Journal of
Marine Science and Engineering

MDPI

Article

Source Apportionment of Soil Heavy Metal(Loid)s in Farmland Using Diverse Models: A Comparative Assessment in the Yellow River Delta

Wei Huang [1], Shuhuan Wang [1], Lu Wang [2,*], Yingqiang Song [1,*], Yue Zhu [1], Hao Yang [3,4], Yingkai Xie [5,6] and Yueming Hu [4]

[1] School of Civil Engineering and Geomatics, Shandong University of Technology, Zibo 255000, China
[2] School of Public Administration, Hainan University, Haikou 570228, China
[3] College of Natural Resources and Environment, South China Agricultural University, Guangzhou 510642, China
[4] College of Tropical Crops, Hainan University, Haikou 570228, China
[5] South China Academy of Natural Resources Science and Technology, Guangzhou 510610, China
[6] Guangdong Youyuan Land Information Technology Co., Ltd., Guangzhou 510610, China
* Correspondence: selinapple@163.com (L.W.); songyingqiang@sdut.edu.cn (Y.S.)

Abstract: The rapid development of industrialization and urbanization has posed serious challenges for coastal farmland ecosystems. Source apportionment of soil heavy metals is an effective way for the detection of non-point source pollution in farmland to help support the high-quality development of coastal agriculture. To this end, 113 surface soil samples were collected in the coastal delta of China, and the contents of As, Cd, Cr, Cu, Ni, Pb, and Zn were determined. A variety of models were integrated to apportion the source of soil heavy metals, including positive matrix factorization (PMF), geographical detector (GD), eXtreme gradient boosting (XGBoost), and structural equation modeling (SEM). The result of PMF models revealed that there was collinearity between various heavy metals, and the same heavy metal may have a mixed source. The XGBoost model analysis indicated that there were significant non-linear relationships between soil heavy metals and source factors. A synergy between air quality and human activity factors was the key source of heavy metal that entered the study area, based on the results of the GD. Furthermore, the input path effect of heavy metals in the soil of the study area was quantified by SEM. The balance of evidence from the above models showed that air quality (SO_2 and NO_2) and factories in the study area had the greatest impacts on Cd, Cr, and Zn. Natural sources were dominant for Pb, while As, Cu, and Ni were contributed by soil parent material and factories. The above results led to the conclusion that there was a cycle path in the study area that continuously promoted the migration and accumulation of heavy metals in farmland soil; that is, the heavy metals discharged during oil exploitation and smelting entered the atmosphere and then accumulated in the farmland soil through precipitation, atmospheric deposition, and other paths. In this study, it is shown that a variety of models can be used to more comprehensively assess the sources of soil heavy metals. This approach can provide effective support for the rapid prevention and decision-making management of soil heavy metal pollution in coastal areas.

Keywords: coastal ecosystems; XGBoost; PMF; SEM; farmland

check for updates

Citation: Huang, W.; Wang, S.; Wang, L.; Song, Y.; Zhu, Y.; Yang, H.; Xie, Y.; Hu, Y. Source Apportionment of Soil Heavy Metal(Loid)s in Farmland Using Diverse Models: A Comparative Assessment in the Yellow River Delta. *J. Mar. Sci. Eng.* **2023**, *11*, 1069. https://doi.org/10.3390/jmse11051069

Academic Editors: Michele Arienzo and Ivan Semenkov

Received: 28 March 2023
Revised: 10 May 2023
Accepted: 16 May 2023
Published: 17 May 2023

1. Introduction

Heavy metal pollution can be a significant threat to coastal delta farmland and seriously threaten soil quality, food security, and human health [1–3]. Given the high cost and difficulties associated with remediation of heavy metal pollution in soil, source apportionment is an important way to identify and help prevent heavy metal pollution in farmland soil [4,5]. Generally, the sources of heavy metals in soil include natural sources (i.e., soil parent material) and anthropogenic sources, such as industrial mining, sewage irrigation, traffic emissions, etc. [6–8]. Due to global industrialization and urbanization over the past few decades, farmland soils have been major receivers of heavy metals from human activities. Rapid quantification of the source types and contributions of these heavy metals is of great significance to ensuring the ecological health of agricultural systems and the safety of agricultural products.

Source apportionment methods for soil heavy metals can be divided into three categories according to the model on which they are based: statistical and geostatistical models, receptor models, and machine learning models. Statistical models are based on mathematical calculations to apportion pollution source types of different soil heavy metals and include principal component analysis (PCA) and cluster analysis (CA) [9–11]. Statistical models can be used to infer several types of sources of soil heavy metals by grouping variables that behave similarly. For example, Zhou et al. [12] used PCA to analyze the sources of soil heavy metals in the old industrial zone in Jiangsu Province of China and found that Cd and Pb belong to anthropogenic inputs, Hg and Cu belong to natural and anthropogenic sources, while As, Pb, Cr, and Ni are from natural sources of soil parent material. However, statistical identification models rely on prior knowledge; they do not have knowledge of pollution sources, and source identification results cannot quantify specific natural or anthropogenic source types. Geostatistical models represent a statistical method that uses the spatial relationships of soil samples [13,14]. In such methods, an ordinary kriging interpolation method is used to obtain the spatial distribution map of soil heavy metals, which can be used to infer the possible sources of soil heavy metals in high concentration areas. Similar to the traditional statistical model, the geostatistical model only subjectively determines that one or more heavy metals may belong to natural or anthropogenic sources, resulting in a collinearity of the sources of various heavy metals.

The receptor model approach is used to analyze the contribution of different sources based on the chemical and physical characteristics of pollutants and can include the fingerprint screening approach [15], chemical mass balance (CMB) [16], positive matrix factorization (PMF) [17], and absolute principal component score/multiple linear regression methods (APCS/MLR) [18]. For example, Wang et al. (2019) analyzed the sources of various heavy metals using isotopic composition and PMF and found that Cd may come from smelting or refinery emissions, while other heavy metals come from both anthropogenic and natural sources [19]. Huang et al. (2018) used APCS/MLR to analyze the sources of heavy metals As, Cd, Hg, and Pb in the soil of a mining area and believed that the lead-zinc industry is the most important anthropogenic emission source, affecting nearly half of Pb and As accumulation and about one-third of Cd accumulation [20]. However, the receptor model also requires prior knowledge to assume that the source of all heavy metals is a contribution matrix of multiple sources. Since the linear relationship between different sources and soil heavy metals is assumed, the receptor model ignores the nonlinear source-sink relationship between sources and soil heavy metals. For example, when the pollution sources around the farmland are not obvious, a single heavy metal in the soil may exist from multiple sources, such as natural input, industrial and mining emissions, and traffic emissions. It is difficult to quantify the contribution of different sources to the same soil heavy metals using the receptor model.

In order to reflect the quantitative contribution of different sources of soil heavy metals, machine learning models have been widely applied, including support vector machines (SVM), random forests (RF), artificial neural networks (ANN), etc. [21–24]. This method essentially establishes a nonlinear fitting relationship between various sources and soil heavy metals and realizes its quantitative contribution by calculating the importance of different sources. The machine learning approach can effectively explain the nonlinear coupling relationship between sources (such as topography, vegetation index, atmosphere, factories, traffic, etc.) and heavy metals [25], and it is widely implemented in source apportionment. However, the machine learning model is subject to the fitting accuracy of the target when calculating the contribution of different sources and does not consider the correlation among various sources.

In fact, there is a complex linear-nonlinear coupling relationship between soil heavy metals and their sources. The source apportionment results obtained using multiple models, when compared with those of a single source apportionment model, show that multiple approaches are effective for source apportionment of heavy metals in soil [26,27]. These relationships include explaining the collinearity between soil heavy metals from multiple perspectives, the synergistic contribution between source factors, and the linear-nonlinear relationship between source factors and heavy metals. In other words, most studies are limited to the results of a single analysis model for source interpretation, failing to fully consider the advantages and limitations of different source apportionment models for a more comprehensive judgment. To consider the above advantages and disadvantages of source apportionment models, we propose a scientific hypothesis: can a balance of evidence approach give a more reasonable and comprehensive assessment of the linear-nonlinear relationship between sources and soil heavy metals? Therefore, we developed diverse models to comprehensively assess sources of soil heavy metals (As, Cd, Cr, Cu, Ni, Pb, and Zn) in coastal delta farmland: (1) to analyze the collinearity of heavy metals in soils using positive matrix factorization (PMF); (2) to detect the single and interactive contributions of topography, vegetation, air quality, and human activity factors for soil heavy metals using a geographical detector (GD); (3) to fit the nonlinear relationship between soil heavy metals and source factors using the eXtreme Gradient Boosting (XGBoost) model; and (4) to quantify the synergistic contribution of source factors of soil heavy metals using the structural equation model (SEM).

2. Materials and Methods

2.1. Study Area

The study area (118°30′–119°9′ E, 37°30′–38°8′ N) is located in the coastal delta of China (Figure 1). The estimated area of the study area is 2.15×10^3 km^2. The annual average temperature, annual average precipitation, and annual sunshine hours are 14.1 °C, 665.3 mm, and 2998.5 h, respectively. The soil parent material in the study area is Yellow River sediment, and the soil type is fluvo-aquic soil. Due to the influence of underground high-salinity diving and human farming activities, soil salinization is serious in the farmlands of the study area. The terrain tilts from southwest to northeast along the Yellow River, and the main crops include wheat, corn, and rice. In this study, the selection of the analyte set (i.e., As, Cd, Cr, Cu, Ni, Pb, and Zn) is based on the Soil Environmental Quality and Risk Control Standard for Soil Contamination of Agricultural Land (GB 15618-2018) (https://www.mee.gov.cn/, (accessed on 15 October 2019)), which was published by the Ministry of Ecology and Environment of the People's Republic of China. In this standard, the heavy metals and risk intervention values involved in soil contamination of agricultural land are specified.

2.2. Sample Collection and Analysis

In the study area, 113 samples were collected from the surface soil (0–20 cm) of the farmland in October 2020. Based on high-resolution remote sensing images, we determined the location of the farmland and evaluated traffic accessibility. We also prepared relevant materials, including gloves, record books, sampling tables, soil drills, soil shovels, sample bags, etc. In each field, 4–6 samples of soil were taken, respectively. After fully mixing, the soil (1 kg) was loaded into a sample bag, and GPS coordinates were recorded. The soil samples were naturally dried indoors and passed through a 200-mesh nylon sieve after being ground. The detailed sampling criteria are based on the Technical Specification for Soil Environmental Monitoring (HJ/T 166-2004) (https://www.mee.gov.cn/, (accessed on 15 October 2019)), which was published by the Ministry of Ecology and Environment of the People's Republic of China.

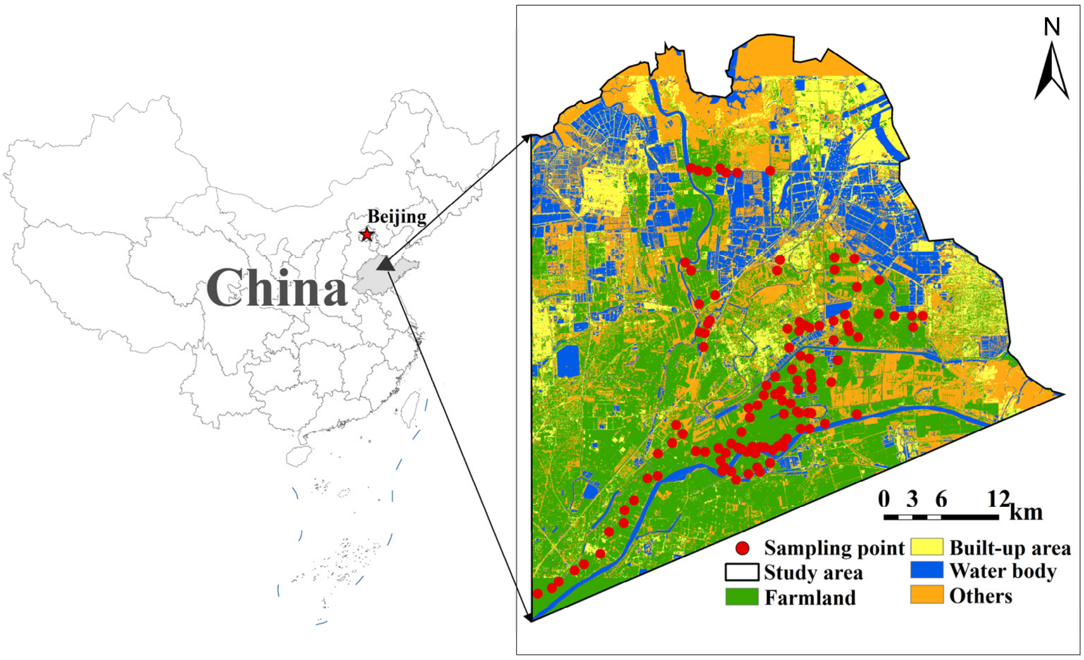

Figure 1. Location of the study area and distribution of soil sampling sites.

Soil samples (0.2 g) were weighed, and 6 mL HNO_3, 2 mL HCl, and 2 mL HF (proportion of 3:1:1) were added in the fume cupboard, and all operations were repeated thrice in parallel. Pretreatment samples were then put into the microwave dissolver (CEM MARS5, Matthews, NC, USA), and time and temperature were set in three stages: 10 min heating to 120 °C for 3 min; 5 min heating to 160 °C for 3 min; and 5 min heating to 190 °C for 40 min. Each batch of samples also includes a GSS-1 standard soil sample and a blank sample (Table S1) in order to ensure the accuracy of the measurement process by ICP-MS. Finally, inductively coupled plasma mass spectrometry (Agilent ICP-MS 7500ce, Santa Clara, CA, USA) was used to quantify As, Cd, Cr, Cu, Ni, Pb, and Zn. The detailed reference is available in two standards published by the Ministry of Ecology and Environment of the People's Republic of China: Soil and Sediment-Determination of Aqua Regia Extracts of 12 Metal Elements-Inductively Coupled Plasma Mass Spectrometry (HJ 803-2016) and Soil and Sediment-Determination of Total Metal Elements-Microwave Assisted Acid Digestion Method (HJ 832-2017) (https://www.mee.gov.cn/, (accessed on 15 October 2019)).

2.3. Auxiliary Data

2.3.1. Terrain

Different terrain factors were extracted based on the digital elevation model (DEM), which was obtained from the shuttle radar topography mission (SRTM) dataset (https://earthexplorer.usgs.gov/, (accessed on 10 January 2020)) with a resolution of 30 m × 30 m. Then, the surface analysis was applied to obtain terrain factors in ArcGIS 10.7 software, including DEM, slope (SLO), and relief (REF).

2.3.2. Vegetation

The Landsat 8 OLI image (24 October 2020) was selected to calculate vegetation factors. The OLI image was processed by radiometric calibration, atmospheric correction, mixed pixel decomposition, and cropping using ENVI 5.3 software. Five vegetation index factors, including the clay index (CI), carbonate index (CAI), normalized difference vegetation index (NDVI), and soil-adjusted vegetation index (SAVI), were calculated from the processed images (Table S2) [28–31], which had a resolution of 30 m × 30 m.

2.3.3. Air Quality

Air quality data from 40 air monitoring stations in Dongying city (annual average in 2020) were used to reflect the impact of the atmosphere on soil heavy metals (Figure S1). The data were from the Dongying air quality real-time publishing system (http://218.58.213.53:8081/dyfb_air/fb_web, (accessed on 10 January 2021)). The air quality factors of 40 monitoring stations were interpolated by ordinary kriging in ArcGIS 10.7 software, and the annual average raster data of NO_2, $PM_{2.5}$, PM_{10}, and SO_2 were obtained with a resolution of 30 m × 30 m.

2.3.4. Human Activity

The point of interest (POI) data was obtained from the Baidu map (https://map.baidu.com, (accessed on 10 May 2021)) using a web crawler method. Factory (FAC) (662), traffic (TRA) (406), and residential (RES) (571) points were obtained to analyze the response relationship between human activities and soil heavy metals. Based on the Chinese GF-1 remote sensing image (http://www.sasclouds.com/chinese/normal/, (accessed on 10 January 2021)) with a resolution of 2 m × 2 m, 701 oil well (OLW) points were acquired by visual interpretation. All point data were analyzed by kernel density in ArcGIS 10.7 software to obtain a spatial distribution map with a resolution of 30 m × 30 m.

In order to avoid the influence of numerical range differences on the accuracy of four source apportionment models, we normalized the values of all factors and soil heavy metals. The normalization formula is as follows [32]:

$$X_i = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (1)$$

where X_i is the rescaled data, and x_{min} and x_{max} denote the minimum and maximum observed data. The normalized auxiliary factors in the study area are shown in Figure 2. The spatial distribution of human activities and air quality factors is clustered, indicating that these areas may cause a strong accumulation of soil heavy metals in the study area.

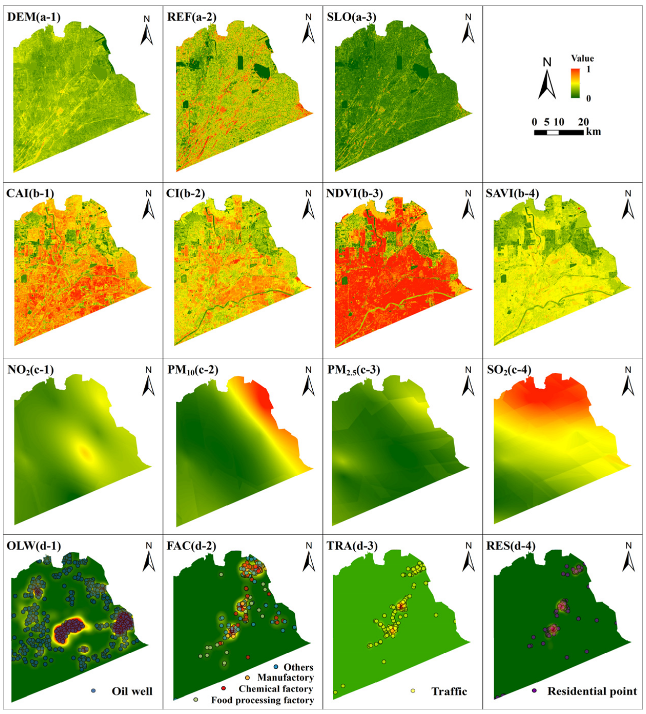

Figure 2. Auxiliary factors in the study area, including (**a-1**) DEM, (**a-2**) REF, (**a-3**) SLO, (**b-1**) CAI, (**b-2**) CI, (**b-3**) NDVI, (**b-4**) SAVI, (**c-1**) NO_2, (**c-2**) PM_{10}, (**c-3**) $PM_{2.5}$, (**c-4**) SO_2, (**d-1**) OLW, (**d-2**) FAC, (**d-3**) TRA, and (**d-4**) RES.

2.4. Source Apportionment Method

2.4.1. PMF

The positive matrix factorization (PMF) model is recommended by the United States Environmental Protection Agency (EPA) (https://www.epa.gov/, (accessed on 10 January 2021)) for matrix analysis of pollution sources based on receptor models [26,33]. It has the advantage of not being limited by the composition of a single pollution source and has been commonly used in the field of pollutant source analysis. The load of soil heavy metals on PMF factors is an important basis for the quantitative distribution of heavy metal sources [34]. The PMF model decomposes the heavy metal element concentration matrix (X) into a source factor score matrix (G), a source factor load matrix (F), and a factor residual matrix (E) by using correlation and covariance matrices, and then determines the source contribution rate of different heavy metals according to prior knowledge. The source apportionment result of the PMF model was calculated in the EPA PMF 5.0 software, and its formula is as follows [35]:

$$X_{ij} = \sum_{k=1}^{p} G_{ik}F_{kj} + E_{ij}, \tag{2}$$

where X_{ij} represents the measured concentration of the j-th heavy metal element at the i-th sampling point, G_{ik} is the relative contribution of the source factor k to the i-th sampling point, F_{kj} is the concentration of the j-th heavy metal element in the source factor k, and E_{ij} is the residual of the j-th element at the i-th sampling point.

The PMF model uses multiple iterations to continuously decompose the sample concentration matrix to obtain the optimal source factor score matrix (G) and source factor load matrix (F) such that the objective function Q is minimized. The objective function Q is defined as:

$$Q = \sum_{i=1}^{n}\sum_{j=1}^{m}\left[\frac{x_{ij} - \sum\limits_{k=1}^{p} G_{ik}F_{kj}}{U_{ij}}\right]^2, \tag{3}$$

where U_{ij} is the uncertainty of X_{ij}, that is, the uncertainty of the concentration of the j-th heavy metal element in the i-th sample. The uncertainty (Unc) formula is calculated from the species-specific method detection limit (MDL) in this study.

When the element concentration is greater than MDL:

$$Unc = \sqrt{(\text{errorfraction} \times \text{concentrations})^2 + (0.5 \times \text{MDL})^2} \tag{4}$$

When the element concentration is less than MDL:

$$Unc = \frac{5}{6} \times \text{MDL} \tag{5}$$

We adopted a robust mode according to the principle that the initial eigenvalue is greater than 1% based on principal component analysis (PCA). By trying to set different factor numbers and operation times, we finally determined that the operation should be conducted 20 times and that the number of factors is 4. When the factor number is 4, Qrobust/Qtrue is in rapid decline, and the residual size is low. Then, the 4-factor scenario was brought into the PMF model for verification. After debugging, the results showed that the objective function Q value was the smallest at the 11th iteration. The categories of each component in the PMF model were strong, and the S/N values were all greater than 8.8.

2.4.2. Geographical Detector

The geographical detector (GD) is applied to detect the spatial heterogeneity of soil heavy metals and reveal the importance of their source factors, including differentiation factor detection and interaction detection [36]. Differentiation factor detection can detect the interpretation of source factors in the spatial variability of soil heavy metals. Interaction detection (i.e., double-factor) can further determine the synergistic effect of dual-source factors on the spatial variability of soil heavy metals. It has five interactive relationships: nonlinear weakening, single-factor nonlinear weakening, double-factor enhancement, independence, and nonlinear enhancement [36]. The double-factor method is the interaction assessment of the two sources, which results in the interaction phenomena of multiple source combinations. In the calculation process of the double-factor method, it can combine multiple source factors in pairs and traverse the spatial correlation of each combination for soil heavy metals. Finally, the synergistic effects of double factors with similar or different sources on soil heavy metals were obtained. The detailed theory of this model is available in previous works [37–39].

2.4.3. XGBoost Model

The XGBoost model is developed from decision tree methodology and has been commonly applied in flash flood risk assessment [40], mineral potential mapping [41], and winter wheat SPAD estimation [42]. The principle of the XGBoost model is to assign an early prediction value to the root of the tree, calculate the residual value of the data set (the difference between the predicted value and the observed value), and then distribute all the residuals to the root of the tree so that it can process sparse data. In order to obtain the optimal solution of the XGBoost model, we use the grid search (GS) algorithm to optimize the hyperparameters of the XGBoost model. GS is used to optimize the XGBoost model by traversing a given combination of hyperparameters. The GS algorithm is simple and widely used, and it is suitable for the adjustment of small-range hyperparameters [43,44]. The source contribution to soil heavy metals was interpreted by calculating the importance score of source factors based on an optimized XGBoost model.

In this study, the XGBoost model was established by using a Python 3.9 programming environment and importing the XGBoost package. The source factors were used as the input variables of the XGBoost model. All samples were randomly divided into a 70% training set and a 30% validation set. The source contribution to soil heavy metals was interpreted by calculating the importance score of source factors, and the accuracy was evaluated based on the coefficient of determination (R^2) [45,46].

$$\mathrm{R}^2 = \frac{\left[\sum_{i=1}^{n}\left(p(x_i) - \overline{p(x_i)}\right)\left(\hat{p}(x_i) - \overline{\hat{p}(x_i)}\right)\right]^2}{\sum_{i=1}^{n}\left(p(x_i) - \overline{p(x_i)}\right)^2 \sum_{i=1}^{n}\left(\hat{p}(x_i) - \overline{\hat{p}(x_i)}\right)^2} \tag{6}$$

In the above formula, $\hat{p}(x_i)$ and $p(x_i)$ are the measured values and predicted values of soil heavy metal contents, respectively, with the XGBoost model, and $\overline{p(x_i)}$ and $\overline{\hat{p}(x_i)}$ are the stationary mean values of $p(x_i)$ and of $\hat{p}(x_i)$ soil heavy metal contents at n sites.

2.4.4. Structural Equation Model

The structural equation model (SEM) is a statistical method for analyzing variables based on their covariance matrix [47]. SEM consists of a structural model and a measurement model. The structural model is used to describe the mathematical relationship between the source factors and the soil heavy metal elements [48], and the measurement model is used to quantify the contribution of the source factors to the soil heavy metals. Its advantages are that it is good at solving the problem of latent variables when dealing with the source analysis of soil heavy metals and can deal with multiple dependent variables at the same time. With SEM, the fitting degree of the whole model can be evaluated by estimating the factor structure and factor relationship [49], and it can be used to efficiently and quickly reveal the complex relationship between soil heavy metals and source factors.

In this study, soil heavy metals, vegetation (VEG), terrain (TER), air quality (AQ), and human activity (HA) were used as latent variables of the SEM, and high driving source factors based on GD screening were used as measurement variables. In order to ensure the objectivity of SEM analysis, the importance of source factors in the XGBoost model is used as the initialization weight of the SEM modeling process. Cronbach's α and KMO were used to verify the accuracy of SEM. Cronbach's α values of heavy metals, VEG, TER, AQ, and HA as latent variables are 0.615, 0.782, 0.676, 0.631, and 0.720, respectively, and the KMO values are greater than 0.55, indicating that the data have good reliability and validity in the modeling process of the structural equation. The SEM model is executed in IBM SPSS Statistics 26 and IBM Amos 26 software.

3. Results

3.1. Statistical Analysis

The results for the statistical analysis of soil heavy metal contents are shown in Table 1, and the average concentrations of As, Cd, Cr, Cu, Ni, Pb, and Zn were 22.3 mg kg^{-1}, 0.36 mg kg^{-1}, 122 mg kg^{-1}, 23.0 mg kg^{-1}, 30.1 mg kg^{-1}, 21.8 mg kg^{-1}, and 43.7 mg kg^{-1}, respectively. The coefficient of variation (CV) of the seven heavy metals is Cd (53%) > Pb (31%) > Zn (28%) > Cr (25%) > Cu (22%) > As (22%) > Ni (19%), which indicates the large numerical dispersion of Cd and Pb. Based on the background (BG) values of Shandong province of China [50], the excessive rates of As, Cd, Cr, Cu, Ni, Pb, and Zn are 100%, 99.1%, 98.2%, 47.8%, 66.4%, 30.1%, and 7.08%, respectively. The statistical results of the numerical normalization of seven heavy metals are shown in Figure 3. For all heavy metals (except As and Cr), the values show they are mostly concentrated in the low-value area, and the numerical structure exhibits a nonstandard distribution. In addition, the mean and median of all heavy metals are far apart, indicating that the spatial variability of heavy metals is strong in the study area.

Table 1. Descriptive statistics of soil heavy metals (mg kg^{-1}).

Elements	Min	Max	Mean	SD	CV (%)	BG	Excessive Rate (%)
As	9.78	31.3	22.3	4.79	22	8.60	100
Cd	0.13	0.91	0.36	0.19	53	0.13	99.1
Cr	53.8	197	123	30.3	25	62.0	98.2
Cu	12.7	38.8	23.0	4.98	22	22.6	47.8
Ni	17.8	45.1	30.1	5.64	19	27.1	66.4
Pb	10.4	45.4	21.8	6.67	31	23.6	30.1
Zn	25.7	86.1	43.7	11.8	28	63.3	7.08

Figure 4 presents the numerical distribution of the vegetation index, which conforms to a normal distribution, and other factors that present left-skewed or right-skewed distributions. The normalized values of RES, TRA, and FAC indicate aggregation in the low-value area (0–0.2). This is because POI is mainly distributed in the city, the farmland sample point is far away from the city, and the corresponding kernel density value is low. The positive linear relationship between similar source factors was significant (CI–SAVI–NDVI, FAC–TRA–RES, and SO_2–$PM_{2.5}$), and there is a strong negative linear relationship between vegetation factor and air quality factor as well as between terrain factor and air quality factor. The linear relationship between terrain, vegetation, human activity, and other factors is not obvious. This indicates that there are strong nonlinear characteristics between different source factors, and inputting these characteristics into different source analysis models may result in the complexity of soil heavy metal sources (multiple homologous or homologous sources).

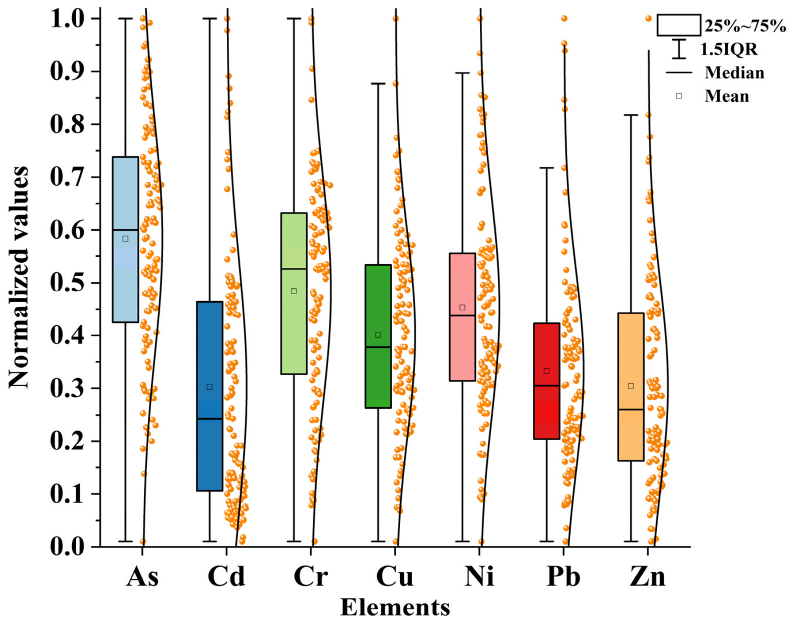

Figure 3. The statistical results of the normalized value for soil heavy metals.

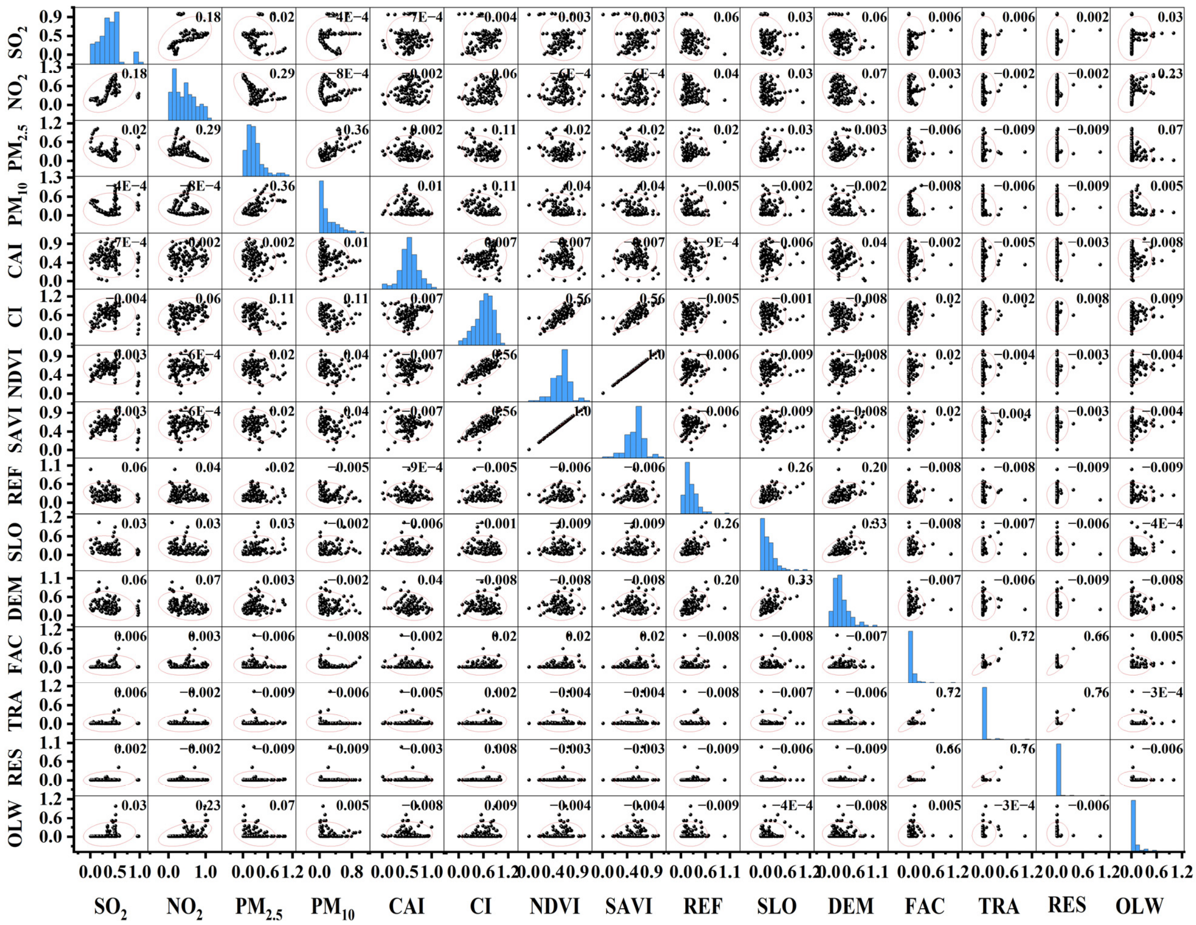

Figure 4. The numerical distribution of normalized values for all factors. The number in the upper right corner of each grid is an adjusted R-square of linear fit.

3.2. Source Apportionment of Heavy Metal in Soil Using the Balance of Evidence Method

In this study, a balance of evidence method is used, including PMF, GD, XGBoost, and SEM, to analyze the sources of heavy metals in soil. The four source apportionment models are found to be complementary, but in different aspects. The PMF can be carried out using a collinearity assessment of soil heavy metals from a statistical perspective. The GD and XGBoost models quantify the driving effect of each source on soil heavy metals, whereas the SEM expands on the accumulation path.

3.2.1. Homology Analysis of Soil Heavy Metals

The source apportionment results of soil heavy metals by the PMF model revealed that the proportions of four pollution source factors are 36.2%, 13.8%, 23.7%, and 26.3%, respectively. Zn (54.3%) and Cu (33.3%), Cd (57.5%) and Cr (36.1%), and Cd (42.5%) and Pb (41.9%) had higher contribution rates in Factor 1, Factor 2, and Factor 3, respectively, while Pb (48.1%), Cr (37.6%), Ni (30.9%), and As (34.8%) had higher contribution rates in Factor 4 (Figure 5a). The distribution trend of high-concentration heavy metals was similar, and they are more likely to have the same source (Figure 5b). The industrial source may be the main source of Zn and Cu in Factor 1 [33]. In Factor 2, Cr could be contributed by various industrial sources, including a range of mining and metallurgical smelting and refining processes that can also result in the release of Cd into the air, water, and soil [51]. Mining and smelting activities also discharge large amounts of dust containing Cd and Pb [52], resulting in accumulation in the soil (Factor 3). Moreover, phosphoric fertilizer and pesticides may contain a large amount of As; Pb was usually considered a sign of transportation; and Ni was influenced by petrochemical and petroleum smelting (Factor 4) [53].

In addition, the factors derived from the PMF are not necessarily indicative of external sources but may, to some extent, reflect intrinsic geochemical associations. For example, Ni, Cr, and As are associated with natural (i.e., lithogenic aluminosilicates) sources, and Fe and Mn oxides adsorb heavy metals (i.e., Cr, Pb, Cu, Cd, Zn, and Ni) in strongly weathered soils [54,55]. The results of PMF models in assessing the sources of heavy metals in soil were quite different, indicating that the subjective source apportionment model has great uncertainty. These complex uncertainties also indicate that the same heavy metal may come from multiple input sources, and a variety of heavy metals may also have the same source. It is necessary to combine the quantitative model to further infer the sources of soil heavy metals.

3.2.2. Source Driving Effect for Soil Heavy Metals

The highest driving factors from the single-factor detection of the GD model for As, Cd, Cr, Cu, Ni, Pb, and Zn were CI (0.70), NO_2 (0.69), SO_2 (0.67), CI (0.61), CI (0.64), CI (0.72), and NO_2 (0.68), respectively (Figure 6a). The driving force of OLW for all heavy metals also exceeds 0.3. On the whole, air quality and vegetation factors have a higher degree of total explanation for the spatial heterogeneity of heavy metals, while terrain factors have the smallest driving force. According to the double-factor co-detection results of the GD model (Figure 6b–h), the driving forces of the double-factor from the same source type (i.e., CI–SAVI, CI–NDVI, CAI–SAVI, and CAI–NDVI) for seven heavy metals are all more than 0.97, much higher than for the single-factor driving contribution. In addition, the synergistic driving effect of different source types of factors on soil heavy metals was also high. For example, the driving forces of vegetation (CI, CAI, SAVI, and NDVI) and air quality factors ($PM_{2.5}$, PM_{10}, SO_2, and NO_2) on soil heavy metals were more than 0.92, and the driving forces of human activity factors (OLW and FAC) and other factors were also high. The results of double-factor synergistic detection indicated that the spatial variation of soil heavy metals in the study area was significantly affected by multiple sources, and the sources had complex mixed input characteristics.

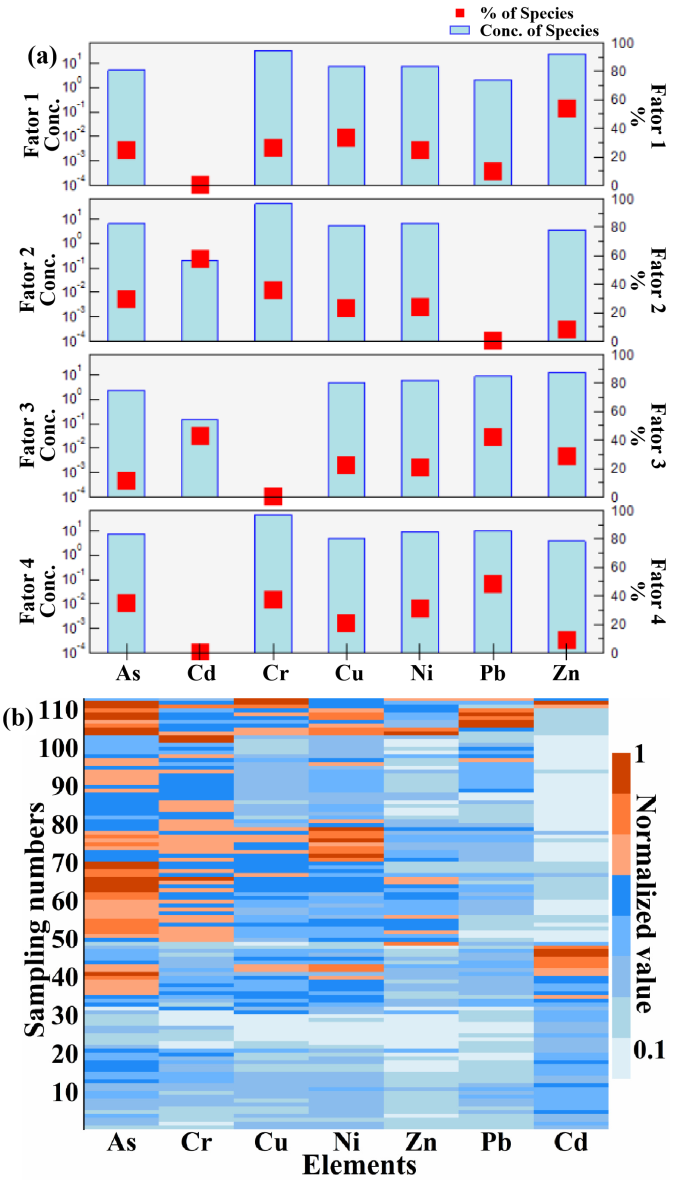

Figure 5. Source apportionment results of (**a**) contribution rates of factors in four sources for soil heavy metals and (**b**) distribution trends of high-concentration heavy metals in 113 samples.

The prediction accuracies (R^2) of the XGBoost model for As, Cd, Cr, Cu, Ni, Pb, and Zn were 57.88%, 60.26%, 63.45%, 59.82%, 58.63%, 62.28%, and 58.56%, respectively (Figure 7). The XGBoost model demonstrates good nonlinear fitting performance between source factors and soil heavy metals and can reflect their nonlinear response relationships. The importance of source factors exhibited that NO_2 was most important for Cr (33%), As (32%), and Ni (30%), FAC contributed the most to the prediction accuracy of Cu (30%), Cr (30%), and Zn (29%), and Pb (18.97%) was influenced by SLO. In addition, the percentage of cumulative importance of source types for seven heavy metals from high to low was air quality (51%), human activity (25%), vegetation index (14%), and terrain (10%).

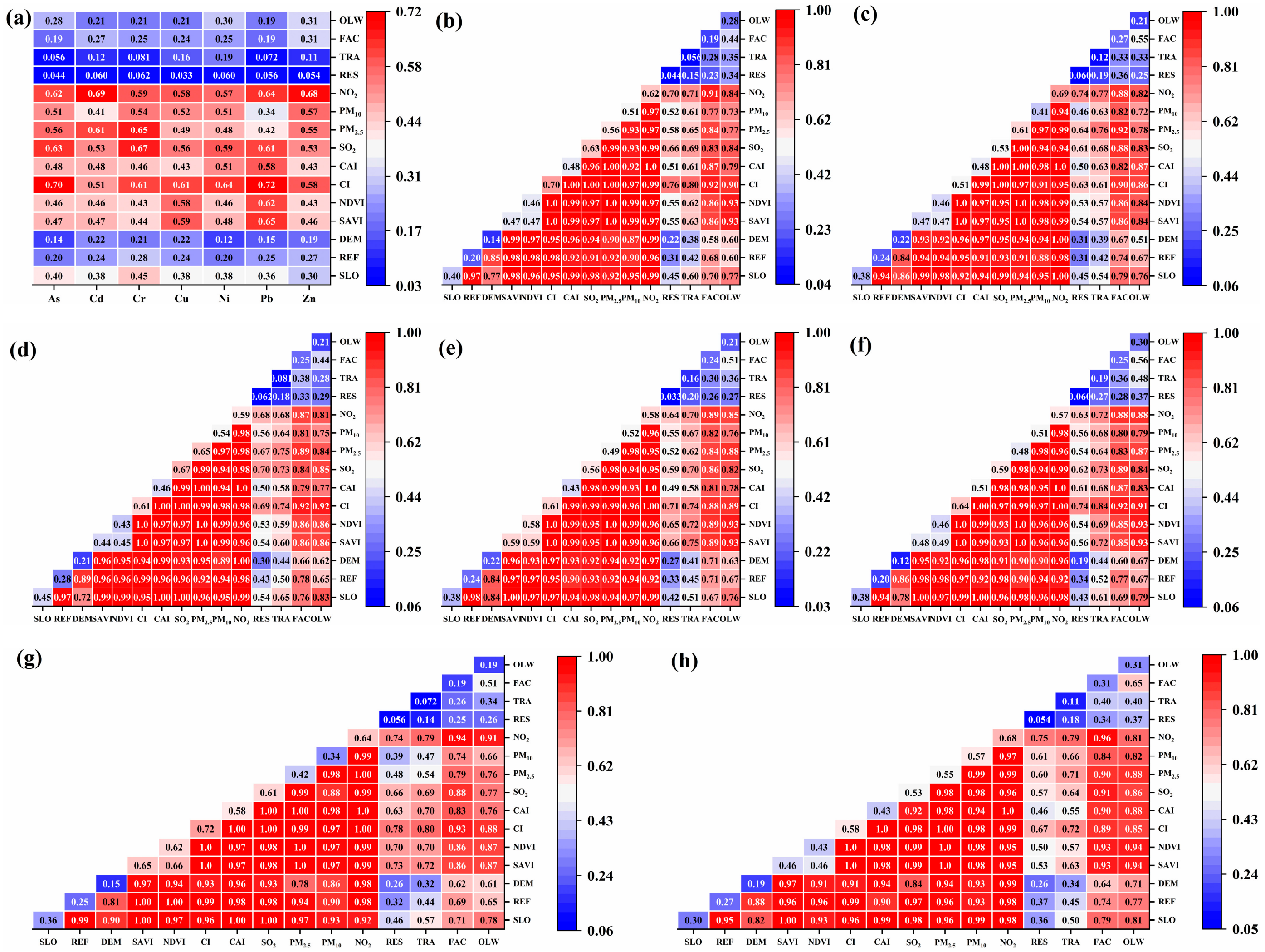

Figure 6. The results of (**a**) single-factor detection analysis and two-factor co-detection for (**b**) As, (**c**) Cd, (**d**) Cr, (**e**) Cu, (**f**) Ni, (**g**) Pb, and (**h**) Zn by geographic detector.

3.2.3. Multi-Source Path Analysis for Soil Heavy Metals

The direct or indirect effects of four sources on heavy metals in soil determined by the SEM model are presented in Figure 8. The loadings of the observed variables (As, Cd, Cr, Cu, Ni, Pb, and Zn) on the endogenous latent variables (i.e., soil heavy metals) were 0.54, 0.53, 0.62, 0.87, 0.86, 0.39, and 0.67, respectively. This difference in loading indicates that there is a homologous or mixed source relationship between soil heavy metals. The direct effect coefficients of exogenous latent variables (i.e., topography, vegetation, air quality, and human activity) on heavy metals were 0.19, 0.15, 0.65, and 0.57, respectively (Figure 8), suggesting that air quality and human activities have a significant direct input effect on the accumulation of heavy metals in farmland soil in the study area.

The largest loading factors of human activity, air quality, topography, and vegetation sources were OLW (0.98), NO_2 (0.92), REF (0.77), and NDVI (0.91), respectively, which are indicative of their significant indirect effects on soil heavy metal accumulation. Air quality had a negative correlation with terrain (−0.35) and vegetation (−0.15), while human activity was positively correlated with terrain (0.09), vegetation (0.12), and air quality (0.14). Furthermore, the highest path effect (0.091) was from human activity to air quality to soil heavy metals, indicating that the coupling of human activity and air quality had a significant indirect effect on the accumulation of soil heavy metals.

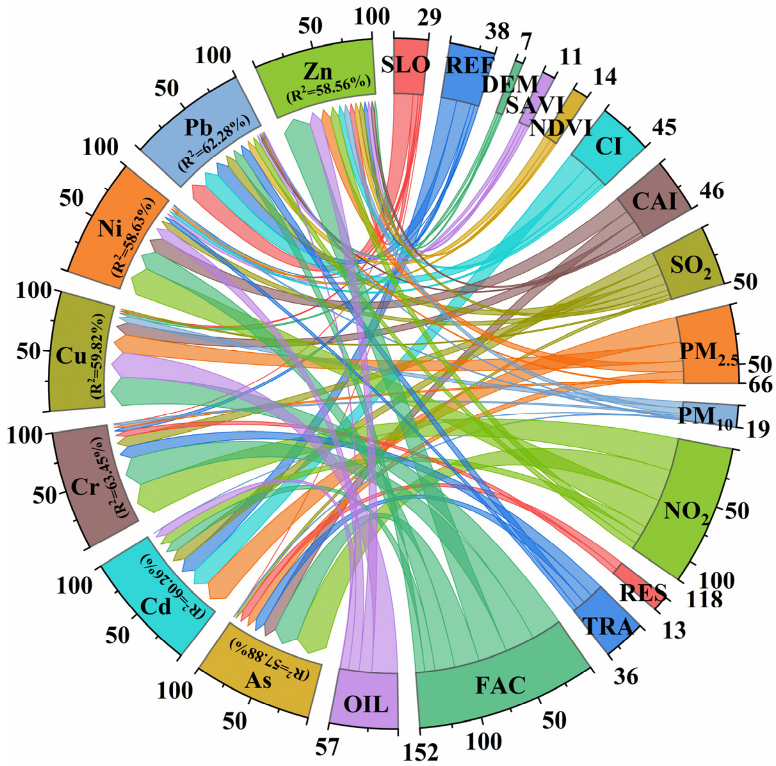

Figure 7. The importance of source factors for soil heavy metals using the XGBoost model.

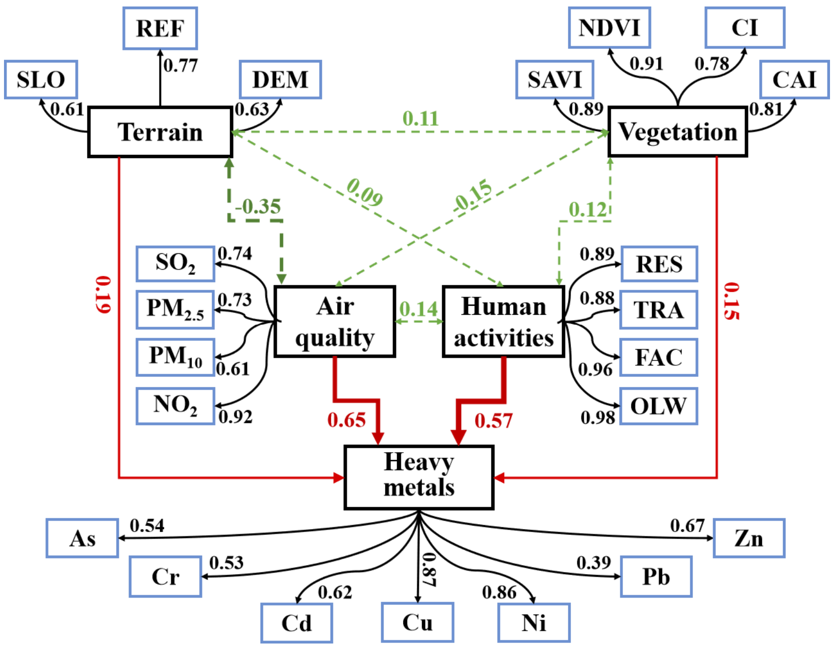

Figure 8. The direct or indirect effects of four sources on soil heavy metals as determined by the SEM model. The thickness of the arrow represents the strength of the relationship (the red solid line indicates relationships between sources and soil heavy metals, and the green dotted line indicates relationships among the four sources). Standardized path coefficients were marked beside the arrows.

4. Discussion

4.1. Response Relationship between Sources and Soil Heavy Metals

The PMF, GD, XGBoost, and SEM results have incomplete consistency, which is caused by the differences in model mechanisms. The PMF is essentially a linear matrix calculation of multiple heavy metals, and it has uncertainty when the potential sources have similar characteristic elements [19]. Moreover, characterizing the potential sources of PMF for each heavy metal is still based on experiential knowledge [56]. This is because the mechanism of PMF suffers from one key shortcoming, namely the assumption that the total content of heavy metals is intrinsically equal to the sum of the contributions from all individual sources [57–59]. There are complementary models among quantitative source apportionment models. Many studies have found that machine learning models (SVM, RF, XGBoost, etc.) are able to accurately reflect the driving effect of a single source on the accumulation of heavy metals in soil [22,23,60,61]. However, they cannot quantify the synergistic importance of multiple factors. Although the GD model was able to analyze the contribution of double-factor synergy to the accumulation of soil heavy metals [62], it cannot obtain the interaction path relationship among source factors. Then the SEM model quantifies the path synergy contribution of multiple source factors for soil heavy metals. Based on the above analysis, although we used four models to analyze the input sources of soil heavy metals, the combination of XGBoost and SEM is the clearest expression of a coherent source apportionment rather than keeping all four models, which can be simpler and more efficient for application in future studies.

The results of quantitative source apportionment of soil heavy metals by GD, XGBoost, and SEM models were similar. The sources that contributed the most to As, Cd, Cr, Cu, Ni, Pb, and Zn were CI and NO_2, NO_2 and SO_2, SO_2 and NO_2, CI and FAC, CI and NO_2, CI and SLO, and NO_2 and FAC, respectively. CI is an effective indicator of soil texture and parent material. Similar studies have shown that the accumulation of As, Cu, Ni, and Pb in soil is largely affected by the natural background [63–65], and its enrichment in soil is also related to the overall topography of the study area and rivers [17]. The study area was located in the coastal delta, with rich vegetation diversity and serious soil salinization [66,67]; that is, the high soil pH environment will inhibit the migration and transformation of heavy metals. Hence, an important reason for the accumulation of As, Cu, Ni, and Pb in the study area may be the dominant input from natural sources.

However, the results of quantitative source apportionment also indicate that NO_2 and SO_2 greatly contribute to soil heavy metals in the study area. On the one hand, NO_2 and SO_2 enter the soil through acid deposition and release heavy metal ions by changing the acid-base environment of the soil, thereby affecting the total content of heavy metals in the soil [68,69]. Some studies have shown that heavy metals are soluble in acidic environments, have higher mobility, and can enter groundwater through media infiltration [70,71]. When the soil pH is alkaline, most of the heavy metals exist in the form of low-solubility salt, and the mobility of heavy metals is low, resulting in the enrichment of heavy metals in the soil [72,73]. On the other hand, some studies have shown that heavy metals in traffic, industrial emissions, and coal combustion are highly correlated with NO_2 and SO_2 [74,75], which indirectly affect the accumulation of heavy metals through precipitation into the soil. As $PM_{2.5}$ enters the soil by atmospheric deposition, the heavy metals and ions it carries will also cause changes to the original pattern of heavy metal accumulation [76–78]. There are some studies showing that particulate matter (PM), SO_2, and NO_2 are highly correlated with heavy metals from industrial emissions [19,26,74,79]. This was identified as one of the key reasons for the input of As, Cu, and Ni by some sources of human activity. Additionally, unreasonable agricultural activities [33] may also play a role in the accumulation and pollution of heavy metals in the soil [80].

However, according to the highest path effect of the SEM model (i.e., human activity –air quality–soil heavy metals (HA–AQ–HMs)), human activities impact the accumulation of soil heavy metals (i.e., As, Cd, Cr, Ni, and Zn) in the study area by promoting air quality pollution. In other words, there is a continuous source of heavy metals deposited in the study area, and we speculate that these sources are oil exploitation and smelting activities. Many studies have found that Cd and Zn are mainly enriched in industrial and mining activities [53,81–83]. Under the influence of the local prevailing wind, dust produced by industrial activities is deposited in the surface soil through atmospheric deposition [84]. The study area was located in China's important petrochemical industry base (i.e., Shengli Oilfield). The heavy metals emitted during oil exploitation and smelting enter the atmosphere and are then input into the soil through precipitation, atmospheric deposition, etc. This cycle path continues to promote the migration and accumulation of heavy metals in farmland soil in the study area.

4.2. Limitations and Implications

In this work, a variety of methods and source variables have been used to analyze the source of heavy metals in farmland soil. However, there are some limitations and implications for future research that still need to be considered. Firstly, the comprehensive and three-dimensional source apportionment of farmland soil should be focused. The influence of soil thickness on heavy metals was considered in soil sampling. Generally speaking, the shallow soil (0–10 cm) of farmland is most affected by human activities. When samples are too deep, the trace element signals are diluted by cleaner material from underneath. Then, more advanced determination methods for soil heavy metal content should be introduced. For example, use a 'pseudo-total' strong acid digestion (e.g., 4 M nitric acid) for trace element determinations. The traditional determination method (through which HF dissolves the silicate minerals) exhibited a large signal from the naturally occurring trace elements that are locked up inside resistant minerals and are not available to the environment. This method gets less of a signal from the anthropogenic components. The usual approach for contaminated land work is 'pseudo-total' or strong acid digestion, which focuses on the material that can be extracted without dissolving all of the silicate minerals. In future work, we could run both HF digestion and pseudo total to see if there is a difference.

Thirdly, more trace elements (i.e., Fe, Me, Ca, Na, etc.) and source variables (i.e., fertilizer, pesticide, total organic carbon, total P, total N, etc.) should be added. Land under farming is subject to a range of direct soil amendments, including fertilizers, manures, and pesticides. In some cases, these are the main sources of soil heavy metals. In particular, Fe and Mn are important soil elements that control the adsorption of other elements, and total organic carbon (TOC) is also important as a controlling phase; total P is an element that provides an indication of phosphate fertilizer addition (which relates to a source of Cd), and both P and total N or nitrate give an idea about the anthropogenic additions through farming. These sources may cause some uncertainties in the source allocation and quantification of soil heavy metals. However, quantitative fertilization and pesticide data are difficult to obtain. For example, in the statistical yearbook of Dongying City in 2020 (Supplementary file) (http://dystjj.dongying.gov.cn/index.html, (accessed on 15 October 2021)), the amount of fertilizer applied in the whole region is only one value, and it is difficult to distinguish the difference in the amount of fertilizer applied in each sample point. Moreover, heavy metals produced by agricultural activities such as fertilizers or pesticides usually require accurate pot or field experiments (Table S3) [85–89], which is a challenge for data-hungry nonlinear model source apportionment of regional-scale soil heavy metals. Finally, more scenario simulations and model sensitivity analyses should be explored. For example, run some model sensitivity analyses by adding and removing some of the key source and receptor components (i.e., with or without Fe and Me) to test what overall effect their presence or absence has on the outcome. More interesting results may be obtained in this case if the analysis had included an anthropogenic farming source (fertilizer, manure, and agrichemical additions) or if the trace element suite had been extended to include major elements and controlling trace elements.

5. Conclusions

In this study, the PMF, GD, XGBoost, and SEM models were integrated to analyze the source of heavy metals in the farmland soil of a coastal delta. The PMF results indicate that heavy metals have the same source, and the same heavy metals also have a mixed source. Upon further combining these with GD, XGBoost, and SEM, it was found that air quality (SO_2 and NO_2) and human activity (FAC) have the highest impact on Cd, Cr, and Zn. Natural sources (SLO) were the main input sources of Pb, while As, Cu, and Ni were elevated by a combination of vegetation (CI and CAI) and human activity (FAC). The results show that a balance of evidence approach can provide a more reasonable and comprehensive assessment of the linear-nonlinear relationship between sources and soil heavy metals. The accumulation of heavy metal content in soil is indirectly affected by the coupling of multiple sources. In other words, the heavy metals discharged from long-term oil exploitation and smelting in the study area enter the atmosphere and are then input into the soil through precipitation, atmospheric deposition, and other ways. It is essential to strengthen emissions control for the sources (i.e., mining and smelting) of heavy metals so the migration and accumulation pathways of heavy metals in farmland soils in the study area can be effectively truncated. Different models of source apportionment have various response mechanisms for soil heavy metals, and the accumulation of heavy metals in soil also has time properties. It is necessary to combine multiple source apportionment models and more source factors with fine time series in future studies and to analyze the accumulation and migration characteristics of time and space for soil heavy metals so as to support the high-quality development of agriculture and early warning systems for soil ecological health in coastal delta regions.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jmse11051069/s1. Figure S1. 40 air quality monitoring stations in the Dongying city. Table S1. The results for the GSS-1 standard soil sample and a blank sample. Table S2. Vegetation indices derived from Landsat 8 OLI image. Table S3. Studies on farming practices for soil heavy metals.

Author Contributions: Conceptualization, Y.S.; methodology, W.H.; software, Y.Z.; validation, L.W. and H.Y.; formal analysis, S.W.; investigation, W.H.; resources, Y.S.; data curation, W.H. and Y.S.; writing—original draft preparation, Y.S.; writing—review and editing, W.H. and Y.S.; visualization, Y.X.; supervision, Y.S.; project administration, Y.S.; funding acquisition, Y.S. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shandong Provincial Natural Science Foundation (ZR2020QD013), and the National Natural Science Foundation of China (U1901601).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Soil Environmental Quality and Risk Control Standard for Soil Contamination of Agricultural Land (GB 15618-2018) comes from the Ministry of Ecology and Environment of the People's Republic of China (https://www.mee.gov.cn/, (accessed on 15 October 2019)). The Technical Specification for Soil Environmental Monitoring (HJ/T 166-2004) comes from the Ministry of Ecology and Environment of the People's Republic of China (https://www.mee.gov.cn/, (accessed on 15 October 2019)). Soil and Sediment-Determination of Aqua Regia Extracts of 12 Metal Elements-Inductively Coupled Plasma Mass Spectrometry (HJ 803-2016) and Soil and Sediment-Determination of Total Metal Elements-Microwave Assisted Acid Digestion Method (HJ 832-2017) come from the Ministry of Ecology and Environment of the People's Republic of China (https://www.mee.gov.cn/, (accessed on 15 October 2019)). The DEM data was obtained from the shuttle radar topography mission (SRTM) (https://earthexplorer.usgs.gov/, (accessed on 10 January 2020)). The air quality data comes from the Dongying air quality real-time publishing system (http://218.58.213.53:8081/dyfb_air/fb_web, (accessed on 10 January 2021)). The point of interest (POI) data was obtained from the Baidu map (https://map.baidu.com, (accessed on 10 May 2021)). The Chinese GF-1 remote sensing image comes from the natural resource satellite remote sensing cloud service platform (http://www.sasclouds.com/chinese/normal/, (accessed on 10 January 2021)). The positive matrix factorization (PMF) model comes from the United States Environmental Protection Agency (EPA) (https://www.epa.gov/, (accessed on 10 January 2021)). The statistical yearbook of Dongying City in 2020 comes from Dongying City Bureau of Statistics (http://dystjj.dongying.gov.cn/index.html, (accessed on 15 October 2021)).

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

References

1. Marchant, B.P.; Saby, N.; Arrouays, D. A survey of topsoil arsenic and mercury concentrations across France. *Chemosphere* **2017**, *181*, 635–644. [CrossRef] [PubMed]
2. Zang, F.; Wang, S.; Nan, Z.; Ma, J.; Zhang, Q.; Chen, Y.; Li, Y. Accumulation, spatio-temporal distribution, and risk assessment of heavy metals in the soil-corn system around a polymetallic mining area from the Loess Plateau, northwest China. *Geoderma* **2017**, *305*, 188–196. [CrossRef]
3. Imseng, M.; Wiggenhauser, M.; Müller, M.; Keller, A.; Frossard, E.; Wilcke, W.; Bigalke, M. The fate of Zn in agricultural soils: A stable isotope approach to anthropogenic impact, soil formation, and soil–plant cycling. *Environ. Sci. Technol.* **2019**, *53*, 4140–4149. [CrossRef] [PubMed]
4. Hu, W.; Huang, B.; He, Y.; Kalkhajeh, Y.K. Assessment of potential health risk of heavy metals in soils from a rapidly developing region of China. *Hum. Ecol. Risk Assess. Int. J.* **2016**, *22*, 211–225. [CrossRef]
5. Wuana, R.A.; Okieimen, F.E. Heavy metals in contaminated soils: A review of sources, chemistry, risks and best available strategies for remediation. *Int. Sch. Res. Not.* **2011**, *2011*, 402647. [CrossRef]
6. Barkett, M.O.; Akün, E. Heavy metal contents of contaminated soils and ecological risk assessment in abandoned copper mine harbor in Yedidalga, Northern Cyprus. *Environ. Earth Sci.* **2018**, *77*, 378. [CrossRef]
7. Kaur, M.; Kumar, A.; Mehra, R.; Mishra, R. Human health risk assessment from exposure of heavy metals in soil samples of Jammu district of Jammu and Kashmir, India. *Arab. J. Geosci.* **2018**, *11*, 411. [CrossRef]
8. Hu, N.; Liu, J.; Huang, P.; Yan, S.; Shi, X.; Ma, D. Sources, geochemical speciation, and risk assessment of metals in coastal sediments: A case study in the Bohai Sea, China. *Environ. Earth Sci.* **2017**, *76*, 309. [CrossRef]
9. Tian, K.; Huang, B.; Xing, Z.; Hu, W. Geochemical baseline establishment and ecological risk evaluation of heavy metals in greenhouse soils from Dongtai, China. *Ecol. Indic.* **2017**, *72*, 510–520. [CrossRef]
10. Zhang, Z.; Wang, J.J.; Ali, A.; DeLaune, R.D. Heavy metals and metalloid contamination in Louisiana Lake Pontchartrain Estuary along I-10 Bridge. *Transp. Res. Part D Transp. Environ.* **2016**, *44*, 66–77. [CrossRef]
11. Liu, R.; Men, C.; Liu, Y.; Yu, W.; Xu, F.; Shen, Z. Spatial distribution and pollution evaluation of heavy metals in Yangtze estuary sediment. *Mar. Pollut. Bull.* **2016**, *110*, 564–571. [CrossRef] [PubMed]
12. Zhou, J.; Feng, K.; Li, Y.; Zhou, Y. Factorial Kriging analysis and sources of heavy metals in soils of different land-use types in the Yangtze River Delta of Eastern China. *Environ. Sci. Pollut. R.* **2016**, *23*, 14957–14967. [CrossRef]
13. Zhang, X.; Wei, S.; Sun, Q.; Wadood, S.A.; Guo, B. Source identification and spatial distribution of arsenic and heavy metals in agricultural soil around Hunan industrial estate by positive matrix factorization model, principle components analysis and geo statistical analysis. *Ecotox. Environ. Safe.* **2018**, *159*, 354–362. [CrossRef] [PubMed]
14. Shit, P.K.; Bhunia, G.S.; Maiti, R. Spatial analysis of soil properties using GIS based geostatistics models. *Model. Earth Syst. Environ.* **2016**, *2*, 107. [CrossRef]
15. Chen, R.; Zhang, Q.; Chen, H.; Yue, W.; Teng, Y. Source apportionment of heavy metals in sediments and soils in an interconnected river-soil system based on a composite fingerprint screening approach. *J. Hazard. Mater.* **2021**, *411*, 125125. [CrossRef]
16. Wang, P.; Li, Z.; Liu, J.; Bi, X.; Ning, Y.; Yang, S.; Yang, X. Apportionment of sources of heavy metals to agricultural soils using isotope fingerprints and multivariate statistical analyses. *Environ. Pollut.* **2019**, *249*, 208–216. [CrossRef] [PubMed]
17. Chai, L.; Wang, Y.; Wang, X.; Ma, L.; Cheng, Z.; Su, L. Pollution characteristics, spatial distributions, and source apportionment of heavy metals in cultivated soil in Lanzhou, China. *Ecol. Indic.* **2021**, *125*, 107507. [CrossRef]
18. Qu, M.; Chen, J.; Huang, B.; Zhao, Y. Enhancing apportionment of the point and diffuse sources of soil heavy metals using robust geostatistics and robust spatial receptor model with categorical soil-type data. *Environ. Pollut.* **2020**, *265*, 114964. [CrossRef]
19. Wang, L.; Gao, S.; Yin, X.; Yu, X.; Luan, L. Arsenic accumulation, distribution and source analysis of rice in a typical growing area in north China. *Ecotox. Environ. Safe.* **2019**, *167*, 429–434. [CrossRef]
20. Huang, Y.; Deng, M.; Wu, S.; Japenga, J.; Li, T.; Yang, X.; He, Z. A modified receptor model for source apportionment of heavy metal pollution in soil. *J. Hazard. Mater.* **2018**, *354*, 161–169. [CrossRef]
21. Jia, X.; Hu, B.; Marchant, B.P.; Zhou, L.; Shi, Z.; Zhu, Y. A methodological framework for identifying potential sources of soil heavy metal pollution based on machine learning: A case study in the Yangtze Delta, China. *Environ. Pollut.* **2019**, *250*, 601–609. [CrossRef] [PubMed]
22. Zhang, B.; Liu, L.; Huang, Z.; Hou, H.; Zhao, L.; Sun, Z. Application of stochastic model to assessment of heavy metal (loid) s source apportionment and bio-availability in rice fields of karst area. *Sci. Total Environ.* **2021**, *793*, 148614. [CrossRef] [PubMed]
23. Zhang, H.; Yin, A.; Yang, X.; Fan, M.; Shao, S.; Wu, J.; Wu, P.; Zhang, M.; Gao, C. Use of machine-learning and receptor models for prediction and source apportionment of heavy metals in coastal reclaimed soils. *Ecol. Indic.* **2021**, *122*, 107233. [CrossRef]
24. Wang, Q.; Xie, Z.; Li, F. Using ensemble models to identify and apportion heavy metal pollution sources in agricultural soils on a local scale. *Environ. Pollut.* **2015**, *206*, 227–235. [CrossRef]
25. Shi, T.; Ma, J.; Zhang, Y.; Liu, C.; Hu, Y.; Gong, Y.; Wu, X.; Ju, T.; Hou, H.; Zhao, L. Status of lead accumulation in agricultural soils across China (1979–2016). *Environ. Int.* **2019**, *129*, 35–41. [CrossRef]
26. Fei, X.; Lou, Z.; Xiao, R.; Ren, Z.; Lv, X. Contamination assessment and source apportionment of heavy metals in agricultural soil through the synthesis of PMF and GeogDetector models. *Sci. Total Environ.* **2020**, *747*, 141293. [CrossRef]
27. Cao, J.; Li, C.; Zhang, L.; Wu, Q.; Lv, J. Source apportionment of potentially toxic elements in soils using APCS/MLR, PMF and geostatistics in a typical industrial and mining city in Eastern China. *PLoS ONE* **2020**, *15*, e0238513.

28. Hengl, T. *A Practical Guide to Geostatistical Mapping*, 2nd ed.; University of Amsterdam: Amsterdam, The Netherlands, 2009.
29. Boettinger, J.L.; Ramsey, R.D.; Bodily, J.M.; Cole, N.J.; Stum, A.K. Landsat spectral data for digital soil mapping. In *Digital Soil Mapping with Limited Data*; Hartemink, A.E., McBratney, A.B., Mendonca Santos, M.L., Eds.; Springer Science: Melbourne, VIC, Australia, 2008; pp. 193–203.
30. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]
31. Rouse, J.W.; Hass, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In Proceedings of the Third ERTS Symposium, Washington, DC, USA, 10–14 December 1973; NASA SP-351. National Aeronautics and Space Administration: Washington, DC, USA, 1973; pp. 309–371.
32. Hassoun, M.H. *Fundamentals of Artificial Neural Networks*; MIT Press: Cambridge, MA, USA, 1995.
33. Xiao, R.; Guo, D.; Ali, A.; Mi, S.; Liu, T.; Ren, C.; Li, R.; Zhang, Z. Accumulation, ecological-health risks assessment, and source apportionment of heavy metals in paddy soils: A case study in Hanzhong, Shaanxi, China. *Environ. Pollut.* **2019**, *248*, 349–357. [CrossRef]
34. Guan, Q.; Zhao, R.; Pan, N.; Wang, F.; Yang, Y.; Luo, H. Source apportionment of heavy metals in farmland soil of Wuwei, China: Comparison of three receptor models. *J. Clean. Prod.* **2019**, *237*, 117792. [CrossRef]
35. EPA. *Positive Matrix Factorization (PMF) 5.0-Fundamentals and User Guide*; US Environmental Protection Agency: Washington, DC, USA, 2014.
36. Wang, J.F.; Xu, C.D. Geodetector: Principle and prospective. *Acta Geogr. Sin.* **2017**, *72*, 116–134.
37. Wang, J.F.; Li, X.H.; Christakos, G.; Liao, Y.L.; Zhang, T.; Gu, X.; Zheng, X.Y. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 107–127. [CrossRef]
38. Wang, J.F.; Hu, Y. Environmental health risk detection with GeogDetector. *Environ. Model. Softw.* **2012**, *33*, 114–115. [CrossRef]
39. Fei, X.; Lou, Z.; Christakos, G.; Ren, Z.; Liu, Q.; Lv, X. The association between heavy metal soil pollution and stomach cancer: A case study in Hangzhou City, China. *Environ. Geochem. Health* **2018**, *40*, 2481–2490. [CrossRef]
40. Ma, M.; Zhao, G.; He, B.; Li, Q.; Dong, H.; Wang, S.; Wang, Z. XGBoost-based method for flash flood risk assessment. *J. Hydrol.* **2021**, *598*, 126382. [CrossRef]
41. Parsa, M. A data augmentation approach to XGboost-based mineral potential mapping: An example of carbonate-hosted Zn Pb mineral systems of Western Iran. *J. Geochem. Explor.* **2021**, *228*, 106811. [CrossRef]
42. Yang, X.; Yang, R.; Ye, Y.; Yuan, Z.; Wang, D.; Hua, K. Winter wheat SPAD estimation from UAV hyperspectral data using cluster-regression methods. *Int. J. Appl. Earth Obs.* **2021**, *105*, 102618. [CrossRef]
43. Alabdullah, A.A.; Iqbal, M.; Zahid, M.; Khan, K.; Amin, M.N.; Jalal, F.E. Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *Constr. Build. Mater.* **2022**, *345*, 128296. [CrossRef]
44. Yan, H.; Yan, K.; Ji, G. Optimization and prediction in the early design stage of office buildings using genetic and XGBoost algorithms. *Build Environ.* **2022**, *218*, 109081. [CrossRef]
45. Fan, J.; Zheng, J.; Wu, L.; Zhang, F. Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models. *Agr. Water Manag.* **2021**, *245*, 106547. [CrossRef]
46. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* **2001**, *106*, 7183–7192. [CrossRef]
47. Yang, R.; Guo, W.; Zheng, J. Soil prediction for coastal wetlands following Spartina alterniflora invasion using Sentinel-1 imagery and structural equation modeling. *Catena* **2019**, *173*, 465–470. [CrossRef]
48. Zhang, Y.; Hou, D.; Xiong, G.; Duan, Y.; Cai, C.; Wang, X.; Li, J.; Tao, S.; Liu, W. Structural equation modeling of PAHs in ambient air, dust fall, soil, and cabbage in vegetable bases of Northern China. *Environ. Pollut.* **2018**, *239*, 13–20. [CrossRef] [PubMed]
49. Zhao, W.; Zhang, R.; Cao, H.; Tan, W. Factor contribution to soil organic and inorganic carbon accumulation in the Loess Plateau: Structural equation modeling. *Geoderma* **2019**, *352*, 116–125. [CrossRef]
50. Wang, Y.; Liu, Z.; Liu, X.; Song, X.; Cui, D. Distribution and potential ecological risk assessment of heavy metals in the topsoil of the Yellow River Delta. *J. Soil Water Conserv.* **2019**, *33*, 8.
51. Yuanan, H.; He, K.; Sun, Z.; Chen, G.; Cheng, H. Quantitative source apportionment of heavy metal (loid) s in the agricultural soils of an industrializing region and associated model uncertainty. *J. Hazard. Mater.* **2020**, *391*, 122244. [CrossRef]
52. Wang, Y.; Guo, G.; Zhang, D.; Lei, M. An integrated method for source apportionment of heavy metal (loid) s in agricultural soils and model uncertainty analysis. *Environ. Pollut.* **2021**, *276*, 116666. [CrossRef]
53. Chai, L.; Wang, Y.; Wang, X.; Ma, L.; Cheng, Z.; Su, L.; Liu, M. Quantitative source apportionment of heavy metals in cultivated soil and associated model uncertainty. *Ecotox. Environ. Saf.* **2021**, *215*, 112150. [CrossRef]
54. Lopez, A.M.; Fitzsimmons, J.N.; Adams, H.M.; Dellapenna, T.M.; Brandon, A.D. A time-series of heavy metal geochemistry in sediments of Galveston Bay estuary, Texas, 2017–2019. *Sci. Total Environ.* **2022**, *806*, 150446. [CrossRef]
55. Burak, D.L.; Fontes, M.P.; Santos, N.T.; Monteiro, L.V.S.; de Sousa Martins, E.; Becquer, T. Geochemistry and spatial distribution of heavy metals in Oxisols in a mineralized region of the Brazilian Central Plateau. *Geoderma* **2010**, *160*, 131–142. [CrossRef]
56. Guan, Q.; Wang, F.; Xu, C.; Pan, N.; Lin, J.; Zhao, R.; Yang, Y.; Luo, H. Source apportionment of heavy metals in agricultural soil based on PMF: A case study in Hexi Corridor, northwest China. *Chemosphere* **2018**, *193*, 189–197. [CrossRef]

57. Wu, J.; Li, J.; Teng, Y.; Chen, H.; Wang, Y. A partition computing-based positive matrix factorization (PC-PMF) approach for the source apportionment of agricultural soil heavy metal contents and associated health risks. *J. Hazard. Mater.* **2020**, *388*, 121766. [CrossRef] [PubMed]
58. Adgate, J.L.; Willis, R.D.; Buckley, T.J.; Chow, J.C.; Watson, J.G.; Rhoads, G.G.; Lioy, P.J. Chemical mass balance source apportionment of lead in house dust. *Environ. Sci. Technol.* **1998**, *32*, 108–114. [CrossRef]
59. Paatero, P.; Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **1994**, *5*, 111–126. [CrossRef]
60. Bhagat, S.K.; Tung, T.M.; Yaseen, Z.M. Heavy metal contamination prediction using ensemble model: Case study of Bay sedimentation, Australia. *J. Hazard. Mater.* **2021**, *403*, 123492. [CrossRef]
61. Li, Y.; Chen, H.; Teng, Y. Source apportionment and source-oriented risk assessment of heavy metals in the sediments of an urban river-lake system. *Sci. Total Environ.* **2020**, *737*, 140310. [CrossRef] [PubMed]
62. Zhao, Y.; Deng, Q.; Lin, Q.; Zeng, C.; Zhong, C. Cadmium source identification in soils and high-risk regions predicted by geographical detector method. *Environ. Pollut.* **2020**, *263*, 114338. [CrossRef]
63. Liu, Q.; Jia, Z.; Li, S.; Hu, J. Assessment of heavy metal pollution, distribution and quantitative source apportionment in surface sediments along a partially mixed estuary (Modaomen, China). *Chemosphere* **2019**, *225*, 829–838. [CrossRef]
64. Fei, X.; Xiao, R.; Christakos, G.; Langousis, A.; Ren, Z.; Tian, Y.; Lv, X. Comprehensive assessment and source apportionment of heavy metals in Shanghai agricultural soils with different fertility levels. *Ecol. Indic.* **2019**, *106*, 105508. [CrossRef]
65. Luo, H.; Wang, Q.; Guan, Q.; Ma, Y.; Ni, F.; Yang, E.; Zhang, J. Heavy metal pollution levels, source apportionment and risk assessment in dust storms in key cities in Northwest China. *J. Hazard. Mater.* **2022**, *422*, 126878. [CrossRef]
66. Liu, L.; Wu, Y.; Yin, M.; Ma, X.; Yu, X.; Guo, X.; Du, N.; Eller, F.; Guo, W. Soil salinity, not plant genotype or geographical distance, shapes soil microbial community of a reed wetland at a fine scale in the Yellow River Delta. *Sci. Total Environ.* **2023**, *856*, 159136. [CrossRef] [PubMed]
67. Li, Y.; Chang, C.; Wang, Z.; Zhao, G. Upscaling remote sensing inversion and dynamic monitoring of soil salinization in the Yellow River Delta, China. *Ecol. Indic.* **2023**, *148*, 110087. [CrossRef]
68. Yamulki, S.; Harrison, R.M.; Goulding, K.W.T.; Webster, C.P. N_2O, NO and NO_2 fluxes from a grassland: Effect of soil pH. *Soil Biol. Bioche.* **1997**, *29*, 1199–1208. [CrossRef]
69. Yang, H.I.; Park, H.J.; Lee, K.S.; Lim, S.S.; Kwak, J.H.; Lee, S.I.; Chang, S.; Lee, S.; Choi, W.J. δ13C, δ15N, N concentration, C/N, and Ca/Al of *Pinus densiflora* foliage in Korean cities of different precipitation pH and atmospheric NO_2 and SO_2 levels. *Ecol. Indicat.* **2018**, *88*, 27–36. [CrossRef]
70. Walker, D.J.; Clemente, R.; Bernal, M.P. Contrasting effects of manure and compost on soil pH, heavy metal availability and growth of *Chenopodium album* L. in a soil contaminated by pyritic mine waste. *Chemosphere* **2004**, *57*, 215–224. [CrossRef]
71. Kumar, R.N.; Nagendran, R. Influence of initial pH on bioleaching of heavy metals from contaminated soil employing indigenous Acidithiobacillus thiooxidans. *Chemosphere* **2007**, *66*, 1775–1781. [CrossRef]
72. Sun, H.; Song, Y.; Liu, W.; Zhang, M.; Duan, T.; Cai, Y. Coupling soil washing with chelator and cathodic reduction treatment for a multi-metal contaminated soil: Effect of pH controlling. *Electrochim. Acta* **2023**, *448*, 142178. [CrossRef]
73. Martínez-Cortijo, J.; Ruiz-Canales, A. Effect of heavy metals on rice irrigated fields with waste water in high pH Mediterranean soils: The particular case of the Valencia area in Spain. *Agr. Water Manag.* **2018**, *210*, 108–123. [CrossRef]
74. Zhang, G.; Bai, J.; Zhao, Q.; Lu, Q.; Jia, J.; Wen, X. Heavy metals in wetland soils along a wetland-forming chronosequence in the Yellow River Delta of China: Levels, sources and toxic risks. *Ecol. Indic.* **2016**, *69*, 331–339. [CrossRef]
75. Mao, H.T.; Wang, X.M.; Wu, N.; Chen, L.X.; Yuan, M.; Hu, J.C.; Chen, Y.E. Temporal and spatial biomonitoring of atmospheric heavy metal pollution using moss bags in Xichang. *Ecotoxicol. Environ. Saf.* **2022**, *239*, 113688. [CrossRef]
76. Vitali, M.; Antonucci, A.; Owczarek, M.; Guidotti, M.; Astolfi, M.L.; Manigrasso, M.; Avino, P.; Bhattacharya, B.; Protano, C. Air quality assessment in different environmental scenarios by the determination of typical heavy metals and Persistent Organic Pollutants in native lichen Xanthoria parietina. *Environ. Pollut.* **2019**, *254*, 113013. [CrossRef] [PubMed]
77. Su, T.; Lin, C.; Lu, S.; Lin, J.; Wang, H.; Liu, C. Effect of air quality improvement by urban parks on mitigating PM2. 5 and its associated heavy metals: A mobile-monitoring field study. *J. Environ. Manag.* **2022**, *323*, 116283. [CrossRef] [PubMed]
78. Li, F.; Yan, J.; Wei, Y.; Zeng, J.; Wang, X.; Chen, X.; Zhang, C.; Li, W.; Chen, M.; Lü, G. PM2. 5-bound heavy metals from the major cities in China: Spatiotemporal distribution, fuzzy exposure assessment and health risk management. *J. Clean. Prod.* **2021**, *286*, 124967. [CrossRef]
79. Yin, X.; Franklin, M.; Fallah-Shorshani, M.; Shafer, M.; McConnell, R.; Fruin, S. Exposure models for particulate matter elemental concentrations in Southern California. *Environ. Int.* **2022**, *165*, 107247. [CrossRef]
80. Wang, F.; Guan, Q.; Tian, J.; Lin, J.; Yang, Y.; Yang, L.; Pan, N. Contamination characteristics, source apportionment, and health risk assessment of heavy metals in agricultural soil in the Hexi Corridor. *Catena* **2020**, *191*, 104573. [CrossRef]
81. Liu, H.; Zhang, Y.; Yang, J.; Wang, H.; Li, Y.; Shi, Y.; Li, D.; Holm, P.E.; Ou, Q.; Hu, W. Quantitative source apportionment, risk assessment and distribution of heavy metals in agricultural soils from southern Shandong Peninsula of China. *Sci. Total Environ.* **2021**, *767*, 144879. [CrossRef]
82. Zhao, K.; Zhang, L.; Dong, J.; Wu, J.; Ye, Z.; Zhao, W.; Ding, L.; Fu, W. Risk assessment, spatial patterns and source apportionment of soil heavy metals in a typical Chinese hickory plantation region of southeastern China. *Geoderma* **2020**, *360*, 114011. [CrossRef]

83. Wang, J.; Su, J.; Li, Z.; Liu, B.; Cheng, G.; Jiang, Y.; Li, Y.; Zhou, S.; Yuan, W. Source apportionment of heavy metal and their health risks in soil-dustfall-plant system nearby a typical non-ferrous metal mining area of Tongling, Eastern China. *Environ. Pollut.* **2019**, *254*, 113089. [CrossRef]
84. Vu, C.T.; Lin, C.; Shern, C.; Yeh, G.; Le, V.G.; Tran, H.T. Contamination, ecological risk and source apportionment of heavy metals in sediments and water of a contaminated river in Taiwan. *Ecol. Indic.* **2017**, *82*, 32–42. [CrossRef]
85. Zeyad, M.T.; Khan, S.; Malik, A. Genotoxic hazard and oxidative stress induced by wastewater irrigated soil with special reference to pesticides and heavy metal pollution. *Heliyon* **2022**, *8*, e10534. [CrossRef]
86. Wang, B.; Chu, C.; Wei, H.; Zhang, L.; Ahmad, Z.; Wu, S.; Xie, B. Ameliorative effects of silicon fertilizer on soil bacterial community and pakchoi (*Brassica chinensis* L.) grown on soil contaminated with multiple heavy metals. *Environ. Pollut.* **2020**, *267*, 115411. [CrossRef] [PubMed]
87. Salem, M.A.; Bedade, D.K.; Al-Ethawi, L.; Al-Waleed, S.M. Assessment of physiochemical properties and concentration of heavy metals in agricultural soils fertilized with chemical fertilizers. *Heliyon* **2020**, *6*, e05224. [CrossRef] [PubMed]
88. Hong, Y.; Li, D.; Xie, C.; Zheng, X.; Yin, J.; Li, Z.; Zhang, K.; Jiao, Y.; Wang, B.; Hu, Y.; et al. Combined apatite, biochar, and organic fertilizer application for heavy metal co-contaminated soil remediation reduces heavy metal transport and alters soil microbial community structure. *Sci. Total Environ.* **2022**, *851*, 158033. [CrossRef] [PubMed]
89. Soleimani, H.; Mansouri, B.; Kiani, A.; Omer, A.K.; Tazik, M.; Ebrahimzadeh, G.; Sharafi, K. Ecological risk assessment and heavy metals accumulation in agriculture soils irrigated with treated wastewater effluent, river water, and well water combined with chemical fertilizers. *Heliyon* **2023**, *9*, e14580. [CrossRef]