Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Disentangling spatial patterns and driving factor interactions of urban soil heavy metals using machine learning with multi-source heterogeneous data: A case study in Nanjing, China

Hanghang Yue
Nanjing University of Information Science and Technology
Taoran Shi
shitaoran@126.com

Nanjing University of Information Science and Technology
Chunfa Wu
Nanjing University of Information Science and Technology
Hong Wang
Nanjing Institute of Landscape Agriculture and Forestry
Jing Shao
Nanjing Institute of Landscape Agriculture and Forestry
Renying Li
Nanjing University of Information Science and Technology

Research Article

Keywords: Urban soil, Heavy metal, Machine learning, Spatial prediction, Structural equation modeling

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11163484/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Disentangling spatial patterns and driving factor interactions of urban soil heavy metals using machine learning with multi-source heterogeneous data: A case study in Nanjing, China

Hanghang Yue[a], Taoran Shi[a,*], Chunfa Wu[a], Hong Wang[b], Jing Shao[b], Renying Li[a,*]

[a]*School of Ecology and Applied Meteorology, Nanjing University of Information Science & Technology, Nanjing, 210044, China*

[b]*Nanjing Institute of Landscape Agriculture and Forestry, Nanjing 210037, China*

E-mail address: shitaoran@126.com (T.R. Shi)

ryli75@163.com (R.Y. Li)

Abstract: To elucidate the spatial differentiation characteristics and driving mechanisms of heavy metals in soils across the Nanjing urban area, this study integrated four machine-learning (ML) algorithms combined with ordinary kriging (OK), local indicators of spatial association (LISA), SHapley Additive exPlanations (SHAP), and structural equation modeling (SEM). Predictive modeling, spatial mapping, and mechanistic interpretations were conducted using heavy metal concentration data from 229 urban soil samples and multi-source environmental covariates. The results revealed that the coefficients of determination (R^2) for the optimal ML models applied to the test set ranged from 0.7221 to 0.9491 across the four heavy metals. Moran's I values for Cd, Pb, and Cu were 0.275, 0.218, and 0.212, respectively, indicating varying degrees of positive spatial autocorrelation. In contrast, the Moran's I value for Zn was only 0.048, suggesting a relatively weak overall spatial autocorrelation with pronounced local patchiness. Spatial predictions derived from two ML–OK coupling strategies revealed a complex spatial pattern characterized by multi-element accumulation in the historic urban core area, localized hotspots in construction-disturbed areas in the southwest, anomalous enrichment in northern areas affected by historical mining or industrial activities, and a transition toward lower concentrations in peripheral areas. SHAP analysis indicated that the spatial distributions of Pb and Zn were primarily associated with multiple heavy metal combined contamination, whereas Cu was influenced by anthropogenic inputs and adsorption-mediated immobilization by soil organic matter and minerals. Cd was more sensitive to competitive interactions among multiple heavy metals, hydrological

transport, and construction disturbance. The explained spatial variance based on SEM of the four heavy metals ranged from 76.8% to 96.3%, further validating the differences among the various elemental driving pathways.

Keywords: Urban soil; Heavy metal; Machine learning; Spatial prediction; Structural equation modeling

1. Introduction

Heavy metal contamination of urban soils represents a critical challenge in urban ecological environments and soil quality management. Heavy metals are characterized by high environmental persistence, degradation resistance, accumulation tendencies, and potential ecotoxicity. Once introduced into soil systems, they may adversely affect ecosystem stability and human health through multiple pathways, including dust resuspension, surface runoff, plant uptake, and human exposure (Long et al., 2021). During rapid urbanization and industrialization, urban soils are affected by the interplay of natural parent materials and anthropogenic activities. Transportation emissions, industrial activities, construction material weathering, urban renewal projects, atmospheric deposition, exogenous soil landfilling, and historically accumulated pollution may contribute to heavy metal accumulation in surface soils (Pan et al., 2025; Shen et al., 2024). Therefore, heavy metal contamination in urban soils is typically characterized by complex sources, pronounced spatial heterogeneity, localized concentration anomalies, and multi-metal combined pollution. Accurately

delineating the spatial distribution patterns of soil heavy metals and identifying their primary driving mechanisms are essential for assessing ecological risk, implementing targeted regional control measures, and executing refined remediation strategies.

Previous studies have extensively investigated soil heavy metal contamination, spatial distribution patterns, and source apportionment. Methods such as the Nemerow integrated pollution index, geoaccumulation index, and potential ecological risk index are widely used to assess the severity of heavy metal contamination and the level of ecological risks (Sheng et al., 2023). Ordinary kriging (OK), inverse distance weighting interpolation, spatial autocorrelation analysis, and local indicators of spatial association (LISA) have been widely applied to characterize the spatial continuity, local clustering patterns, and contamination hotspots of heavy metals (Ma et al., 2024). Additionally, principal component analysis, positive matrix factorization, absolute principal component scores–multiple linear regression, geological detectors, and isotopic tracing have been employed to identify potential pollution sources and their major influencing factors (Shao et al., 2024). These studies have established the foundation for understanding heavy metal contamination patterns in urban soils; however, they have predominantly focused on assessing contamination levels, identifying hotspot areas, and estimating the contribution of potential sources. Consequently, further research is required to achieve high-precision continuous spatial predictions by integrating multiple-source environmental covariates, including soil properties, topographic and hydrological conditions, remote sensing information, land use, and human activity intensity. Moreover, because heavy metals exhibit distinct

sources, migration abilities, and soil geochemical behaviors, their spatial distribution mechanisms can be highly element-specific. Therefore, a comparative analysis across multiple elements within a unified predictive and interpretative framework is necessary.

With advances in remote sensing, geographic information systems, and environmental big data, digital soil mapping has provided novel technical pathways for the spatial prediction of soil heavy metals. Because machine learning (ML) methods can handle high-dimensional variables, nonlinear relationships, and interaction effects, they have been widely utilized for predicting soil properties and spatial heavy metal mapping (Li et al., 2024a; Sun et al., 2024; Zhao et al., 2023). Models such as support vector machines (SVM), support vector regression (SVR), gradient-boosted trees, and neural networks can capture relationships between heavy metal concentrations and multi-source covariates, including climate, topography, hydrology, soil properties, remote sensing information, and anthropogenic activities, thereby improving the accuracy of spatial predictions. However, predictions from single ML models rely primarily on environmental covariates and may fail to capture the spatial autocorrelation structure among sampling locations that remains unexplained by these covariates. This limitation may result in insufficient spatial continuity in the predictions or limited capability to correct local anomalies. Conversely, geostatistical methods, such as OK, can improve local predictions by exploiting spatial autocorrelation. However, their ability to characterize environmental driving processes and nonlinear relationships is limited. Accordingly,

integrating ML predictions with residual OK correction can simultaneously leverage information from multisource covariates and the residual spatial dependence structure, thereby enhancing the continuity and reliability of heavy metal concentration predictions in urban soils (Han et al., 2025).

Beyond spatial prediction accuracy, model interpretability remains a critical challenge in current ML applications to soil heavy metal contamination. Although most ML models exhibit strong predictive performance, their "black-box" nature hinders their in-depth interpretation of the underlying contamination processes. While some studies have addressed the identification of variable importance, most remain confined to predictor importance ranking. Consequently, they lack a systematic analysis of directional factor influences, nonlinear thresholds, two-factor interaction effects, and structured driving pathways. In recent years, interpretable ML approaches, such as SHapley Additive exPlanations (SHAP), have been employed to quantify the magnitude and direction of individual environmental variable contribution to model prediction outcomes, providing an effective tool for interpreting complex models (Li et al., 2024c; Yan and Yang, 2025). Two-dimensional response surface analysis or partial dependence analysis can further elucidate the nonlinear interactions and threshold effects among key factors. Structural equation modeling (SEM) identifies direct and indirect effects of individual factors, as well as covariance relationships among different factors through path analysis. Therefore, integrating ML-based spatial prediction, SHAP-based interpretability analysis, dual-factor interaction identification, and SEM provides a comprehensive framework to systematically

111 elucidate the mechanisms underlying heavy metal spatial heterogeneity in urban soils.
112 This integrated approach operates across three dimensions: spatial prediction, variable
113 interpretation, and driving-path identification.

114 To address these challenges, this study investigated heavy metals in the urban
115 soils of Nanjing by integrating multi-source environmental data to establish a unified
116 framework for spatial prediction and driving mechanism identification. The specific
117 objectives were: (1) to develop spatial prediction models for soil heavy metals; (2) to
118 construct two mapping methods integrating ML with OK to improve the continuity
119 and reliability of spatial predictions of urban soil heavy metals; (3) to identify the
120 primary factors influencing individual heavy metals and their nonlinear interactions
121 based on SHAP analysis and bivariate response analysis; and (4) to elucidate the
122 driving paths of exogenous inputs, soil matrix regulation, hydrological migration, and
123 interelement interactions driving the spatial differentiation of heavy metals using
124 SEM. These findings provide methodological guidance and a scientific basis for
125 identifying heavy metal contamination in urban soils, implementing spatially targeted
126 management, and advancing refined soil environmental governance.

127 **2. Materials and methods**

128 2.1. Study area and sample collection

129 Nanjing is located in the Ningzhen hilly region along the lower reaches of the
130 Yangtze River (31°14′–32°37′ N, 118°22′–119°14′ E). The area has a subtropical
131 monsoon climate characterized by four distinct seasons, with an annual average
132 temperature of approximately 15.4 °C and annual precipitation of approximately

1,106 mm.

This study focused on the six central urban districts of Nanjing, including Xuanwu, Gulou, Qinhuai, Jianye, Qixia, and Yuhuatai (Fig. 1). This area is characterized by intensive urban development, dense human activities, and the combined influences of industrial operations, transportation emissions, and historical land development disturbances, making it ideal for investigating the spatial heterogeneity of heavy metals within urban soils (Nanjing Municipal Bureau Statistucs).

A GIS-assisted grid-based sampling scheme was employed during 2018–2019 to select 229 sampling sites across green spaces in the central urban area of Nanjing (Fig. 1). Sampling sites were prioritized within representative green space units in each grid cell, while maintaining consistent surrounding environments and green space types across sites within the same grid cell. Soil sampling, sample processing, and quality control procedures were conducted based on the technical specifications for soil environmental monitoring (HJ/T 166-2004). Sample processing and physicochemical property analyses are described in the Supplementary Materials.

Fig. 1. Study areas and sampling site locations.

2.2. Research methods

2.2.1. ML model

In this study, four ML models, namely, Multi-layer perceptron (MLP), SVM,

Histogram-based gradient boosting (HistGradientBoosting), and SVR were selected. They represent distinct modeling paradigms, including neural networks, maximum margin learning, gradient-boosted trees, and kernel regression, and can thus characterize complex nonlinear relationships and environmental response mechanisms involved in soil heavy metal contamination from different theoretical perspectives. Specifically, MLP employs multi-layer nonlinear mappings and adaptive parameter learning, enabling it to effectively capture high-order coupling interactions Zn, Cd, Cu, and Pb concentrations driven by the combined effects of multiple environmental factors. Hist Gradient Boosting efficiently implements gradient-boosted decision trees and accelerates the training process through histogram-based binning while effectively capturing nonlinear splits and interaction effects among variables. SVM exhibits strong generalization capability in high-dimensional, small-sample settings; it relies on the principle of structural risk minimization, effectively addressing the challenges of contamination status discrimination in complex environmental covariates. In contrast, SVR is a kernel-based learning method designed for continuous variable prediction. It can characterize nonlinear functional relationships between environmental factors and heavy metal concentrations while controlling model complexity and exhibits good robustness to noise interference.

2.2.2. SHAP-based model interpretation

Grounded in game-theoretic Shapley values, SHAP quantitatively characterizes the marginal contribution of an individual feature to the model output (Lundberg and Lee, 2017). In this study, SHAP was employed to interpret the direction and

177 contribution rate of environmental influence on predicted soil heavy metal
178 concentrations across Nanjing at both the global and local levels.

179 2.2.3. Local spatial autocorrelation

180 The LISA decompose global spatial autocorrelation into local spatial units,
181 identifying neighborhood-scale spatial clusters and outliers, and testing for significant
182 high-value clusters, low-value clusters, or spatial anomalies between each site and its
183 neighbors (Anselin, 1995). Based on the attribute values of sampling sites and their
184 neighbors, LISA categorizes local spatial associations into four types: high–high (HH),
185 low–low (LL), high–low (HL), and low–high (LH).

186 Given that urban soil heavy metals commonly exhibit distinct local hotspots,
187 strong spatial heterogeneity, and multisource input superposition, this study applied
188 the LISA method to characterize the local spatial association patterns of Cd, Cu, Pb,
189 and Zn following global spatial autocorrelation analysis. Notably, the obtained LISA
190 cluster map was derived from sampling site neighborhood relationships; its significant
191 points represent point-scale local spatial associations instead of continuous areal
192 contamination extents. Accordingly, LISA outputs identify local hotspots, cold spots,
193 and spatial outliers, whereas characterizing continuous spatial patterns of heavy
194 metals requires integrated assessment via subsequent spatial prediction results.

195 2.2.4. Moran's I Index

196 Global Moran's I is a classic statistic for assessing the overall spatial
197 autocorrelation of variables and detecting clustered, dispersed, or random
198 distributions of soil heavy metal concentrations across the study area. Herein, it was

adopted to test global spatial dependence of Cd, Cu, Pb and Zn, whereas the Moran scatter plot was used to examine relationships between observed values and their spatial lags, supporting the interpretation of LISA cluster patterns and spatial mapping. Suitable for quantifying correlations between variable observations and their spatial lags, Moran's I further facilitates the identification of local clusters and outliers among spatial units.

2.2.5. ML–geostatistical fusion prediction

Existing studies indicate that the fusion model combining ML with OK generally outperforms both ML or conventional geostatistical methods alone and can effectively enhance the accuracy of the spatial prediction of soil pollution and the reliability of spatial mapping (Wang et al., 2024). In this study, the Full-domain Regular Grid-based ML and OK interpolation method (FRG-ML–OK) and ML coupled with OK interpolation method (ML–OK) were adopted to conduct spatial prediction mapping of soil heavy metals. FRG-ML–OK involves global prediction followed by spatial smoothing correction, whereas ML–OK involves trend fitting followed by residual interpolation.

The spatial distribution of heavy metals in urban soil typically exhibits two characteristics: (i) large-scale and relatively continuous spatial variability influenced by factors such as natural background conditions, land use, and urban development; and (ii) localized high-value anomalies due to the impact of regional pollution sources such as transportation emissions and industrial activities. Therefore, this study developed two coupled modeling approaches to compare how fusion mechanisms

221 characterize high-value patches, low-value boundaries, local anomalous sites, and the
222 overall regional gradient.

223 The sampling sites within the study area were mainly distributed across urban
224 green spaces, parks, road intersections, residential areas, schools, and urban
225 development areas; therefore areas such as water bodies, mountain edges, and areas
226 with sparse sampling sites involved higher extrapolation uncertainties. The clipped
227 boundary of the study area was used as a spatial prediction mask. This treatment does
228 not alter the scope of the research object but confines interpolation and mapping to an
229 effective spatial support domain with high sampling point coverage and meaningful
230 environmental covariates, thereby reducing the uncertainty of boundary extrapolation
231 and prediction in areas devoid of sampling sites (Meyer and Pebesma, 2021).

232 2.2.6. Structural equation modeling

233 SEM is a multivariate statistical method that combines factor and path analyses,
234 making it suitable for testing complex relationships between variables. This method
235 can simultaneously handle latent and observed variables and evaluate the matching
236 degree between the theoretical model and sample data through goodness-of-fit
237 indicators, thereby validating the rationality of the research hypotheses.

238 The workflow diagram summarizes the integrated analytical framework used in
239 this study (Fig. 2).

240

241 **Fig. 2.** Algorithmic workflow diagram of the study.

242

2.3. Data sources and processing

In this study, a spatial prediction framework for heavy metals was established based on multisource heterogeneous data. The original multisource data were primarily derived from public databases (summarized in the Supplementary Material), including remote sensing spectral indices, topographic factors, soil properties, meteorological factors, and human activity intensity (Text S4, Table S1). Given the influence of pH, OM, clay particles, and ion competition on the adsorption, migration, and bioavailability of heavy metals, a series of indicator combinations has been developed to enhance the ability of these features to characterize pollution processes (Tipping et al., 2011).

Continuous variables were subjected to Z-score standardization, and collinearity was screened using the Pearson's correlation coefficient and variance inflation factors (VIFs) (Fig. S1). A total of 30 feature variables were selected for subsequent modeling (Table S2).

2.4. Model performance evaluation

All ML models used in this study were implemented using Python 3.12.3. The dataset was randomly divided into training and test datasets in an 8:2 ratio. The models were developed using the training set and corresponding feature variables. The predicted and observed values from the validation set were compared, and the prediction accuracy was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) for the validation set (Shi et al., 2022). The calculation formulas for R^2,

RMSE, MAE, and MAPE are as follows:

$$R^2=1-\frac{\sum_{i=1}^{N}(\hat{z}_i-z_i)^2}{\sum_{i=1}^{N}(z_i-\bar{z})^2}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{z}_i-z_i)^2}$$

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|\hat{z}_i-z_i|$$

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{z}_i-z_i}{z_i}\right|\times 100\%$$

where, N denotes the number of sampling sites; z_i and $\hat{z}_i$ denote the observed and predicted values at sampling site i, respectively; and $\bar{z}$ denotes the mean of the observed values.

3. Results and Discussion

3.1. Model performance for metal concentration predictions

In this study, Cd, Cu, Zn, and Pb concentrations were used as response variables, and ML regression models were constructed based on multisource environmental factors. The predictive performance of the model was comprehensively evaluated using the training set, test set, and 5-fold cross-validation. The prediction results of the optimal model for each element are illustrated in Fig. 3.

For Zn, the optimal model was the SVR model, with R^2 values of 0.9621 and 0.9491 for the training and test sets, and corresponding MAPE values of 3.08% and 3.09%, respectively. The 5-fold cross-validation yielded an R^2 of 0.9516 and a MAPE

283 of 3.45% (Fig. 3a). The model exhibited relatively stable performance across the
284 training, test, and cross-validation sets, and the predicted values were mainly
285 distributed near the 1:1 line, indicating that SVR had a strong ability to characterize
286 the continuous variation of Zn concentrations. These results indicate that SVR can
287 effectively handle the nonlinear response relationship between Zn concentration and
288 environmental variables using kernel function mapping (Smola and Schölkopf, 2004).

289 For Pb, the optimal model was the MLP model, with R^2 values of 0.8416 and
290 0.7221 for the training and test sets, respectively, and corresponding MAPE values of
291 9.16% and 12.70%. The 5-fold cross-validation yielded an R^2 value of 0.7757 and
292 MAPE of 10.81% (Fig. 3b). Compared with Cd and Zn, the predictive accuracy of the
293 Pb model was relatively low, indicating that the spatial distribution of Pb
294 concentrations was more difficult to explain using the environmental variables
295 selected in this study. This may be related to the complex sources of Pb, pronounced
296 local point-source inputs, and strong spatial heterogeneity (Zeng et al., 2023). The
297 MLP captured the nonlinear patterns of Pb concentration variation to a certain extent;
298 however, the relatively high error in the test set indicated that Pb prediction was
299 affected by anomalously high values, localized pollution inputs, and uneven spatial
300 distribution of samples (Breiman, 2001).

301 For Cu, the optimal model was the SVM model, with R^2 values of 0.8114 and
302 0.7684 for the training and test sets, respectively, and corresponding MAPE values of
303 3.66% and 4.78%. The 5-fold cross-validation yielded an R^2 of 0.7061 and a MAPE
304 of 4.68% (Fig. 3c). Overall, the SVM captured the main trend in concentration

variation relatively well; however, its R^2 values were lower than those of Cd and Zn, indicating that the spatial variation in Cu concentrations was relatively more difficult to explain. Combined with the scatterplot distributions, predicted and observed Cu values were generally consistent, although minor underestimation persisted in certain high-value ranges. This discrepancy indicates that the model's ability to characterize localized high values could be further improved.

For Cd, the Hist Gradient Boosting Regressor exhibited the best predictive performance, with R^2 values of 0.9270 and 0.9283 for the training and test sets, respectively, and corresponding MAPE values of 4.04% and 5.16%, respectively. The 5-fold cross-validation yielded an R^2 value of 0.8153 and MAPE value of 6.73% (Fig. 3d). The results for the training and test sets were relatively close, indicating that the model had good fitting and generalization performances. The relatively high predictive accuracy of the Cd model indicates that the gradient boosting framework can effectively characterize the potential nonlinear relationships and interaction effects between Cd concentrations and multisource environmental factors.

Overall, the ML models developed in this study can effectively accommodate the spatial distribution characteristics of heavy metals in the soil. Among them, Cd and Zn exhibited relatively high predictive accuracy, indicating relatively stable statistical response relationships between their concentration variations and selected environmental factors. The Cu model showed moderate performance and reflected its overall variation trend, whereas the Pb model exhibited relatively greater predictive uncertainty, suggesting that its spatial distribution may be controlled by more complex

localized pollution processes. The optimal models differed among heavy metals, indicating that the response relationships between individual elements and environmental variables differed in terms of nonlinearity degree, noise structure, and feature effects mode. The relatively consistent performance of the models on the training and test sets indicates good generalization ability, supporting their application in subsequent spatial distribution prediction of soil heavy metals (Friedman, 2001).

Fig. 3. Actual vs. predicted concentrations of heavy metals using the optimal ML models.

3.2 Spatial distribution characteristics of heavy metals

3.2.1 Local LISA analysis and Moran scatter plots

To identify the local clustering types and spatial heterogeneity of soil heavy metals in the study area, LISA and Moran's I were used to analyze the spatial distribution patterns of Cu, Zn, Pb, and Cd (Fig. 4). The results revealed that the Moran's I values for Cu, Pb, and Cd were 0.212, 0.218, and 0.275, respectively, indicating a certain degree of positive spatial autocorrelation for all three elements. In contrast, the Moran's I value of Zn was only 0.048, indicating a weak overall spatial autocorrelation and the most prominent local patchiness and dispersion. Overall, the four elements were predominantly distributed as insignificant spatial units, indicating that soil heavy metals within the study area did not exhibit large-scale continuous clusters, but rather a spatial pattern characterized by the coexistence of local hotspots,

local cold spots, and a small number of spatial outliers.

The spatial distribution of Zn was characterized by pronounced patchiness and dispersion (Fig. 4a). HH clusters were mainly located in the core area of the old district at the junction of the southern Gulou District, southwestern Xuanwu District, and northern Qinhuai District, and extended partially into the central and eastern margins of the Jianye District. Spatially, these areas correspond to zones of intensive human activity, such as Xinjiekou, Zhujiang Road, and the northern side of Confucius Temple. These areas are characterized by high road density, frequent commercial activities, and intensive construction and redevelopment. Tire wear, corrosion of galvanized materials, road dust, and construction operations may have contributed to the local enrichment of Zn. LL clusters were mainly scattered across the southern Qinhuai District and northern Yuhuatai District, and were interspersed with HH clusters, indicating that the spatial pattern of Zn did not form a stable, large-scale continuous cluster. The clustering proportions showed that non-significant areas accounted for 93.0% of the Zn distribution, whereas HH, LL, and HL clusters accounted for relatively small proportions, further indicating that the spatial distribution of Zn was strongly influenced by local inputs and exhibited weak overall spatial autocorrelation. Dust Cu, Pb, and Zn from major roads in Nanjing exhibited significant spatial differences, indicating that traffic emissions can lead to the formation of spatially discrete, non-continuous heavy metal inputs at the urban scale, which is consistent with the local patchy distribution pattern of Zn observed in this study (Liu et al., 2014).

The LISA cluster map of Pb showed that HH clusters were mainly scattered at the boundary between the southern Gulou District and the northwestern Qinhuai District, as well as in the adjacent areas of the southern Qinhuai District and the northern Yuhuatai District, exhibiting distinct local hotspot features (Fig. 4e). These areas are located along the main roads in the old district, historically developed zones, and areas where construction and renewal activities are frequent. Local high Pb concentrations were more likely associated with local sources, such as transportation emissions, legacy industrial activities, and intensive construction disturbances. The LL clusters were mainly distributed in the eastern Xuanwu District and south-central Qixia District, indicating that the Pb concentrations were relatively low and exhibited distinct spatial continuity. The clustering proportions showed that insignificant areas accounted for approximately 91.3%, LL clusters accounted for 5.2%, and HH clusters accounted for 2.6%. This indicates that although Pb exhibited a certain degree of positive spatial correlation, the area of high Pb concentrations was limited, manifested as local accumulation in the core area of the old district and in the transportation and commercial activities mixed-use zone.

The LISA clustering results for Cu showed that HH clusters were mainly distributed at the boundary between the northeastern Gulou District and the northern Xuanwu District, as well as near the urban core area between the southern Xuanwu and northern Qinhuai Districts (Fig. 4i). Geographically, these areas correspond to the vicinity of Xuanwu Lake, Central Road, Longpan Road transportation corridor, and the high-density urban development zone of the old district, and are characterized by

high traffic volumes and a highly mixed use of commercial and residential land. Vehicle braking wear, road dust resuspension, aging of construction materials, and urban development and renovation may be important sources of local Cu enrichment. LL clusters were mainly distributed in the central-southern and eastern fringe areas of the Qixia District (Fig. 4i), indicating relatively low Cu concentrations and a certain degree of spatial continuity. According to the clustering proportions, insignificant areas accounted for 92.6%, whereas HH and LL clusters accounted for 1.7% and 5.2%, respectively, indicating that although local Cu hotspots were present, insignificant spatial units dominated. This result is largely consistent with those of previous studies on the spatial distribution of heavy metals in greenland soils across the main urban area of Nanjing. Specifically, high-value areas for Cu, Zn, and Pb were located in the northeastern part of the main urban area, whereas the high-value areas for Cd were distributed in the southwestern and northwestern parts (Zhang et al., 2018).

The spatial distribution pattern of Cd differed markedly from those of Cu, Zn, and Pb, characterized by a few prominent local high-value areas and a relatively high proportion of low-value clusters. HH clusters were located in the southwestern part of the study area, particularly in the southwestern Jianye District and the adjacent northwestern area of the Yuhuatai District (Fig. 4m), spatially corresponding to the southern Hexi and areas bordering the northern Yuhuatai District. This area has been strongly influenced by rapid urban development, intensive land use, road traffic emissions, and inputs of external soil or construction fill material, which may have resulted in anomalous Cd enrichment at a few local sampling sites. Compared to the

other three elements, Cd had the lowest proportion of HH clusters (0.4%), indicating that its high values did not represent a continuous band-like distribution but were more likely driven by a small number of outlier points. LL clusters were mainly distributed in the southern Qinhuai District, northern Yuhuatai District, and some parts of the Qixia District. The proportion of LL clusters was 8.7%, higher than those of Cu, Zn, and Pb, indicating that low-Cd areas exhibited relatively stronger spatial continuity in some developed and peripheral urban areas. The clustering proportions showed that non-significant areas accounted for 90.0% of the Cd distribution, LL clusters accounted for 8.7%, and HH clusters accounted for only 0.4%, indicating that anomalous Cd enrichment was more likely driven by a small number of local high-value points rather than by the formation of a large-scale continuous pollution belt.

Overall, the variations in the spatial clustering scale and clustering intensity among different heavy metals reflect the typical distribution characteristics of urban soil heavy metals under the combined influences of urban development, transportation activities, and exogenous pollution inputs.

Fig. 4. Spatial distribution and global-local spatial autocorrelation patterns of Zn, Pb, Cu, and Cd, the LISA cluster map (a, e, i, m), concentration distribution map (b, f, j, n), Moran's I scatter plot (c, g, k, o), and pie chart of cluster type proportions for specific elements (d, h, l, p) are presented in order.

437 3.2.2. Prediction of spatial distribution

438 To improve the reliability of spatial predictions of soil heavy metals, two ML–
439 OK coupling strategies were employed for spatial distribution mapping. The results
440 from the two types of coupled mapping methods showed that the major high- and
441 low-value areas of heavy metals were generally spatially consistent, indicating good
442 stability of the prediction results (Fig. 5). The ML–OK yielded a relatively smooth
443 output, better highlighting the continuous variation trends at the regional scale (Fig.
444 5b, d, f, h). In contrast, FRG-ML–OK was more sensitive to local high-value patches,
445 boundary transition zones, and small-scale spatial heterogeneity, more clearly
446 reflecting the concentration differences among various functional units within the
447 central urban area (Fig. 5a, c, e, g).

448 The first spatial pattern was characterized by accumulation in the central urban
449 area, primarily distributed in the southern Gulou, southwestern Xuanwu, and northern
450 Qinhuai Districts, and the area at the junction of these three districts. This spatial
451 pattern corresponded to the core areas of the old city and the dense major
452 thoroughfare zones, including Xinjiekou, Zhujiang Road, Central Road, Changjiang
453 Road, the northern side of the Confucius Temple, Hanzhong Gate, and Shuixi Gate.
454 This area was characterized by high or moderately high concentrations of Zn, Pb, and
455 Cu, indicating that transportation, commercial activities, road dust resuspension,
456 weathering of aging building materials, demolition and renovation, and intensive
457 urban renewal exert cumulative effects on multiple heavy metals. In particular, Zn
458 was more likely to form patchy high-concentration areas along road-dense zones and

urban renewal areas, reflecting the source inputs of tire wear, corrosion of galvanized material, and road dust (Blok, 2005). High Pb concentrations were more concentrated in the core urban transportation and commercial mixed-use areas, suggesting that its spatial pattern was strongly influenced by historical transportation emissions and long-term deposition (Wu et al., 2010). Cu is influenced by traffic braking wear, weathering and corrosion of metallic building materials, abrasion of metal road ancillary facilities, and the input of particulate matter during construction and renewal processes (Pamuru et al., 2022). Consequently, relatively high Cu concentrations were observed in both the core area of the old district and certain renewed areas. This study further demonstrated that different elements within the same high-concentration area of the central urban region exhibit distinct input mechanisms and spatial patch configurations (Jeong et al., 2022).

The second spatial pattern was characterized by construction-related disturbance in the southwest, mainly distributed in the southwestern Jianye District, northwestern Yuhuatai District, and adjacent urban development areas, spatially corresponding to southern Hexi, the Olympic Sports Center, Shuangzha, Youfang Bridge, Xishan Bridge, Saihong Bridge, the Ningwu Railway, and urban expressway. The most prominent feature of this region was the relatively high or second-highest concentrations of Cd and Cu, whereas the responses of Zn and Pb were weaker than those observed in the core area of the old district. Cd was relatively sensitive to exogenous inputs and soil disturbance. In areas with intensive construction activities, the mixing of exogenous soil, backfill soil, and the original surface soil, as well as the

resuspension and migration of soil particles and road dust, may facilitate the local input and spatial enrichment of Cd, thereby generating localized, high-value anomalies (Yang et al., 2020). The elevated Cu concentrations in this reach may be attributable to a combination of emissions and wear from non-exhaust traffic emissions, and particulate matter inputs from construction and demolition activities (Huang et al., 2022). Therefore, the southwestern construction disturbance area represents an important supplement to the conventional understanding of heavy metal accumulation in old urban centers, demonstrating that rapid urbanization and newly developed areas may also give rise to high-concentration heavy metal patches (Yadav and Raman, 2021).

The third spatial pattern corresponded to areas affected by historical mining or industrial impacts in the northern regions with high or second-high Cd concentrations, and spatially corresponded to Swallow Rock, Yaohua Gate, Qixia Mountain, and the Muyan Riverside Scenic Area. This pattern was likely associated with the combined effects of historical mining, industrial activities, and atmospheric deposition. The Qixia Mountain area has a history of Pb–Zn mining activity. Cd commonly occurs as an associated element in Pb–Zn ores and their minerals. Waste rock, tailings, dust deposition, and surface runoff generated by mining may result in the enrichment of elements such as Cd, Pb, and Zn in the surrounding soils (Li et al., 2024b). The LISA results did not indicate continuous and significant HH clusters in this region; instead, such clustering appears as a surface distribution of high or second-high concentrations on the continuous spatial prediction map (Fig. 5g, h). This difference suggests that the

503 northern anomalies may be influenced by a combination of point-source and boundary
504 effects, alongside the proximity structure of the sampling sites. This finding suggests
505 that relying solely on local spatial autocorrelation analysis may underestimate
506 potential risks in boundary or sparsely sampled regions, and continuous spatial
507 prediction can provide additional details for identifying local anomalies (Wang et al.,
508 2023b).

509 The fourth spatial pattern represents an outer peripheral ecological low-value
510 transition zone, mainly distributed across the eastern Xuanwu District, Zijin Mountain,
511 Zhongshan Mountain National Park, eastern and northeastern Qixia District, Xianlin
512 District, parts of southern Yuhuatai District, and peripheral low-intensity construction
513 zones. In these areas, mountainous, ecological, and residential land accounted for
514 relatively high proportions, whereas road density, commercial activity intensity,
515 industrial disturbance, and development intensity were relatively low. Consequently,
516 the heavy metal values were moderate to low or low. This low-value distribution
517 pattern was largely consistent with the distribution of LL clusters based on the LISA
518 results. However, low-value areas should not be regarded as entirely risk-free. Outlier
519 high values may still occur in certain road, construction boundaries, or historical
520 activity areas. Therefore, subsequent risk identification should be conducted by
521 integrating local sample data with land use patterns.

522 Compared with the point-scale LISA results, the continuous mapping results
523 further revealed the spatial extension directions of heavy metal distribution and the
524 potential source boundaries. The results of the two mapping approaches elucidated the

spatial zonal superimposed relationships among different source-related processes, rather than simply identifying which administrative district has the highest concentration of a particular element. Overall, the soil heavy metals in the study area did not exhibit a simple decreasing pattern from the urban center to the periphery. Instead, it formed a composite spatial pattern characterized by the coexistence of accumulation in the central urban area, disturbance from southwestern development activities, legacy effects of mining or industrial activities in the northern region, and a peripheral ecological low-value transition zone. This understanding provides a spatial basis for subsequent pollution source apportionment, delineation of priority control zones, and refined risk assessments.

Fig. 5. Spatial prediction of heavy metal concentrations based on the FRG-ML–OK model (a, c, e, g) and ML–OK model (b, d, f, h) for Zn, Pb, Cu, and Cd.

3.3. Factors influencing the distribution of heavy metals

3.3.1. Importance analysis of influencing factors

To further quantify the contributions of various environmental factors to the spatial differentiation of soil heavy metals, SHAP values were calculated based on the optimal ML model for each element, and the mean absolute SHAP value was used to characterize the global importance of each factor. The SHAP method can quantify the relative contribution of individual variables to model prediction outcomes and reveal whether the variable promotes or inhibits the predicted heavy metal concentrations

based on the sign of its SHAP value. Overall, the dominant factors varied significantly among the heavy metals (Fig. 6).

The dominant factors for Zn were CZAI, CPI, and ZCAI, followed by the soil buffering capacity index, clay–organic carbon coupling index, OMCC, VHMEC, and EC. This result indicates that the spatial distribution of Zn is mainly controlled by competitive adsorption and combined pollution under multi-metal coexistence conditions and is also jointly regulated by the soil buffering system, OM components, and ion environment. The relatively high contributions of CZAI, CPI, and ZCAI indicate a strong geochemical coupling between Zn and Cu and Cd and demonstrate that Zn enrichment is frequently associated with urban combined pollution processes. In the six districts of Nanjing's central area, the Zn input process primarily involves vehicle tire wear, corrosion of galvanized components, resuspension of road dust, construction-related disturbances, aging of metallic materials, and localized industrial or logistics activities. The core urban areas, including the Gulou District, Qinhuai District, southwestern Xuanwu District, northern Yuhuatai District, actively developing Hexi area of Jianye District, and industrial and transportation complex zones in southern Qixia District, are likely to be areas with high Zn accumulation. The importance of SBCI, COCI, OM, OMCC, EC, and pHMI indicates that the migration and transformation of Zn were not entirely determined by exogenous inputs. Zn has higher mobility than Pb and Cu, and the adsorption and desorption of Zn are highly sensitive to pH, ionic strength, clay minerals, and OM content, and the competitive effect is particularly pronounced under conditions of multi-metal coexistence (Yu et

al., 2023). Higher OM and clay content can enhance Zn retention through complexation, adsorption, and colloidal binding. Studies have revealed that with increasing soil OM content, the proportion of Zn bound in humic acid within the immobilized Zn fraction increases, whereas the proportion of nonspecifically adsorbed Zn decreases, indicating that soil OM can affect the binding states and stability of Zn (Fan et al., 2016). Elevated EC may alter the ion competition environment (Wang et al., 2023a), thereby affecting the desorption and migration of Zn, whereas pH directly influences Zn speciation and bioavailability. A study on the long-term evolution of urban soils in Nanjing revealed that Cu, Zn, and Pb exhibited continuous accumulation in areas with intensive urban commercial and industrial activities, indicating that the impact of anthropogenic activity intensity is reflected by indices such as CPI and VHMEC on Zn distribution (Zhang et al., 2005).

The influencing factors for Pb were ranked in descending order as CPI, Cu-Zn antagonism index, Zn-Cd antagonism index, solar radiation, and $PM_{2.5}$, all of which exhibited relatively high mean absolute SHAP values, followed by DIA, EC, and silt content. Among these, CPI contributed the most to Pb, indicating that the spatial distribution of Pb was primarily controlled by composite pollution. Considering the urban functional patterns of the six central districts of Nanjing, the old urban core areas, including the Gulou District, Qinhuai District, and the southwestern part of the Xuanwu District, are characterized by high traffic density, intensive commercial activity, and a long history of urban development. Road dust, historical transportation emissions, dust from demolition and renovation activities, and urban renewal-related

disturbances have resulted in the long-term accumulation of Pb. Pb exhibits strong retention and historical accumulation in urban soils; therefore, its high-value distribution often reflects not only the current emission intensity but also the legacy effects of previous transportation, industrial, and construction activities. The high importance of CZAI and ZCAI indicates that Pb does not migrate or accumulate in isolation, but rather is influenced by competitive adsorption, synergistic input, and geochemical partitioning processes under multi-metal coexistence conditions alongside other elements such as Cu, Zn, and Cd. In urban soils, Pb can undergo stable adsorption, complexation, or co-precipitation with organic matter, clay minerals, Fe-Mn oxides, and carbonate components. However, when metals such as Cu, Zn, and Cd coexist, different metal ions compete for limited adsorption sites, thereby affecting the immobilization, release, and spatial redistribution of Pb. Multi-metal competitive adsorption experiments have demonstrated that the adsorption selectivity of Pb in soils is generally higher than that of Cu, Zn, and Cd. Moreover, its competitive behavior is controlled by pH and cation exchange capacity (CEC). The contributions of solar radiation, $PM_{2.5}$, DIA, EC, and SSC further indicate that the spatial distribution of Pb was simultaneously regulated by atmospheric deposition, proximity to industrial activities, soil ion environment, and the carrier effect of fine particulate matter. $PM_{2.5}$ can serve as an integrated indicator of traffic emissions, industrial fumes and dust, and resuspended particulate matter. The spatial proximity to industrial activities represented by the DIA may reflect the effects of industrial legacies and localized emissions. EC reflects the ionic strength and salinity conditions of the soil

613 solution, which may alter the adsorption–desorption equilibrium of Pb. A higher SSC
614 indicates a higher proportion of fine soil particles, which facilitates the enrichment of
615 Pb in the fine particle fraction. Studies of surface soils across various urban functional
616 zones in Nanjing have shown more pronounced enrichment of Cu, Zn, and Pb in
617 industrial and traffic zones. Furthermore, the overall ecological risk ranking indicated
618 that industrial areas posed a higher ecological risk than transportation zones. These
619 findings are consistent with the relatively high contributions of CPI, $PM_{2.5}$, and DIA
620 observed in this study (Jiao et al., 2022)

621 The key driving factors for Cu were CZAI and CPI, followed by Mn, AP, OMCC,
622 and pHMI. The high contribution of CPI indicates that Cu was affected by the
623 regional combination contamination of heavy metals. Within the six central districts
624 of Nanjing, elevated Cu concentrations may be associated with traffic braking wear,
625 wear of motor vehicle components, metal processing and maintenance activities,
626 aging of construction materials, disturbances caused by urban renewal, and localized
627 industrial legacy effects. The importance of Mn, OMCC, pHMI, and AP indicates that
628 the environmental behavior of Cu is regulated by soil physicochemical conditions. Mn
629 oxides or Mn-related mineral components may promote the adsorption and
630 immobilization of Cu through surface complexation, co-precipitation, and colloidal
631 binding (Hem et al., 1989). OM-clay minerals can influence the migration and
632 retention behavior of Cu (Chotzen et al., 2016). AP may affect Cu availability through
633 phosphate precipitation, surface complexation, or alterations in the electrical
634 properties of soil colloids (Doelsch et al., 2024). pHMI reflects the regulation of Cu

635 mobility through changes in acid–alkaline conditions. In general, pH increases favor
636 the adsorption and immobilization of Cu, whereas soil acidification may increase its
637 release risks. Studies on Nanjing urban soils have demonstrated significant
638 correlations between Cu and Zn or Pb, as well as a coupled relationship between Cu
639 and soil properties such as organic carbon, pH, clay content, and CEC, indicating that
640 its spatial differentiation was driven by both anthropogenic inputs and constrained by
641 soil matrix conditions (Zhang et al., 2019). In addition, soil competition adsorption
642 and desorption experiments have demonstrated that Cu typically has a strong
643 adsorption and immobilization capacity in soils; however, its environmental behavior
644 is relatively sensitive to OM composition and pH changes. This explains the relatively
645 high contributions of OMCC and pHMI observed in the present study.

646 The most critical factor for Cd was CZAI, followed by the soil erosion–
647 precipitation interaction index, groundwater level, distance to the nearest residential
648 area, wind speed, ZCAI, and soil erodibility (K). Compared with Pb, Cu, and Zn, the
649 dominant factors for Cd distribution reflect hydrological transport, near-surface
650 disturbance, and local input, indicating that the spatial distribution of Cd was more
651 readily controlled by rainfall leaching, erosion transport, groundwater level
652 fluctuations, and localized pollution sources. The high contribution of CZAI indicates
653 a coexisting coupling relationship between Cd and elements such as Cu and Zn. This
654 relationship may be associated with the shared origin of multiple metals, for instance,
655 the simultaneous presence of various metals in mining legacy sites, industrial dust,
656 road dust, and externally introduced fill soils. However, this may also indicate that Cd

is strongly affected by competitive adsorption in multi-metal systems. When ions such as Cu and Zn occupy surface adsorption sites on soil colloids, OM, or oxides, the mobility and availability of Cd can increase. The relatively high contributions of SEPI, GWL, K, and WS further indicate that the spatial heterogeneity of Cd is closely related to hydrological processes and surface redistribution. The adsorption capacity of Cd in soils is generally weaker than that of Pb and Cu; however, Cd exhibits higher mobility and is readily affected by processes such as rainfall leaching, surface runoff, soil erosion, groundwater level fluctuations, and wind-induced resuspension and deposition. The southwestern part of Jianye District and the northwestern part of Yuhuatai District have experienced high levels of urban development and land-use intensification. External backfilling, construction-related dust, road dust, and construction disturbances may enhance the local input and redistribution of Cd. Historically abandoned mines are present in the northern Qixia District and around Qixia Mountain. Cd is frequently associated with Pb-Zn ore and tailings, which may represent an important source of localized high Cd concentrations in the environment. The contribution of DRA indicates that domestic emissions, road traffic, building maintenance, and urban greening management in and around residential areas may also contribute to Cd input; however, their impacts are more likely to manifest as localized disturbances rather than widespread and continuous pollution. Previous studies have indicated that Cd is one of the most mobile heavy metals in soil–groundwater systems, and its release and redistribution are significantly affected by pH, redox conditions, OM, groundwater fluctuations, and atmospheric deposition

679 (Kubier et al., 2019). Therefore, the high contributions of SEPI, GWL, and WS
680 observed in this study indicate that rainfall leaching, erosion transport, and
681 near-surface dispersion and deposition may be important processes underlying the
682 formation of spatial heterogeneity in Cd distribution (Liu et al., 2025). Furthermore,
683 DRA and NA500 predominantly reflect the effects of domestic emissions and
684 construction disturbance.

685 In summary, Zn and Pb reflect multi-metal pollution from common sources, Cu
686 is simultaneously influenced by exogenous inputs and soil physicochemical
687 conditions, and Cd is more sensitive to hydrological and disturbances. The spatial
688 distribution patterns and dominant controlling factors observed across these various
689 heavy metal differed markedly. The SHAP results thus provide a robust foundation for
690 identifying the primary drivers of these variations across the study area.

691

692 **Fig. 6.** SHAP-based interpretation of the optimal ML models for Zn (a), Pb (b), Cu (c),
693 and Cd (d).

694

695 3.3.2. Analysis of influencing factor interactions

696 To further elucidate the nonlinear coupling relationships among environmental
697 factors, two-dimensional partial dependence plots were used to analyze the joint
698 effects of the dominant variable combinations on the prediction.

699 For Zn, the interactions among the three variable groups CPI–CZAI, ZCAI–
700 pHMI, and OMCC–AP exhibited distinct nonlinear responses, indicating that the

predicted Zn concentration was affected by the combined effects of pollution load, multiple metal coexistence relationships, pH-dependent migration and immobilization conditions, and the OM–P relationship. The CPI–CZAI interaction interface showed that the predicted Zn concentration increased markedly when CPI and CZAI increased simultaneously (Fig. 7a). When CPI ranged from 45–55 and CZAI ranged from 3.0–4.4, the predicted Zn concentration was within the high-value range of 105.05–118.19 mg/kg. When CPI ranged from 28 to 36 and CZAI ranged from 1.4 to 2.2, the predicted Zn concentration was within the low-value range of 39.33–65.61 mg/kg. The ZCAI–pHMI interaction interface revealed that high Zn concentrations were mainly distributed in areas where both ZCAI and pHMI were elevated. When ZCAI ranged from 240 to 275 and pHMI ranged from 15.5 to 18.0, the predicted Zn concentration reached 98.23–110.09 mg/kg. When ZCAI ranged from 90 to 140 and pHMI ranged from 11.5 to 14.5, the predicted Zn concentration decreased to 38.93–62.65 mg/kg (Fig. 7b). Therefore, the combined influence of multi-metal coexistence and pH-dependent environment affects Zn retention. The OMCC–AP interaction interface showed that when OMCC ranged from 8 to 14 and AP was 20–35 mg/kg, the predicted Zn concentration reached 98.40–105.97 mg/kg. When OMCC < 4 or AP < 10 mg/kg, the predicted Zn concentration was within the lower range of 60.52–75.67 mg/kg (Fig. 7c). This interaction indicates that OM components can provide retention sites for Zn through complexation, adsorption, and colloidal binding, whereas an appropriate level of available P may further enhance Zn immobilization in the solid phase by participating in surface coordination, metal–phosphate binding, or by

723 altering the surface charge of soil colloids (He et al., 2022). Compared to the
724 independent effect of a single factor, the coupling between OMCC and AP provides a
725 better explanation for the localized accumulation of Zn under specific soil chemical
726 conditions. pH is an important factor controlling Zn adsorption and retention, whereas
727 the regulatory role of OM on Zn adsorption and migration is strongly pH dependent
728 (Pérez-Novo et al., 2011). In addition, P inputs can significantly influence Zn
729 adsorption capacity in acidic soils, thereby affecting its mobility and speciation (Cao
730 et al., 2009).

731 For Pb, the interactions of the three variable pairs CPI–ZCAI, EC–CZAI, and
732 $PM_{2.5}$–DIA exhibited pronounced nonlinear response characteristics. The CPI–ZCAI
733 interaction showed that the predicted Pb concentration generally increased with
734 increasing CPI, and the increase was more pronounced when the ZCAI values were
735 lower (Fig. 7d). When CPI > 45 and ZCAI < 120, the predicted Pb concentration
736 reached a relatively high level, indicating that Pb was more prone to accumulate under
737 conditions of relatively high composite pollution load and weak Zn-Cd antagonistic
738 constraints (Salazar et al., 2012). The EC–CZAI interaction plot showed that high Pb
739 concentrations (59.46–64.62 mg/kg) mainly occurred in areas with relatively high EC
740 (500–700 μS/cm) and low CZAI values (1.5–2.0) (Fig. 7e). However, when CZAI
741 exceeded 3.0, the predicted Pb concentration decreased overall, falling within the
742 range of 33.62–43.96 mg/kg. Elevated EC may alter the surface charge of soil colloids,
743 metal ion exchange processes, and adsorption–desorption equilibria, thereby
744 enhancing the local enrichment response of Pb. The $PM_{2.5}$-DIA interaction indicated

that Pb exhibited a pronounced nonlinear response to atmospheric particulate matter inputs and distance from industrial sources. When $PM_{2.5}$ ranged from 43 to 45 μg/m³ and DIA ranged from 1,000 to 3,500 m, the predicted Pb concentration was 47.40–53.31 mg/kg (Fig. 7f). In contrast, when $PM_{2.5}$ was below 40 μg/m³ and DIA ranged from 3,500 to 5,000 m, the predicted Pb concentration decreased to 17.82–29.65 mg/kg. Atmospheric particulate matter input may represent an important exogenous contribution to the enrichment of Pb, and the distance from industrial sources further regulates the local accumulation intensity of Pb by influencing pollutant dispersion and deposition conditions. Previous studies have shown that Pb can migrate among $PM_{2.5}$, street dust, and surface soils and is closely associated with urban anthropogenic emissions (Hu et al., 2014).

For Cu, the CPI–CZAI, AK–Mn, and AP–OMCC interactions exhibited pronounced nonlinear response characteristics. The CPI–CZAI interaction plot showed that the predicted Cu concentration generally increased with increasing CPI, with a more pronounced increase under low-to-moderate CZAI conditions. When the CPI ranged from 45 to 55 and the CZAI ranged from 1.3 to 2.7, the predicted Cu concentration (36.32–38.89 mg/kg) was relatively high. However, when CPI was <35 and CZAI was >3.0, the predicted Cu concentration (23.48–28.62 mg/kg) was relatively low (Fig. 7g). This result indicates that a high combined pollution load is an important factor contributing to the high Cu response, whereas CZAI may further regulate the accumulation intensity of Cu under multi-metal coexistence. The AK–Mn interaction plot showed that Cu exhibited nonlinear responses to soil nutrients and Mn

767 oxides (Fig. 7h). AK reflects soil nutrient supply and anthropogenic activities, such as
768 fertilization and landscape management. Its interaction with Mn indicates that the
769 spatial distribution of Cu was influenced not only by pollution inputs but also by the
770 soil adsorption matrix conditions. The AP–OMCC interaction further revealed that
771 when OMCC ranged from 12 to 16 and AP ranged from 15 to 30 mg/kg, the predicted
772 Cu concentration was within the higher range of 35.42–37.37 mg/kg (Fig. 7i). Cu is a
773 typical organophilic element that readily forms stable complexes with humic
774 substances and organic functional groups, such as carboxyl and phenolic hydroxyl
775 groups. AP may indirectly regulate Cu immobilization and migration by altering the
776 electrical properties of soil colloids, promoting metal–phosphate precipitation, or
777 modifying the rhizosphere chemical environment.

778 For Cd, the variable group ZCAI–CZAI interaction showed that when ZCAI
779 ranged from 75 to 130 and CZAI ranged from 3.3 to 4.5, the predicted Cd
780 concentration was 0.63–0.78 mg/kg. However, when ZCAI increased to 170–275 and
781 CZAI ranged from 1.5 to 3.0, the predicted Cd concentration decreased to 0.34–0.49
782 mg/kg (Fig. 7j). This indicates that Cd exhibited a sensitive nonlinear response to
783 changes in the multi-metal coexistence relationships. Zn and Cu may alter the
784 immobilization capacity and mobility of Cd by competing for limited adsorption sites
785 on soil OM, clay minerals, and Fe-Mn oxides, thereby affecting Cd accumulation.
786 When competitive adsorption alters the distribution of metals on soil colloid surfaces,
787 Cd is more prone to release, migration, or localized re-enrichment. The Mn-$PM_{2.5}$
788 interaction interface indicated that Cd exhibited a pronounced interactive response to

atmospheric particulate matter input and Mn-related adsorption media. When $PM_{2.5}$ ranged from 42 to 45 μg/m³ and Mn ranged from 600 to 850 mg/kg, the predicted Cd concentration was 0.60–0.71 mg/kg. When the $PM_{2.5}$ ranged from 38.5 to 40.5 μg/m³ and Mn ranged from 400 to 600 mg/kg, the predicted Cd concentration decreased to 0.38–0.49 mg/kg (Fig. 7k). This indicates that Cd inputs may be associated with the deposition of fine particulate matter, road dust resuspension, and industrial dust deposition. Mn oxides can serve as important adsorption carriers for Cd, influencing its retention and redistribution in topsoil. The K–DWS interaction interface showed that high Cd concentrations mainly occurred in areas close to water systems and with moderate K levels. When DWS was <500 m and K ranged from 0.0365 to 0.0385 $(t \cdot hm^2 \cdot h)/(hm^2 \cdot MJ \cdot mm)$, the predicted Cd concentration was 0.58–0.66 mg/kg. However, when DWS > 1,300 m and K > 0.039 $(t \cdot hm^2 \cdot h)/(hm^2 \cdot MJ \cdot mm)$, the predicted Cd concentration decreased to 0.42–0.50 mg/kg (Fig. 7l). This result indicates that areas close to water systems or experiencing strong hydrological disturbances may influence the spatial distribution of Cd through surface runoff, erosion, and transport, sediment redistribution, and groundwater level fluctuations. Compared to Pb and Cu, Cd has higher mobility and is more prone to migration, transformation, and local enrichment under conditions of rainfall leaching, soil erosion, and increased hydrological connectivity. Multi-process modeling studies have demonstrated that regional Cd accumulation is primarily driven by the combined effects of atmospheric deposition and flooding/hydrological inputs (Song et al., 2026), which is consistent with the interaction patterns identified in this study.

The interactive response characteristics of the four heavy metals indicate that the spatial distribution of soil heavy metals in the study area is driven by the coupling of three mechanisms: exogenous pollution inputs, soil matrix regulation, and intermetal competition/antagonism. Specifically, CPI, $PM_{2.5}$, and DIA characterize the combined pollution load, atmospheric particulate matter deposition, and proximity to industrial sources, serving as important indicators for the intensity of exogenous pollution inputs. EC, AP, OMCC, Mn, K, and DWS regulate the immobilization, release, and migration of heavy metals by modifying soil ion strength, adsorption carriers, organic complexation, erosion susceptibility, and hydrological redistribution processes. CZAI and ZCAI reflect competitive adsorption, co-input, or antagonistic effects under conditions of multiple metal coexistence. The dominant interaction pattern associated with different elements exhibited significant variations, indicating distinct environmental response processes for various heavy metals in soil. This also suggests that the prevention and control of soil heavy metal pollution in the six central districts of Nanjing should be implemented through zoned and classified management strategies according to the elemental sources, migration capacities, and soil matrix conditions.

Fig. 7. Two-dimensional response surfaces of pairwise interactions between key influencing factors for Zn (a, b, c), Pb (d, e, f), Cu (g, h, i), and Cd (j, k, l) prediction.

3.4. SEM-based interpretation of driving pathways for heavy metal spatial

differentiation

Based on the SHAP importance rankings and bivariate interaction analysis, SEM was constructed for Pb, Cu, Cd, and Zn to validate the reliability of the ML-based interpretation results described above. The models further distinguished the direction and intensity of the direct effects of individual factors on the target elements and covariation relationships among variables, thereby revealing the differentiated driving mechanisms underlying the distribution of heavy metals in Nanjing soils.

Overall, all four SEMs exhibited high explanatory power (Fig. 8). Based on the variance explained values (R^2) for Pb, Cu, Cd, and Zn, the models explained 82.6, 76.8, 84.1, and 96.3% of their spatial variation, respectively. Among them, the Zn model had the highest explanatory power, suggesting that there was the most robust structural relationship between its spatial distribution and the selected environmental factors. Cd and Pb ranked second, suggesting that although both exhibited strong local heterogeneity, their dominant controlling pathways could still be effectively captured by the models. The Cu model had relatively low explanatory power indicating that the spatial variation of Cu was simultaneously influenced by exogenous inputs and soil matrix conditions, and that certain additional influencing factors may not have been considered.

The path analysis results from the Zn SEM showed that CPI, CZAI, and ZCAI all exerted significant positive effects on Zn, with path coefficients of 0.502 ($p < 0.001$), 0.493 ($p < 0.001$), and 0.203 ($p < 0.001$), respectively, indicating that the spatial enrichment of Zn was simultaneously controlled by the combined pollution

and competitive or antagonistic relationships among multiple metals. SBCI exerted the strongest positive effect on Zn, with a path coefficient of 1.017 ($p < 0.01$), indicating that enhanced soil buffering capacity significantly promoted the retention or accumulation of Zn in the soil. COCI exerted a significant negative effect on Zn, with a path coefficient of −0.959 ($p < 0.01$). An increase in COCI may lead to a more stable soil structure, greater abundance of colloidal adsorption sites, or relatively weaker anthropogenic disturbances, thereby reducing the anomalous migration of Zn and the formation of high-concentration patches (Tang et al., 2024). Both VHMEC and EC exhibited significant positive effects on Zn, with path coefficients of 0.056 ($p < 0.05$) and 0.052 ($p < 0.01$), respectively. OMCC exerted a significant negative effect on Zn, with a path coefficient of −0.088 ($p < 0.05$). The direct pathways of AP and pHMI did not reach statistical significance, indicating that their effects on Zn may be exerted indirectly through interactions with OMCC, EC, or metal antagonism variables. P inputs can alter the adsorption and desorption kinetics of Zn in acidic soils. The covariance pathways further showed a significant positive correlation between CPI and CZAI (r = 0.462, $p < 0.001$), a significant positive correlation between AP and OMCC (r = 0.327, $p < 0.001$), and a significant negative correlation between ZCAI and OMCC (r = −0.231, $p < 0.001$). These results indicate that the spatial distribution of Zn was not determined by a single pathway but rather resulted from a strongly coupled relationship among combined pollution, multi-metal competition, OM complexation capacity, and P levels.

The SEM for Pb showed that its spatial distribution was mainly controlled by the

combined influence of composite pollution and interactions among various metals. CPI exerted a significant positive effect on Pb, with a standardized path coefficient of 0.883 ($p < 0.001$), representing the strongest positive direct path in the Pb model. This result is consistent with the conclusion from the SHAP analysis, in which CPI made the greatest contribution to Pb, indicating that the spatial accumulation of Pb exhibited a synergistic enrichment pattern under enhanced multi-metal composite pollution. CZAI and ZCAI both exerted significant negative effects on Pb, with standardized path coefficients of −0.487 ($p < 0.001$) and −0.386 ($p < 0.001$), respectively, indicating that metal antagonism or competition associated with Cu–Zn and Zn–Cd significantly inhibited the accumulation intensity of Pb. These three elements are common divalent metal ions that exhibit competitive adsorption behavior on soil colloids, OM, carbonates, and Fe-Mn oxides. Pb typically has a higher partition coefficient and stronger adsorption affinity; however, its adsorption behavior is still affected by the competitive effects of coexisting metal ions such as Cd, Cu, and Zn (Bradl, 2004). $PM_{2.5}$ exerted a significant positive effect on Pb, with a path coefficient of 0.080 ($p < 0.05$), indicating that atmospheric particulate matter deposition contributed to the accumulation of Pb in the surface soils. DIA exerted a significant negative effect on Pb, with a path coefficient of −0.094 ($p < 0.01$), indicating that the closer the location to industrial sources, the more readily Pb accumulated. This result corroborates the conclusion from the bivariate interaction analysis that higher predicted Pb concentrations occurred under conditions of high $PM_{2.5}$ and low DIA values. EC, SSC, and SR exerted significant negative effects on

Pb, with path coefficients of −0.108 ($p < 0.001$), −0.110 ($p < 0.01$), and −0.096 ($p < 0.01$), respectively. These results indicate that soil electrochemical conditions, particle size composition, and surface thermal or radiation environment may influence the adsorption, migration, and immobilization of Pb. Covariance pathway analysis demonstrated a significant positive correlation between $PM_{2.5}$ and DIA ($r = 0.280$, $p < 0.001$) and a positive correlation between EC and CZAI ($r = 0.147$, $p < 0.05$), indicating that exogenous inputs of particulate matter, proximity to industrial sources, and metal competition were not entirely independent but collectively constituted a composite driving system for the spatial differentiation of Pb.

Path analysis results from the Cu SEM showed that CPI exerted a significant positive effect on Cu with a path coefficient of 0.791 ($p < 0.001$). The higher the combined pollution load, the greater the likelihood that Cu will be enriched by the coexistence of multiple pollution factors, common source inputs, and synergistic accumulation. CZAI exerted a significant negative effect on Cu, with a path coefficient of −0.753 ($p < 0.001$), and the magnitude of this effect was comparable to that of the CPI. Competitive adsorption–desorption studies have demonstrated that under acidic soil conditions, Cu has a higher adsorption affinity than Zn for active sites on soil colloids. Furthermore, the coexistence of Cu^{2+} significantly inhibits the adsorption of Zn^{2+}, indicating a pronounced competitive relationship between the two metals for adsorption sites in the soil solid phase (Covelo et al., 2007). AP, OMCC, and Mn had a significant positive effect on Cu, with path coefficients of 0.140 ($p < 0.001$), 0.103 ($p < 0.01$), and 0.079 ($p < 0.05$), respectively. Although the direct

effects of pHMI and AK on Cu did not reach statistical significance, they may indirectly influence Cu speciation through their synergistic interactions with OMCC, AP, or Mn. Covariance results showed a significant positive correlation between CPI and CZAI ($r = 0.462$, $p < 0.001$), as well as a significant positive correlation between OMCC and AP ($r = 0.327$, $p < 0.001$), indicating that the spatial accumulation of Cu was closely associated with the composite pollution load and Cu-Zn coexistence relationship. Conversely, the OM complexation environment and P level exhibited a synergistic variation pattern. Based on the SHAP analysis, these results further indicate that the spatial enrichment of Cu is more likely to occur in areas characterized by strong composite pollution inputs, pronounced Cu–Zn competition relationship, and significant OM complexation, P effects, and adsorption by Mn oxides.

The path analysis results from the Cd SEM showed that ZCAI exerted a significant negative effect, with a path coefficient of −1.063 ($p < 0.001$), whereas CZAI exerted a significant positive effect, with a path coefficient of 0.499 ($p < 0.001$). Furthermore, a significant positive correlation was observed between ZCAI and CZAI scores ($r = 0.683$, $p < 0.001$). These results indicate that the spatial accumulation of Cd depends on the competitive and antagonistic effects of Cu and Zn. Mn had a significant positive effect, with a path coefficient of 0.167 ($p < 0.001$). Mn oxides possess a large specific surface area and strong surface activity, enabling them to regulate Cd speciation in soils via adsorption, co-precipitation, and redox processes. DWS exerted a significant negative influence, with a path coefficient of −0.059 ($p <$

943 0.05). Areas near water systems may be more susceptible to the impacts of surface
944 runoff, leaching, and sediment redistribution. Cd has high mobility and bioavailability,
945 and its migration in soil-water environments is influenced by a combination of pH,
946 redox conditions, organic complexation, and groundwater chemistry. Consequently, it
947 responds more readily to changes in external hydrological and chemical conditions
948 than Pb, Cu, and Zn (Zhao et al., 2015). The direct pathways of GWL, SEPI, K, WS,
949 DRA, and $PM_{2.5}$ did not reach statistical significance; however, they exerted indirect
950 effects. Consequently, within the six central districts of Nanjing, the spatial
951 differentiation of Cd is dominated by multi-metal competition and is jointly regulated
952 by immobilization by Mn oxides and the migration and redistribution processes
953 associated with water systems.
954
955 **Fig. 8.** Direct, indirect, and covariance relationships among key influencing factors
956 controlling Zn (a), Pb (b), Cu (c), and Cd (d) concentrations in urban soils using SEM.
957
958 **4. Conclusions and perspective**
959 The optimal models for Cd, Cu, Pb, and Zn were HistGradientBoosting, SVM,
960 MLP, and SVR, respectively. Cd exhibited the strongest positive spatial
961 autocorrelation, and its spatial pattern was characterized by both local anomalous
962 enrichment and contiguous low-value areas. Both Pb and Cu exhibited a certain
963 degree of positive spatial autocorrelation, and the HH clusters were located in the core
964 area of the old district and in areas with dense transportation and construction

activities. Zn exhibited the weakest spatial autocorrelation, and the HH and LL clusters were scattered and interlaced, respectively. ML–OK placed greater emphasis on continuous variation trends at the regional scale, whereas FRG-ML–OK was more sensitive to local high-value patches, boundary transitions, and small-scale spatial heterogeneity. Pb was primarily driven by combined pollution and atmospheric deposition. Cu was jointly influenced by anthropogenic inputs and the adsorption and immobilization of OM and minerals in the sediment. Cd was governed by multi-metal competition and migration effects associated with the hydrological pathways. Zn was collectively dominated by the combined pollution load, soil buffering capacity, and multi-metal coupling.

In the future, higher-resolution remote sensing imagery, street view data, detailed land use data, and road traffic intensity data can be incorporated with soil profile sampling and long-term monitoring, to improve the characterization of spatial differentiation, vertical migration, and temporal variation of heavy metals. It is essential to develop a modeling framework that integrates predictive accuracy, spatial extrapolation capability, and mechanistic interpretability. Strengthening the coupling of ML, geostatistical, and environmental process models should also be prioritized to enhance the reliability of spatial distribution prediction and attribution analysis for soil heavy metals within complex urban environments. Finally, based on the high-value areas and key driving factors identified in this study, future research should integrate isotope tracing, sequential extraction, mineralogical analysis, and microscale characterization to further verify the source contributions, speciation, and

migration and transformation mechanisms of heavy metals. This approach will advance soil heavy metal research from the identification of spatial patterns toward a deeper elucidation of process-based mechanisms and the implementation of precise risk management strategies.

CRediT authorship contribution statement

Hanghang Yue: Writing–original draft, Methodology, Formal analysis, Data curation. **Taoran Shi:** Writing–review & editing, Methodology, Supervision, Funding acquisition. **Chunfa wu:** Validation, Software. **Hong Wang:** Investigation, Funding acquisition, Data curation, Resources. **Jing Shao:** Investigation, Data curation. **Renying Li：** Methodology, Investigation, Resources.

Funding

We gratefully acknowledge financial support from the support received by Taoran Shi from the Natural Science Foundation of China (No. 42207438).

Appendix A. Supplementary data

Supplementary data to this article can be found online at

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare no competing interests.

References

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93-115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.

Blok, J. (2005). Environmental exposure of road borders to zinc. *Science of the Total Environment*, *348*(1-3), 173-190. DOI: https://doi.org/10.1016/j.scitotenv.2004.12.073.

Bradl, H. B. (2004). Adsorption of heavy metal ions on soils and soil constituents. *Journal of Colloid and Interface Science*, *277*(1), 1-18. https://doi.org/10.1016/j.jcis.2004.04.005.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324.

Cao, X., Wahbi, A., Ma, L., Li, B., & Yang, Y. (2009). Immobilization of Zn, Cu, and Pb in contaminated soils using phosphate rock and phosphoric acid. *Journal of Hazardous Materials*, *164*(2-3), 555-564. https://doi.org/10.1016/j.jhazmat.2008.08.034.

Chotzen, R. A., Polubesova. T., Chefetz, B., & Mishael, Y. G. (2016). Adsorption of soil-derived humic acid by seven clay minerals: A systematic study. *Clays and Clay Minerals*, *64*(5), 628-638. https://doi.org/10.1346/CCMN.2016.064027.

Covelo, E. F., Vega, F. A., & Andrade, M. L., 2007. Competitive sorption and

desorption of heavy metals by individual soil components. *Journal of Hazardous Materials*, *140*(1-2), 308-315. https://doi.org/10.1016/j.jhazmat.2006.09.018.

Doelsch, E., Le Bars, M. L., Etschmann B., Formentini, T., Legros, S., Levard, C., Chaurand, P., Basile-Doelsch, I., Rose, J., & Brunetti, G. (2024). Bulk and mapping speciation analyses unveil the pattern and heterogeneity of Cu species during organic waste treatment. *Environmental Science & Technology, 58*(32), 14439-14449. https://doi.org/10.1021/acs.est.4c02887.

Fan, T. T., Wang, Y. J., Li, C. B., He, J. Z., Gao, J., Zhou, D. M., Friedman, S. P., & Sparks, D. L. (2016). Effect of organic matter on sorption of Zn on soil: Elucidation by Wien effect measurements and EXAFS spectroscopy. *Environmental Science & Technology, 50*(6), 2931-2937. DOI: https://doi.org/10.1021/acs.est.5b05281.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, *29*(5), 1189-1232. DOI: https://doi.org/10.1214/aos/1013203451.

Han, Z. Y., Wang, J. Y., Liao, X. Y., & Yang, J. (2025). Accurate prediction of spatial distribution of soil heavy metal in complex mining terrain using an improved machine learning method. *Journal of Hazardous Materials*, *491*, 137994. DOI: https://doi.org/10.1016/j.jhazmat.2025.137994.

He, N., Hu, L., Jiang, C. Y. Z., & Li, M. K. (2022). Remediation of chromium, zinc, arsenic, lead and antimony contaminated acidic mine soil based on *Phanerochaete chrysosporium* induced phosphate precipitation. *Science of the*

Total Environment, *850*, 157995. https://doi.org/10.1016/j.scitotenv.2022.157995.

Hem, J. D., Lind, C. J., & Roberson, C. E. (1989). Coprecipitation and redox reactions of manganese oxides with copper and nickel. *Geochimica et Cosmochimica Acta*, *53*(11), 2811-2822. https://doi.org/10.1016/0016-7037(89)90159-2.

Hu, X., Sun, Y. Y., Ding, Z. H., Zhang, Y., Wu, J. C., Lian, H. Z., & Wang, T. J. (2014). Lead contamination and transfer in urban environmental compartments analyzed by lead levels and isotopic compositions. *Environmental Pollution*, *187*, 42-48. https://doi.org/10.1016/j.envpol.2013.12.025.

Huang, C. S., Zhang, L. J., Meng, J. L., Yu, Y. J., Qi, J. Y., Shen, P., Li, X., Ding, P., Chen, M. B., & Hu, G. C. (2022a). Characteristics, source apportionment and health risk assessment of heavy metals in urban road dust of the Pearl River Delta, South China. *Ecotoxicology and Environmental Safety*, *236*, 113490. https://doi.org/10.1016/j.ecoenv.2022.113490.

Huang, G. X., Wang, X. H., Chen, D., Wang, Y. P., Zhu, S. X., Zhang, T., Liao, L., Tian, Z., & Wei, N. (2022b). A hybrid data-driven framework for diagnosing contributing factors for soil heavy metal contaminations using machine learning and spatial clustering analysis. *Journal of Hazardous Materials*, *437*, 129324. https://doi.org/10.1016/j.jhazmat.2022.129324.

Jeong, H, Ryu, J. S., & Ra, K. (2022). Characteristics of potentially toxic elements and multi-isotope signatures (Cu, Zn, Pb) in non-exhaust traffic emission sources. *Environmental Pollution*, *292*, 118339.

https://doi.org/10.1016/j.envpol.2021.118339.

Jiao, S., Chen, Z. H., Yu, A. H, & Chen, H. H. (2022). Evaluation of the heavy metal pollution ecological risk in topsoil: a case study from Nanjing, China. *Environmental Earth Sciences*, *81*, 532. https://doi.org/10.1007/s12665-022-10659-4.

Kubier, A., Wilkin, R.T., & Pichler, T. (2019). Cadmium in soils and groundwater: a review. *Applied Geochemistry*, *108*, 104388. https://doi.org/10.1016/j.apgeochem.2019.104388.

Li, K., Guo, G. H., Zhang, D. G., Lei, M., & Wang, Y. Y. (2024a). Accurate prediction of spatial distribution of soil potentially toxic elements using machine learning and associated key influencing factors identification: a case study in mining and smelting area in southwestern China. *Journal of Hazardous Materials*, *478*, 135454. https://doi.org/10.1016/j.jhazmat.2024.135454.

Li, L. L., Wu, B., Guo, S. H., Hu, E. Z., Zhang, Y. L., Sun, L. X., & Li, S. Q. (2024b). Multipath diffusion process and spatial accumulation simulation of Cd in lead-zinc mining areas. *Journal of Hazardous Materials*, *465*, 133461. https://doi.org/10.1016/j.jhazmat.2024.133461.

Li, W. F., Jiang, Y. F., Ye, Y. C., Guo, X., & Shi, Z. (2024c). Spatiotemporal interpretable mapping framework for soil heavy metals. *Journal of Cleaner Production*, 468, 143101. https://doi.org/10.1016/j.jclepro.2024.143101.

Liu, E.F., Yan, T., Birch, G., & Zhu, Y. X. (2014). Pollution and health risk of potentially toxic metals in urban road dust in Nanjing, a mega-city of China.

Science of the Total Environment, *476-477*, 522-531. https://doi.org/10.1016/j.scitotenv.2014.01.055.

Liu, N., Wang, X., Chen, M., & Pan, Y. X. (2025). Migration and transformation of Pb and Cd in the unsaturated zone induced by water table fluctuations. *Journal of Hazardous Materials*, *492*, 138188. https://doi.org/10.1016/j.jhazmat.2025.138188.

Long, Z., Zhu, H., Bing, H. J., Tian, X., Wang, Z. G., Wang, X. F., & Wu, Y. H. (2021). Contamination, sources and health risk of heavy metals in soil and dust from different functional areas in an industrial city of Panzhihua City, Southwest China. *Journal of Hazardous Materials*, *420*, 126638. https://doi.org/10.1016/j.jhazmat.2021.126638.

Lundberg, S. M., & Lee, S. I., (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30. Curran Associates, Inc., 4768-4777. https://doi.org/10.48550/arXiv.1705.07874.

Ma, X. D., Guan, D. X., Zhang, C. S., Yu, T., Li, C., Wu, Z. L., Li, B., Geng, W. D., Wu, T. S., & Yang, Z. F. (2024). Improved mapping of heavy metals in agricultural soils using machine learning augmented with spatial regionalization indices. *Journal of Hazardous Materials*, *478*, 135407. https://doi.org/10.1016/j.jhazmat.2024.135407.

Meyer, H., & Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, *12*, 1620-1633. https://doi.org/10.1111/2041-210X.13650.

1119 Nanjing Municipal Bureau Statistucs. Nanjing Statistical Yearbook in 2024.
1120 https://tjj.nanjing.gov.cn/material/njnj_2024/.
1121 Pamuru, S.T., Forgione, E., Croft, K., Kjellerup, B. V., & Davis, A. P. (2022).
1122 Chemical characterization of urban stormwater: Traditional and emerging
1123 contaminants. *Science of the Total Environment*, *813*, 151887. DOI:
1124 https://doi.org/10.1016/j.scitotenv.2021.151887.
1125 Pan, Y. J., Sha, A. M., Han, W. J., Liu, C., Liu, G. W. C., Welsch, E., Zeng, M., Xu, S.
1126 S., Zhao, Y., Tian, S., Li, Y. Y., Deng, R., Zhang, X., Shi, H. H., Cui, Y., Huang,
1127 C. S., & Peng, H. X. (2025). Identifying spatial drivers of soil heavy metal
1128 pollution risk integrating positive matrix factorization, machine learning, and
1129 multi-scale geographically weighted regression. *Journal of Hazardous Materials*,
1130 *485*, 136841. DOI: https://doi.org/10.1016/j.jhazmat.2024.136841.
1131 Pérez-novo C., Bermúdez-couso A., López-periago E., Fernández-Calviño, D., &
1132 Arias-Estévez, M. (2011). Zinc adsorption in acid soils: Influence of phosphate.
1133 *Geoderma*, *162*(3-4), 358-364. DOI:
1134 https://doi.org/10.1016/j.geoderma.2011.03.008.
1135 Salazar, M. J., Rodriguez, J. H., Nieto, G. L., & Pignata, M. L. (2012). Effects of
1136 heavy metal concentrations (Cd, Zn and Pb) in agricultural soils near different
1137 emission sources on quality, accumulation and food safety in soybean [*Glycine*
1138 *max* (L.) Merrill]. *Journal of Hazardous Materials*, *233-234*, 244-253.
1139 https://doi.org/10.1016/j.jhazmat.2012.07.026.
1140 Shao, F. L., Li, K. N., Ouyang, D., Zhou, J. W., Luo, Y. T., & Zhang, H. B. (2024).

Sources apportionments of heavy metal(loid)s in farmland soils close to industrial parks: Integrated application of positive matrix factorization (PMF) and cadmium isotopic fractionation. *Science of the Total Environment*, *924*, 171598. https://doi.org/10.1016/j.scitotenv.2024.171598.

Shen, C., Huang, S. F., Wang, M., Wu, J., Su, J. H., Lin, K. F., Chen, X. R., He, T. H., Li, Y., Sha, C. Y., & Liu, M. (2024). Source-oriented health risk assessment and priority control factor analysis of heavy metals in urban soil of Shanghai. *Journal of Hazardous Materials*, *480*, 135859. https://doi.org/10.1016/j.jhazmat.2024.135859.

Sheng, Y., Wang, Z. Q., & Feng, X., (2023). Potential ecological risk and zoning control strategies for heavy metals in soils surrounding core water sources: A case study from Danjiangkou Reservoir, China. *Ecotoxicology and Environmental Safety*, *252*, 114610. https://doi.org/10.1016/j.ecoenv.2023.114610.

Shi, S. W., Hou, M. Y., Gu, Z. F., Jiang, C., Zhang, W. Q., Hou, M. Y., Li, C. X., & Xi, Z. L. (2022). Estimation of heavy metal content in soil based on machine learning models. *Land*, *11*(7), 1037. https://doi.org/10.3390/land11071037.

Smola, A.J., & Schölkope, B., 2004. A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.

Song, Y. T., Wang, M. E., Xie, T., Wang, Y. H., Chen, X. Y., & Chen, W. P. (2026). Atmospheric deposition and flood inputs dominate paddy soil cadmium

accumulation in Southern China: a multi-process model coupling input and output pathways. *Journal of Hazardous Materials*, *502*, 140953. https://doi.org/10.1016/j.jhazmat.2025.140953.

Sun, Y. Q., Lei, S. G., Zhao, Y. B., Wei, C., Yang, X. C., Han, X. T., Li, Y. Y., Xia, J. N., & Cai, Z. (2024). Spatial distribution prediction of soil heavy metals based on sparse sampling and multi-source environmental data. *Journal of Hazardous Materials*, *465*, 133114. https://doi.org/10.1016/j.jhazmat.2023.133114.

Tang, L., Liu, J., Xiang, C., Gao, W. Y., Chen, Z. S., Jiang, J., Guo, J. K., & Xue, S. G. (2024). Colloid mobilization and transport in response to freeze-thaw cycles: insights into the heavy metal(loid)s migration at a smelting site. *Journal of Hazardous Materials*, *480*, 135959. https://doi.org/10.1016/j.jhazmat.2024.135959.

Tipping, E., Lofts, S., & Sonke, J. E., 2011. Humic ion-binding model VII: a revised parameterisation of cation-binding by humic substances. *Environmental Chemistry*, *8*(3), 225-235. https://doi.org/10.1071/EN11016.

Wang, H. J., Zhao, M. L., Huang, X. M., Song, X. Y., Cai, B. Y., Tang, R., Sun, J. X., Han, Z. L., Yang, J., Liu, Y. F., & Fan, Z. Q. (2024). Improving prediction of soil heavy metal(loid) concentration by developing a combined co-kriging and geographically and temporally weighted regression (GTWR) model. *Journal of Hazardous Materials*, *468*, 133745. https://doi.org/10.1016/j.jhazmat.2024.133745.

Wang, R., Liu, T., Lu, C. Y., Zhang, Z. C., Guo, P. R., Jia, B. B., Hao, B. H., Wang, Y.

C., & Guo, W., (2023a). Bioorganic fertilizers improve the adaptability and remediation efficiency of *Puccinellia distans* in multiple heavy metals-contaminated saline soil by regulating the soil microbial community. *Journal of Hazardous Materials*, *448*, 130982. https://doi.org/10.1016/j.jhazmat.2023.130982.

Wang, X. Y., Liu, E., Yan, M. X., Zheng, S. W., Fan, Y., Sun, Y. X., Li, Z. J., & Xu, J. L. (2023b). Contamination and source apportionment of metals in urban road dust (Jinan, China) integrating the enrichment factor, receptor models (FA-NNC and PMF), local Moran's index, Pb isotopes and source-oriented health risk. *Science of the Total Environment*, *878*, 163211. https://doi.org/10.1016/j.scitotenv.2023.163211.

Wu, J., Edwards, R., He, X. E., Liu, Z., & Kleinman, M. (2010). Spatial analysis of bioavailable soil lead concentrations in Los Angeles, California. *Environmental Research*, *110*(4), 309-317. https://doi.org/10.1016/j.envres.2010.02.004.

Yadav K., & Raman R. S., (2021). Size-segregated chemical source profiles and potential health impacts of multiple sources of fugitive dust in and around Bhopal, central India. *Environmental Pollution*, *284*, 117385. https://doi.org/10.1016/j.envpol.2021.117385.

Yan, Y.B., & Yang, Y. (2025). Revealing the synergistic spatial effects in soil heavy metal pollution with explainable machine learning models. *Journal of Hazardous Materials*, *482*, 136578. https://doi.org/10.1016/j.jhazmat.2024.136578.

Yang, S., Liu, J., Bi, X., Ning, Y. Q., Qiao, S. Y., Yu, Q. Q., & Zhang, J. (2020). Risks

related to heavy metal pollution in urban construction dust fall of fast-developing Chinese cities. *Ecotoxicology and Environmental Safety*, *197*, 110628. https://doi.org/10.1016/j.ecoenv.2020.110628.

Yu, H. J., Li, C. C., Yan, J., Ma, Y. Q., Zhou, X. Y., Yu, W. J., Kan, H. Y, Meng, Q., Xie, R. S., & Dong, P. (2023). A review on adsorption characteristics and influencing mechanism of heavy metals in farmland soil. *Reviews in RSC Advances*, *13*, 3505-3519. https://doi.org/10.1039/D2RA07095B.

Zeng, W. B., Wan, X. M., Gu, G. Q., Lei, M., Yang, J. & Chen, T. B. (2023). An interpolation method incorporating the pollution diffusion characteristics for soil heavy metals: taking a coke plant as an example. *Science of The Total Environment*, *857*, 159698. https://doi.org/10.1016/j.scitotenv.2022.159698.

Zhang G.L., Yang, F.G., Zhao Y.G., Zhao, W. J., Yang, J. L., & Gong, Z.T. (2005). Historical change of heavy metals in urban soils of Nanjing, China during the past 20 centuries. *Environment International*, *31*(6), 913-919. https://doi.org/10.1016/j.envint.2005.05.035.

Zhang, J.Y., Liu, X.D., Pang, S.D., & Yu, Y. C. (2018). Pollution assessment of heavy metals in urban green land soils in Nanjing City based on the matter-element extension model. *Research of Environmental Sciences*, *31*(9): 1572-1579. (in Chinese) https://doi.org/10.13198/j.issn.1001-6929.2018.03.39.

Zhang, J. Y., Yu, F., Pang, S. D., & Yu, Y. C. (2019). Spatial distribution and pollution assessment of potentially toxic elements in urban forest soil of Nanjing, China. *Polish Journal of Environmental Studies*, *28*(4), 3015-3024.

https://doi.org/10.15244/pjoes/92812.

Zhao. F. J., Ma, Y. B., Zhu, Y. G., Tang, Z., & McGrath, S. P. (2015). Soil contamination in China: current status and mitigation strategies. *Environmental Science & Technology*, *49*(2), 750-759. https://doi.org/10.1021/es5047099.

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

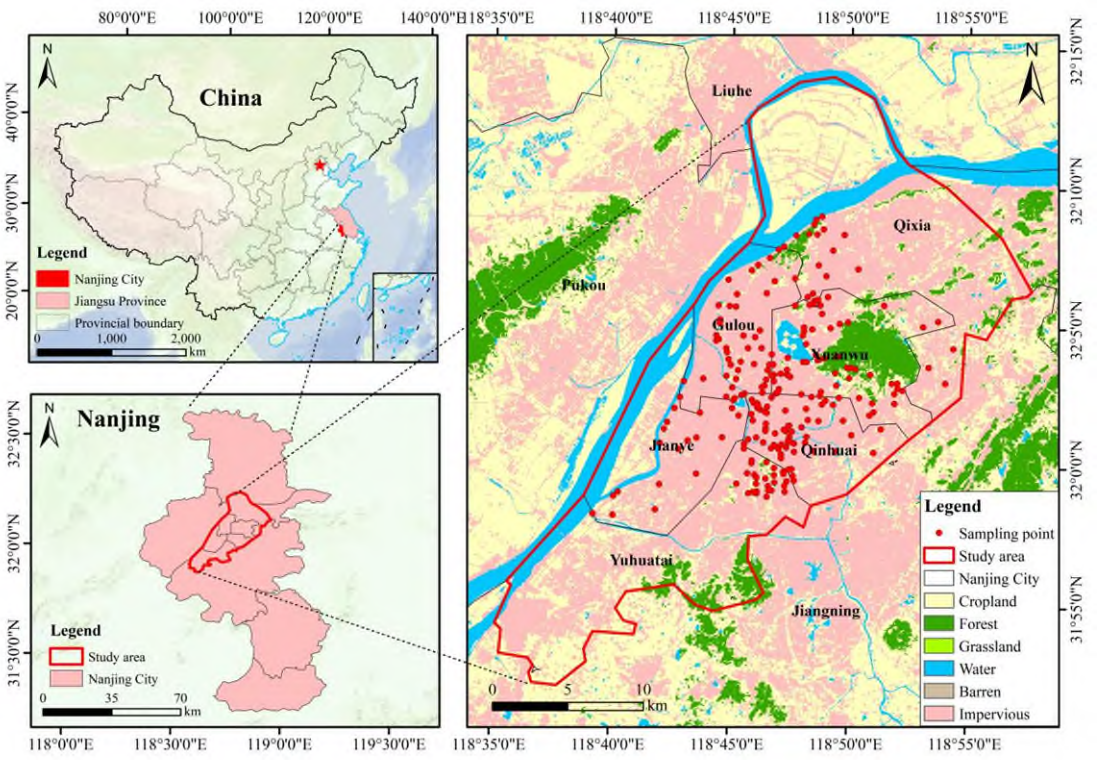


Fig. 1. Study areas and sampling site locations.

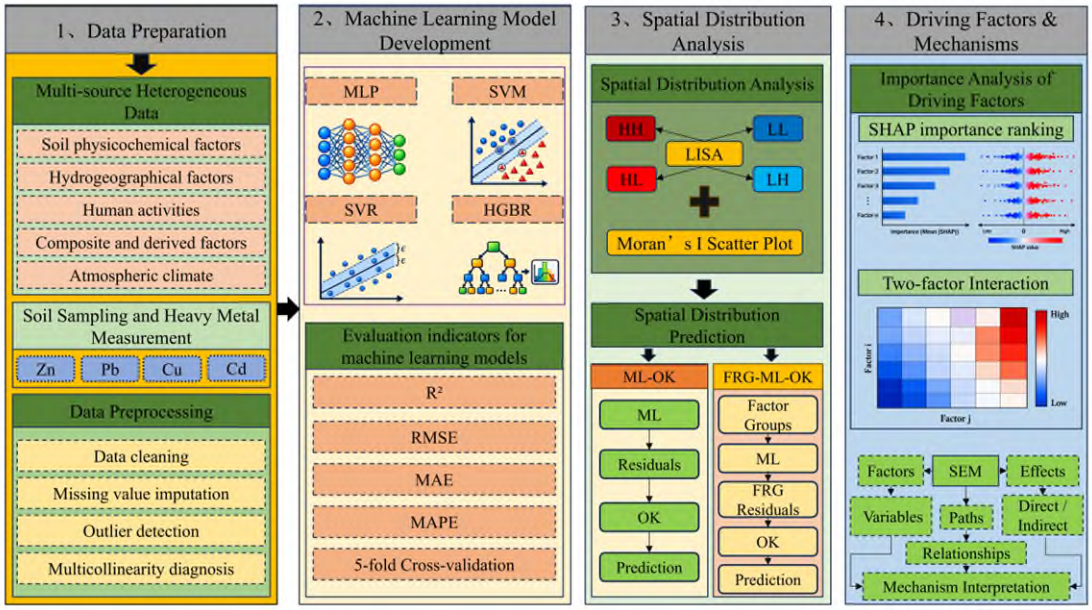


Fig. 2. Algorithmic workflow diagram of the study.

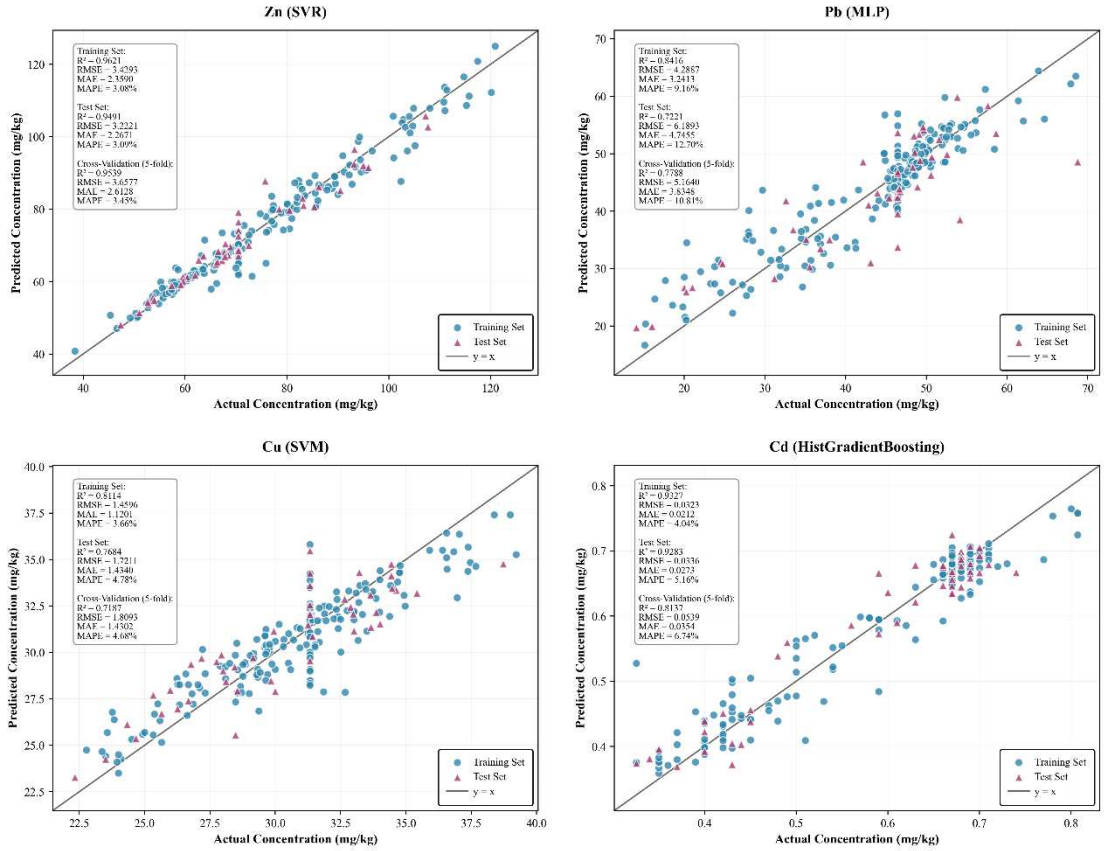


Fig. 3. Actual vs. predicted concentrations of heavy metals using the optimal ML models. The ideal 1:1 line (gray solid) is shown to facilitate comparison.

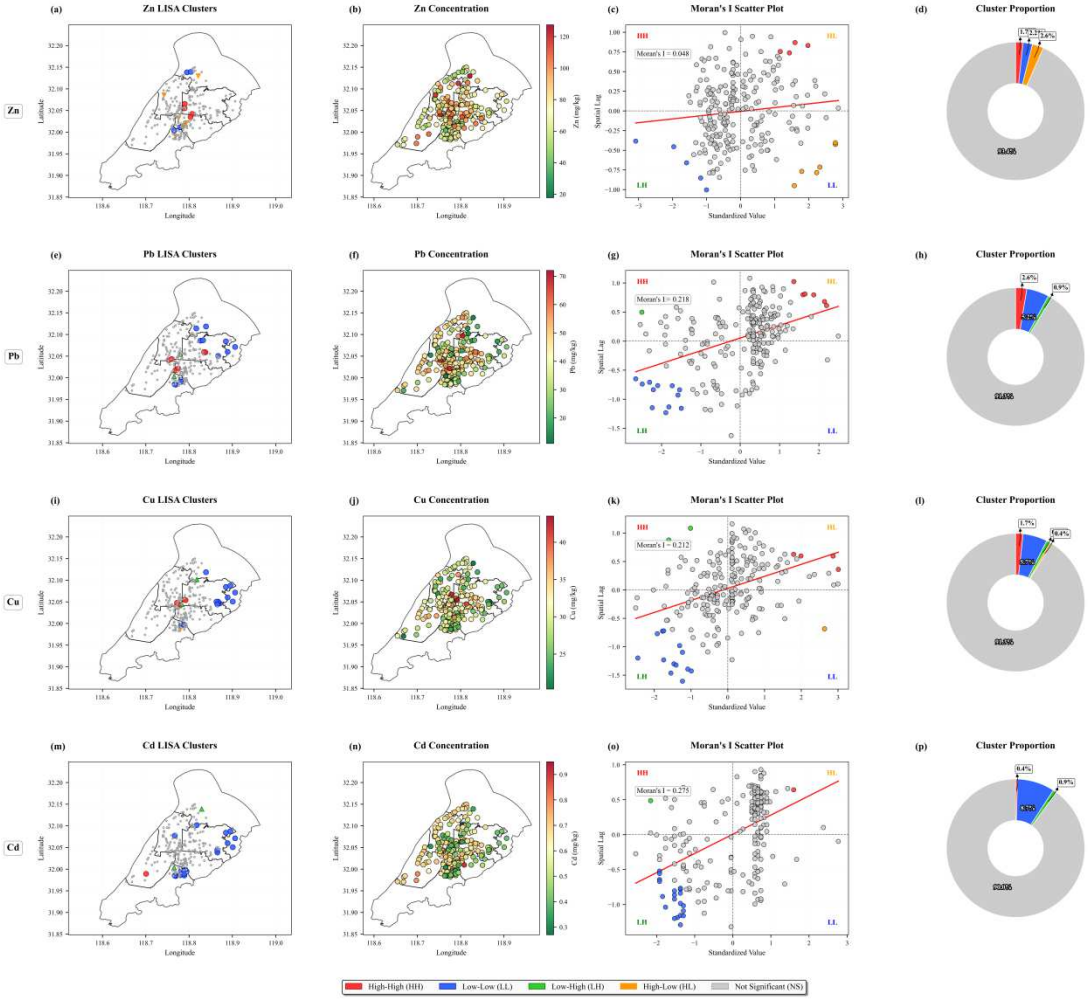


Fig. 4. Spatial distribution and global–local spatial autocorrelation patterns of Zn, Pb, Cu, and Cd, the LISA cluster map (a, e, i, m), concentration distribution map (b, f, j, n), Moran's I scatter plot (c, g, k, o), and pie chart of cluster type proportions for specific elements (d, h, l, p) are presented in order.

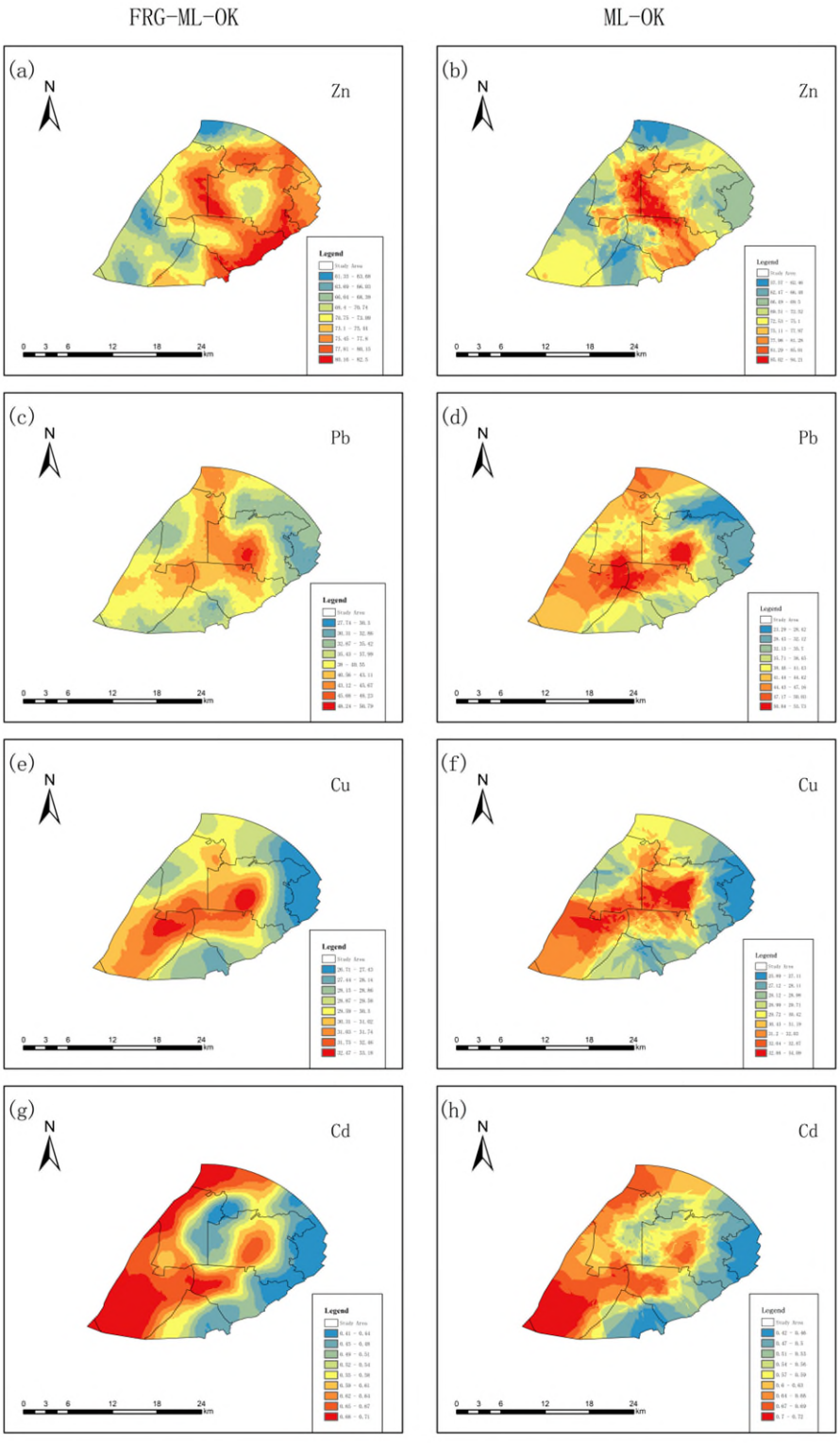


Fig. 5. Spatial prediction of heavy metal concentrations based on FRG-ML-OK model (a, c, e, g) and ML-OK model (b, d, f, h) for Zn, Pb, Cu, and Cd.

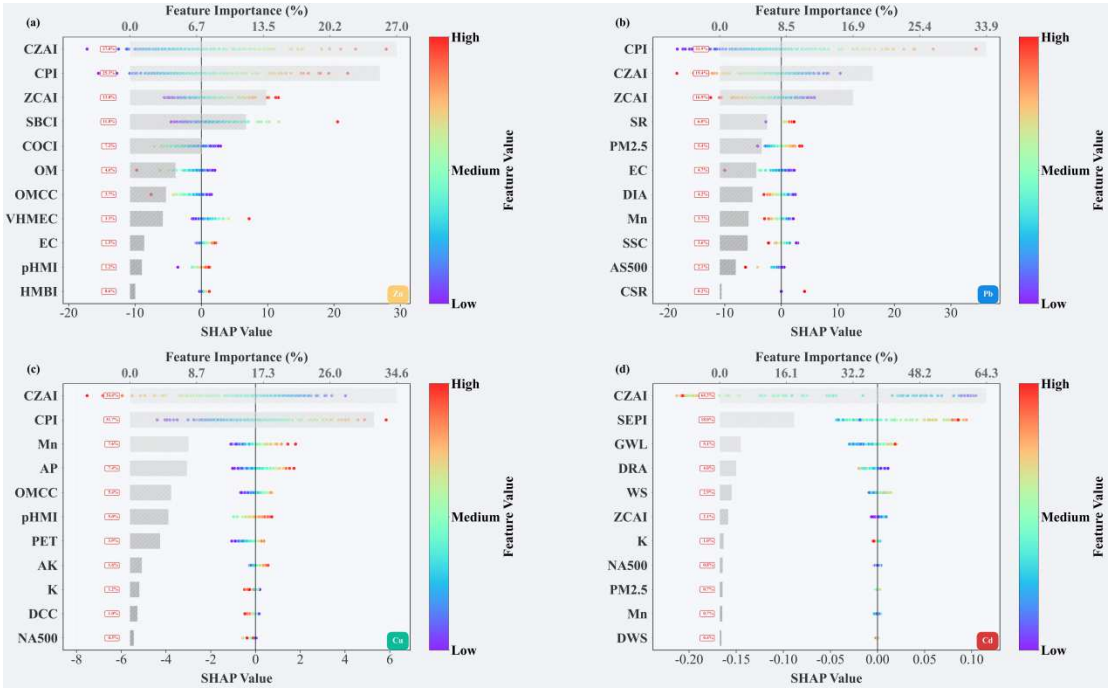


Fig. 6. SHAP-based interpretation of the optimal ML models for Zn (a), Pb (b), Cu (c), and Cd (d). Grey horizontal bars represent relative feature importance (%, from mean absolute SHAP values), the colored scatter points indicate the SHAP values of individual samples. The color gradient represents the feature value, ranging from low to high.

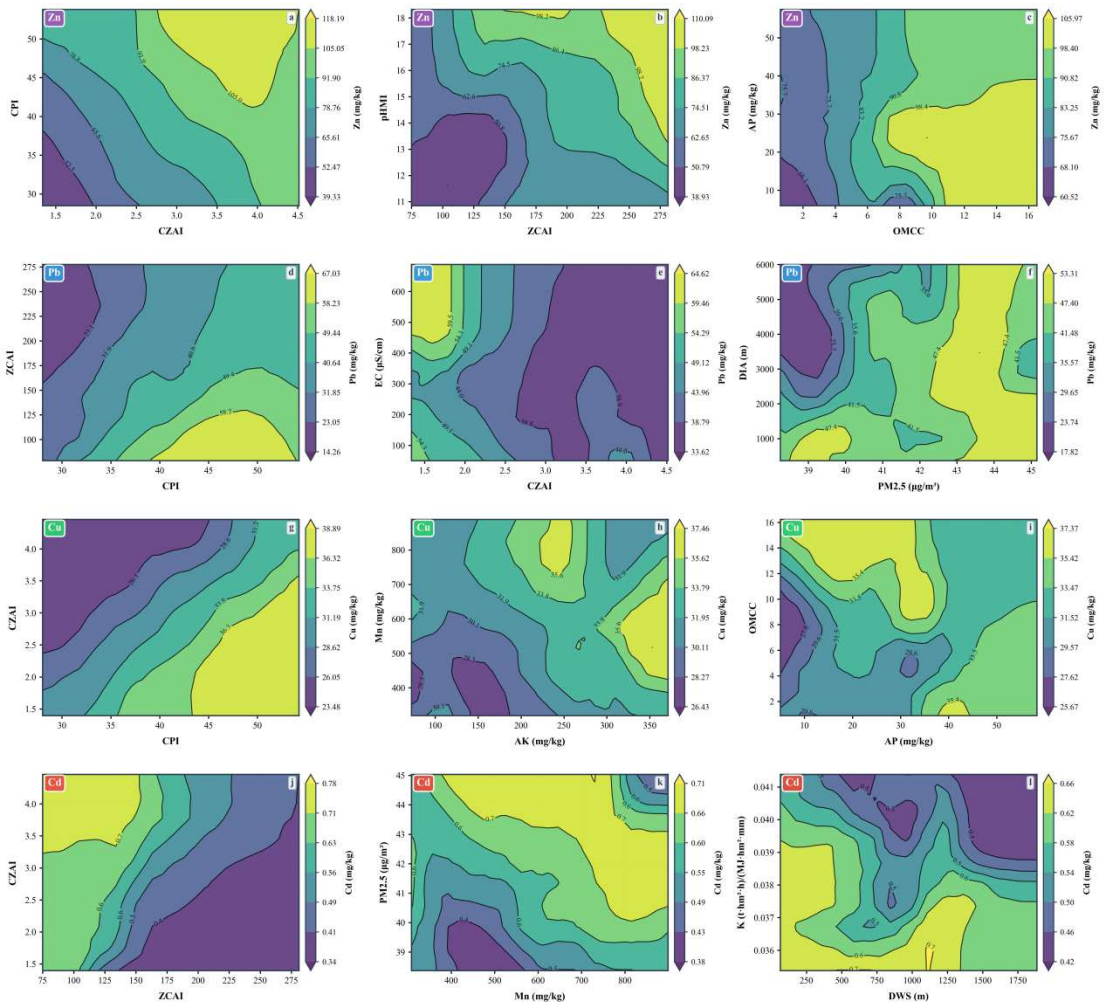


Fig. 7. Two-dimensional response surfaces of pairwise interactions between key influencing factors for Zn (a, b, c), Pb (d, e, f), Cu (g, h, i), and Cd (j, k, l) prediction. The contour plots show the joint effects of selected variable pairs on the predicted concentrations of the corresponding heavy metals, with color gradients and contour lines indicating concentration levels.

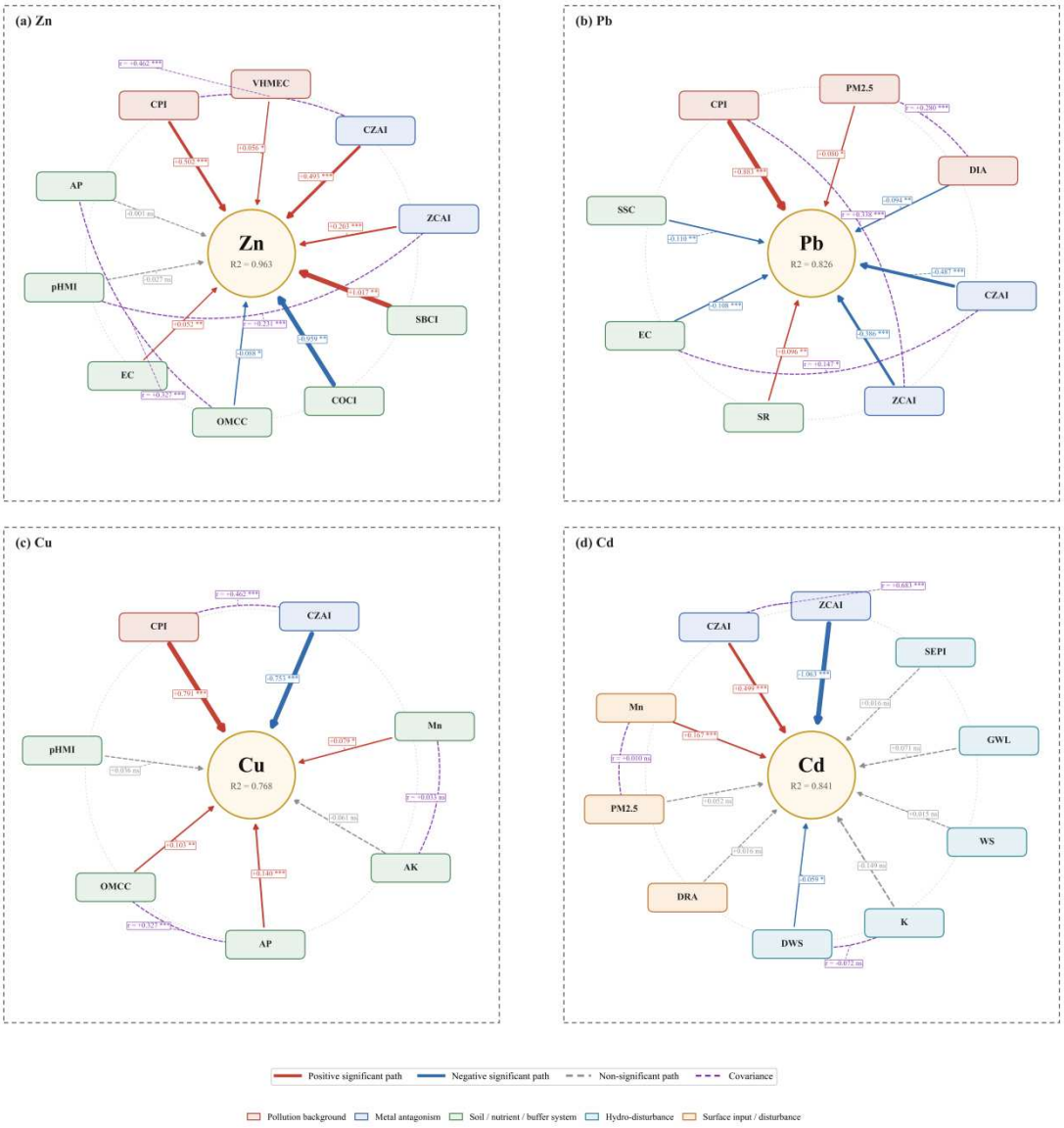


Fig. 8. Direct, indirect and covariance relationships among key influencing factors controlling Zn (a), Pb (b), Cu (c), and Cd (d) concentrations in urban soils using SEM. Rectangles represent observed explanatory variables, and central circles represent the target heavy metal concentrations. The R^2 values shown within the circles indicate the explanatory power of each SEM. Red solid arrows represent significant positive effects, blue solid arrows represent significant negative effects, grey dashed arrows represent non-significant paths, and purple dashed lines represent covariance relationships between explanatory variables. Standardized path coefficients are labelled along the corresponding paths. "*", "**" and "***" represent the significant level $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively. Variable categories are distinguished by node colors, including pollution load or multi-metal enrichment indicators, metal antagonism indicators, soil nutrient and buffering properties, hydrological disturbance factors, and urban development activities.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarydata.docx