reports financial support was provided by Ministry of Science, Technology and Innovation, Colombia. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by funds from the Ministry of Science, Technology and Innovation, Colombia, through the call 785 National doctorate.

Acknowledgements

This work was supported by funds from the Ministry of Science, Technology and Innovation, Colombia, through the call 785 National doctorate.

Data availability

Data will be made available on request.

References

Ahmed, B., Islam, S., Quraishi, S.B., Alam, Md.N.E., Ahsan, M.S., Kabir, A., 2024. A probabilistic risk assessment of heavy metal in water and sediment: an industrially affected urban river in Bangladesh. Water Environ. Res. 96 (8). https://doi.org/10.1002/wer.11097.

Bi, Z., Sun, J., Xie, Y., Gu, Y., Zhang, H., Zheng, B., Ou, R., Liu, G., Li, L., Peng, X., Gao, X., Wei, N., 2024. Machine learning-driven source identification and ecological risk prediction of heavy metal pollution in cultivated soils. J. Hazard. Mater. 476, 135109. https://doi.org/10.1016/j.jhazmat.2024.135109.

Boughandjioua, S., Laouacheria, F., Azizi, N., 2023. Machine learning algorithms investigation for urban drainage decision systems: overview. 2023. In: International Conference on Decision Aid Sciences and Applications (DASA), pp. 306–313. https://doi.org/10.1109/DASA59624.2023.10286621.

Choi, J., Jeong, H., Choi, K.-Y., Hong, G., Yang, D., Kim, K., Ra, K., 2020. Source identification and implications of heavy metals in urban roads for the coastal pollution in a beach town, busan, korea. Mar. Pollut. Bull. 161. https://doi.org/10.1016/j.marpolbul.2020.111724.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ Psychol Meas 20 (1), 37–46.

EPA., U. (2007). Method 3051a (sw-846): microwave assisted acid digestion of sediments, sludges, and oils.

Fatmi, B., Hazzab, A., Rahmani, A., Ghenaim, A., 2024. Examining temporal trends in heavy metal levels to analyze sediment pollution dynamics in the Saida urban watershed (N-W Algeria). Water Environ. Res. 96 (8). https://doi.org/10.1002/wer.11084.

Galarza-Molina, S., 2017. Decision-making Tool For the Operation of a Stormwater Harvesting System. https://doi.org/10.11144/Javeriana.10554.34609 [Doctoral dissertationPontificia Universidad Javeriana].

García-Gaines, R.A., & Frankenstein, S. (2015). Uscs and the usda soil classification system: development of a mapping scheme.

Gopal, V., Krishnamurthy, R.R., Vignesh, R., Nathan, C.S., Anshu, R., Kalaivanan, R., Mohana, P., Magesh, N.S., Bharath, K.M., Bessa, A.Z.E., Abdelrahman, K., Abioui, M., 2023. Assessment of heavy metal contamination in the surface sediments of the vedaranyam coast, southern india. Regional Studies Marine Sci. 65, 103081. https://doi.org/10.1016/j.rsma.2023.103081.

Guimarães, L., Teixeira, F., Pereira, J., Becker, B., Oliveira, A., Lima, A., Verόl, A., Miguez, M., 2021. The challenges of urban river restoration and the proposition of a framework towards river restoration goals. J. Clean. Prod. 316, 128330. https://doi.org/10.1016/j.jclepro.2021.128330.

Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. Kernlab - an s4 package for kernel methods in r. J. Stat. Softw. 11, 1–20.

Kwon, S.H., Kim, J.H., 2021. Machine learning and Urban drainage systems: state-of-the-art review. Water (Basel) 13 (24), 3545. https://doi.org/10.3390/w13243545.

Ltd, M.P. (2024, September). Mastersizer 3000e. https://www.malvernpanalytical.com/en/products/productrange/mastersizer-range/mastersizer-3000e.

Ma, H., Ding, F., Wang, Y., 2022. A novel multi-innovation gradient support vector machine regression method. ISA Trans. https://doi.org/10.1016/j.isatra.2022.03.006.

Nakazawa, M. (2021). Fmsb: functions for medical statistics book with some demographic data. https://cran.r-project.org/package=fmsb.

Nawrot, N., Wojciechowska, E., Mohsin, M., Kuittinen, S., Pappinen, A., Rezania, S., 2021. Trace metal contamination of bottom sediments: a review of assessment measures and geochemical background determination methods. Minerals 11 (8), 872. https://doi.org/10.3390/min11080872.

Nie, S., Chen, H., Sun, X., An, Y., 2024. Spatial distribution prediction of soil heavy metals based on random forest model. Sustainability 16 (11), 4358. https://doi.org/10.3390/su16114358.

Peterson, B.G., & Carl, P. (2020). Performanceanalytics: econometric tools for performance and risk analysis. https://cran.rproject.org/package=PerformanceAnalytics.

Pimiento, M.A., 2023. Experimental and Statistical Modeling Methodologies for Improving Settled Sediment Management in Urban Stormwater Systems. Pontificia Universidad Javeriana. https://doi.org/10.11144/Javeriana.10554.64118.

Pimiento, M.A., Duque, V., Torres, A., 2023. Urban stormwater sediment risk assessment from drainage structures in Bogotá, Colombia. Environ. Sci.: Water Res. Technol. 9 (12), 3269–3280. https://doi.org/10.1039/D2EW00746K.

Pimiento, M.A., Rivera, D.S., Borrero, J.A.L., Abello, A.E.T., 2018. Relationship between hydrological, physical and chemical characteristics present in rainwater runoff sediments. Ing. Compet. 20, 27. https://doi.org/10.25100/iyc.v20i2.5846.

Rauch, J.N., Pacyna, J.M, 2009. Earth's global Ag, Al, Cr, Cu, Fe, Ni, Pb, and Zn cycles. Global Biogeochem. Cycles 23 (2), 1–16. https://doi.org/10.1029/2008GB003376.

Rentachintala, L.R.N.P., Reddy, M.G.M., Mohapatra, P.K., 2022. Urban stormwater management for sustainable and resilient measures and practices: a review. Water Sci. Technol. 85 (4), 1120–1140. https://doi.org/10.2166/wst.2022.017.

Rietveld, M., Clemens, F., Langeveld, J., 2022. The mismatch between long-term monitoring data and modelling of solids wash-off to gully pots. Urban Water J. 19, 183–194. https://doi.org/10.1080/1573062X.2021.1986079.

Rumuri, R., Ramkumar, T., Vasudevan, S., Gnanachandrasamy, G., 2023. Enrichment of heavy metals as function of salinity and pH of estuarine sediments, South East Coast of India. Geology, Ecol. Landscapes 7 (3), 212–220. https://doi.org/10.1080/24749508.2021.1952764.

Sałata, A., Dąbek, L., 2017. Methods of assessment of stormwater sediments quality. In: E3S Web of Conferences, 17. https://doi.org/10.1051/e3sconf/20171700080.

Sanabria-Morales, J.A., Rivera-Soler, L., Cabra-Soto, A.D.P., Jiménez-Rojas, C.A., Torres-Camacho, J.N., Pimiento, M.A., Duque-Pardo, V., Torres, A., 2021. Risk assessment of runoff sediments in an experimental catchment in Bogotá related to hydrological and granulometric characteristics. Producción + Limpia 15, 92–108. https://doi.org/10.22507/pml.v15n2a5.

Skorbiłowicz, E., Ofman, P., Skorbiłowicz, M., Sidoruk, M., Tarasiuk, U., 2024. Geochemical assessment of heavy metal distribution in Bug River sediments, Poland: the impacts of urbanization and agricultural practices. Water (Basel) 16 (11), 1573. https://doi.org/10.3390/w16111573.

Team, R.C. (2023). R: a language and environment for statistical computing.

Wang, G., Liu, Y., Tao, W., Zhao, X., Li, X., 2019. Reflection of concentrations of polybrominated diphenyl ethers in health risk assessment: a case study in sediments from the metropolitan river, North China. Environ. Pollut. 247, 80–88. https://doi.org/10.1016/j.envpol.2019.01.041.

Wang, L., Liu, D., Sun, Y., Zhang, Y., Chen, W., Yuan, Y., Hu, S., Li, S., 2024. Machine learning-based analysis of heavy metal contamination in Chinese lake basin sediments: assessing influencing factors and policy implications. Ecotoxicol. Environ. Saf. 283, 116815. https://doi.org/10.1016/j.ecoenv.2024.116815.

Winston, R.J., Witter, J.D., Tirpak, R.A., 2023. Measuring sediment loads and particle size distribution in road runoff: implications for sediment removal by stormwater control measures. Sci. Total Environ. 902, 166071. https://doi.org/10.1016/j.scitotenv.2023.166071.

Yan, H., Zhu, D.Z., Loewen, M.R., Zhang, W., Yang, Y., Zhao, S., van Duin, B., Chen, L., Mahmood, K., 2024. Particle size distribution of total suspended sediments in urban stormwater runoff: effect of land uses, precipitation conditions, and seasonal variations. J. Environ. Manage. 365, 121467. https://doi.org/10.1016/j.jenvman.2024.121467.

Yang, H.J., Kang, T.-W., Choi, B., Hwang, S.H., Shin, D., Park, W.-P., 2022. Potential sources of heavy metals in sediments of an urban–agricultural watershed and relationship with land use using a statistical approach. Sustainability 14 (15), 9444. https://doi.org/10.3390/su14159444.

Yin, G., Chen, X., Zhu, H., Chen, Z., Su, C., He, Z., Qiu, J., Wang, T., 2022. A novel interpolation method to predict soil heavy metals based on a genetic algorithm and neural network model. Sci. Total Environ. 825, 153948. https://doi.org/10.1016/j.scitotenv.2022.153948.

Zhang, X., Wang, Y., Guo, S., Li, H., Liu, J., Zhang, Z., Yan, L., Tan, C., Yang, Z., Guo, X., 2020. Concentration and speciation of trace metals and metalloids from road-deposited sediments in urban and rural areas of Beijing, China. J. Soils Sediments 20, 3487–3501. https://doi.org/10.1007/s11368-020-02702-3.

Zhao, W., Ma, J., Liu, Q., Dou, L., Qu, Y., Shi, H., Sun, Y., Chen, H., Tian, Y., Wu, F., 2023. Accurate prediction of soil heavy metal pollution using an improved machine learning method: a case study in the Pearl River Delta, China. Environ. Sci. Technol. 57 (46), 17751–17761. https://doi.org/10.1021/acs.est.2c07561.