Journal of Hazardous Materials 517 (2026) 143639

Contents lists available at ScienceDirect

Journal of Hazardous Materials

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

Spatial patterns of soil heavy metal(loid)s outside the boundaries of China's iron and steel enterprises during 2015–2025: Environmental drivers, mechanistic insights, and policy implications

Xin Xu [a,b], Zheng Zhang [a,b], Ziang Ma [a,b], Peng Wang [a,b], Xiaoqing Zhang [a,b], Xin Bo [a,b,*]

[a] Department of Environmental Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, PR China
[b] BUCT Institute for Carbon-Neutrality of Chinese Industries, Beijing 100029, PR China

HIGHLIGHTS

- Cd showed the highest exceedance rate among the heavy metals.
- A machine learning approach identified key drivers influencing soil heavy metal levels.
- SHAP-GAM analysis was employed to identify tipping points of critical variables for each heavy metal.
- Geographic detector analysis was performed to quantify spatial heterogeneity and the influence intensity of principal factors.
- Policy implications were proposed based on steel plant emission performance and heavy metal risk indicators.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Iron and steel industry
Heavy metal(loid) pollution
Drivers
Generalized additive models
Emission-risk assessment

ABSTRACT

Long-term spatiotemporal patterns of soil heavy metal(loid) contamination outside the boundaries of China's iron and steel enterprises have not been adequately characterized at the national scale. To address this gap, a long-term database of soil heavy metal(loid) concentrations during 2015–2025 was established to investigate environmental drivers, mechanistic processes, and policy implications at the macroscale. According to the proportion of samples exceeding China's 2018 agricultural land risk control standards, the exceedance rates of Cu, Zn, Ni, Pb, Cr, Cd, As, and Hg were 3.82%, 2.38%, 1.34%, 0.52%, 0.83%, 12.50%, 0.62%, and 1.03%, respectively. Both environmental and anthropogenic factors shaped these patterns, while topography and soil properties regulated metal accumulation through hydrothermal redistribution, organic matter dynamics, and adsorption. Climate variables, especially temperature, precipitation, and wind speed, were the dominant controls, and thresholds of ~15 °C were identified for As, Cd, and Pb. Vegetation coverage predominantly affected Hg, whereas socio-economic factors and industrial emissions significantly modulated the distributions of Cu, Hg, Ni, and Zn. Provincial analysis showed uneven emission performance: relatively low performance occurred in

* Corresponding author at: Department of Environmental Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, PR China.
E-mail address: boxin@buct.edu.cn (X. Bo).

https://doi.org/10.1016/j.jhazmat.2026.143639
Received 12 June 2026; Received in revised form 5 August 2026; Accepted 13 September 2026
Available online 15 September 2026
0304-3894/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Shanghai and Guangdong, medium performance was found in most northern and central–southern provinces, and comparatively high performance was concentrated in several eastern and northeastern regions. Contamination patterns differed clearly among the heavy metals, with Cd and Hg recognized as the dominant pollutants. Risk-index evaluation indicated greater ecological risks in northern and northeastern areas. These findings support pollution control and sustainable development in the iron and steel industry.

1. Introduction

Among environmental receptors, soil is a major medium affected by emissions from the iron and steel industry and represents a key interface among industrial activity, ecosystem security, and human health [1]. As an essential component of ecosystem processes and agricultural production, soil supports nutrient cycling and crop development but remains particularly susceptible to heavy metal contamination [2, 3]. Heavy metals are persistent, highly toxic, and prone to bioaccumulation; consequently, their long-term buildup in soil can substantially impair ecological functions and reduce the quality and yield of agricultural products [4, 5]. These metals may also enter the food chain and thereby pose potential risks to human health. Soil contamination by heavy metals has therefore become a major environmental concern worldwide.

Continuing industrialization and urbanization in China have increased heavy metal concentrations in soil through several processes, including industrial emissions, atmospheric deposition [6–8], and solid-waste accumulation. Among industrial sources, the iron and steel industry is considered an important contributor because of its high energy consumption, substantial emissions, and large-scale production [9, 10]. The effects of iron and steel enterprises on heavy metal contamination in surrounding off-site soils show marked spatial heterogeneity arising from the combined effects of regional natural and socio-economic factors. Heavy metal distributions in soil are determined not only by direct inputs such as industrial emissions [11, 12] but also by geological and climatic conditions, soil properties, ecological functions, socio-economic conditions, and industrial activities. Previous studies at broader spatial scales have described the effects of natural environmental factors on total heavy metal concentrations in topsoil [3, 13, 14], emphasizing the varied roles of atmospheric deposition, climate, topography, human activity, and soil geochemical properties in metal accumulation [15–18]. Nevertheless, most studies have focused on a single group of factors, such as soil properties or specific anthropogenic activities [19–21], and few have integrated multiple natural and socio-economic drivers within one analytical framework. Studies of the multifactorial drivers of heavy metal distributions in off-site soils around steel enterprises remain especially limited. Critical thresholds of environmental variables that may produce “biphasic effects” are also insufficiently explored [22]. Moreover, environmental backgrounds differ considerably among industrial sites, potentially producing regional variation in the mechanisms controlling off-site heavy metal accumulation. Nationwide investigations of off-site soil heavy metals around steel plants from an interactive, multifactorial perspective remain scarce, limiting the understanding of spatial heterogeneity and the development of region-specific management strategies. Two scientific questions therefore require attention. First, the dominant drivers of heavy metal distributions in off-site soils around steel enterprises at the national scale, together with their relative contributions, have not been systematically clarified. Second, the mechanisms, interactions, and critical pathways through which natural and anthropogenic factors jointly shape these distributions remain poorly understood. Resolving these questions is scientifically and practically important for establishing an integrated framework of environmental drivers and improving targeted prevention and control of soil environmental risks.

Machine learning (ML) methods have been widely used to predict spatial distributions of soil heavy metals, characterize spatial heterogeneity, and identify major driving factors because they can reveal patterns in complex environmental data [23–25]. Despite their strong predictive performance with high-dimensional multivariate data, conventional ML models provide limited information on how individual variables affect sample-specific predictions or interact with other environmental factors [5, 26]. The game theory-based Shapley Additive Explanations (SHAP) framework [27] has recently received increasing attention because it improves model interpretability and explicitly quantifies interactions among environmental factors. SHAP provides a unified metric for attributing model predictions to individual explanatory variables at the sample level [28, 29], enabling factor-specific effects to be identified and offering an approach for detecting critical thresholds in soil heavy metal concentrations. However, SHAP mainly interprets model predictions and quantifies individual-variable contributions; it has a limited ability to directly characterize complex nonlinear responses and environmental thresholds between continuous predictors and heavy metal concentrations. Generalized additive models (GAMs), by contrast, use smooth functions to flexibly model nonlinear relationships between predictors and response variables, allowing response inflection points and potential environmental thresholds to be identified [22]. Integration of SHAP and GAM therefore enables both the identification and quantification of principal drivers of heavy metal distributions in off-site soils around steel enterprises and a clearer interpretation of their nonlinear responses and threshold effects. The combined framework extends the analysis from driver identification to mechanism interpretation and provides a more complete theoretical basis for explaining the formation of these spatial patterns.

Against this background, representative iron and steel enterprises across China were examined, and a comprehensive database of off-site soil heavy metal concentrations was constructed using diverse natural and anthropogenic factors. In total, 48 variables were assembled, covering geological and climatic conditions, soil properties, ecological functions, socio-economic indicators, and industrial activity, thereby forming a multisource framework of potential drivers. Three objectives were pursued: (1) to delineate the nationwide spatial distributions of off-site soil heavy metals around steel plants and quantitatively evaluate the contributions and threshold effects of natural and socio-economic drivers for individual metals; (2) to clarify nonlinear interactions and mechanistic pathways among multiple drivers; and (3) to determine the principal pathways through which the iron and steel industry affects surrounding soil environments. The findings should advance understanding of the accumulation mechanisms of industrially derived heavy metals in soil and offer scientific support for environmental management, land-use planning, and national soil pollution-control policies in the steel sector.

2. Materials and methods

2.1. Data collection and pretreatment

Iron and steel enterprises were defined as industrial entities involved in ironmaking, steelmaking, and steel rolling, while upstream mining companies and other raw-material suppliers were excluded. Soil monitoring data from the peripheries of iron and steel enterprises in China during 2015–2025 were systematically compiled from environmental impact assessment reports available on the official websites of provincial ecological and environmental departments (detailed sources are listed in Table S1) and from relevant literature. The resulting dataset included enterprise names, geographic locations, soil sampling dates, sampling-point identifiers, site coordinates, and corresponding pollutant

concentrations. The monitored pollutants were As, Cd, Cr, Cu, Hg, Ni, Pb, and Zn. All samples originated from the topsoil layer (0–20 cm), maintaining comparability and representativeness among sampling locations and providing a reliable basis for later analyses of contamination patterns and environmental implications. Data collection and quality-control procedures are described in Text S1.

2.2. Covariate collection

2.2.1. Natural environmental covariates

The present study included six categories of environmental covariates: geological conditions, climatic conditions, soil features, ecological functions, socio-economic factors, and industrial activity (Table S2). Climatic data, including annual average temperature, annual average precipitation, annual average relative humidity, and annual average wind speed, were acquired from the Resource and Environment Science and Data Center (https://www.resdc.cn/). The variables were extracted for every sampling location based on its geographic coordinates using the Extract Values to Points tool in ArcGIS. The ranges of AT, AP, ARH, and AWS were 2.43–24.65 °C, 138.64–2263.91 mm, 50.84–83.37%, and 1.20–2.78 m/s, respectively. Altitude and ground slope were obtained from the 90 m × 90 m Digital Elevation Model (DEM) offered by the Resource and Environment Science and Data Center, and both were extracted for each sampling location using the spatial analysis tools in ArcGIS. The study area revealed notable topographic variation, with altitude ranging from –5.00–2174.00 m and slope from 0.00° to 16.70°. Soil properties, including soil pH, organic matter content, cation exchange capacity, gravel, clay, sand, and silt content, and soil texture class data, were obtained from the high-resolution national soil information grid basic attributes dataset of China (2010–2018) (http://soil.geodata.cn/data/datadetails.html?dataguid=36810085119113)[30, 31]. Among these, the soil texture classes primarily included sandy loam, loam, silty loam, clay loam, silty clay loam, silty clay, and clay. These variables were extracted using ArcGIS according to the geographic coordinates of each sampling location, and the raster pixel values were converted into actual values using the corresponding scale factors. The normalized difference vegetation index (NDVI) data were obtained from NASA's open-access Earthdata platform as an indicator of vegetation coverage across the study area (https://www.earthdata.nasa.gov/) [32], and NDVI values were extracted according to the geographic coordinates of each sampling location and converted to actual values using the corresponding scale factors.

2.2.2. Anthropogenic covariates

Anthropogenic activity factors mainly comprised gross domestic product, population density, road network density, proportion of secondary industry, total electricity consumption, and fertilizer application; these data were obtained from statistical yearbooks, government notices, and other public datasets [33–35]. Population density and road network density were extracted for each sampling location from the corresponding raster datasets in ArcGIS using geographic coordinates. The ranges of gross domestic product, population density, and road network density were 50.00–67345.00 × 10^4 yuan/km^2, 8.00–6890.00 people/km^2, and 0.00–10.90 km/km^2, respectively. The proportion of secondary industry, total electricity consumption, and fertilizer application ranged from 9.36 to 77.83%, 57500.00–3052971.00 ten thousand kilowatt-hours (kWh), and 188.00–104673.50 t, respectively. Anthropogenic covariates also included human activities associated with heavy metal pollution. Industrial indicators for iron and steel enterprises included distance to emission sources, product output, and pollutant emissions. Product-output data were obtained mainly from Emission Permit Implementation Reports [36], the National Bureau of Statistics (https://data.stats.gov.cn/), industrial yearbooks, and on-site surveys. Emissions of SO_2, NO_x, PM_{10}, $PM_{2.5}$, BC, OC, EC, CO, VOCs, dioxin, and fluoride were primarily derived from the atmospheric pollutant emission inventory for China's iron and steel industry; the compilation methods have been reported previously [37, 38]. The emission inventory of trace elements, including Cd, Cr, Ni, As, Pb, Cu, Mn, Se, V, Zn, Sb, Sn, Fe, and Hg, was compiled following Tian et al. [39]. Distance to emission sources was calculated with the Euclidean Distance tool in ArcGIS.

2.3. Machine learning model

2.3.1. Data preprocessing and input feature selection

Soil heavy metal concentrations were converted to uniform units and standardized to maintain consistency and comparability. Z-score normalization was applied to both input and output parameters before their incorporation into the ML model [40, 41]. The input features were standardized using the StandardScaler function in Scikit-Learn (version 1.3.2), as expressed in Eq. (1).

$$X_i' = \frac{(x_i - \mu)}{\sigma} \tag{1}$$

where X_i' and x_i are the standardized and original values of the input variables, respectively, and μ and σ are the mean and standard deviation of the sample data, respectively.

Potential multicollinearity among the explanatory variables was evaluated by calculating the variance inflation factor (VIF) before model development. Variables with VIF values greater than 5 were considered to have relatively high multicollinearity. The observed collinearity mainly arose from shared industrial production processes and synergistic emission characteristics (Table S3). Because the variables represented distinct environmental information, all explanatory variables were retained for subsequent machine learning analyses.

2.3.2. Input feature selection

Spearman's rank correlation coefficient was employed to analyze correlations between different variables, and the correlation coefficient (ρ) was calculated using Eq. (2):

$$\rho_s = \frac{\sum_{i=1}^{n}\left(R_{X_i} - \overline{R}_X\right)\sum_{i=1}^{n}\left(R_{Y_i} - \overline{R}_Y\right)}{\sqrt{\sum_{i=1}^{n}\left(R_{X_i} - \overline{R}_X\right)^2}\sqrt{\sum_{i=1}^{n}\left(R_{Y_i} - \overline{R}_Y\right)^2}} \tag{2}$$

where $\rho_s$reflects the Spearman's rank correlation coefficient measuring the monotonic association between a feature and the target variable or between two target variables. R_{X_i} and R_{Y_i} denote the ranks of the i-th observation in the respective variables, while $\overline{R}_X$ and $\overline{R}_Y$ denote the mean ranks of the input feature and the output target, respectively. The coefficient ranges from –1–1, with 0 indicating no monotonic relationship and values approaching –1 or 1 indicating strong negative or positive monotonic associations, respectively.

During the correlation analysis, Spearman's rank correlation coefficients were determined among input features and between input features and soil heavy metal concentrations to identify variables conveying similar information. Input features were screened and optimized by collectively considering environmental and physical relevance, data completeness, and relationships with the target variables based on the correlation results. This feature selection strategy based on Spearman rank correlation retained key environmental and anthropogenic information while reducing multicollinearity among input variables, thereby enhancing the stability and generalization performance of the machine learning models and yielding a robust set of input features for later driver identification using ML and SHAP analyses.

2.3.3. ML model development and optimization

A grid-search method was used to optimize the principal hyperparameters of six machine learning models: random forest (RF), multilayer perceptron (MLP), decision tree (DT), support vector regression

(SVR), extreme gradient boosting (XGBoost), and gradient boosting decision tree (GBDT) (Text S2–S7 and Table S4). The optimized parameters included n_estimators and max_depth for RF; tree depth and node-splitting parameters for DT; kernel function, regularization parameter (C), and ε for SVR; hidden-layer architecture, optimization algorithm, regularization coefficient, and maximum iterations for MLP; n_estimators, max_depth, learning rate, minimum child weight, and split gain threshold (gamma) for XGBoost; and n_estimators, learning rate, and max_depth for GBDT. After the optimal combinations were determined, the dataset was randomly partitioned into training and testing subsets at a ratio of 80% and 20%, respectively. Ten-fold cross-validation was conducted during model training to assess generalization and robustness. The procedure was repeated ten times, and the mean prediction accuracy across all iterations was used as the final performance estimate. Models were developed in Python 3.13.5 using the Scikit-learn package.

Three metrics were selected to evaluate model performance and generalizability: root-mean-square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). RMSE represents the sample standard deviation of the difference between predicted and observed values and reflects dispersion, with smaller values indicating better predictive accuracy. MAE measures the average absolute error between predicted and observed values, and lower values indicate greater model accuracy. The coefficient of determination (R^2) measures goodness of fit and represents the proportion of total variance explained by the regression model. Values approaching 1 indicate better fitting performance. RMSE, MAE, and R2 are defined in Eqs. (3), (4), and (5), respectively [42, 43].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5}$$

where y_i and $\bar{y}$ denote the observed value of the i-th sample and the mean of the actual values of the total samples, respectively; $\hat{y}_i$ represents the predicted value; and n denotes the number of samples.

The model with the best overall performance was chosen as the optimal model for later interpretation and analysis.

2.3.4. Evaluation of model reliability

To evaluate whether the integrated dataset provides sufficient support for optimal model development and to assess the stability of model performance as the training sample size increases, a learning curve analysis was conducted. Specifically, 20%, 40%, 60%, 80%, and 100% of the complete dataset were randomly sampled to generate training datasets of different sizes. For each sampling proportion, 15 independent random samplings were performed using different random seeds. Each sampled dataset was then randomly divided into training and testing sets at a ratio of 8:2, and an optimal model was developed accordingly. Model performance was evaluated using the Pearson correlation coefficient (R) between the predicted and observed values in the testing set. For each sampling proportion, the mean and standard deviation of the testing R values obtained from the 15 repeated runs were calculated to quantify both predictive performance and model stability. Finally, learning curves were constructed to characterize the variation in model performance with increasing training sample size, thereby evaluating whether model performance gradually converged to a stable level and whether the current sample size was sufficient to support reliable optimal model development.

2.4. Model interpretation

The differences in performance of the model between the training and test datasets were typically within an acceptable range; however, variations were found among individual algorithms and target heavy metals, reflecting differences in the ability of machine learning models to capture complex environmental relationships and generalize beyond the training data (Table S5). In particular, the random forest (RF) model displayed comparatively high stability across most heavy metal predictions. By contrast, some individual decision-tree (DT) models revealed comparatively poor performance on the test dataset, with low R^2values for Zn and Cd predictions (0.32 and 0.28, respectively), suggesting that individual tree models have a limited ability to capture complex nonlinear environmental interactions. Moreover, GBDT, XGBoost, and MLP models also displayed certain degrees of training–test performance discrepancies for some heavy metals, suggesting that predictive performance is influenced not only by algorithm structure but also by differences in sample distributions, the complexity of environmental variables, and potentially unobserved factors [41]. Collectively, although the predictive performance of some models decreased from the training to the test dataset, most optimized models maintained comparatively stable predictive capabilities and effectively characterized the complex relationships between environmental factors and soil heavy metal concentrations.

Based on the comparative evaluation of model performance, the RF model achieved the highest overall predictive accuracy among the tested algorithms and was therefore selected as the optimal model for subsequent analyses. To further evaluate whether the integrated dataset provided sufficient support for RF model development and to assess the stability of RF performance with increasing training sample size, a learning curve analysis was conducted. The analysis revealed distinct patterns in the predictive performance of the RF models for different heavy metals as the training sample size increased (Fig. S1). Overall, the Pearson correlation coefficient (R) between the predicted and observed values in the testing set generally increased with increasing training sample size and gradually approached a stable level for most heavy metals. For As and Zn, model performance improved rapidly with increasing sample size and reached a plateau, indicating that the integrated dataset was sufficient to support robust model development. For Cr, Cu, and Pb, model performance continued to improve as the training sample size increased, suggesting that additional training samples could further enhance the generalization capability of the models. In contrast, the performance of the Cd and Ni models exhibited noticeable fluctuations across different sample sizes, indicating relatively limited stability. For Hg, the testing-set R values remained consistently low regardless of the training sample size, suggesting that its spatial distribution may be jointly influenced by environmental factors not included in the current model and measurement uncertainties. Overall, the learning curve analysis demonstrated that the integrated dataset was adequate for developing RF models for most heavy metals and further confirmed that the current sample size was sufficient to support reliable model training.

An integrated modeling and interpretation framework was subsequently established to quantify the relative contributions of anthropogenic and natural environmental factors to soil heavy metal concentrations surrounding iron and steel enterprises. The selected optimal machine learning model was used to capture nonlinear relationships between predictors and response outcomes. Shapley additive explanation (SHAP) analysis was then applied to interpret the model results and evaluate the contribution of each variable [27].

For every predictor i, the Shapley value (φ_i) reflects the average marginal contribution of that variable to the output of the model across all possible feature combinations. The formulation is expressed in Eq. (6):

$$\varphi_i = \sum_{S \subseteq F\{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!}[f(S \cup \{i\}) - f(S)] \quad (6)$$

F represents the full set of covariates, S represents any subset that excludes variable i, and $f(s)$ indicates the prediction from the model derived from subset S. SHAP values were calculated using the treeshap and shapviz packages (https://github.com/ModelOriented/shapviz), following the conversion of caret-trained models into SHAP-compatible formats. The importance of variables and their distribution patterns were presented using beeswarm plots and ranking diagrams produced by the *sv_importance* function.

Generalized additive models (GAMs) were fitted to actual covariate values (independent variables) and their SHAP values (dependent variables) using the mgcv package (https://CRAN.R-project.org/package=mgcv) to further examine nonlinear responses of the major influencing variables. SHAP values were estimated at 1000 evenly spaced points over the predictor range. Zero crossings of the GAM response curves were numerically identified by the bisection method and defined as tipping points [44, 45]. These points represent thresholds at which the marginal effects of covariates on predicted metal concentrations change direction. Regions with SHAP values greater than zero or less than zero were classified as positive- or negative-effect regimes, respectively.

Finally, pairwise interactions among the three most important covariates for each heavy metal were visualized using the *sv_dependence* function. This analysis enabled clearer interpretation of synergistic and antagonistic interactions between anthropogenic and natural factors controlling patterns of metal accumulation in soil.

2.5. Geographic detector

The geographic detector is a commonly used spatial statistical approach for identifying spatial heterogeneity and quantifying how strongly geographic factors explain the spatial distribution of a given attribute [46, 47]. Its underlying principle assumes that if an explanatory variable has a significant influence on the dependent variable, the two should exhibit similar spatial distribution patterns. The detector is calculated as follows:

$$q = 1 - SSW/SST = 1 - \sum_{h=1}^{L} N_h \sigma_h^2 / N\sigma^2 \quad (7)$$

where q is the explanatory strength of a factor, ranging within [0,1]; larger values indicate stronger spatial heterogeneity of Y and higher explanatory capacity of factor X for the spatial pattern of Y. $h = 1, \ldots, L$ is the stratification of factor X. N_h and N represent the number of units within layer h and across the entire study area, respectively. σ_h^2 and σ^2 the variance of the dependent variable Y within stratum h and over the whole region. *SSW* and *SST* represent the sum of within-layer variance and the total regional variance, respectively. The factor detector is used to quantify the spatial differentiation of Y and to evaluate the explanatory ability of a single factor X for its spatial variation, expressed by the q value. The interaction detector is employed to assess the joint effects among factors, specifically to examine whether the joint effect of two factors ($X_{1\text{and}}\ X_2$) enhances or weakens their capacity to explain the spatial heterogeneity of Y. The categories of factor interactions are listed in Table S6.

2.6. Entropy-Weighted TOPSIS Method

The entropy-weighted technique for order preference by similarity to ideal solution (entropy-weighted TOPSIS) was adopted for a comprehensive assessment of the air pollutant emission levels of steel enterprises in this work. This method enables objective assignment of indicator weights based on information entropy and supports a reliable multi-criteria ranking of enterprises according to their emission performance. Moreover, spatial weighting factors were introduced to account for regional clustering or synergistic effects [48], thereby enhancing the spatial interpretation of the findings. The detailed procedures are described in Text S8.

2.7. Soil heavy metal pollution and ecological risk assessment

2.7.1. Data weighting method

Initially, the soil heavy metal content data collected from published studies were preprocessed. Since multiple studies were conducted within the same province, the annual average concentrations of soil heavy metals for each province were calculated according to the number of sampling points reported in the respective studies:

$$C_{jt}^i = \frac{\sum_{l=1}^{n}(C_{lt}^i \times N_{lt}^i)}{\sum_{l=1}^{n} N_{lt}^i} \quad (8)$$

where C_{jt}^idenotes the average content of heavy metal $i$province $j$during year t, C_{lt}^ireflects the content of heavy metal i reported in the l-th study for the same year, and $N_{lt}^i$is the number of sampling points.

Time weights (w_t)represent the relative importance of individual years, and their accurate determination is essential for later analyses [49]. Entropy (I) measures information content, with greater values indicating lower informational contribution. To ensure maximum extraction of sample information, entropy should be maximized under a given time (λ) when determining time weights. On this basis, the objective function and constraints for the time weight vector are defined as follows [50]:

$$I = \max\left(-\sum_{t=1}^{p} w_t \ln w_t\right) \quad (9)$$

Subject to:

$$\lambda = \max\left(-\sum_{t=1}^{p}(w_t \times \ln w_t\right) \quad (10)$$

$$\text{s.t.}\lambda = \sum_{t=1}^{p} \frac{p-t}{p-1} \times w_t \quad (11)$$

$$\sum_{t=1}^{p} w_t = 1, \quad w_t \in [0,1] \quad (12)$$

where I is the entropy, w_t is the weight of the t-th time series, p is the number of years, and λ reflects the importance of data from individual years. The closer λ is to 1, the greater the importance of earlier data; whereas the closer λ is to 0, the more emphasis is placed on recent data [50]. For the assessment and management of soil pollution by heavy metals, recent data were regarded as more relevant than earlier data, so the value of λ was set to 0.1. The time-dependent weights were determined using Python version 3.13.5 and are presented in Table S7.

Next, the average content C_{jt}^i in different sampling years for each province was calculated based on the time weights. The average concentration for 11 years (covering 2015–2025) $C_j^i$was calculated as follows:

$$C_j^i = \sum_{t=1}^{p} C_{jt}^i \times w_t \quad (13)$$

2.7.2. Pollution and ecological risk assessment

The contamination factor (CF) and pollution load index (PLI) were employed to assess the pollution status of individual soil heavy metals. The CF measures the extent of accumulation of heavy metals by comparing measured concentrations with regional background values [51], and was calculated as follows:

$$C_f^i = \frac{C_i}{B_i} \quad (14)$$

where C_f^i denotes CF, C_i reflects the measured concentration of element i (mg/kg), and $B_i$reflects its geochemical background concentration (mg/kg; see Table S8). The classification criteria for CF are summarized in Table S9.

The PLI combines multiple trace metal elements within the study area to assess the overall pollution level of topsoil and was computed using Eqs. (14) and (15) [52].

$$PLI = \sqrt[n]{C_f^1 \times C_f^2 \times \cdots \times C_f^n} \tag{15}$$

whereC_f^i is the contamination factor index. The classification criteria for PLI are listed in Table S10. The potential ecological risk index (RI) [51] accounts for the concentration, toxicity level, and ecological sensitivity of elements and is widely used to evaluate potential risks to ecosystems caused by soil heavy metals. RI was calculated using Eq. (16):

$$RI = \sum E_r^i = \sum T_r^i \times C_f^i = \sum T_r^i \frac{C_i}{C_b^i} \tag{16}$$

where RI represents the overall potential ecological risk index for all trace metals; E_r^i represents the ecological risk index of an individual element; T_r^i is the corresponding toxic response coefficient; C_f^i denotes the contamination factor of the element; C_i is the measured concentration (mg/kg); and $B_i$is the background concentration (mg/kg) (see Table S8). The T_r^i values for Cu, Cr, Ni, Zn, Pb, Cd, As, and Hg are 5, 2, 5, 1, 5, 30, 10, and 40, respectively. The classification criteria for RI are presented in Table S11.

2.8. Statistical Analysis

Statistical analyses were performed in SPSS 25.0 (IBM Corporation, Armonk, NY, USA). Data distributions were evaluated using the Shapiro–Wilk test, with $p > 0.05$ indicating normality. Because the data were not completely normally distributed, Spearman correlation coefficients were used to assess relationships among the variables and observed metal(loid) concentrations. Distance matrices were then constructed for natural environmental covariates, anthropogenic covariates, and heavy metal concentrations to investigate the overall coupling between multidimensional environmental heterogeneity and spatial variation in metal concentrations. Mantel and partial Mantel tests quantified correlations among these matrices and evaluated independent effects after spatial distance was controlled. Direct inclusion of the complete set of natural and anthropogenic covariates would have made the analysis overly complex and difficult to interpret. Therefore, the drivers of each heavy metal were ranked by SHAP values from the machine learning models, and the ten most influential variables were selected for geographic detector analysis. Entropy-weighted TOPSIS was used to comprehensively evaluate air pollutant emissions from iron and steel enterprises and compare provinces. At the provincial scale, the contamination factor (CF), pollution load index (PLI), and risk index were calculated to evaluate soil heavy metal contamination, with each province used as the basic statistical unit.

3. Results and discussion

3.1. Overall characteristics of heavy metals in soils surrounding iron and steel enterprises

The descriptive statistics for HM concentrations in soils outside the boundaries of iron and steel enterprises are presented in Table 1. The mean concentrations of Cu, Zn, Ni, Pb, Cr, Cd, As, and Hg were 35.64, 91.56, 39.30, 33.93, 62.70, 0.44, 7.82, and 0.72 mg/kg, respectively. The point exceedance rates of the eight HMs were 3.82%, 2.38%, 1.34%, 0.52%, 0.83%, 12.50%, 0.62%, and 1.03%, respectively, relative to China's 2018 agricultural land risk control standards. The CVs of Cu, Ni, Cd, and Hg were 2.32, 1.32, 2.54, and 14.67, respectively, indicating marked spatial variability of these four HMs and potential point-source contamination. A low CV indicates that the element is primarily affected by the soil background conditions and processes occurring naturally, whereas a high CV suggests a substantial contribution from anthropogenic sources [2, 53]. In this study, except for Cu, Ni, Cd, and Hg, which showed the highest CVs, the CVs of the other four HMs ranged from 0.66 to 0.89, indicating that they may still be affected by anthropogenic pollution sources.

The Mantel test identified correlations among natural environmental covariates, anthropogenic covariates, and heavy metals in soils beyond the boundaries of iron and steel plants (Fig. 1 and Fig. S2). In particular, Cu was primarily controlled by geological conditions (altitude, ground line gradient), soil properties (soil organic matter, clay content, sand content), and total electricity consumption, while distance to emission sources indicated a significant role of human activities ($p < 0.01$). In particular, soil organic matter and clay minerals enhance Cu immobilization by providing abundant high-affinity adsorption sites through mechanisms including organic complexation, surface complexation, and cation exchange, thus reducing Cu mobility in soils [54]. Meanwhile, Cu-containing particulate matter emitted from industrial activities can be transported through the atmosphere and later deposited onto soils; therefore, the distance to emission sources can provide an effective indicator for evaluating the contribution of anthropogenic emissions to Cu enrichment [8]. Zn was strongly affected by climate conditions (annual average precipitation, annual average relative humidity, annual average wind speed), soil properties (soil organic matter, clay content, sand content), population density, and distance to emission sources. The enrichment and spatial pattern of Zn in soils are collectively controlled by soil properties and environmental transport processes. Soil organic matter and particle composition control Zn retention through adsorption and binding processes, while precipitation supports Zn migration and leaching by enhancing hydrological transport [55]. Meanwhile, wind speed influences the long-range atmospheric transport and deposition of Zn-containing particulate matter, thus affecting external Zn inputs to soils [56]. Ni was affected by annual average temperature, annual average wind speed, soil properties (soil organic matter, sand content, silt content), atmospheric pollutants (NO_x and SO_2), carbonaceous

Table 1
Descriptive statistical results of different heavy metals [a].

	Cu	Zn	Ni	Pb	Cr	Cd	As	Hg
min (mg/kg)	3.00	8.30	0.04	0.03	10.00	0.01	0.01	0.00
max (mg/kg)	1560.00	980.00	860.00	407.00	324.00	16.00	51.60	292.00
mean (mg/kg)	35.64	91.56	39.30	33.93	62.70	0.44	7.82	0.72
SD (mg/kg)	82.75	81.46	51.75	29.78	41.69	1.11	6.01	10.62
CV	2.32	0.89	1.32	0.88	0.66	2.54	0.77	14.67
Chinese soil criteria 2018 pH > 7.5 (mg/kg)	100.00	300.00	190.00	170.00	250.00	0.60	25.00	3.40
point over standard rate	3.82	2.38	1.34	0.52	0.83	12.50	0.62	1.03

[a] min, minimum value; max, maximum value; mean, mean value; SD, standard deviation; and CV, coefficient of variation. The dimension is one. The average pH of the 968 sites sampled was 7.68, and the corresponding soil standard was therefore selected.

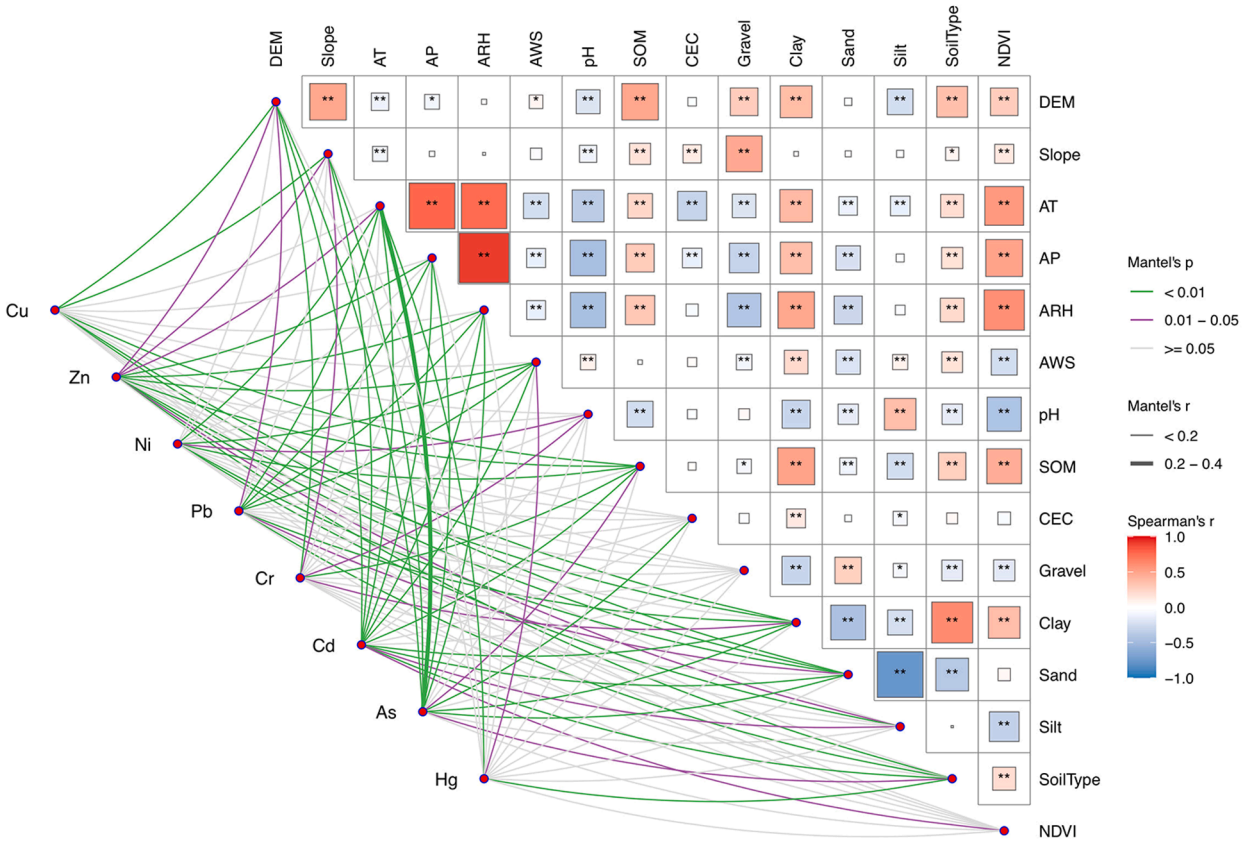


Fig. 1. Mantel test between soil heavy metals and natural environmental covariates. The upper right area represents the Spearman correlation among covariates. DEM, altitude; Slope, ground line gradient; AT, annual average temperature; AP, annual average precipitation; ARH, annual average relative humidity; AWS, annual average wind speed; pH, soil acid and alkaline pH; SOM, soil organic matter; CEC, cation exchange capacity; Gravel, gravel content; Clay, clay content; Sand, sand content; Silt, silt content; SoilType, soil type; NDVI, normalized differential vegetation index.

particulates, and several trace elements (Cd, Cr, Ni, As, Pb, Cu, Mn, Se, V, Zn, Sb, Sn, Fe, and Hg) ($p < 0.01$), suggesting that atmospheric deposition represents an important source of Ni. Ni is widely associated with fine particulate matter emitted from industrial combustion and metallurgical processes [57]. These Ni-bearing particles can undergo long-range atmospheric transport and later enter soils through dry and wet deposition pathways [58]. Thus, the significant correlations between Ni and atmospheric pollutants, including NO_x, SO_2, and carbonaceous particulate matter, further indicate that atmospheric deposition represents an important pathway contributing to regional soil Ni enrichment. Pb was mainly controlled by climate conditions (annual average temperature, annual average precipitation, annual average relative humidity, annual average wind speed) and soil properties (clay content and soil type) ($p < 0.01$). Climate variables influence Pb enrichment by affecting the atmospheric transport and deposition processes of Pb-bearing particles, and by controlling soil water conditions and associated processes of migration. By contrast, clay content and soil type control Pb retention through their effects on adsorption sites, surface binding capacity, and immobilization mechanisms, thus affecting Pb enrichment in soils [59]. Cr was significantly affected by altitude, soil organic matter, and population density, and showed positive associations with NO_x, SO_2, carbonaceous particulates, and trace elements (Cd, Cr, Ni, As, Pb, Cu, Mn, Se, V, Zn, Sb, Sn, Fe, Hg) ($p < 0.01$), demonstrating the strong role of atmospheric deposition. The migration and speciation of Cr in soils are mainly controlled by soil geochemical properties. Soil organic matter and Fe/Mn oxides have important roles in controlling the transformation among individual Cr fractions by providing reactive surfaces and affecting Cr binding processes [60]. Meanwhile, atmospheric deposition of particulate matter from industrial emissions provides additional external inputs of Cr to soils, further contributing to Cr enrichment. Cd was affected by geological conditions (altitude, slope), climate conditions (annual average temperature, annual average precipitation, annual average relative humidity), soil properties (soil organic matter, sand content), and atmospheric pollutants (NO_x, SO_2, and carbonaceous particulates) ($p < 0.01$). Topographic and climatic variables control Cd migration and transformation processes, thus affecting its spatial variability. Meanwhile, soil organic matter and mineral components determine Cd adsorption and stabilization through their effects on binding sites and retention capacity [59]. Additionally, atmospheric emissions of pollutants can enhance external Cd inputs through deposition processes, further promoting localized Cd enrichment in soils [61]. As exhibited a more complex pattern, being affected by geological conditions (altitude, ground line gradient), climate conditions (annual average temperature, annual average precipitation, annual average relative humidity), soil properties (soil organic matter, sand content), socio-economic variables (gross domestic product, population density, total electricity consumption), and trace elements (Cd, Cr, Ni, As, Pb, Cu, Mn, Se, V, Zn, Sb, Sn, Fe, Hg) ($p < 0.01$). These results show that As enrichment in soils is controlled by the joint effects of natural environmental processes and anthropogenic activities [62–64]. By contrast, Hg was mainly controlled by soil type and population density ($p < 0.01$). Collectively, the spatial pattern of individual heavy metals reflects the joint effects of natural background conditions and anthropogenic inputs, with atmospheric pollutants and multi-element co-deposition playing especially important roles for Ni, Cr, Cd, and As.

Spearman correlation analysis was performed to assess the effects of natural and anthropogenic covariates on heavy metal concentrations

(Fig. S3 and Fig. S4). The geological variables showed different patterns: altitude and ground line gradient were strongly and positively correlated with Cu, Zn, Cr, and Cd concentrations ($p < 0.05$), whereas altitude alone was negatively correlated with Hg ($p < 0.05$). Among the climatic variables, annual average temperature, precipitation, and relative humidity were significantly and positively associated with Zn, Pb, Cd, and As ($p < 0.05$), while annual average wind speed was negatively associated with Cu, Zn, Cd, and As ($p < 0.05$). Soil pH was positively related to Cr but negatively related to Pb, Cd, As, and Hg ($p < 0.05$). Soil organic matter showed positive relationships with most metals, including Cu, Zn, Pb, Cr, Cd, As, and Hg ($p < 0.05$). Gravel content was positively associated with Cu, Cr, and Cd ($p < 0.05$) and negatively associated with As ($p < 0.05$). Clay content had significant positive correlations with Cu, Zn, Ni, Pb, Cd, As, and Hg ($p < 0.05$), while sand content had significant negative correlations with all heavy metals ($p < 0.05$). Silt content was positively associated with Zn, Ni, and As ($p < 0.05$). Soil type and NDVI were also positively related to Zn, Ni, Pb, As, and Hg ($p < 0.05$). Significant pairwise correlations were additionally found among the soil heavy metals.

Gross domestic product (GDP) and population density (POP) were significantly and positively associated with Cu, Pb, Cd, and As ($p < 0.05$), but negatively associated with Zn ($p < 0.05$). Road network density (DRN) was positively correlated with Cu and Cr ($p < 0.05$) and negatively correlated with Ni and Hg ($p < 0.05$). The proportion of secondary industry (PSI), total electricity consumption (TEC), and fertilizer application (FA) showed negative relationships with Cu, Zn, Ni, Cd, and As ($p < 0.05$). Product output (PO), an indicator of industrial activity, was negatively correlated with Cu, Zn, Ni, Pb, and Cd ($p < 0.05$). Distance to emission sources (DES) was positively correlated with Zn and Hg but negatively correlated with Cu. Emissions from iron and steel enterprises, denoted by abbreviations ending in "E" for the corresponding pollutants, were negatively related to most heavy metals, including Cu, Zn, Pb, Cr, and Cd. By contrast, carbonaceous pollutants such as PM_{10}, BC, OC, and EC were significantly and positively associated with As ($p < 0.05$), and most gaseous pollutants were significantly and positively associated with Hg in soil ($p < 0.05$). Although iron and steel enterprises are an important anthropogenic source, soil heavy metal concentrations reflect long-term cumulative inputs from many natural and human sources. At the national scale, relationships between emissions from these enterprises and certain soil heavy metals therefore may not be simply positive; instead, they represent statistical associations produced by multiple natural and anthropogenic drivers.

3.2. *Effects of natural environmental covariates on soil heavy metal concentrations surrounding iron and steel enterprises*

The concentrations of different heavy metals around steel enterprises were controlled by multiple interacting factors (Fig. 2 and Fig. S5), which can be grouped into four dominant mechanisms: meteorological-topographic control, soil physicochemical control, emission-source control, and agricultural composite influence. These metals exhibit both shared characteristics and distinct patterns across categories. For example, the findings indicate that for As and Cd, the dominant influencing variables include AT, AWS, and topography- or climate-related factors such as elevation (DEM) and precipitation. This pattern implies that their deposition processes strongly depend on atmospheric dispersion capacity, wet deposition intensity, and terrain-induced blocking effects. Therefore, As and Cd can be regarded as representative metals whose spatial deposition is jointly governed by meteorological conditions and topographic constraints. Zn and Ni partially exhibit similar patterns, with AWS contributing prominently to their SHAP values, indicating that atmospheric transport remains a key mechanism controlling their distribution. Conversely, the spatial distributions of Cr, Pb, and Cu are primarily controlled by soil properties, where variables such as soil organic matter, pH, clay content, and slope are consistently identified as key predictors, indicating that these metals are typical soil adsorption- and surface process-dominated elements. This pattern reflects a strong tendency for adsorption, complexation, and fixation by organic matter and mineral particles, indicating that their spatial variation is largely governed by surface processes and the adsorption capacity of soil. Another group of metals, such as Hg, exhibits a distinct emission source-dominated pattern. The top three variables for Hg are indicators associated with anthropogenic emissions or industrial structure (e.g., HgE, PSI, FE), indicating that its spatial differentiation directly corresponds to variations in industrial emission intensity and energy structure. Ni is further affected by fertilizer application (FA) and Cd emissions (CdE), indicating a compound source mechanism jointly driven by agricultural inputs and industrial releases. Taken together, these findings highlight systematic differences among heavy metals in terms of source characteristics, transport processes, and soil adsorption behavior: meteorological and topographic conditions determine deposition pathways at the atmosphere–surface interface; soil properties regulate retention and mobility potential; and industrial activities control the magnitude and spatial heterogeneity of external inputs. Together, these factors produce multifaceted heavy metal exposure patterns identified in areas surrounding steel enterprises. Furthermore, generalized additive models were applied to the identified factors and their corresponding SHAP values to further characterize their effects across different value ranges.

3.2.1. *Geological conditions*

Topographic variables influence soil As and Cr levels by controlling surface hydrothermal redistribution [22, 65]. In this research, topography had a significant effect on soil As and Cr levels (Fig. 3c and Fig. S6a, $p < 0.001$). Notably, altitude was the dominant variable controlling the variability of soil As and Cr levels, showing significant negative associations. Additionally, altitude showed negative effects on As content when it exceeded the threshold of 100.16 m (Fig. S6a, R^2=0.21, $p < 0.001$). Notably, the SHAP-GAM analysis revealed distinct elevational threshold responses only for As and Cr, whereas no significant elevation-dependent patterns were observed for the other heavy metals. This finding suggests that the responses of soil heavy metals to elevation are highly element-specific rather than exhibiting a universal altitudinal gradient effect [66]. Previous studies have shown that elevation does not directly control heavy metal levels in soils; instead, it indirectly controls their transport and redistribution by affecting surface runoff, soil erosion, sediment deposition, and local hydrological conditions [67]. In addition, pedogenesis and the degree of soil weathering have been recognized as the primary mechanisms controlling the distribution of heavy metals along elevational gradients, whereas soil organic matter and pH generally contribute less to variations in heavy metal levels in surface soils [68]. For example, in strongly podzolized soils, heavy metals are prone to leaching from the surface layer and subsequently accumulate in illuvial horizons. Conversely, at higher elevations, where leaching intensity is generally reduced, heavy metals are more likely to accumulate in surface soils [68]. Therefore, the effect of elevation on the distribution of heavy metals in soil is primarily mediated through its regulation of pedogenic processes and leaching dynamics, and the strength of these indirect effects varies substantially among individual heavy metals [66, 68].

3.2.2. *Climate conditions*

Climate, as a dominant control on pedogenic processes, regulates mineral weathering, elemental transformation, and transport pathways, thereby controlling the levels and spatial distribution of heavy metals in soils [14, 60, 69]. Climatic variables were found to be the primary drivers of soil As and Cd concentrations (Fig. S5), and their effect sizes exceeded those of other environmental variables in this work. These findings suggested that climatic conditions exerted a dominant influence on the concentrations of these two heavy metals.

Among the climatic variables, AT was the most influential factor affecting total As and Cd concentrations in soil (Fig. 3a & 3b, $p < 0.001$).

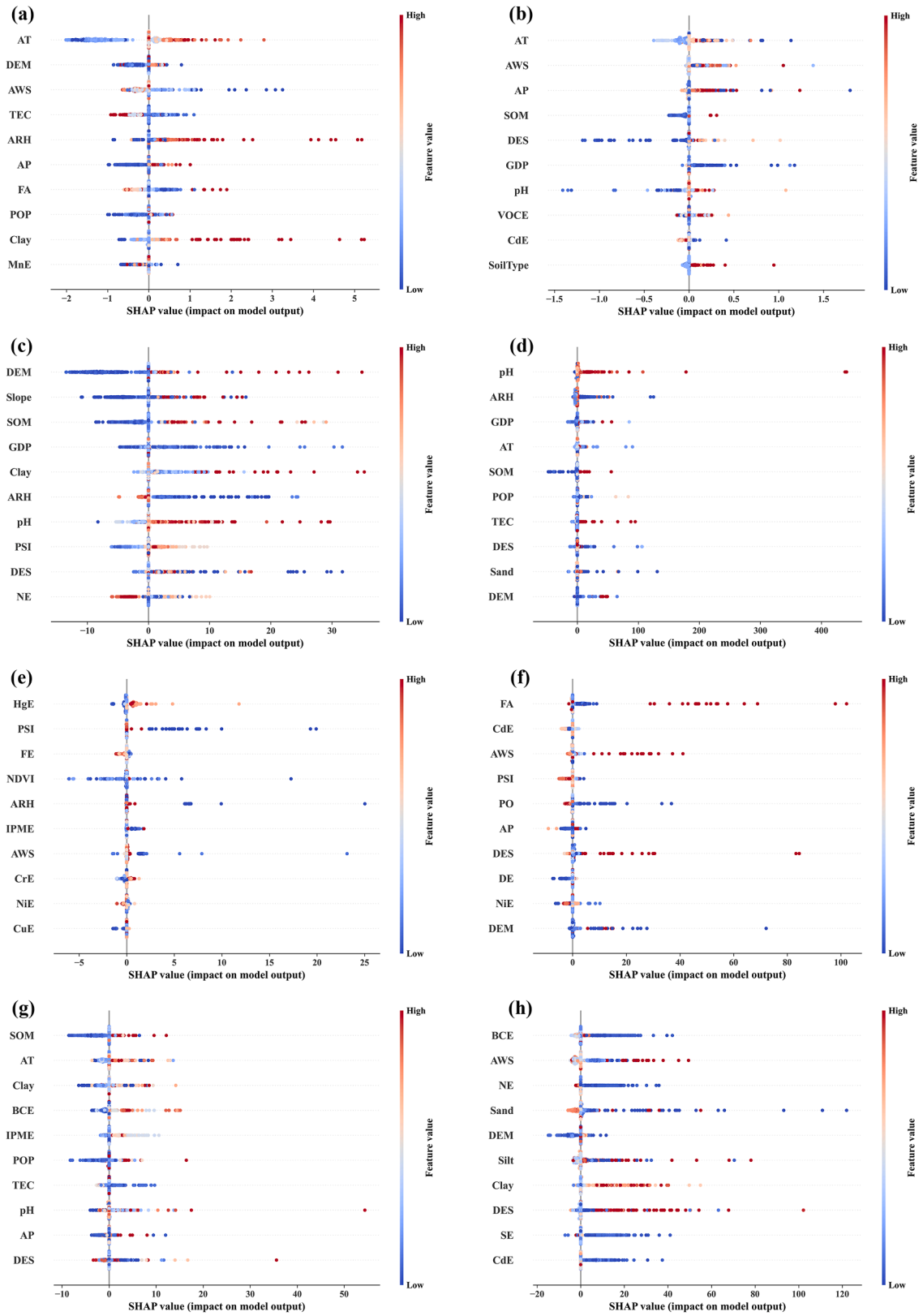


Fig. 2. Shapley additive explanation (SHAP) values of the 10 leading environmental variables for the contents of (a) As, (b) Cd, (c) Cr, (d) Cu, (e) Hg, (f) Ni, (g) Pb, and (h) Zn. The values were derived from random forest models incorporating natural environmental covariates and anthropogenic covariates for every metal. DEM, altitude; Slope, ground line gradient; AT, annual average temperature; AP, annual average precipitation; ARH, annual average relative humidity; AWS, annual average wind speed; pH, soil acidity and alkalinity; SOM, soil organic matter; Clay, clay content; Sand, sand content; Silt, silt content; SoilType, soil type; NDVI, normalized differential vegetation index; GDP, gross domestic product; POP, population density; PSI, proportion of secondary industry; TEC, total electricity consumption; FA, fertilizer application; PO, product output; DES, distance to emission sources. Abbreviations ending with "E" indicate emissions of the respective pollutants.

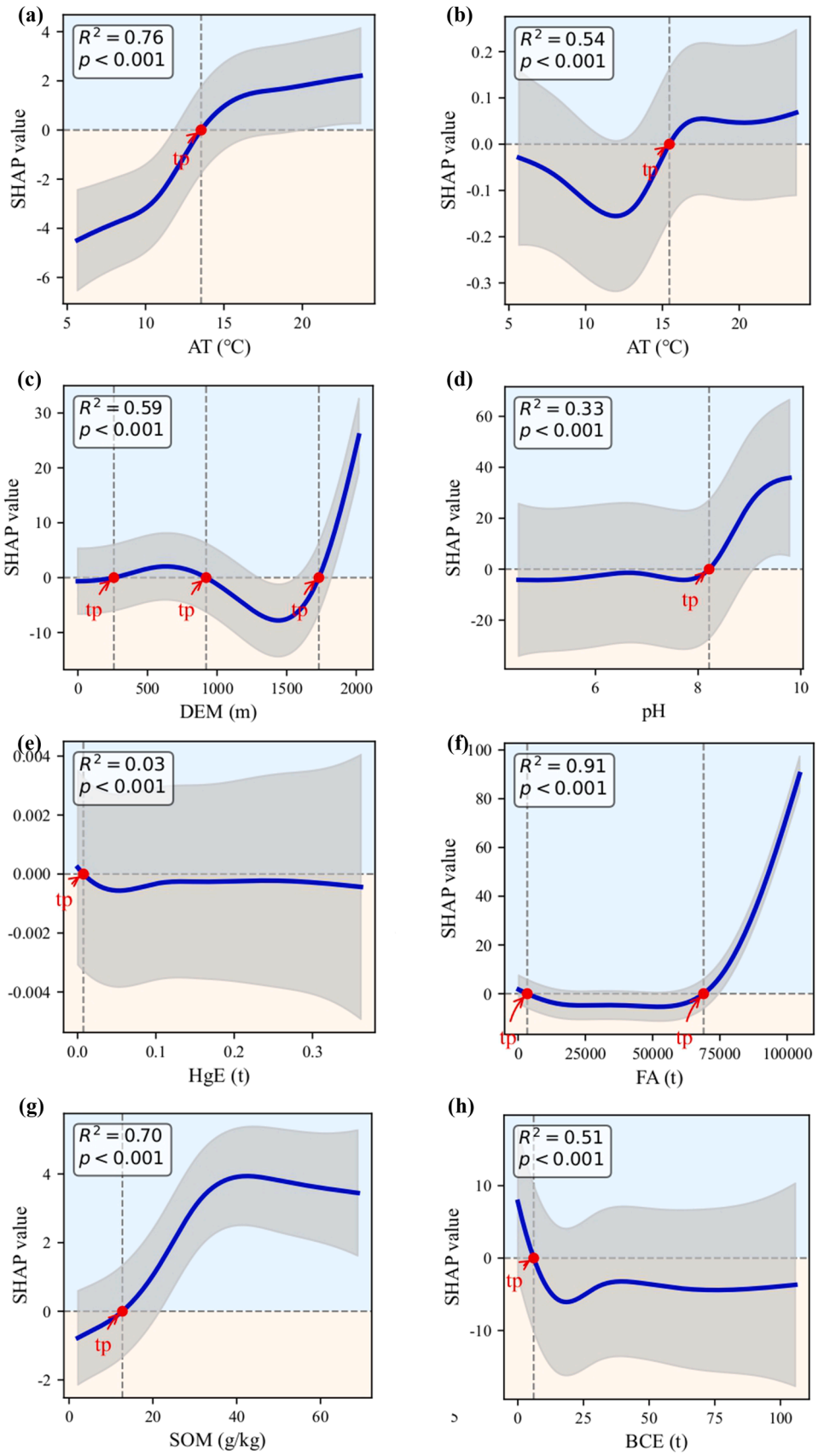


Fig. 3. Generalized additive model-based analysis of Shapley additive explanation (SHAP) values for the most influential factors associated with (a) As, (b) Cd, (c) Cr, (d) Cu, (e) Hg, (f) Ni, (g) Pb, and (h) Zn. The y-axis represents the SHAP values. The light pink and blue shaded areas indicate negative and positive effects on metal concentrations, respectively. AT, annual average temperature; DEM, altitude; pH, soil acid and alkaline; HgE, Hg emissions; FA, fertilizer application; SOM, soil organic matter; BCE, BC emissions.

For As, temperature had a negative effect below 13.60 °C and a positive effect above this threshold (Fig. 3a, R^2=0.76, $p < 0.001$). Cd showed a similar pattern: temperature had a negative effect below 15.51 °C but a positive effect above this threshold (Fig. 3b, R^2=0.54, $p < 0.001$). Temperature was also the second most important factor affecting Pb concentrations in soil. As observed for As and Cd, the effect on Pb was negative below 14.31 °C and positive above this threshold (Fig. S6g, R^2=0.67, $p < 0.001$). Notably, the temperature thresholds for these three metals were all close to 15 °C, showing highly consistent tipping-point behavior. More specifically, under relatively low mean annual temperature (MAT) conditions, temperature showed a negative association with the accumulation of soil As, Cd, and Pb. This pattern may reflect the higher temperature sensitivity of soil organic matter decomposition at low temperatures. Previous studies have demonstrated that the temperature sensitivity (Q_{10}) of soil organic matter decomposition generally decreases with increasing temperature, indicating that microbial decomposition is more strongly regulated by temperature under cooler conditions [70, 71]. As a result, within the lower temperature range, increasing temperature may enhance biogeochemical cycling through stimulated microbial activity and organic matter turnover, thereby reducing the accumulation of heavy metals in surface soils. The sensitivity of microbial processes to further warming appeared to decline, whereas prolonged exposure to higher temperatures may have enhanced geochemical processes such as mineral weathering, organic matter mineralization, and redox reactions when the mean annual temperature exceeded approximately 15 °C. Additionally, the temperature response of soil organic matter mineralization is jointly regulated by substrate quality and substrate availability rather than temperature alone [72], which may further stimulate soil biogeochemical cycling. In addition, previous studies have demonstrated that mean annual temperature (MAT) is a key driver governing the speciation of soil heavy metals and exhibits marked nonlinear threshold responses, indicating that temperature can influence heavy metal migration and transformation by regulating mineral weathering, organic matter decomposition, adsorption–desorption processes, and leaching dynamics [22, 60]. As, Cd, and Pb consistently showed a statistical response threshold at approximately 15 °C, indicating a similar nonlinear response pattern to temperature among these elements in the current study. However, it should be noted that this threshold reflects a nonlinear inflection point identified by the statistical model, reflecting the statistical relationship between environmental variables and soil heavy metal concentrations rather than a geochemically defined critical temperature. Therefore, the mechanistic interpretations proposed here are intended as plausible hypotheses supported by existing literature, and the underlying processes require further verification through controlled experiments and long-term field observations.

Annual average wind speed (AWS) was also an important control on As, Cd, Ni, and Zn concentrations (Fig. S5). For As, AWS had a positive effect below 2.02 m/s and a negative effect above this threshold (Fig. S7a, $R^2 = 0.56$, $p < 0.001$). Greater wind speed may enhance the dilution and transport of particulate-bound As, thereby reducing surface deposition and soil accumulation. Cd showed a more complex nonlinear response. AWS had a positive effect at 1.33–1.86 m/s, a negative effect at 1.86–2.40 m/s, and again a significant positive effect above 2.40 m/s (Fig. S6b, $R^2 = 0.36$, $p < 0.001$). This pattern was probably related to atmospheric transport and deposition: moderate wind favored horizontal dispersion and dilution, whereas stronger wind may have enhanced downward transport from higher atmospheric layers or promoted resuspension and redeposition [73]. For Ni, AWS had a positive effect below 1.79 m/s and a negative effect at 1.79–2.59 m/s; this negative effect became more evident above 2.59 m/s (Fig. S7f, R^2=0.42, $p < 0.001$). Zn showed comparable thresholds at 1.87 and 2.64 m/s, and its response resembled that of Ni (Fig. S6h, R^2= 0.62, $p < 0.001$). Thus, different wind-speed regimes regulated migration and deposition at the atmosphere–soil interface in distinct ways. Low wind favored particulate deposition and local accumulation, moderate wind enhanced dispersion and dilution, and high wind probably induced secondary transport or long-range redistribution, producing marked nonlinear responses in soil heavy metal concentrations.

Annual average precipitation (AP) had dual effects on metal cycling: greater mean annual precipitation enhanced weathering and metal inputs [60, 69], while also increasing losses through runoff and leaching [74, 75]. AP showed a nonlinear threshold effect on Cd concentrations (Fig. S7b, R^2=0.20, $p < 0.001$). The effect on Cd accumulation was positive below 566.01 mm and above 1515.38 mm but negative at 566.01–1515.38 mm. Below 566.01 mm, Cd inputs from weathering exceeded leaching losses, promoting accumulation in soil. Above 1515.38 mm, Cd again showed a positive response despite the expected increase in leaching. This response may reflect enhanced weathering of parent materials and Cd inputs, moisture-stimulated microbial adsorption and recycling, adsorption and complexation by soil organic matter and Fe-Mn oxides, and local topographic or depositional conditions favoring surface retention. Low and extremely high precipitation therefore promoted Cd enrichment through different mechanisms, whereas intermediate precipitation enhanced leaching and reduced Cd concentrations. Annual relative humidity (ARH) also showed a threshold effect on Cu: the effect was positive below 59.30% and negative above this threshold (Fig. S6d, R^2=0.36, $p < 0.001$). This pattern probably reflected the combined effects of humidity on mineral weathering, soil-solution transport, and microbial activity [76, 77]. Under low humidity, limited moisture restricted Cu leaching and migration, allowing Cu released by weathering to accumulate. Above the threshold, greater moisture promoted Cu migration or loss through leaching, complexation, and competition for adsorption sites.

3.2.3. Soil properties

Soil pH, soil organic matter (SOM), and clay content directly affected Cr, Cu, and Pb concentrations. SOM showed threshold-dependent switching effects on Cr and Pb. Concentrations were suppressed below the respective SOM thresholds of 10.64 g/kg for Cr and 12.89 g/kg for Pb but increased above these thresholds (Fig. S7c, $R^2 = 0.63$, $p < 0.001$; Fig. 3g, $R^2 = 0.70$, $p < 0.001$). Greater SOM may provide additional binding sites for Cr and Pb, reducing their mobility and leaching through enhanced complexation and thereby promoting soil accumulation.

Clay content also showed a threshold effect on Pb. Pb concentrations were suppressed below 229.47 g/kg clay but increased above this threshold (Fig. S7g, $R^2 = 0.62$, $p < 0.001$), probably because more adsorption sites on clay minerals promoted Pb fixation and enrichment [68, 69]. Soil pH had a pronounced threshold effect on Cu. The effect was negative below pH 8.22 and positive above this threshold (Fig. 3d, $R^2 = 0.33$, $p < 0.001$), likely because Cu speciation and its interactions with soil components vary with pH [78].

Under acidic to weakly alkaline conditions (pH < 8.22), Cu occurred mainly as divalent ions, experienced strong competitive adsorption with Fe-Mn oxides and clay minerals, and showed relatively higher solubility, making it more susceptible to migration through leaching or complexation processes, thereby resulting in lower soil Cu contents [79, 80]. Conversely, under more alkaline conditions (pH > 8.22), deprotonation of organic matter and mineral surfaces provided more negatively charged adsorption sites, while Cu more readily formed hydroxyl complexes or basic precipitates, enhancing its adsorption and fixation and leading to increased accumulation in soils [81]. Taken together, soil pH, SOM, and clay content regulated variations in Cr, Pb, and Cu contents by controlling key processes such as adsorption, complexation, and precipitation, and these effects showed distinct nonlinear threshold features.

3.2.4. Ecological functions

In this research, the normalized difference vegetation index (NDVI) was applied as a key indicator of ecological functions to represent regional vegetation cover and its regulatory capacity in controlling migration and transformation of pollutants [5]. The results indicated

that NDVI appeared only among the ten leading variables affecting the spatial pattern of Hg and ranked fourth in importance (Fig. S5). Hg showed higher sensitivity to vegetation cover, possibly because of its strong volatility and re-deposition characteristics compared with other heavy metals. Vegetation can influence atmospheric Hg deposition and redistribution through foliar interception, surface adsorption, and air-vegetation exchange processes [82]. As a result, NDVI showed greater explanatory power for predicting Hg but was not identified as a dominant variable for other heavy metals, highlighting the specific role of vegetation-mediated regulation in the air-soil cycling of Hg.

3.3. Effects of anthropogenic covariates on soil heavy metal concentrations surrounding iron and steel enterprises

3.3.1. Social economy

Following analysis of the natural environmental controls, the effects of socio-economic factors on heavy metal concentrations in soils around iron and steel enterprises were examined. GDP, fertilizer application, and the proportion of secondary industry (PSI) significantly affected Cu, Hg, and Ni concentrations (Fig. S4, $p < 0.05$). GDP was the third most important factor controlling Cu (Fig. S7d). It had no significant effect below 27320.33×10^4 yuan/km^2 but had a positive effect above this threshold (Fig. S7d, $R^2 = 0.30$, $p < 0.001$). High-GDP regions generally had more developed economies, more advanced industrial structures, and greater industrialization, potentially associated with concentrated or large-scale steel enterprises. Mature heavy-industry systems involve greater energy consumption, raw-material use, and production intensity, together with flue-gas emissions, storage of Cu-containing solid waste, and particulate release during transport, all of which may increase Cu inputs to surrounding soils [83]. PSI, the second most influential factor for Hg, had a U-shaped relationship with Hg concentrations. The effect was positive below 22.33% and above 65.75% but negative at 22.33%–65.75%. At low PSI, industrialization had begun while pollution-control infrastructure remained weak, allowing emissions from light and dispersed industrial sources to accumulate in soil [84, 85]. At high PSI, concentrated heavy industry and greater energy and coal use increased Hg emissions and soil accumulation. Regions with moderate PSI generally underwent industrial restructuring, including the closure or technological upgrading of highly polluting sectors and improved pollution control, resulting in lower Hg emissions and a negative effect.

Fertilizer application showed a significant positive effect on Ni concentrations in soil above a fertilizer application level of 69072.89 t (Fig. 3f, $R^2 = 0.91$, $p < 0.001$), indicating that Ni and co-occurring trace metals from fertilization exceeded soil buffering capacity and, together with industrial emissions, led to notable Ni enrichment. When fertilizer application ranged within 3616.36–69072.89 t, the effect was negative, potentially because moderate fertilization improved soil physicochemical properties and increased organic matter content, thus promoting Ni binding with organic matter or mineral particles and reducing its bioavailable concentration [86]. Taken together, these results showed that human activities increased heavy metal loads through industrial emissions and that agricultural management practices also significantly influenced the distribution of Ni in soil.

3.3.2. Industrial activity

Indicators of industrial activity, including emissions of different air pollutants or trace elements, were major determinants of concentrations of heavy metals (such as Hg and Zn) in soils (Fig. 3) [4, 87]. HgE, the most important factor affecting soil Hg contents, displayed a negative correlation when it exceeded 0.0087 t (Fig. 3e, $R^2 = 0.03$, $p < 0.001$), possibly because of localized negative effects arising from limited soil accumulation capacity or environmental regulation under low emission conditions. FE, another key factor affecting Hg content in soil, revealed threshold effects at 1.64 t and 47.54 t, respectively (Fig. S7e, $R^2 = 0.17$, $p < 0.001$). Notably, FE was positively correlated with Hg content in soil only within the range of 1.64–47.54 t, which may have enhanced Hg enrichment by altering soil properties or affecting Hg mobility [88]. Its effect on Hg content became negative, possibly because of excessive fluoride reducing the adsorption capacity of soils for Hg or enhancing its mobility and subsequent loss when FE exceeded 47.54 t. Taken together, although HgE and FE explained a limited proportion of total variance in Hg content in soil (as suggested by relatively low R2 values), statistically significant p-values indicated that their local effects and trends were reliable. The identified nonlinear threshold effects revealed multifaceted regulatory mechanisms through which industrial emissions influenced soil Hg distribution.

For Zn, BCE and NE were two important influencing factors. Their effects became negative above thresholds of 6.34 t and 527.69 t, respectively (Fig. 3h and Fig. S7h), consistent with the Spearman correlation results (Fig. S4). This reversal did not indicate that the emissions suppressed Zn accumulation; rather, it reflected a nonlinear relationship between emission intensity and the contribution of Zn inputs. Below these thresholds, increases in BCE and NE generally corresponded to larger metal inputs or more intensive activity and thus greater Zn accumulation. Above the thresholds, further emission increases were no longer accompanied by corresponding Zn accumulation, indicating a diminishing marginal contribution or weaker association under the combined effects of multiple sources and processes. CdE had a comparable effect on Ni: above 0.16 t, the effect became negative (Fig. S6f, $p < 0.001$). This trend did not indicate that Cd emissions suppressed Ni accumulation but instead reflected a substantial reduction in the explanatory power of CdE for regional Ni inputs or migration under high-emission conditions.

3.4. Interactions among important factors and mechanistic insights into soil heavy metal concentrations surrounding iron and steel enterprises

The present study further investigated pairwise interactions among principal factors affecting soil heavy metal concentrations using the SHAP method based on a random forest model (Fig. 4, Figs. S8 and S9). Collectively, these findings suggested co-regulated mechanisms in which atmospheric processes, topographic features, soil physicochemical properties, and anthropogenic emissions collectively controlled accumulation of heavy metals in the environment. Taking climate factors as an example, individual combinations of annual average temperature with altitude, annual average wind speed, and annual average precipitation revealed greater SHAP variability for most metals (e.g., As, Cd, Pb) under moderate environmental gradients. This indicated that deposition, transport, or dilution processes of metals were most sensitive to external disturbances within this range. Interactions among factors regulating specific soil heavy metal forms also displayed distinct patterns. High elevation, low annual average wind speed, or comparatively low annual average precipitation were associated with enhanced deposition or limited dispersion, thereby affecting enrichment levels of As, Cd, and Zn. Interactions between geological conditions and soil properties (e.g., elevation-slope, SOM–clay content) revealed distinct threshold characteristics for metals such as Cr and Pb, suggesting that surface structure and physicochemical conditions collectively controlled their retention or mobility. Coupling between anthropogenic factors (e.g., black carbon emissions, NO_x emissions, Cd emissions, and industrial structure) and meteorological conditions revealed mixed positive and negative responses for Hg, Ni, and Zn, suggesting that intensity of emissions and diffusion background together determined the direction and magnitude of surface accumulation. Collectively, the integrated interaction effects of individual heavy metals suggested a consistent pattern: spatial distribution of each metal could not be explained independently by a single factor, and their environmental behavior was typically driven jointly by several variables, with sensitivity peaking under moderate environmental gradients.

The preceding analysis established a machine learning framework for predicting heavy metal concentrations and identified the dominant

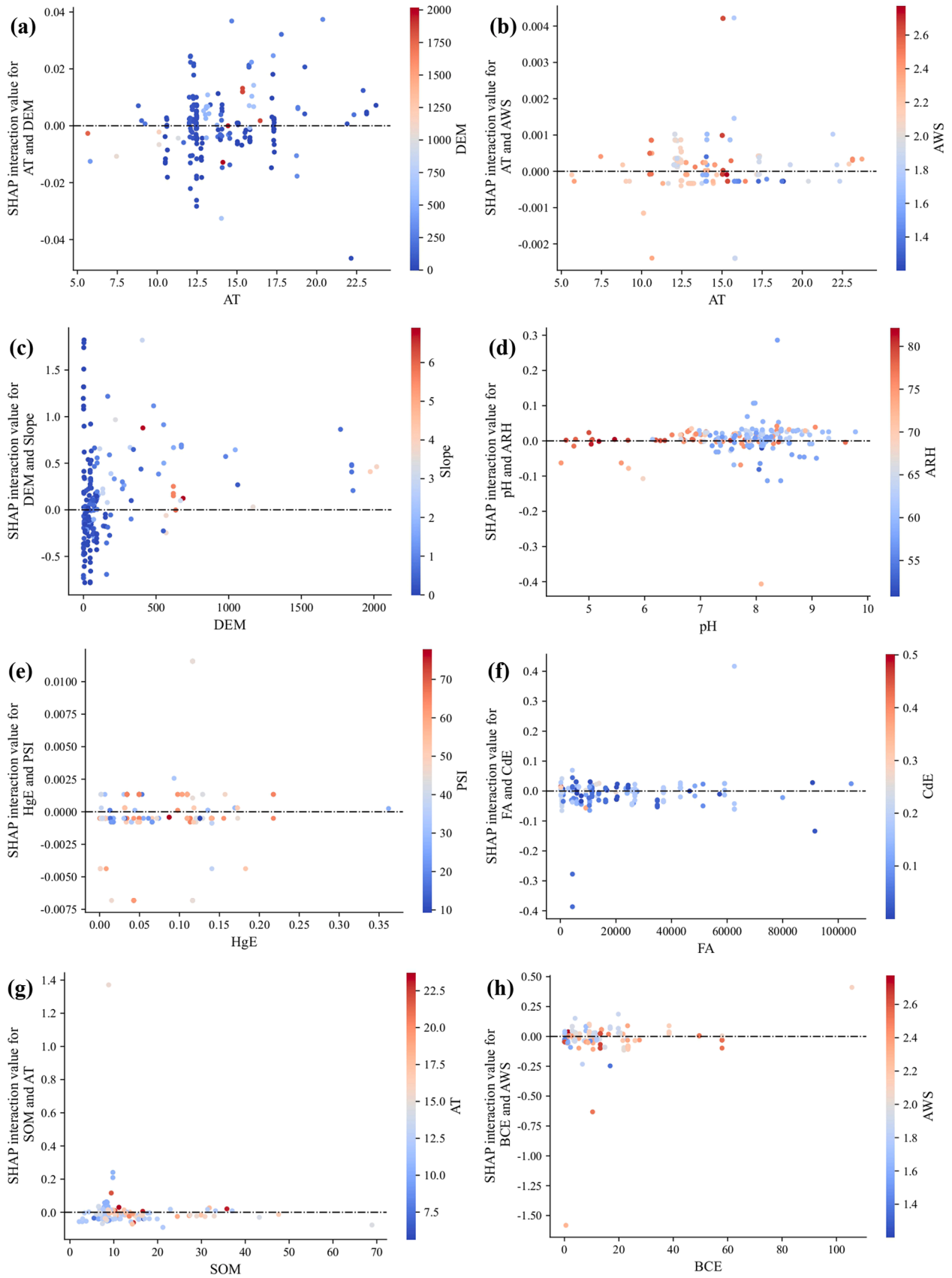


Fig. 4. Shapley additive explanation (SHAP) values of the interaction between the two most important factors for (a) As, (b) Cd, (c) Cr, (d) Cu, (e) Hg, (f) Ni, (g) Pb, and (h) Zn. The y-axis in these figures represents SHAP values. AT, annual average temperature; DEM, altitude; AWS, annual average wind speed; Slope, ground line gradient; pH, soil acidity and alkalinity; ARH, annual average relative humidity; HgE, Hg emissions; PSI, proportion of secondary industry; FA, fertilizer application; CdE, Cd emissions; SOM, soil organic matter; BCE, BC emissions.

natural and anthropogenic covariates using SHAP values. Generalized additive models (GAMs) were then used to characterize nonlinear responses and potential thresholds of the major drivers. However, SHAP and GAM contributions are based mainly on statistical model inference and do not directly represent the spatial heterogeneity of the drivers or their explanatory power for spatial patterns in heavy metal concentrations. Geographic detector analysis was therefore used to evaluate the spatial explanatory power (q value) of the major factors and their interactions. This analysis validated the robustness of the machine learning results and improved understanding of the mechanisms underlying the spatial distributions of heavy metals.

Considering the large number of natural environmental covariates and anthropogenic covariates, direct inclusion of all variables could make the analyses overly complex and difficult to interpret. Thus, for every heavy metal, SHAP values from the machine learning models were employed to rank driving factors, and the ten most influential variables were chosen for geographic detector analysis. Subsequently, single-factor q values of principal variables were determined for every metal, and pairwise interactions were then examined to reveal spatial heterogeneity and combined influences on heavy metal concentration patterns. This approach integrated importance inferred from statistical modeling with a spatial heterogeneity perspective, thereby improving the mechanistic understanding of heavy metal spatial distributions.

For the 10 principal SHAP-ranked drivers of heavy metal spatial distributions in soils around steel enterprises, geographic detector interaction analysis revealed that spatial heterogeneity of all examined metals was generally governed by multi-factor coupling, with nonlinear enhancement as the dominant interaction mode (Fig. S10). For most pairwise combinations, the interaction explanatory power substantially exceeded that of individual factors and generally exceeded max(A, B), indicating that meteorological conditions (e.g., AT, AWS, AP, ARH), soil physicochemical properties (e.g., pH, TEC, clay, SOM, sand, silt, DEM, slope), geological conditions, socio-economic factors (e.g., POP, PSI), and industrial activity indicators (e.g., DE, IPME, CuE, CdE, NiE, HgE) produced a cumulative amplifying effect on spatial distributions of heavy metals. Certain metals displayed two-factor enhancement effects under specific factor combinations. For instance, Cu showed a marked increase in spatial explanatory power under the combined effect of temperature (AT) and altitude (DEM); Zn was notably sensitive to combinations of black carbon emissions (BCE)–sand content (Sand) and NO_x emissions (NE)–altitude (DEM); Ni revealed notable effects under the interaction between fertilizer application (FA) and Cd emissions (CdE); Pb revealed enhancement for TEC–AP, As for AT–ARH and AP–POP, and Hg for HgE–NiE and PSI–AWS combinations. These findings showed that interactions among specific factors could substantially enhance their explanatory power for spatial distributions of heavy metals.

In particular, spatial patterns of individual heavy metals were clearly controlled by coupled effects of several sources. Cu, Zn, and Ni were mainly driven by joint effects of meteorological conditions, soil background conditions, and anthropogenic emissions, with their spatial heterogeneity demonstrating strong nonlinear additive effects. Distributions of Pb, Cr, and Cd largely depended on interactions among soil properties, geological conditions, and regional human activities, indicating a typical multidimensional coupling mechanism. The enrichment and transport of As and Hg were affected by complex nonlinear interactions among meteorological variables, soil background conditions, land use, and emission sources. Collectively, the formation mechanisms of all examined heavy metals displayed pronounced multidimensional interactions and nonlinear characteristics, suggesting that single variables could not adequately explain their spatial patterns and that these distributions should be interpreted from the perspective of coupled driving forces.

3.5. Emission performance of iron and steel enterprises and provincial-scale soil heavy metal contamination assessment

Based on the province-level weighted average score S_p, emission-control performance of the iron and steel industry was divided into three levels: high performance ($S_p \geq 0.80$), medium performance ($0.50 < S_p < 0.80$), and low performance ($S_p \leq 0.50$). These scores represent the relative performance of enterprises across multiple pollutants and should not be interpreted as direct measures of the environmental or health hazards of individual pollutants. They should therefore be regarded as indicators of comprehensive emission performance rather than environmental or health risk. Shanghai, Guangdong, and Shaanxi were classified as low-performance provinces, indicating considerable scope for improving pollution control, operational stability of abatement facilities, and compliance with emission standards. Medium-performance provinces occurred mainly in northern China, including Hebei, Shanxi, and Inner Mongolia, and in central and southern China, including Hunan and Henan (Table S12). These provinces have established a foundation for emission control but retain structural potential for improvement. High-performance provinces were concentrated in eastern China, including Anhui, Shandong, Jiangxi, and Zhejiang, and in northeastern China, including Jilin and Heilongjiang. These regions generally had more advanced emission-reduction measures, wider adoption of pollution-control technologies, and better production-process management, indicating comparatively strong governance and effective mitigation. The performance distribution demonstrates regional heterogeneity in emission control across China's iron and steel industry and supports spatially differentiated analyses and targeted policy interventions.

Sample-based results showed that soil heavy metal contamination around iron and steel enterprises differed among provinces (Table S13). Contamination factors (CF) and the pollution load index (PLI) were therefore calculated. As shown in Fig. 5, the CF values revealed distinct spatial patterns among the metals. Cd and Hg were the dominant pollutants: Cd reached high pollution levels ($CF \geq 6$) in Shanxi, Zhejiang, and Ningxia, while Hg showed markedly high values in Qinghai, Ningxia, Guangdong, and Guizhou. Pb reached a considerable pollution level ($3 \leq CF < 6$) in Shaanxi, Gansu, and Liaoning, whereas Ni reached high pollution ($CF \geq 6$) in Jilin and clearly exceeded the levels in other regions. Cu and Zn generally showed moderate pollution, although Cu reached a considerable level in Hebei and Xinjiang. Cr and As contamination was usually low, with most provinces remaining at low to moderate levels.

Fig. 6a presents the provincial distribution of the pollution load index for eight heavy metals in soils around iron and steel enterprises. Overall, six provinces were classified as "unpolluted" (Class 0), and twenty-two were classified as "slightly polluted" (Class I). Within Class I, comparatively high PLI values occurred in the northwestern provinces of Ningxia and Gansu, the northern provinces of Inner Mongolia and Shanxi, the northeastern province of Liaoning, the central-southern provinces of Hunan and Guangdong, and the eastern province of Zhejiang.

The RI index was used to evaluate the ecological risks of eight heavy metals in soils around iron and steel enterprises across regions (Fig. 6b). RI values showed clear spatial differences among provinces. Ecological risk was comparatively high in northern and northeastern regions, including Hebei, Shanxi, Tianjin, Jilin, and Heilongjiang, whereas most eastern and southwestern provinces had moderate risk. Qinghai, Shanghai, Chongqing, Guangxi, Shandong, and Xinjiang showed low ecological risk. The high-risk provinces in northern and northeastern China were mainly traditional steel-producing areas, while ecological risk in northwestern and southwestern China was generally moderate. These patterns demonstrate regional heterogeneity in the potential ecological risks of soil heavy metals in China and provide a scientific reference for region-specific pollution control and ecological

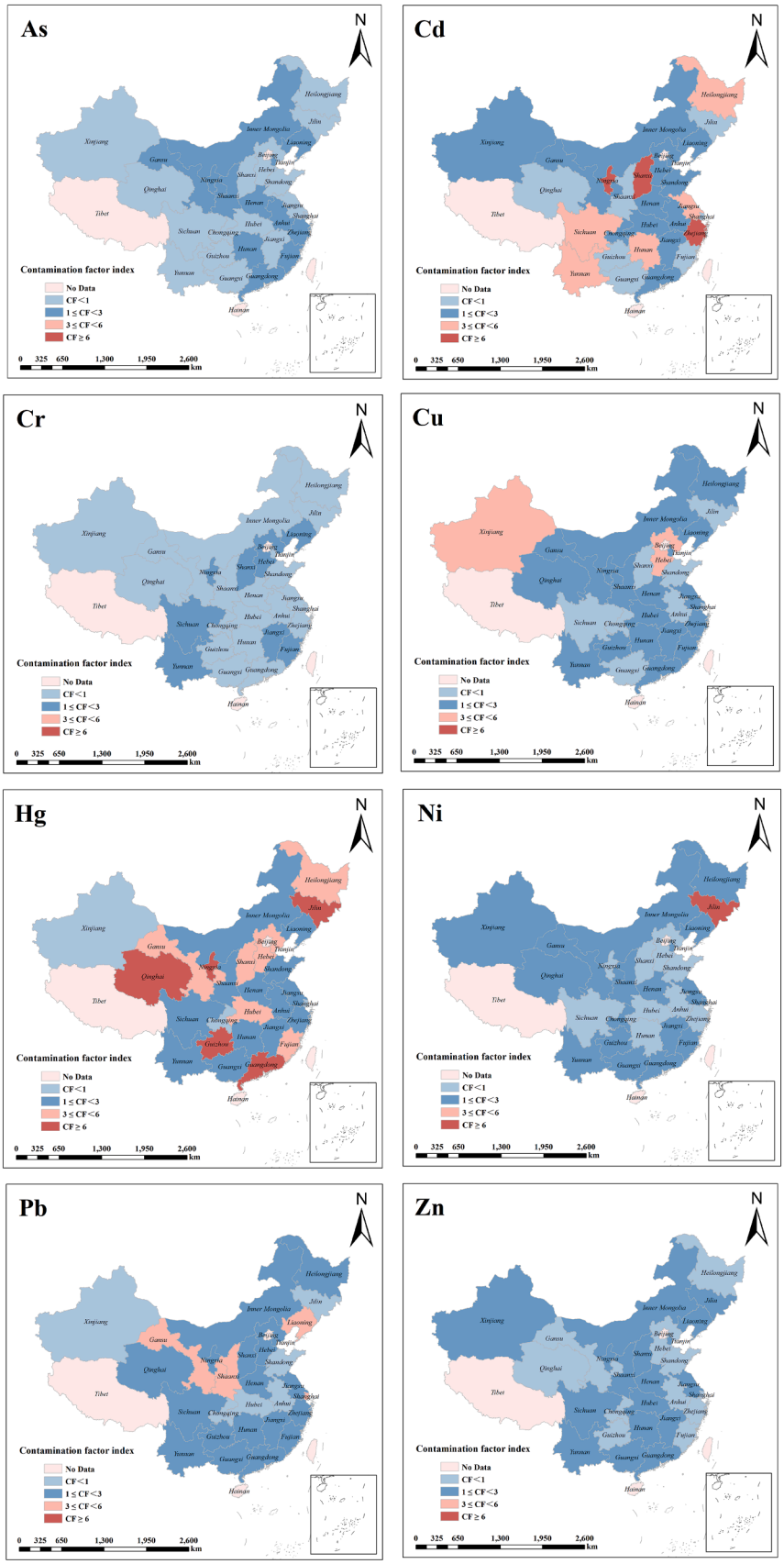


Fig. 5. Spatial distribution of eight trace metal(loid)s in soils surrounding iron and steel enterprises based on contamination factor (CF) index assessment.

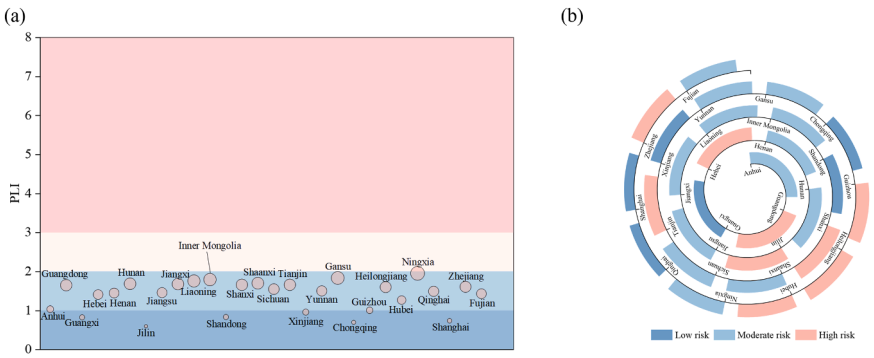


Fig. 6. PLI assessment (a) and potential RI assessment (b) of trace metal(loid)s in soils around iron and steel enterprises.

management.

3.6. Policy implications

Clear spatial differences were found in emission-control performance of the iron and steel industry, soil heavy metal contamination, and related ecological risks in China, providing a basis for differentiated and targeted policy measures. First, emission performance was closely associated with contamination levels in soils around iron and steel enterprises. In low-performance provinces, including Shanghai, Guangdong, and Shaanxi, contamination factors (CF) for Pb, Hg, and other metals were comparatively high. Hg reached a high pollution level in Guangdong, while Pb reached a considerable level in Shaanxi and Shanghai, indicating insufficient pollution-control capacity and inadequate operational stability of end-of-pipe treatment facilities. Provinces with medium performance, including Hebei, Shanxi, Inner Mongolia, Hunan, and Henan, had established a foundation for emission reduction; however, Cd, Hg, and Pb contamination in some areas remained moderate to considerable and occasionally reached high levels. For example, Cd pollution was high in Shanxi and considerable in Hunan, while Hg pollution was considerable in Hebei and Shanxi. Further improvement in governance structures and application of control technologies is therefore still required.

High-performance provinces, including Anhui, Shandong, Jiangxi, Zhejiang, Jilin, and Heilongjiang, generally showed lower soil heavy metal contamination, and most elements remained at low to moderate levels. Ecological risks in Anhui, Jiangxi, and Shandong were mainly low to moderate, reflecting comparatively effective ultra-low-emission retrofits, pollution-control technologies, and production-process management. Existing achievements should be maintained in these regions while green process innovation and cleaner production continue to be promoted. By contrast, Jilin, Heilongjiang, and Zhejiang showed high ecological risk, mainly because of the intensified effects of highly toxic Cd and Hg. Risk management in these provinces should therefore shift from total-quantity control toward targeted regulation of high-risk elements. Second, policy measures should account for regional differences and adopt governance strategies suited to local performance and contamination levels. In high-pollution and high-risk areas, emission-control capacity, soil monitoring, and remediation should be strengthened, and efficient emission-reduction and cleaner-production technologies should be promoted to control both sources and environmental risks. Source-oriented strategies should also reflect the actual processes used by iron and steel enterprises. Emissions from sintering, coking, and blast-furnace operations require particular attention, together with improved operation and maintenance of dust-collection systems, tighter control of fugitive emissions, and broader use of cleaner-production technologies. In moderate-risk regions, current governance should be maintained while technological structures, management efficiency, monitoring of major pollutants, and source control are improved. Low-risk regions should focus on ecological protection and pollution prevention to protect soil quality and ecosystem integrity from new emission pressures. Finally, a regionally differentiated integrated framework should combine industry emission performance, soil heavy metal contamination, and ecological risk within a unified assessment system. Integration of emission mitigation with soil ecological risk management can support graded and precise policies, sustainable development of iron and steel enterprises, and regional ecological security.

3.7. Limitations

Several limitations may affect the overall comprehensiveness of the findings. First, the analysis was based on publicly available soil monitoring and enterprise emission data, limiting spatial coverage to 28 provinces. Although this may reduce national representativeness, the included regions encompass substantial environmental heterogeneity and are sufficient to capture the dominant spatial patterns of soil heavy metal contamination. Second, the machine learning framework combined with SHAP analysis effectively identified influential drivers and their relative importance, but the results depend on the availability, completeness, and accuracy of the input variables. Relevant factors omitted from the dataset may therefore introduce uncertainty into the estimated contribution of each driver. Nonlinear responses and potential tipping points were investigated using SHAP-GAM, while pairwise interaction analysis and geographic detector methods were used to characterize synergistic effects and the spatial explanatory power of major factors. Because these analyses were based mainly on cross-sectional observations, the identified thresholds, interactions, and spatial differentiation patterns should be interpreted as statistical associations rather than causal relationships. Larger, long-term monitoring datasets and independent validation data should be incorporated in future research to evaluate the stability, robustness, and transferability of nonlinear thresholds across temporal scales and environmental settings. Soil heavy metal contamination and ecological risk were assessed from measured concentrations using common indices, including the contamination factor, pollution load index, and potential ecological risk index. These indices do not directly account for dynamic soil processes, long-term accumulation, or diffuse nonpoint sources, which may affect the accuracy of contamination and risk assessment. Atmospheric transport processes, including wind direction, atmospheric dispersion, and dry and wet deposition, were also not explicitly incorporated. Consequently, the enterprise emission variables reflected statistical relationships between industrial emission intensity and soil heavy metal concentrations rather than actual pollution exposure at individual sampling locations. Future work should expand spatial coverage through larger and denser soil monitoring datasets and should develop region-specific models for different climatic zones, soil types, and topographic conditions. These efforts would clarify the mechanisms of heavy metal migration and accumulation under varied environmental settings, reveal regional differences in the importance of individual

drivers, and improve the regional applicability of the proposed framework.

In addition, atmospheric dispersion or regional chemical transport models should be integrated with high-resolution meteorological data to simulate pollutant transport, dispersion, and deposition more realistically. Environmental and socioeconomic variables with finer spatial and temporal resolutions should also be combined with long-term monitoring datasets and multisource environmental information to characterize the spatial exposure relationship between industrial emissions and soil heavy metal accumulation more accurately. Future studies should integrate process-level emission inventories with continuous emission-monitoring data from steel enterprises to quantify the contributions of different production processes to off-site soil heavy metal accumulation. This approach would identify critical emission stages, improve source-specific attribution of soil contamination, and provide a stronger scientific basis for more effective and targeted pollution prevention and control. These improvements would enhance the physical interpretability and predictive reliability of the model and provide deeper insight into the drivers and ecological risks of soil heavy metal pollution near steel enterprises.

4. Conclusions

Natural and anthropogenic covariates affecting heavy metal contamination in soils around iron and steel enterprises were systematically identified. SHAP analysis was integrated with generalized additive models (GAMs) to determine tipping points for individual factors and reveal nonlinear threshold effects. Heavy metal concentrations near steel plants were jointly controlled by natural and anthropogenic factors and showed clear nonlinear responses. Topography and soil properties affected metal accumulation through hydrothermal redistribution, organic matter dynamics, and adsorption. Climate variables, especially temperature, precipitation, and wind speed, were dominant controls, with consistent tipping points of ~15 °C for As, Cd, and Pb. Vegetation coverage mainly affected Hg, whereas socio-economic conditions and industrial emissions significantly modulated the distributions of Cu, Hg, Ni, and Zn, often with nonlinear or diminishing responses at high input levels. These findings demonstrate the complex, threshold-dependent interactions through which environmental and human drivers shape heavy metal accumulation and spatial patterns in soil, providing a scientific basis for targeted pollution control and risk management.

Pairwise interactions among major drivers of soil heavy metals were further evaluated using SHAP values from random forest models. Climate variables, including temperature (tipping point of about 15 °C), wind speed (1.33–2.64 m/s), and precipitation (566–1515 mm), showed marked SHAP variability under moderate gradients, indicating sensitive metal deposition and migration. Interactions between topography and soil, such as elevation–slope and SOM–clay, and between anthropogenic and climatic factors, such as Hg emissions–Ni emissions and PSI–AWS, substantially affected specific metals. Most factor pairs enhanced explanatory power nonlinearly and often exceeded the maximum single-factor q-values obtained by Geodetector analysis. For example, Cu accumulation was strongly amplified by the interaction between average temperature and elevation, Zn by bulk density and sand content, and Ni by fine soil fraction and Cd emissions. The spatial heterogeneity of As, Cd, Pb, Cr, Cu, Zn, Ni, and Hg was therefore governed by multifactorial, threshold-dependent interactions, highlighting the importance of coupled natural and anthropogenic drivers in risk assessment and management.

At the provincial level, emission performance of China's iron and steel industry was classified as high ($S_p \geq 0.80$), medium ($0.50 < S_p < 0.80$), or low ($S_p \leq 0.50$). Low performance occurred in Shanghai and Guangdong, while most northern and central-southern provinces showed moderate performance. Comparatively high performance was found in several eastern provinces, including Anhui, Shandong, Jiangxi, and Zhejiang, and in northeastern provinces such as Jilin and Heilongjiang. Contamination patterns varied considerably among heavy metals. Cd and Hg were the most prominent pollutants; Pb reached a considerable pollution level in Shaanxi, Gansu, Liaoning, and other provinces, whereas Ni reached a high level in Jilin Province. Cu and Zn generally showed moderate pollution, while Cr and As remained relatively low. Based on PLI, about 21% of the provinces were classified as the "unpolluted" level (Class 0), and about 79% were categorized as "slightly polluted" level (Class I). RI assessment indicated elevated ecological risks in northern and northeastern regions. These findings provide a quantitative basis for region-specific pollution mitigation and risk management. Precision management should focus on principal drivers and high-risk areas, with optimized emission-control and soil-monitoring systems used to reduce heavy metal contamination and potential ecological risks and promote sustainable development of the iron and steel industry.

Environmental implication

The mechanisms through which natural and anthropogenic covariates influence heavy metals in soils outside the boundaries of iron and steel enterprises were systematically examined, revealing nonlinear and threshold-dependent responses to coupled drivers. The entropy-weighted TOPSIS method was also used to evaluate atmospheric pollutant emission performance among provinces and to assess soil heavy metal pollution and ecological risks. The findings offer a scientific basis for identifying high-risk areas, implementing precise emission management, and coordinating multifactor governance. They also provide methodological support for green and low-carbon development in the steel industry and for regional environmental management.

Statement of significance

The mechanisms through which natural and anthropogenic covariates influence heavy metals in soils outside the boundaries of iron and steel enterprises were systematically examined, revealing nonlinear and threshold-dependent responses to coupled drivers. The entropy-weighted TOPSIS method was also used to evaluate atmospheric pollutant emission performance among provinces and to assess soil heavy metal pollution and ecological risks. The findings offer a scientific basis for identifying high-risk areas, implementing precise emission management, and coordinating multifactor governance. They also provide methodological support for green and low-carbon development in the steel industry and for regional environmental management.

CRediT authorship contribution statement

Xin Xu: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Zheng Zhang:** Software, Methodology, Data curation. **Ziang Ma:** Software, Methodology, Conceptualization. **Peng Wang:** Software, Methodology, Conceptualization. **Xiaoqing Zhang:** Data curation. **Xin Bo:** Writing – review & editing, Supervision, Resources, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by grants from the Fundamental Research Funds for the Central Universities (buctrc202133).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2026.143639.

Data availability

Data will be made available on request.

References

[1] Chen, L., Zhou, M., Wang, J., Zhang, Z., Duan, C., Wang, X., Zhao, S., Bai, X., Li, Z., Li, Z., Fang, L., 2022. A global meta-analysis of heavy metal (loid) s pollution in soils near copper mines: Evaluation of pollution level and probabilistic health risks. Sci Total Environ 835, 155441. https://doi.org/10.1016/j.scitotenv.2022.155441.

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

[3] Hou, D., Jia, X., Wang, L., McGrath, S.P., Zhu, Y.-G., Hu, Q., Zhao, F.-J., Bank, M.S., O'Connor, D., Nriagu, J., 2025. Global soil pollution by toxic metals threatens agriculture and human health. Science 388, 316–321. https://doi.org/10.1126/science.adr5214.

[4] Hou, D., O'Connor, D., Igalavithana, A.D., Alessi, D.S., Luo, J., Tsang, D.C.W., Sparks, D.L., Yamauchi, Y., Rinklebe, J., Ok, Y.S., 2020. Metal contamination and bioremediation of agricultural soils for food safety and sustainability. Nat Rev Earth Env 1, 366–381. https://doi.org/10.1038/s43017-020-0061-y.

[5] Yan, Y., Yang, Y., 2025. Revealing the synergistic spatial effects in soil heavy metal pollution with explainable machine learning models. J Hazard Mater 482, 136578. https://doi.org/10.1016/j.jhazmat.2024.136578.

[6] He, Y., Yang, H., Fang, R., Wen, Z., Gong, P., Xia, R., Zhou, J., Zhou, J., 2026. Soil properties regulate the fate and accumulation of atmospherically deposited trace metals in rice across six major growing regions of China. Environ Sci Technol 60, 19597–19610. https://doi.org/10.1021/acs.est.6c04551.

[7] Xia, R., Xu, X., Zeng, Z., Sun, Y., Cui, H., Zhou, J., Zhou, J., 2026. Cadmium isotopes reveal dominant atmospheric deposition pathways in agroecosystems. Environ Sci Technol 60, 4159–4170. https://doi.org/10.1021/acs.est.5c16224.

[8] Ma, X., Sha, Z., Li, Y., Si, R., Tang, A., Fangmeier, A., Liu, X., 2024. Temporal-spatial characteristics and sources of heavy metals in bulk deposition across China. Sci Total Environ 926, 171903. https://doi.org/10.1016/j.scitotenv.2024.171903.

[9] Wu, X., Meng, J., Liang, X., Sun, L., Coffman, D.M., Kontoleon, A., Guan, D., 2025. Technological pathways for cost-effective steel decarbonization. Nature 647, 93–101. https://doi.org/10.1038/s41586-025-09658-9.

[10] Zhang, J., Shen, H., Chen, Y., Meng, J., Li, J., He, J., Guo, P., Dai, R., Zhang, Y., Xu, R., Wang, J., Zheng, S., Lei, T., Shen, G., Wang, C., Ye, J., Zhu, L., Sun, H.Z., Fu, T.-M., Yang, X., Guan, D., Tao, S., 2023. Iron and Steel Industry Emissions: A Global Analysis of Trends and Drivers. Environ Sci Technol 57, 16477–16488. https://doi.org/10.1021/acs.est.3c05474.

[11] Li, J., Ma, H., Yu, H., Feng, L., Xia, X., He, S., Chen, X., Zhao, Q., Wei, L., 2024. Effect and potential mechanisms of sludge-derived chromium, nickel, and lead on soil nitrification: Implications for sustainable land utilization of digested sludge. J Hazard Mater 466, 133552. https://doi.org/10.1016/j.jhazmat.2024.133552.

[12] Zhang, S., Zhang, J., Niu, L., Chen, Q., Zhou, Q., Xiao, N., Man, J., Ma, J., Wei, C., Zhang, S., Luo, Y., Yao, Y., 2024. Escalating arsenic contamination throughout Chinese soils. Nat Sustain 7, 766–775. https://doi.org/10.1038/s41893-024-01341-7.

[13] Moreno-Jiménez, E., Maestre, F.T., Flagmeier, M., Guirado, E., Berdugo, M., Bastida, F., Dacal, M., Díaz-Martínez, P., Ochoa-Hueso, R., Plaza, C., Rillig, M.C., Crowther, T.W., Delgado-Baquerizo, M., 2023. Soils in warmer and less developed countries have less micronutrients globally. Glob Change Biol 29, 522–532. https://doi.org/10.1111/gcb.16478.

[14] Moreno-Jiménez, E., Plaza, C., Saiz, H., Manzano, R., Flagmeier, M., Maestre, F.T., 2019. Aridity and reduced soil micronutrient availability in global drylands. Nat Sustain 2, 371–377. https://doi.org/10.1038/s41893-019-0262-x.

[15] Meharg, A.A., Meharg, C., 2021. The Pedosphere as a Sink, Source, and Record of Anthropogenic and Natural Arsenic Atmospheric Deposition. Environ Sci Technol 55, 7757–7769. https://doi.org/10.1021/acs.est.1c00460.

[16] Nriagu, J.O., 1989. A global assessment of natural sources of atmospheric trace metals. Nature 338, 47–49. https://doi.org/10.1038/338047a0.

[17] Liu, X., Tournassat, C., Grangeon, S., Kalinichev, A.G., Takahashi, Y., Marques Fernandes, M., 2022. Molecular-level understanding of metal ion retention in clay-rich materials. Nat Rev Earth Env 3, 461–476. https://doi.org/10.1038/s43017-022-00301-z.

[18] Imseng, M., Wiggenhauser, M., Keller, A., Müller, M., Rehkämper, M., Murphy, K., Kreissig, K., Frossard, E., Wilcke, W., Bigalke, M., 2018. Fate of Cd in Agricultural Soils: A Stable Isotope Approach to Anthropogenic Impact, Soil Formation, and Soil-Plant Cycling. Environ Sci Technol 52, 1919–1928. https://doi.org/10.1021/acs.est.7b05439.

[19] Yang, L., Li, S., Wen, T., Meng, F., Chen, G., Qian, X., 2020. Influence of ferrous-metal production on mercury contamination and fractionation in farmland soil around five typical iron and steel enterprises of Tangshan, China. Ecotox Environ Safe 188, 109774. https://doi.org/10.1016/j.ecoenv.2019.109774.

[20] Xu, X., Zhang, Z., Wang, P., Ma, Z., Wang, Z., Zhao, Z., Wang, Q., Bo, X., 2025. Integrated model for heavy metal source tracing and multi-factor quantification, with risk assessment: A case study of a typical mega iron and steel enterprise in China. Ecotox Environ Safe 305, 119180. https://doi.org/10.1016/j.ecoenv.2025.119180.

[21] Lin, B., Wang, X., 2015. Carbon emissions from energy intensive industry in China: Evidence from the iron & steel industry. Renew Sust Energy Rev 47, 746–754. https://doi.org/10.1016/j.rser.2015.03.056.

[22] Bai, Z., Li, F., Zhang, S., Li, T., Wang, G., Xu, X., Pan, X., Zhong, Q., Zhou, W., Pu, Y., Jia, Y., 2025. Shapley additive explanations revealed the effects of climate, topography, anthropogenic activities and geochemical properties on soil cobalt, nickel and vanadium contents. J Hazard Mater 495, 139070. https://doi.org/10.1016/j.jhazmat.2025.139070.

[23] Sun, Y., Zeng, Y., Zhang, B., Zhao, X., Wu, F., Zhao, Y., Hao, L., Zhao, X., 2026. Disentangling anthropogenic inputs from natural heterogeneity: An adaptive machine learning-Geostatistics workflow for soil heavy metal risk assessment. Environ Chem Ecotoxicol 8, 2047–2060. https://doi.org/10.1016/j.enceco.2026.05.001.

[24] Yi, S., Li, X., Chen, W., 2026. An integrated framework coupling process driven model and machine learning to simulate soil heavy metal (loid) s accumulation and predict regional risk. J Hazard Mater 505, 141515. https://doi.org/10.1016/j.jhazmat.2026.141515.

[25] Wang, M., Zhao, W., Wu, X., Yang, A., Chen, Y., Qu, Y., Ma, J., Wu, F., 2026. Advanced three-dimensional prediction model based on stable machine learning for soil pollution: A case study from a contaminated site in Southern China. J Hazard Mater 494, 138561. https://doi.org/10.1016/j.jhazmat.2025.138561.

[26] Wang, Q., Li, C., Hao, D., Xu, Y., Shi, X., Liu, T., Sun, W., Zheng, Z., Liu, J., Li, W., Liu, W., Zheng, J., Li, F., 2023. A novel four-dimensional prediction model of soil heavy metal pollution: Geographical explanations beyond artificial intelligence "black box. J Hazard Mater 458, 131900. https://doi.org/10.1016/j.jhazmat.2023.131900.

[27] Lundberg, S.M., Lee, S., 2017. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing System. California, USA. Pub Place: Curran Associates Inc, Long Beach.

[28] Wu, J., Huang, C., 2025. Machine learning-supported determination for site-specific natural background values of soil heavy metals. J Hazard Mater 487, 137276. https://doi.org/10.1016/j.jhazmat.2025.137276.

[29] Song, X., Sun, Y., Wang, H., Huang, X., Han, Z., Shu, Y., Wu, J., Zhang, Z., Zhong, Q., Li, R., Fan, Z., 2025. Uncovering soil heavy metal pollution hotspots and influencing mechanisms through machine learning and spatial analysis. Environ Pollut 370, 125901. https://doi.org/10.1016/j.envpol.2025.125901.

[30] Liu, F., Wu, H., Zhao, Y., Li, D., Yang, J.-L., Song, X., Shi, Z., Zhu, A.-X., Zhang, G.-L., 2021. Mapping high resolution national soil information grids of China. Sci Bull 67, 328–340. https://doi.org/10.1016/j.scib.2021.10.013.

[31] Liu, F., Zhang, G.-L., Song, X., Li, D., Zhao, Y., Yang, J., Wu, H., Yang, F., 2020. High-resolution and three-dimensional mapping of soil texture of China. Geoderma 361, 114061. https://doi.org/10.1016/j.geoderma.2019.114061.

[32] Didan, K., 2015. MOD13A3 MODIS/Terra vegetation Indices Monthly L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed 2025-12-21 from https://doi.org/10.5067/MODIS/MOD13A3.006.

[33] Xu, X., 2017. China's GDP spatial distribution kilometer grid data set. Data registration and publishing system of Resources and Environmental Science Data Center of Chinese Academy of Sciences (http://www.resdc.cn/DOI) 2017.DOI: 10.12078/2017121102.

[34] Xu, X., 2017. China's population spatial distribution kilometer grid data set. Data registration and publishing system of Resources and Environment Science Data Center of Chinese Academy of Sciences (http://www.resdc.cn/DOI) 2017.DOI: 10.12078/2017121101.

[35] Li, Y., Zhang, D., Zhang, X., Peng, J., Shang, M., 2024. A dataset of 1km Grid Road network density in China [DS/OL]. V6. Science Data Bank, [2026-7-26]. https://www.scidb.cn/en/detail?dataSetId=de7e723223f846c3a116c64af291f4a6.

[36] Ministry of Ecology and Environment of the People's Republic of China. National pollutant emission permit management information platform (in Chinese), Ministry of Ecology and Environment of the People's Republic of China. https://permit.mee.gov.cn/permitExt/defaults/defaultindex!getInformation.action (accessed December 21, 2025).

[37] Bo, X., Jia, M., Xue, X., Tang, L., Mi, Z., Wang, S., Cui, W., Chang, X., Ruan, J., Dong, G., Zhou, B., Davis, S.J., 2021. Effect of strengthened standards on Chinese ironmaking and steelmaking emissions. Nat Sustain 4, 811–820. https://doi.org/10.1038/s41893-021-00736-0.

[38] Xu, X., Ma, Z., Zhang, Z., Qu, J., Bo, X., Zhang, Y., 2025. Substantial Emission Reductions from Chinese Iron- and Steelmaking after the Introduction of Ultralow Emission Standards during the Prepandemic and Midpandemic Periods. ACS ES&T Air 2, 2550–2565. https://doi.org/10.1021/acsestair.5c00197.

[39] Yang, J., Tian, H., Fu, Z., Bai, X., Wang, K., Liu, W., Lu, Y., Zhou, Y., Zhao, H., Cui, J., Du, Q., 2025. Historical inventories and future scenarios of multiple hazardous air pollutants (HAPs) emissions from the iron and steel production industry in China. Sci Total Environ 958, 178051. https://doi.org/10.1016/j.scitotenv.2024.178051.

[40] Tian, Y., Zhang, Y., 2022. A comprehensive survey on regularization strategies in machine learning. Inform Fusion 80, 146–166. https://doi.org/10.1016/j.inffus.2021.11.005.

[41] Du, Z., Sun, X., Zheng, S., Wang, S., Wu, L., An, Y., Luo, Y., 2024. Optimal biochar selection for cadmium pollution remediation in Chinese agricultural soils via optimized machine learning. J Hazard Mater 476, 135065. https://doi.org/10.1016/j.jhazmat.2024.135065.
[42] He, B., Zhu, X., Cang, Z., Liu, Y., Lei, Y., Chen, Z., Wang, Y., Zheng, Y., Cang, D., Zhang, L.J.E.S., 2023. Interpretation and Prediction of the CO_2 Sequestration of Steel Slag by Machine Learning. Environ Sci Technol 57, 17940–17949.
[43] Yang, C., Dong, H., Chen, Y., Wang, Y., Fan, X., Tham, Y.J., Chen, G., Xu, L., Lin, Z., Li, M., Hong, Y., Chen, J., 2023. Machine Learning Reveals the Parameters Affecting the Gaseous Sulfuric Acid Distribution in a Coastal City: Model Construction and Interpretation. Environ Sci Technol Lett 10, 1045–1051. https://doi.org/10.1021/acs.estlett.3c00170.
[44] Jia, Y., Hu, X., Kang, W., Dong, X., 2024. Unveiling Microbial Nitrogen Metabolism in Rivers using a Machine Learning Approach. Environ Sci Technol 58, 6605–6615. https://doi.org/10.1021/acs.est.3c09653.
[45] Qiu, Y., Niu, J., Zhang, C., Chen, L., Su, B., Zhou, S., 2025. Interpretable machine learning reveals transport of aged microplastics in porous media: Multiple factors co-effect. Water Res 274, 123129. https://doi.org/10.1016/j.watres.2025.123129.
[46] Lin, F., Pan, K., Li, J., Wu, C., Cai, Q., Xue, H., Li, M., Huang, R., 2025. Assessment of influential factors and risk area prediction of scrub typhus in Gannan region, China, based on ecological niche modelling and geographical detector. Acta Trop 268, 107726. https://doi.org/10.1016/j.actatropica.2025.107726.
[47] Lin, M., Liu, X., Deng, H., Lin, Z., 2024. Quantitative Analysis of Heavy Metal Pollution Sources, Influencing Factors, and Delineation of Pollution Risk Areas in Industrial Soil: A Combined Application of Receptor Models and GeoDetector. ACS EST Engg 4, 562–571. https://doi.org/10.1021/acsestengg.3c00391.
[48] Li, Z., Luo, Z., Wang, Y., Fan, G., Zhang, J., 2022. Suitability evaluation system for the shallow geothermal energy implementation in region by Entropy Weight Method and TOPSIS method. Renew Energ 184, 564–576. https://doi.org/10.1016/j.renene.2021.11.112.
[49] Li, F., Yan, J., Wei, Y., Zeng, J., Wang, X., Chen, X., Zhang, C., Li, W., Chen, M., Lü, G., 2021. $PM_{2.5}$-bound heavy metals from the major cities in China: Spatiotemporal distribution, fuzzy exposure assessment and health risk management. J Clean Prod 286, 124967. https://doi.org/10.1016/j.jclepro.2020.124967.
[50] Guo, Y., Yao, Y., Yi, P., 2007. Method and application of dynamic comprehensive evaluation. Syst Eng-Theory & Pract 27, 154–158. https://doi.org/10.1016/S1874-8651(08)60060-5.
[51] Hakanson, L., 1980. An ecological risk index for aquatic pollution control. A sedimentological approach. Water Res 14, 975–1001. https://doi.org/10.1016/0043-1354(80)90143-8.
[52] Tomlinson, D.L., Wilson, J.G., Harris, C.R., Jeffrey, D.W., 1980. Problems in the assessment of heavy-metal levels in estuaries and the formation of a pollution index. Helgol Meeresunters 33, 566–575. https://doi.org/10.1007/BF02414780.
[53] Azizi, K., Ayoubi, S., Nabiollahi, K., Garosi, Y., Gislum, R., 2022. Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. J Geochem Explor 233, 106921. https://doi.org/10.1016/j.gexplo.2021.106921.
[54] Wang, G., Cissé, G., Staunton, S., 2024. Changes in chemical fractionation of copper and zinc in soil as a function of incubation moisture content and organic matter amendments. Chemosphere 351, 141198. https://doi.org/10.1016/j.chemosphere.2024.141198.
[55] He, J., Li, C., Tan, X., Peng, Z., Li, H., Luo, X., Tang, L., Wei, J., Tang, C., Yang, W., Jiang, J., Xue, S., 2024. Driving factors for distribution and transformation of heavy metals speciation in a zinc smelting site. J Hazard Mater 471, 134413. https://doi.org/10.1016/j.jhazmat.2024.134413.
[56] Wei, T., Dong, Z., Li, F., Kang, S., Qin, X., 2025. Quantifying the distribution and origins of aerosol zinc across the Northern Hemisphere using stable zinc isotopes: A review. J Hazard Mater 491, 137828. https://doi.org/10.1016/j.jhazmat.2025.137828.
[57] Yang, Q., Liu, G., Falandysz, J., Yang, L., Zhao, C., Chen, C., Sun, Y., Zheng, M., Jiang, G., 2024. Atmospheric emissions of particulate matter-bound heavy metals from industrial sources. Sci Total Environ 947, 174467. https://doi.org/10.1016/j.scitotenv.2024.174467.
[58] He, Y., Yang, H., Fang, R., Wen, Z., Gong, P., Xia, R., Zhou, J., Zhou, J., 2026. Soil Properties Regulate the Fate and Accumulation of Atmospherically Deposited Trace Metals in Rice across Six Major Growing Regions of China. Environ Sci Technol 60, 19597–19610. https://doi.org/10.1021/acs.est.6c04551.
[59] Li, Y., Wang, K., Dötterl, S., Xu, J., Garland, G., Liu, X., 2024. The critical role of organic matter for cadmium-lead interactions in soil: Mechanisms and risks. J Hazard Mater 476, 135123. https://doi.org/10.1016/j.jhazmat.2024.135123.
[60] Bai, Z., Li, T., Zhang, S., Wang, G., Xu, X., Zhou, W., Pan, X., Pu, Y., Jia, Y., Yang, Z., Long, L., 2024. Effects of climate and geochemical properties on the chemical forms of soil Cd, Pb and Cr along a more than 4000 km transect. J Hazard Mater 467, 133746. https://doi.org/10.1016/j.jhazmat.2024.133746.
[61] Tian, K., Liang, Q., He, Y., Ma, J., Zhao, T., Wu, Q., Wu, Q., Hu, W., Huang, B., Khan, K.S., Teng, Y., 2024. Quantitative assessment of Cd sources in rice grains through Cd isotopes and MixSIAR model in a typical e-waste dismantling area of Southeast China. Sci Total Environ 954, 176217. https://doi.org/10.1016/j.scitotenv.2024.176217.
[62] Kong, S., Cai, D., Shao, Y., Wei, X., Yi, Z., Root, R.A., Chorover, J., 2024. Identification of key factors and mechanism determining arsenic mobilization in paddy soil-porewater-rice system. J Hazard Mater 479, 135684. https://doi.org/10.1016/j.jhazmat.2024.135684.
[63] Vandeuren, A., Pereira, B., Van de Casteele, C., Sonnet, P., Delmelle, P., 2025. Processes controlling geogenic arsenic distribution in soils formed from iron-rich sedimentary rocks. J Hazard Mater 483, 136610. https://doi.org/10.1016/j.jhazmat.2024.136610.
[64] Wang, J., Shaheen, S.M., Swertz, A.-C., Liu, C., Anderson, C.W.N., Fendorf, S., Wang, S.-L., Feng, X., Rinklebe, J., 2024. First Insight into the Mobilization and Sequestration of Arsenic in a Karstic Soil during Redox Changes. Environ Sci Technol 58, 17850–17861. https://doi.org/10.1021/acs.est.4c02989.
[65] Duan, X., Zhang, G., Rong, L., Fang, H., He, D., Feng, D., 2015. Spatial distribution and environmental factors of catchment-scale soil heavy metal contamination in the dry-hot valley of Upper Red River in southwestern China. CATENA 135, 59–69. https://doi.org/10.1016/j.catena.2015.07.006.
[66] Li, R., Bing, H., Wu, Y., Zhou, J., Xiang, Z., 2018. Altitudinal patterns and controls of trace metal distribution in soils of a remote high mountain, Southwest China. Environ Geochem Health 40, 505–519. https://doi.org/10.1007/s10653-017-9937-2.
[67] Xue, W., Wang, C., Pan, S., Zhang, C., Huang, Y., Liu, Z., 2024. Effects of elevation and geomorphology on cadmium, lead and chromium enrichment in paddy soil and rice: A case study in the Xiangtan basin of China. Sci Total Environ 912, 168613. https://doi.org/10.1016/j.scitotenv.2023.168613.
[68] Magnani, A., Ajmone-Marsan, F., D'Amico, M., Balestrini, R., Viviano, G., Salerno, F., Freppaz, M., 2018. Soil properties and trace elements distribution along an altitudinal gradient on the southern slope of Mt. Everest, Nepal. CATENA 162, 61–71. https://doi.org/10.1016/j.catena.2017.11.015.
[69] Liu, N., Li, X., Chen, P., Yuan, W., Wang, D., Wang, X., 2025. Climate and vegetation controlling accumulation and translocation of heavy metals in water tower regions of Qinghai-Tibet Plateau. J Hazard Mater 484, 136752. https://doi.org/10.1016/j.jhazmat.2024.136752.
[70] Davidson, E., Janssens, I., 2006. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. Nature 440, 165–173. https://doi.org/10.1038/nature04514.
[71] Hamdi, S., Moyano, F., Sall, S., Bernoux, M., Chevallier, T., 2013. Synthesis analysis of the temperature sensitivity of soil respiration from laboratory studies in relation to incubation methods and soil conditions. Soil Biol Biochem 58, 115–126. https://doi.org/10.1016/j.soilbio.2012.11.012.
[72] Zhang, S., Wang, M., Xiao, L., Guo, X., Zheng, J., Zhu, B., Luo, Z., 2024. Reconciling carbon quality with availability predicts temperature sensitivity of global soil carbon mineralization. Proc Natl Acad Sci 121, e2313842121. https://doi.org/10.1073/pnas.2313842121.
[73] Qin, M., Jin, Y., Peng, T., Zhao, B., Hou, D., 2023. Heavy metal pollution in Mongolian-Manchurian grassland soil and effect of long-range dust transport by wind. Environ Int 177, 108019. https://doi.org/10.1016/j.envint.2023.108019.
[74] Liu, L., Ouyang, W., Wang, Y., Tysklind, M., Hao, F., Liu, H., Hao, X., Xu, Y., Lin, C., Su, L., 2020. Heavy metal accumulation, geochemical fractions, and loadings in two agricultural watersheds with distinct climate conditions. J Hazard Mater 389, 122125. https://doi.org/10.1016/j.jhazmat.2020.122125.
[75] Zhang, L., Wang, B., Wu, W., Wang, C., Cheng, H., Duan, X., 2024. Enhanced health risk of soil heavy metal exposure following an extreme rainstorm under climate change. Sci Total Environ 954, 176409. https://doi.org/10.1016/j.scitotenv.2024.176409.
[76] Li, Y., Li, L., Han, Y., Shi, J., He, J., Cheng, S., Liu, H., Zhang, B., 2022. Soil indigenous microorganisms alleviate soluble vanadium release from industrial dusts. J Hazard Mater 434, 128837. https://doi.org/10.1016/j.jhazmat.2022.128837.
[77] Qi, Y., Gong, Y., Wu, F., Lu, Y., Cheng, W., Huang, F., Yu, H., 2022. Coupled variations in V-Fe abundances and isotope compositions in latosols: Implications for V mobilization during chemical weathering. Geochim Cosmochim Ac 320, 26–40. https://doi.org/10.1016/j.gca.2021.12.028.
[78] Temminghoff, E.J., Van der Zee, S.E., de Haan, F.A., 1997. Copper Mobility in a Copper-Contaminated Sandy Soil as Affected by pH and Solid and Dissolved Organic Matter. Environ Sci Technol 31, 1109–1115. https://doi.org/10.1021/es9606236.
[79] Arabi, Z., Rinklebe, J., El-Naggar, A., Hou, D., Sarmah, A.K., Moreno-Jiménez, E., 2021. Im)mobilization of arsenic, chromium, and nickel in soils via biochar: A meta-analysis. Environ Pollut 286, 117199. https://doi.org/10.1016/j.envpol.2021.117199.
[80] Shen, T., Tang, Y., Lu, X.-Y., Meng, Z., 2018. Mechanisms of copper stabilization by mineral constituents in sewage sludge biochar. J Clean Prod 193, 185–193. https://doi.org/10.1016/j.jclepro.2018.05.071.
[81] Han, R., Gao, Y., Jia, Y., Wang, S., 2024. Heterogeneous precipitation behavior and mechanism during the adsorption of cationic heavy metals by biochar: Roles of inorganic components. J Hazard Mater 480, 136322. https://doi.org/10.1016/j.jhazmat.2024.136322.
[82] Fu, X., Zhang, H., Liu, C., Zhang, H., Lin, C., Feng, X., 2019. Significant Seasonal Variations in Isotopic Composition of Atmospheric Total Gaseous Mercury at Forest Sites in China Caused by Vegetation and Mercury Sources. Environ Sci Technol 53, 13748–13756. https://doi.org/10.1021/acs.est.9b05016.
[83] Cai, B., Tang, R., Wang, H., Sun, J., Zhao, M., Huang, X., Song, X., Han, Z., Fan, Z., 2024. Impact of economic development on soil trace metal(loid)s pollution: A case study of China. Environ Pollut 344, 123319. https://doi.org/10.1016/j.envpol.2024.123319.
[84] Elliott, R.J.R., Sun, P., Zhu, T., 2017. The direct and indirect effect of urbanization on energy intensity: A province-level study for China. Energy 123, 677–692. https://doi.org/10.1016/j.energy.2017.01.143.
[85] Zhou, X., Wang, X., 2019. Impact of industrial activities on heavy metal contamination in soils in three major urban agglomerations of China. J Clean Prod 230, 1–10. https://doi.org/10.1016/j.jclepro.2019.05.098.

[86] Li, J., Xu, Y., Zhang, Y., Liu, Z., Gong, H., Fang, W., Ouyang, Z., Li, W., Xu, L., 2024. Quantifying the mitigating effect of organic matter on heavy metal availability in soils with different manure applications: A geochemical modelling study. Ecotox Environ Safe 276, 116321. https://doi.org/10.1016/j.ecoenv.2024.116321.

[87] Vega, A.S., Arce, G., Rivera, J.I., Acevedo, S.E., Reyes-Paecke, S., Bonilla, C.A., Pastén, P., 2022. A comparative study of soil metal concentrations in Chilean urban parks using four pollution indexes. Appl Geochem 141, 105230. https://doi.org/10.1016/j.apgeochem.2022.105230.

[88] Xu, L., Liu, N., Pei, S., Jiang, F., Zhao, R., Li, N., Cai, Y., 2025. Organosiloxanes in the photovoltaic production industry: their sources, mass loads, and actual contribution to mercury methylation in wastewater. Environ Sci Technol 59, 13983–13991. https://doi.org/10.1021/acs.est.5c02698.