



Article

Pollution Risk Prediction for Cadmium in Soil from an Abandoned Mine Based on Random Forest Model

Jie Cao ¹ , Zhaohui Guo ^{1,*}, Yongjun Lv ², Man Xu ², Chiyue Huang ¹ and Huizhi Liang ¹

¹ School of Metallurgy and Environment, Central South University, Changsha 410083, China

² Linxiang Station of Yueyang Ecology and Environment Monitoring Center, Linxiang 414300, China

* Correspondence: zhguo@csu.edu.cn

Abstract: It is highly uncertain as to the potential risk of toxic metal(loid)s in abandoned mine soil. In this study, random forest was used to predict the risk of cadmium pollution in the soils of an abandoned lead/zinc mine. The results showed that the random forest model is stable and precise for the pollution risk prediction of toxic metal(loid)s. The mean of Cd, Cu, Tl, Zn, and Pb was 6.02, 1.30, 1.18, 2.03, and 2.08 times higher than the soil background values of China, respectively, and their coefficients of variation were above 30%. As a case study, cadmium in the mine soil had “slope” hazard characteristics while the ore sorting area was the major source area of cadmium. The theoretical values of the random forest model are similar to the practical values for the ore sorting area, metallogenic belt, riparian zone, smelting area, hazardous waste landfill, and mining area. The potential risk of soil Cd in the ore sorting area, metallogenic belt, and riparian zone are extremely high. The tendency of pollution risk migrates significantly both from the ore sorting area to the smelting area and the mining area, and to the hazardous waste landfill. The correlation of soil pollution risk is significant between the mining area, the smelting area, and the riparian zone. The results suggested that the random forest model can effectively evaluate and predict the potential risk of the spatial heterogeneity of toxic metal(loid)s in abandoned mine soils.

Keywords: abandoned lead/zinc mine; random forest model; toxic metal(loid)s; cadmium; potential pollution risks



Citation: Cao, J.; Guo, Z.; Lv, Y.; Xu, M.; Huang, C.; Liang, H. Pollution Risk Prediction for Cadmium in Soil from an Abandoned Mine Based on Random Forest Model. *Int. J. Environ. Res. Public Health* **2023**, *20*, 5097. <https://doi.org/10.3390/ijerph20065097>

Academic Editor: Paul B. Tchounwou

Received: 22 February 2023
Revised: 8 March 2023
Accepted: 13 March 2023
Published: 14 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Metal(loid) pollution exists in a variety of environmental mediums such as soil, water, gas, etc. [1,2]. Public poisoning due to metal pollution in the soil has been occurring worldwide [3]. The prevention of metal pollution in the soil is one of the most difficult problems to solve, which is related to the fact that metal(loid) pollution in the soil is difficult to detect and can accumulate over time [4,5]. The high accumulation of metal(loid)s within the soil will result in soil pollution that is highly regional in nature [6]. Soil pollution from mines is a typical case. Metal(loid) pollution from mines is caused by functional activities such as mining, mineral processing, smelting, etc. The areas that undertake these activities are called functional areas [7,8]. The functional area of the mine, as the source of various pollution, is an important indicator of pollution prevention. The level of pollution risk of metal(loid)s within functional areas is inconsistent, and the level of pollution risk of metal(loid)s depends on the functional behavior of different functional areas, which will lead to the spatial heterogeneity of functional areas [9,10]. In order to prevent the harmful effects of metal pollution in mine soils on public health, it makes sense to carry out risk prediction work, which is essential for the risk identification of metal(loid) pollution sources in mines and for obtaining information on the characteristics of metal(loid) pollution sources.

The phenomenon of cross-pollution between functional areas indicates that there is a major source of pollution risk in the functional areas, which requires an urgent response [11,12]. The accurate identification of pollution risk sources by classical statistical

methods means using traditional full-scale sampling [13,14]. However, the disadvantage of the traditional method is the long time it takes to complete and the high expense, which is not an optimal choice for mining companies. The spatial heterogeneity of the pollution distribution of metals in soils from mines usually leads to a lot of complex data that are often difficult to interpret via traditional monitoring methods, such that metal(loid)s data from functional areas are difficult to analyze accurately and are insufficient to reveal the potential pollution risk of metal pollutants in complex environmental conditions (mines) [15]. The methods for pollution prediction, such as the human health risk assessment model [16], UNMIX model [17], or geographic information system model [18], have been established. However, compared with the above methods, random forest (RF) is considered as one of the more effective methods to provide spatial assessment and prediction, which has the advantages of requiring a small sample size, not being affected by a complex environment, and allowing researchers to dig deeper into the underlying data [19]. Wang et al. (2023) [20] applied a random forest model and land use regression model to compare the results of concentration data of six metals (Pb, Cd, Cr, As, Hg, and Zn) in agricultural soils for their advantages and disadvantages, and the validation results proved that the use of RF model is more suitable for the prediction of metal contents in agricultural soils. Azizi et al. (2023) [21] predicted the spatial distribution of some metals (Ni, Fe, Cu, and Mn) in western Iran using environmental covariates and random forest. The results demonstrate that random forest can use easily available environmental data to predict the large-scale areas under study, which is essential for decision-making on the sustainable management of environmental problems. However, to our knowledge, there are few studies that apply random forest to complex areas such as mines. Therefore, based on the above study, a proposal based on random forest for the prediction of the pollution risk of metal(loid)s in mine soils can be further proposed.

Then, for complex soil pollution, such as mine soil pollution, the RF model will be the key to obtain comprehensive information on the characteristics of toxic metal(loid) pollution sources. Therefore, the study aims (1) to evaluate the feasibility of the random forest model for the identification of the potential sources and risk characteristics of toxic metal(loid)s in soil from an abandoned lead/zinc mine, and (2) using cadmium as a case, extrapolate the potential pollution risk of various toxic metal(loid)s in the mine soil.

2. Materials and Methods

2.1. Study Area

The abandoned lead/zinc mine is located in central southern China, with a longitude of $113^{\circ}18'$ and latitude of $29^{\circ}24'$, and is characterized by red soil formed from quaternary laterite, slate, and shale. Due to the long-term direct discharge of industrial wastewater and the disordered stacking of waste slag, the historical legacy of the lead/zinc mine sites is one of serious pollution. There is a river running through the whole area of the mine. The flow of the river is mainly influenced by the amount of rainfall, switching to flood or dry periods with the change in seasons, and the river is the main surface runoff. Under the action of long-term water flow migration, a large number of toxic metal(loid)s are deposited in the soil and the riverbed in the vicinity of lead/zinc mine. There are seven functional areas in the mine. The ore sorting area is the area where physical and chemical measures are applied to the ore to obtain the needed ingredients for smelting or other industries. The riparian zone is the area on either side of the river–land interface until the influence of the river disappears. The hazardous waste landfill is the storage area for solid waste and industrial waste. The mining area is an area engaged in ore mining. The smelting area is the industrial area where ore calcination and refining are carried out. The tailings area is the area where the tailings or other industrial wastes after ore sorting are deposited. The metallogenic belt is a geological unit of mineral resources with potential for mineralization (Figure 1).

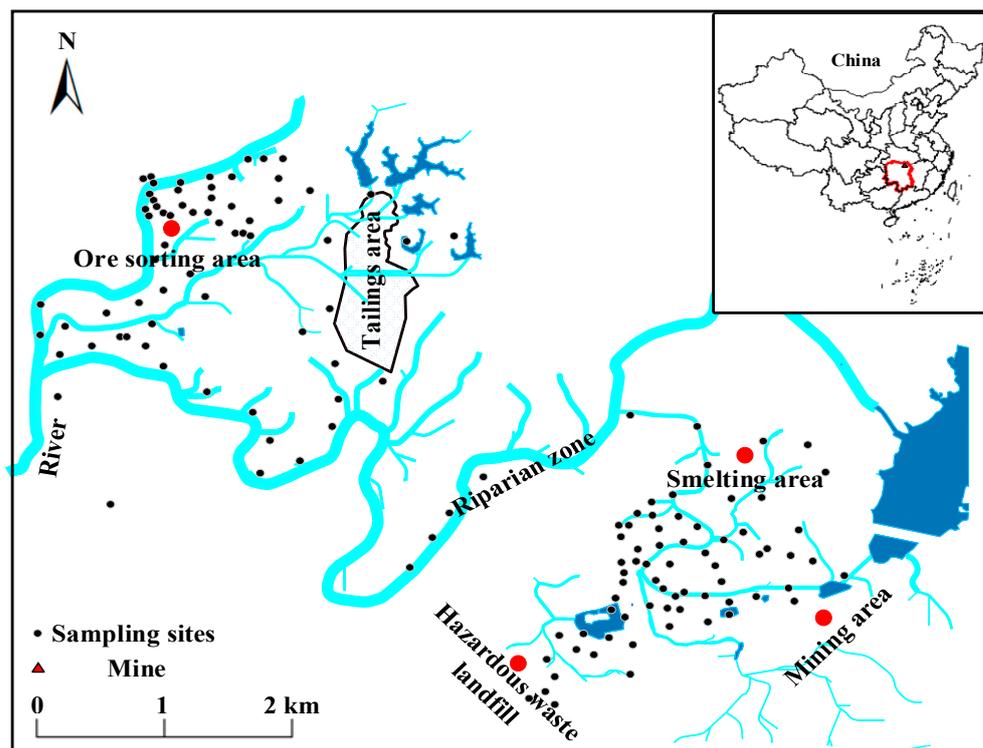


Figure 1. Study area and sampling sites.

2.2. Sampling and Analysis

Soil sampling in the abandoned lead/zinc mine was designed according to the strategy of combining points, lines, and surfaces along with the river, and the screening results of pollution identification. The soil sampling was carried out at sites at intervals of about 300 m on the side flowing through the lead/zinc mine, which had to be arranged according to the direction of the mine hole and slope, combined with the flow direction of surface runoff in the mine, and two control sites had to be set in each local area of blank control. All of soil samples were collected via mechanical drilling or excavation with stainless steel shovels. A total of 147 soil samples were collected and data for ten metal(loid)s (As, Cd, Cr, Cu, Hg, Mn, Sb, Tl, Pb, and Zn) were determined, with 147 pieces of data for individual metal(loid)s and a total of 1470 pieces of data obtained.

Each of the 147 soil samples was placed indoors for air-drying for 7 days. The samples were ground and then filtered through a 2 mm sieve and prepared for use. Soil samples were digested with mixed acid (HCl-HNO₃-HClO₄). The procedure was as follows: 1.0 g of soil sample was mixed with 5.0 mL of concentrated nitric acid (HNO₃), 3.0 mL of concentrated hydrochloric acid (HCl), and 2.0 mL of concentrated perchloric acid (HClO₄), and digested using microwave at 160 °C for 2 h.

Inductively coupled plasma mass spectrometry (ICP, iCAP 7600, Thermo Scientific, 81 Wyman Street, Waltham, MA, USA) was used to determine the concentrations of As, Cd, Cr, Cu, Hg, Mn, Sb, Tl, Pb, and Zn in the digested solution. The experiments were carried out on the reagent blank group and repeated soil samples to check the accuracy of the experimental method and data. The recovery rate was 100 ± 10%. The analytical method was tested using the national first-class soil standard material (HJ 25.1-2019) of the People's Republic of China. The background values (As:13.6, Cd:0.081, Cr:71.4, Cu:25.4, Hg:0.087, Sb:1.58, Tl:0.61, Pb:27.3, and Zn:88.6) for soil metal(loid)s were based on the CNEMC (China National Environmental Monitoring Center), Beijing, China, 1990 [22].

2.3. Modeling of Random Forest

A random forest is a non-parametric model that iteratively classifies or regresses data to find the best split point, generates N decision trees, and finally uses a voting mechanism in the forest to determine the output. Random forest are characterized by randomly selected features and samples, allowing each tree in the forest to have similarities and differences. The bootstrap method is used to randomly draw k new sets of self-help samples with put-back from the original training dataset, and from this, k classification regression trees are constructed, and each undrawn sample forms K out-of-bag data ($OOB_1, OOB_2, \dots, OOB_k$). Given n features, n features are randomly selected at each tree node. The feature with the highest classification power is chosen for node splitting by calculating the amount of information contained in each feature. The impact of each feature on the model's accuracy rate is directly measured during feature selection. The prediction accuracy OA_1, OA_2, \dots, OA_K for k out-of-bag datasets is obtained by inputting each out-of-bag dataset into the corresponding decision tree. For the assessment index f_i , the index values under this assessment index are randomly replaced in all out-of-bag datasets, while the other index values remain unchanged, resulting in the new out-of-bag datasets $OOB_1^i, OOB_2^i, \dots, OOB_k^i$. After random replacement of the evaluation indicator f_i , $OA_1^i, OA_2^i, \dots, OA_K^i$, they are input into the decision tree corresponding to step to determine the prediction accuracy of the out-of-bag dataset.

The random forest model is based on the background of big data information and is not affected by complex environments. In this work, 147 soil samples were randomly divided into modeling set and validation set, this step was conducted to build and validate the model in the same batch of data to improve its validity. A total of 130 samples were used as the modeling set to build the model and 17 samples were used as the validation set to illustrate the feasibility of the model after the validation of the data samples. With toxic metal(loid) content in the soil as the dependent variable, the model was established using random forest, and the prediction was made according to the validation set. By referencing the sklearn library, it was possible to construct classifier objects, training sample sets, predicted values, and complete evaluation, in four steps. The stability and accuracy of the model were analyzed by the coefficient of determination (R^2). When R^2 was closer to one, the fitting effect of the regression equation was better and the model was more stable.

3. Results and Discussion

3.1. Pollution Characteristics of Toxic Metal(loid)s in the Mine Soil

The difference between the minimum and maximum values of 10 metal(loid)s (As, Cd, Cr, Cu, Hg, Mn, Sb, Tl, Zn, and Pb) in the soil of the abandoned lead/zinc mine is too high and thus affects the mean values (Table 1). The median values for Cd, Cu, Tl, Zn, and Pb were 6.02, 1.30, 1.18, 2.03, and 2.08 times higher than the background values, respectively, indicating that five types of toxic metal(loid) (Cd, Cu, Tl, Zn, and Pb) pollution exist in the soil [23]. The coefficient of variation of these data provides a better indication of the degree of dispersion compared to standard deviation, and the coefficients of variation for the 10 metals (As, Cd, Cr, Cu, Hg, Mn, Sb, Tl, Pb, and Zn) were above 15%, with Cd being the highest value, indicating a higher level of risk for toxic metal(loid)s in the soil. According to the above analysis, Cd is considered to be the main contaminant at risk in abandoned mine [24].

The variability of cadmium pollution distribution in the lead/zinc mine soil is great, with severe pollution in the ore sorting area and metallogenic belt. The degree of soil pollution shows a decreasing trend centered on the ore sorting area and metallogenic belt (Figure 2), and there is some correlation between the sources of pollution, which is related to human activities such as early and unreasonable exploitation of mineral resources and the random stacking of tailings [25,26]. Moreover, there may be some correlation between the ore sorting area, metallogenic belt, and riparian zone, according to the division of the water system; there is a high pollution area block in the upstream ore sorting area, while the metallogenic belt and riparian zone are located downstream, and there is an obvious

decreasing trend of the pollution block from the ore sorting area to the metallogenic belt and riparian zone. Therefore, based on the flow direction of the water system and the gradually decreasing spatial content, it is inferred that the ore sorting area and the metallogenic belt and riparian zone may have mutual pollution [27]. This can be attributed to the activities of mineral processing, smelting, and solid disposal, and the chaos of wastewater treatment and drainage that severely pollutes the surrounding soil. The area heavily polluted by cadmium is mainly situated in the ore sorting area, where cadmium in the soil is the source of pollution, and the pollution from the ore sorting area to the mining area shows a gradually decreasing trend, which is related to unreasonable exploitation of mineral resources at an early stage and randomly stacking residues [28]. Secondly, cadmium pollution from hazardous waste landfill and smelting areas has a more serious impact on the soil, as well as in mining areas under the movement of surface water and groundwater.

Table 1. Characteristics of the content of toxic metal(loid)s in the soil of the abandoned mine.

Metal(loid)s	Minimum	Median	Maximum	Mean	SD ^a	CV ^b (%)	BV ^c
As	1.31	10.00	37.5	11.23	6.61	58.89	13.6
Cd	0.039	0.49	5.98	0.91	1.22	133.32	0.081
Cr	29.8	70.00	220	72.37	17.47	24.13	71.4
Cu	4.9	33.00	204	40.91	28.13	68.76	25.4
Hg	0.012	0.082	0.52	0.11	0.09	79.97	0.087
Mn	148	484.00	2230	528.53	287.40	54.38	/
Sb	0.23	1.54	6.73	1.80	1.09	60.94	1.58
Tl	0.47	0.72	2.15	0.77	0.24	31.88	0.61
Zn	49.8	180.00	1320	269.69	241.52	89.56	88.6
Pb	ND ^d	56.7	289.00	77.12	62.75	81.37	27.3

^a SD: Standard deviation; ^b CV: coefficient of variation; ^c BV: background value ^d ND: not detected, and not participating in validation.

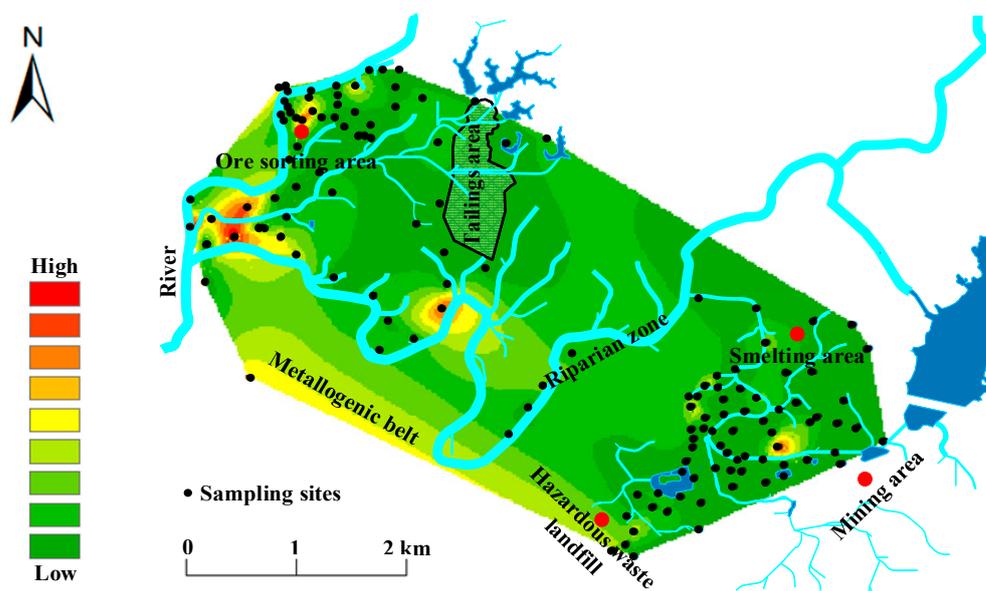


Figure 2. Spatial distribution of cadmium pollution in the soil from an abandoned lead/zinc mine.

3.2. Validation of Random Forest Model

In this study, based on the content of 10 elements in 130 groups, a total of 1300 pieces of data was substituted into the random forest model, and the feature values were selected according to a quarter of the total number of sample feature variables. Through the feasibility check, R^2 of the elements (As, Cd, Cr, Cu, Hg, Mn, Sb, Tl, Pb, and Zn) were all found to exceed 0.95, the theoretical value of the model was very consistent with

a practical value (Pearson’s $r > 0.98$), and the feasibility of constructing the model was passed (Figure 3).

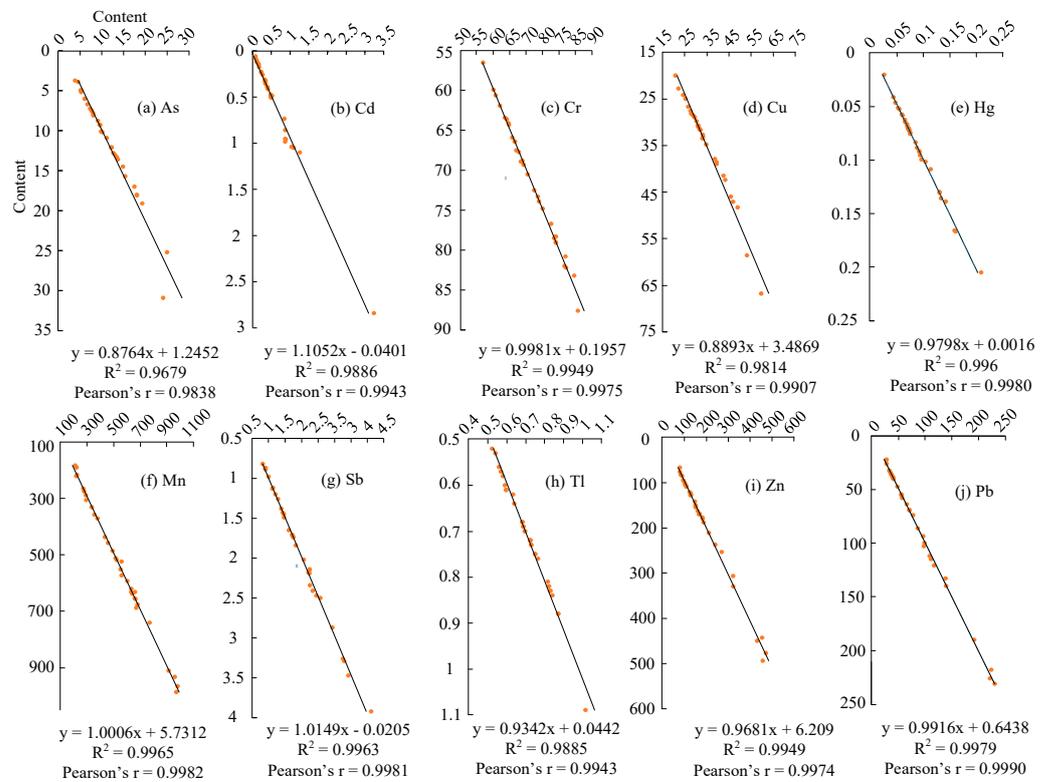


Figure 3. Feasibility verification of random forest model.

Unconstructed groups of 17 pieces of data were selected as the validation of the prediction model, and groups of 17 pieces of data were substituted into the random forest model to obtain the theoretical values of cadmium and compare them with the practical values. The R^2 was as high as 0.81, and the pattern of similarity was essentially the same, which demonstrated the superiority of the random forest model in the cadmium prediction task. The mean error was less than 1%, indicating a high degree of coherence between the two data sets [29], and the median error was below 10% (Table 2), indicating a strong similarity between the two data sets. Results from all three indicators show that the random forest model has an obvious recognition capability and high accuracy for predicting toxic metal(loid)s [30,31]. When data cannot normally be obtained due to environmental and geological factors in the regional soil, the random forest model can still achieve high predictive precision [32,33]. The results show that the random forest model is effective at predicting soil cadmium levels, confirming the science and the advance of the random forest prediction model.

Table 2. Verification of model construction.

ID	1	2	3	4	5	6	7	8	9
Practical value	1.48	3.35	5.64	0.94	2.95	2.49	1.18	0.63	0.89
Theoretical value	1.47	3.63	4.34	1.2	2.4	2.41	1.8	1	1.18
ID	10	11	12	13	14	15	16	17	
Practical value	2.37	1.67	1.72	1.07	3.49	0.64	5.54	0.53	
Theoretical value	2.67	1.86	1.8	1.41	2.81	1.12	4.49	1.23	

3.3. Spatial Risk Prediction of Cadmium Pollution in the Mine Soil

Pollution risk assessment in soils is an important tool for environmental prevention and control [34]. According to the theoretical values of cadmium, the high-risk area of the mine for cadmium is the ore sorting area, which shows abnormally high values, and the distribution of the surrounding soil shows a progressive decline in risk, which indicates that production behavior has impacted the accumulation of cadmium in the mine soil, and steps should be taken in advance to intercept and control it to avoid a progressive increase in the environmental risk of the surrounding soil [35]. The metallogenic belt and riparian zone are adjacent to the ore sorting area and have similar tendency of pollution risk for cadmium, confirm that the ore sorting area is a significant potential soil pollution source in the metallogenic belt and riparian zone. Therefore, effective soil control measures should be taken from the sources of pollution in the ore sorting area in order to eliminate pollution caused by the migration of soil pollution through the water stream, thus reducing the level of pollution risk in the metallogenic belt and riparian zone. Furthermore, the variability of cadmium in different functional areas confirms the spatial heterogeneity of the distribution of cadmium pollution in the lead/zinc mine soil [36].

The mine pollution hazard is characterized by a “slope” from the ore sorting area to the metallogenic belt (Figure 4). The strong correlation between the risk of soil pollution in the mining area, smelting area, and the riparian zone means that cadmium in riparian zone soils can migrate into the river and pollute the soils in mining and smelting areas. As a result, the riparian zone should be blocked to mitigate the level of soil risk in the mining area and smelting areas.

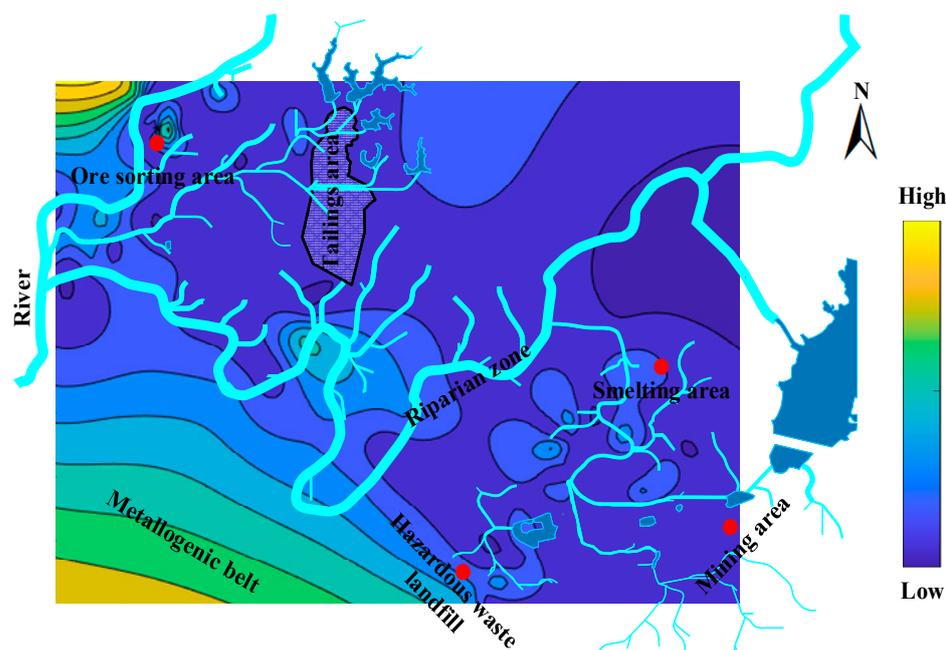


Figure 4. Spatial risk distribution of cadmium pollution in the soil from an abandoned lead/zinc mine.

3.4. Validation of Spatial Pollution Risk Prediction for Cadmium

The practical values of toxic metal(loid)s in 147 sets of soil samples were used as modeling sets in a random forest model. Cadmium was used as an independent validation set (a total data of 147), and the prediction model of the cadmium content in soil and a spatial prediction model was constructed based on the modeling set. Finally, based on the measured cadmium content, the content and spatial prediction were verified [37].

The optimum precision in inversion modelling was high, where the cadmium R^2 was 0.75, and the pattern of adjustment was essentially the same. Of these, both groups of data with mean and median error exhibited a high degree of coincidence, and the errors in the

two data sets were less than 5% (Figure 5), which indicates that the similarity between the two data sets was extremely high [38,39]. It has been demonstrated that the random forest model has good stability and high inversion accuracy in the cadmium forecasting task. In conclusion, the random forest model has a clear recognition capability and a high precision in the prediction of cadmium from toxic metal(loid)s.

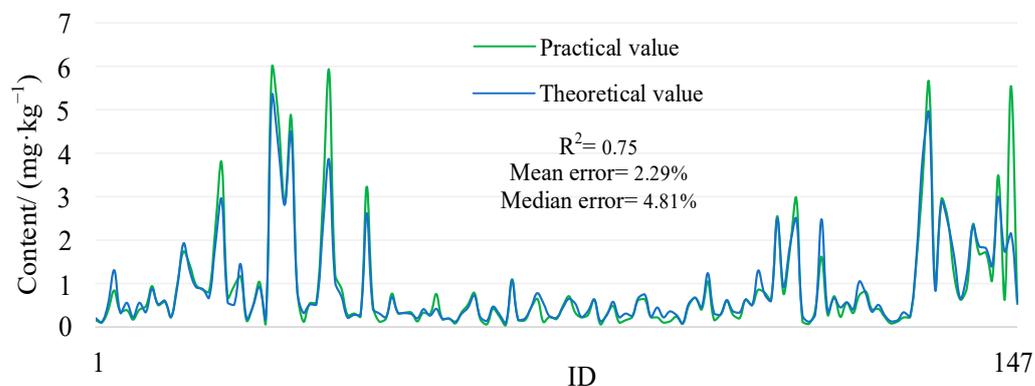


Figure 5. Comparison of risk values for spatial pollution of cadmium in soil from abandoned lead/zinc mine.

The theoretical values of the spatial risk distribution of cadmium in soil are similar to the practical values. According to a cadmium spatial pollution risk in soil comparison between Figure 6a,b, the risks for the ore sorting area (Focus 1), metallogenic belt (Focus 2), and riparian zone (Focus 3) are high. The risk for the smelting area, mining area, and hazardous waste landfill (Focus 4) are low. The similarity between Focus 1 to 4 is very high, and the ore sorting area is still the highest risk area among them, indicating that the functional behavior of the ore sorting area has a serious impact on the spatial distribution of cadmium in soil. The diffusion trend of cadmium pollution risk is similar and migrates obviously from the ore sorting area to the smelting area and mining area, and hazardous waste landfill, which proves that the ore sorting area has caused serious pollution to the downstream soil of the lead/zinc mine. The results suggest that the random forest model has shown great stability in cross-validation and a strong capacity for generalization and a high predictive precision in independent validation. The predictive model was very precise and stable, and the theoretical values were valid. The high-risk area of cadmium in the soil is located in the ore sorting area and should be carried out to prevent the pollution from migrating with surface water and groundwater.

In the future, soil risk prediction for complex areas such as mines can be predicted using random forest at a large scale. From an effectiveness point of view, the current results basically meet the needs, which will help the government and enterprises to carry out the identification of risk sources. Based on this work, the ore sorting area is a high-risk area for cadmium in abandoned lead/zinc mines. Therefore, the risk prediction of priority polluting metals can be conducted separately from the risk prediction of lead/zinc mines, and, finally, the high-risk areas of metals can be integrated to obtain the prevention and control areas that should be given priority.

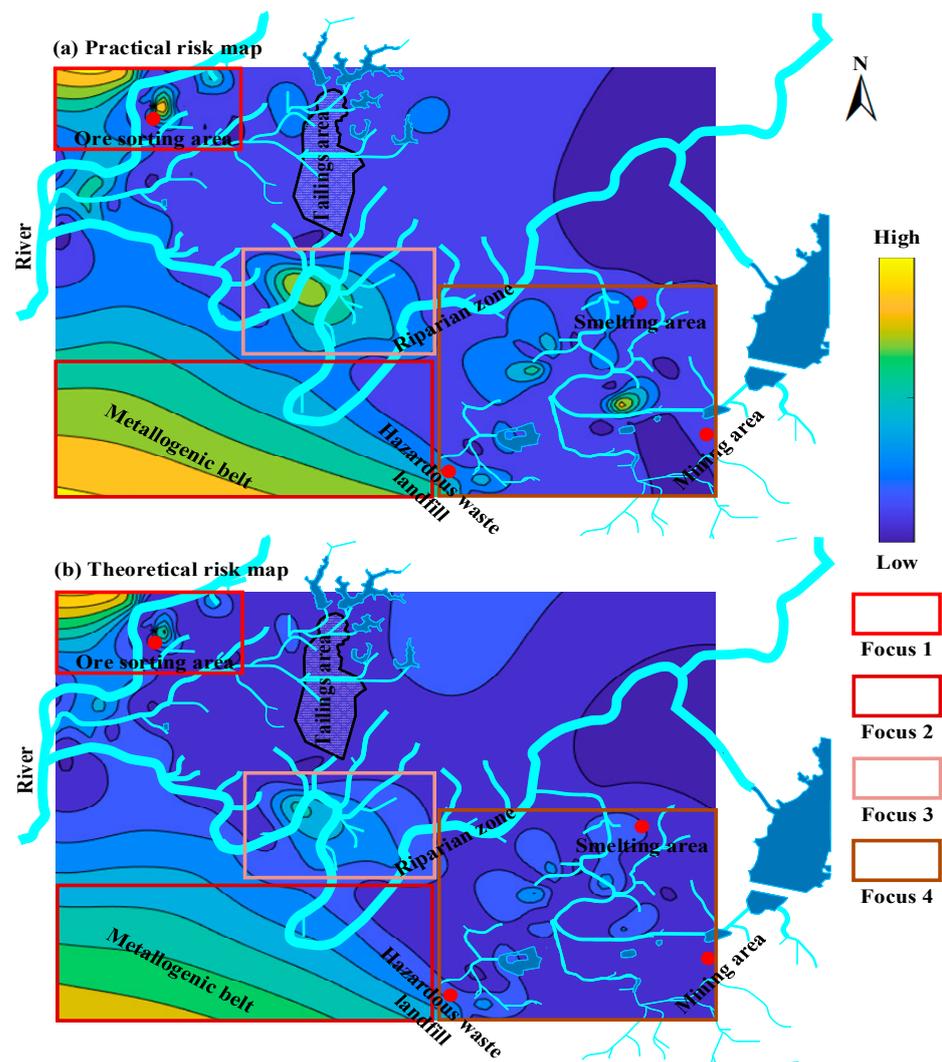


Figure 6. Spatial distribution of cadmium pollution risk in soil of abandoned lead/zinc mine.

4. Conclusions

(1) A random forest model can identify the unequal distribution of toxic metal(loid) data in soil, and effectively predict the spatial heterogeneity and the potential pollution risk of soil cadmium in a heavily polluted area from an abandoned mine.

(2) According to random forest model, the theoretical values of As, Cd, Cr, Cu, Hg, Mn, Sb, Tl, Pb, and Zn in the mine soil corresponded perfectly to the practical values and may be used to predict the contents of toxic metal(loid)s (As, Cd, Cr, Cu, Hg, Mn, Sb, Tl, Pb, and Zn).

(3) The random forest model is universal for the spatial prediction of toxic metal(loid) pollutants under complex environmental conditions. The ore sorting area is the source of the pollution risk for cadmium, and the level of risk shows a downward trend from the ore sorting area to the smelting area, mining area, and hazardous waste landfill.

Author Contributions: J.C.: Data curation, formal analysis, writing—original draft, and data analysis. Z.G.: Conceptualization, methodology, funding acquisition, and reviewing and editing. Y.L.: Reviewing and investigation. M.X.: Reviewing and investigation. H.L.: Data analysis. C.H.: Sample collection and investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (2018YFC1800400).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Küçüksümbül, A.; Akar, A.T.; Tarcan, G. Source, degree and potential health risk of metal(loid)s contamination on the water and soil in the Söke Basin, Western Anatolia, Turkey. *Environ. Monit. Assess.* **2022**, *194*, 6. [\[CrossRef\]](#)
2. Dinis, M.d.L.; Fiúza, A.; Futuro, A.; Leite, A.; Martins, D.; Figueiredo, J.; Góis, J.; Vila, M.C. Characterization of a mine legacy site: An approach for environmental management and metals recovery. *Environ. Sci. Pollut. Res.* **2020**, *27*, 10103–10114. [\[CrossRef\]](#)
3. Barlow, N.L.; Bradberry, S.M. Investigation and monitoring of heavy metal poisoning. *J. Clin. Pathol.* **2023**, *76*, 82–97. [\[CrossRef\]](#)
4. Tian, H.; Huang, C.; Wang, P.; Wei, J.; Li, X.; Zhang, R.; Ling, D.; Feng, C.; Liu, H.; Wang, M.; et al. Enhanced elimination of Cr(VI) from aqueous media by polyethyleneimine modified corn straw biochar supported sulfide nanoscale zero valent iron: Performance and mechanism. *Bioresour. Technol.* **2023**, *369*, 128452. [\[CrossRef\]](#)
5. Huang, C.; Zeng, G.; Huang, D.; Lai, C.; Xu, P.; Zhang, C.; Cheng, M.; Wan, J.; Hu, L.; Zhang, Y. Effect of Phanerochaete chrysosporium inoculation on bacterial community and metal stabilization in lead-contaminated agricultural waste composting. *Bioresour. Technol.* **2017**, *243*, 294–303. [\[CrossRef\]](#)
6. Shi, J.; Du, P.; Luo, H.; Wu, H.; Zhang, Y.; Chen, J.; Wu, M.; Xu, G.; Gao, H. Soil contamination with cadmium and potential risk around various mines in China during 2000–2020. *J. Environ. Manag.* **2022**, *310*, 114509. [\[CrossRef\]](#)
7. Qi, M.; Wu, Y.; Zhang, S.; Li, G.; An, T. Pollution Profiles, Source Identification and Health Risk Assessment of Heavy Metals in Soil near a Non-Ferrous Metal Smelting Plant. *Int. J. Environ. Res. Public Health* **2023**, *20*, 1004. [\[CrossRef\]](#)
8. Xiao, L.; Zhou, Y.; Huang, H.; Liu, Y.J.; Li, K.; Li, M.Y.; Tian, Y.; Wu, F. Application of Geostatistical Analysis and Random Forest for Source Analysis and Human Health Risk Assessment of Potentially Toxic Elements (PTEs) in Arable Land Soil. *Int. J. Environ. Res. Public Health* **2020**, *17*, 9296. [\[CrossRef\]](#)
9. Seo, J.W.; Hong, Y.S. Comparative Evaluation of Heavy Metal Concentrations in Residents of Abandoned Metal Mines. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6280. [\[CrossRef\]](#)
10. Zhou, X.Y.; Wang, X.R. Impact of industrial activities on heavy metal contamination in soils in three major urban agglomerations of China. *J. Clean. Prod.* **2019**, *230*, 1–10. [\[CrossRef\]](#)
11. Neeraj, A.; Hiranmai, R.Y.; Iqbal, K. Comprehensive assessment of pollution indices, sources apportionment and ecological risk mapping of heavy metals in agricultural soils of Raebareli District, Uttar Pradesh, India, employing a GIS approach. *Land. Degrad. Dev.* **2023**, *34*, 173–195. [\[CrossRef\]](#)
12. Long, Z.; Zhu, H.; Bing, H.; Tian, X.; Wang, Z.; Wang, X.; Wu, Y. Contamination, sources and health risk of heavy metals in soil and dust from different functional areas in an industrial city of Panzhihua City, Southwest China. *J. Hazard. Mater.* **2021**, *420*, 126638. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Wang, H.; Zhang, H.; Xu, R.K. Heavy metal pollution characteristics and health evaluation of farmland soil in a gold mine slag area of Luoyang in China. *Int. J. Agric. Biol. Eng.* **2021**, *14*, 213–221. [\[CrossRef\]](#)
14. Han, J.; Mammadov, Z.; Kim, M.; Mammadov, E.; Lee, S.; Park, J.; Mammadov, G.; Elovzat, G.; Ro, H.M. Spatial distribution of salinity and heavy metals in surface soils on the Mugan Plain, the Republic of Azerbaijan. *Environ. Monit. Assess.* **2021**, *193*, 95. [\[CrossRef\]](#)
15. Taghizadeh-Mehrjardi, R.; Fathizad, H.; Ali Hakimzadeh Ardakani, M.; Sodaiezhadeh, H.; Kerry, R.; Heung, B.; Scholten, T. Spatio-Temporal Analysis of Heavy Metals in Arid Soils at the Catchment Scale Using Digital Soil Assessment and a Random Forest Model. *Remote Sens.* **2021**, *13*, 1698. [\[CrossRef\]](#)
16. Rashid, A.; Ayub, M.; Ullah, Z.; Ali, A.; Sardar, T.; Iqbal, J.; Gao, X.; Bundschuh, J.; Li, C.; Khattak, S.A.; et al. Groundwater Quality, Health Risk Assessment, and Source Distribution of Heavy Metals Contamination around Chromite Mines: Application of GIS, Sustainable Groundwater Management, Geostatistics, PCAMLR, and PMF Receptor Model. *Int. J. Environ. Res. Public Health* **2023**, *20*, 2113. [\[CrossRef\]](#)
17. Yu, E.; Liu, H.; Dinis, F.; Zhang, Q.; Jing, P.; Liu, F.; Ju, X. Contamination Evaluation and Source Analysis of Heavy Metals in Karst Soil Using UNMIX Model and Pb-Cd Isotopes. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12478. [\[CrossRef\]](#)
18. Kumar, P.; Dipti, Kumar, S.; Singh, R.P. Severe contamination of carcinogenic heavy metals and metalloid in agroecosystems and their associated health risk assessment. *Environ. Pollut.* **2022**, *301*, 118953. [\[CrossRef\]](#)
19. Yaseen, Z.M. An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: Review, challenges and solutions. *Chemosphere* **2021**, *277*, 130126. [\[CrossRef\]](#)
20. Wang, H.; Yilihamu, Q.; Yuan, M.; Bai, H.; Xu, H.; Wu, J. Prediction models of soil heavy metal(loid)s concentration for agricultural land in Dongli: A comparison of regression and random forest. *Ecol. Indic.* **2020**, *119*, 106801. [\[CrossRef\]](#)
21. Azizi, K.; Ayoubi, S.; Nabiollahi, K.; Garosi, Y.; Gislum, R. Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. *J. Geochem. Explor.* **2022**, *233*, 106921. [\[CrossRef\]](#)

22. China National Environmental Monitoring Centre (CNEMC). *The Element Background Values of Chinese Soil*; Chinese Environmental Science Press: Beijing, China, 1990.
23. Zhang, Y.; Guo, R.; Li, Y.; Qin, M.; Zhu, J.; Ma, Z.; Ren, Y. Concentrations, distribution, and risk assessment of endosulfan residues in the cotton fields of northern Xinjiang, China. *Environ. Geochem. Health*. **2022**, *44*, 4063–4075. [[CrossRef](#)] [[PubMed](#)]
24. Cai, Z.; Lei, S.; Zhao, Y.; Gong, C.; Wang, W.; Du, C. Spatial Distribution and Migration Characteristics of Heavy Metals in Grassland Open-Pit Coal Mine Dump Soil Interface. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4441. [[CrossRef](#)] [[PubMed](#)]
25. Liao, H.W.; Jiang, Z.C.; Zhou, H.; Qin, X.Q.; Huang, Q.B.; Zhong, L.; Pu, Z.G. Dissolved Heavy Metal Pollution and Assessment of a Karst Basin around a Mine, Southwest China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 14293. [[CrossRef](#)]
26. Cao, J.; Xie, C.Y.; Hou, Z.R. Spatiotemporal distribution patterns and risk characteristics of heavy metal pollutants in the soil of lead–zinc mines. *Environ. Sci. Eur.* **2022**, *34*, 27. [[CrossRef](#)]
27. Wang, Z.; Bing, H.; Zhu, H.; Wu, Y. Fractions, Contamination and Health Risk of Cadmium in Alpine Soils on the Gongga Mountain, Eastern Tibetan Plateau. *Bull. Environ. Contam. Toxicol.* **2021**, *106*, 86–91. [[CrossRef](#)] [[PubMed](#)]
28. Cao, J.; Xie, C.Y.; Hou, Z.R. Ecological evaluation of heavy metal pollution in the soil of Pb-Zn mines. *Ecotoxicology* **2022**, *31*, 259–270. [[CrossRef](#)]
29. Jia, X.L.; Fu, T.T.; Hu, B.F.; Shi, Z.; Zhou, L.Q.; Zhu, Y.W. Identification of the potential risk areas for soil heavy metal pollution based on the source-sink theory. *J. Hazard. Mater.* **2020**, *393*, 122424. [[CrossRef](#)]
30. Guleria, A.; Singh, R.; Chakma, S.; Birke, V. Ecological and human health risk assessment of chromite ore processing residue (COPR) dumpsites in Northern India: A multi-pathways based probabilistic risk approach. *Process. Saf. Environ.* **2022**, *163*, 405–420. [[CrossRef](#)]
31. Zhou, W.; Yang, H.; Xie, L.J.; Li, H.R.; Huang, L.; Zhao, Y.P.; Yue, T.X. Hyperspectral inversion of soil heavy metals in Three-River Source Region based on random forest model. *Catena* **2021**, *202*, 105222. [[CrossRef](#)]
32. Guo, F.; Xu, Z.; Ma, H.H.; Liu, X.J.; Yang, Z.; Tang, S.Q. A Comparative Study of the Hyperspectral Inversion Models Based on the PCA for Retrieving the Cd Content in the Soil. *Spectrosc. Spect. Anal.* **2021**, *41*, 1625. [[CrossRef](#)]
33. Jia, X.; Cao, Y.; O'Connor, D.; Zhu, J.; Tsang, D.C.W.; Zou, B.; Hou, D. Mapping soil pollution by using drone image recognition and machine learning at an arsenic-contaminated agricultural area. *Environ. Pollut.* **2021**, *270*, 116281. [[CrossRef](#)] [[PubMed](#)]
34. Huang, S.; Shao, G.F.; Wang, L.Y.; Tang, L.N. Spatial Distribution and Potential Sources of Five Heavy Metals and One Metalloid in the Soils of Xiamen City, China. *Bull. Environ. Contam. Toxicol.* **2019**, *103*, 308–315. [[CrossRef](#)]
35. Liu, G.; Zhou, X.; Li, Q.; Shi, Y.; Guo, G.L.; Zhao, L.; Wang, J.; Su, Y.Q.; Zhang, C. Spatial distribution prediction of soil As in a large-scale arsenic slag contaminated area based on an integrated model and multi-source environmental data. *Environ. Pollut.* **2020**, *267*, 115631. [[CrossRef](#)]
36. Shi, X.; Ren, B. Predict three-dimensional soil manganese transport by HYDRUS-1D and spatial interpolation in Xiangtan manganese mine. *J. Clean. Prod.* **2021**, *292*, 125879. [[CrossRef](#)]
37. Guo, Z.H.; Zhang, Y.X.; Xu, R.; Xie, H.M.; Xiao, X.Y.; Peng, C. Contamination vertical distribution and key factors identification of metal(loid)s in site soil from an abandoned Pb/Zn smelter using machine learning. *Sci. Total Environ.* **2023**, *856*, 159264. [[CrossRef](#)] [[PubMed](#)]
38. Li, X.Y.; Geng, T.; Shen, W.J.; Zhang, J.R.; Zhou, Y.Z. Quantifying the influencing factors and multi-factor interactions affecting cadmium accumulation in limestone-derived agricultural soil using random forest (RF) approach. *Ecotox. Environ. Safe.* **2021**, *209*, 111773. [[CrossRef](#)] [[PubMed](#)]
39. Liu, X.; Shi, H.; Bai, Z.; Zhou, W.; Liu, K.; Wang, M.; He, Y. Heavy metal concentrations of soils near the large opencast coal mine pits in China. *Chemosphere* **2020**, *244*, 125360. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.