Beyond mere pollution source identification: Determination of land covers emitting soil heavy metals by combining PCA/APCS, GeoDetector and GIS analysis

Yong Yang\textsuperscript{a}, Xue Yang\textsuperscript{a}, Mingjun He\textsuperscript{b}, George Christakos\textsuperscript{b,c,⁎}

\textsuperscript{a} College of Resources & Environment, Huazhong Agricultural University, Wuhan, China
\textsuperscript{b} Ocean College, Zhejiang University, Zhoushan, China
\textsuperscript{c} Geography Department, San Diego State University, CA, USA

\textbf{A R T I C L E   I N F O}

Keywords:
Soil heavy metal
Source apportionment
Land cover
PCA/APCS
GeoDetector

\textbf{A B S T R A C T}

In this work, we propose a method that is not limited in the identification of the type of pollution source but it also suggests the land covers that emit heavy metals into the surrounding soils by introducing a three-stage procedure, as follows: (a) the Principal Component Analysis/Absolute Principal Component Scores technique is applied to the spatial distribution of soil heavy metal accumulations to identify the type of source that is responsible for soil heavy metal accumulation, (b) based on the spatial distribution of the principal component scores and on four selected driving factors (land cover, distance to mine or smelter, distance to road, and topographic elevation), the Geographical Detector model was used to identify the effect intensity of the driving factors on soil heavy metal accumulation and obtain the landscape type of pollutant sources, and (c) GIS analysis (buffer and overlap analysis) was performed on the principal component scores around the suspected land covers linked to the landscape type of pollutant sources to determine the land covers that, in fact, emit heavy metals into the surrounding soils. Based on the proposed approach, four mining and metallurgy land or land groups were determined to be the actual sources of soil heavy metal pollution in Daye city, Hubei Province, China. Lastly, a Multiple Linear Regression model with decay function was proposed to quantify the contributions of previously identified pollution sources to soil heavy metal accumulation. It was found that the HuangJin mountain quarry, the Tonglu mountain cooper mine (together with some related mineral processing and smelting enterprises), the Lion mountain mining and mineral processing base, and the large Oujia mountain mine are the four sources that contributed 3.2%, 34.3%, 8.3%, and 44% of the total soil heavy metal accumulations in the study area.

1. Introduction

During the past several decades, soil heavy metal pollution has been a significant environmental issue because of its toxicity and difficult degradation. Thus, an increasing number of studies have focused on the investigation, spatial distribution, risk assessment and source apportionment of regional soil heavy metals (Chen et al., 2005; Guo et al., 2012; Xu et al., 2017; Kong et al., 2018). Among them, the spatial distributions generated by interpolation methods (e.g., Kriging techniques, Olea, 1999) have revealed the spatial characteristics or patterns of heavy metal distributions in soils, and the results consistently showed that the high soil heavy metal concentrations are linked to the surrounding pollution landscape.

Moreover, the source apportionment methods (e.g., the diffusion and receptor models) can identify the main pollution sources of soil heavy metals, such as agricultural sources, and industrial and traffic emissions (Pathak et al., 2015; Liang et al., 2017; Deng et al., 2018). Furthermore, some quantitative source apportionment methods, such as principal component analysis with multiple linear regression (PCA-MLR) and positive matrix factorization (PMF) can quantify the contribution (in %) of the possible sources of soil heavy metals concentration (Sofowote et al., 2008; Cao et al., 2011). Many of the results of soil heavy metal source apportionment have shown that the identified pollution sources are associated with anthropogenic activities or pollution emission landscapes, including roads, factories, and farmlands (Kheir et al., 2014; Zhang et al., 2009; Qiu et al., 2015). For example, according to the results of Men et al (2018), traffic-related exhaust, coal combustion, metal manufacture and the use of pesticides with fertilizers...
accounted for 34.47%, 25%, 25%, and 14.88%, respectively, of the concentration of heavy metals in the road dust of Beijing city, China. Our results of long-term monitoring of heavy metals in the soils of the Wuhan industrial zone indicated that 82.5% of soil heavy metal concentration increment during 2010–2014 was attributed to industrial/agricultural activities sources based on the PCA-MLR method (see, Yang et al., 2017). However, the above quantitative source apportionment methods used in the existing studies can only point out the types of pollution sources, but cannot identify the actual landscapes related to pollutant sources. Also, the existing methods could not quantify the contribution of different pollution source landscapes to the accumulation of heavy metals in soil. In other words, it is unknown whether a specific land cover, belonging to a type of pollution source identified by source apportionment methods, has really emitted heavy metals into the surrounding soils. It is also unknown which land cover can emit more heavy metals into the surrounding soils than others.

In view of the above considerations, the aim of this paper is to improve the identification of the land covers responsible for heavy metals emissions based on the results of quantitative source apportionment of soil heavy metals accumulations, and to compare the contributions of these land covers on heavy metals emissions. Insight will be gained by means of a real-world case study concerned with soil heavy metals in Daye city soils (Hubei Province, China). It is anticipated that the methods introduced in this work could also improve the subsequent implementation of effective control and management measures of soil heavy metal emissions.

2. Materials and methods

2.1. Study area and dataset

Daye city (latitude 29°40′—30°15′ N, longitude 114°31′—115°20′ E), with rich mineral resources and long history of metal mining activities, is located in the middle of Hubei Province, China. In September 2016, a total of 213 soil samples (0–20 cm depth) were collected in the main urban area of Daye. The concentrations of Cd, Co, Cr, Cu, Mn, Ni, Pb and Zn in those soil samples were measured with inductively coupled plasma mass spectrometry (ICP-MS) after preprocessing (Yang et al., 2016). These concentrations, which may vary across space and/or in time, are mathematically modeled as random fields (Christakos, 2017) that account for in situ uncertainties and lawful variations of heavy metal distributions (the term “lawful” means that the heavy metal distribution obeys a physical law of change across space-time). Moreover, the data of land cover of Daye city were manually interpreted using contemporaneous Google earth data. The spatial distributions of soil samples and land cover are shown in Fig. 1.

2.2. Data analysis and spatial distribution maps

The flowchart of the various components of the proposed approach is shown in Fig. 2. Standard statistical analysis was initially conducted to obtain the statistical characteristics of the soil heavy metal concentrations and accumulations. In this study, the accumulations of soil heavy metals were defined as the difference between the measured value and the corresponding background value. The background value of each heavy metal is obtained from the soil background value of Hubei Province (hence, it is the average value of heavy metal in the soil under the natural environment of the whole Province). In some places the concentration of heavy metals in the soil may be higher or lower than this background value. Accordingly, if the concentration of soil heavy metal at a location is lower than this background value, the accumulations of soil heavy metal will be less than zero.

Specifically, the reasoning behind the application of the background values of Hubei Province to the present study is as follows: Based on 4000 soil profiles in China, the background values of 12 elements (including Cu, Pb, Zn and Cd) were obtained by CNEMC (1990). In Hubei Province, 91 classic soil profiles were used to obtain the background values of heavy metals in soils, as follows: First, the sampling units were divided according to soil types in Hubei Province. Second, in every sampling unit, the soil profile was selected at the site based on the obvious characteristics of soil type: relatively flat and stable terrain, good vegetation coverage, without human disturbance. Third, the arithmetic mean value of each soil element was used as the background value of the Province. It is noteworthy that the background values proposed by CNEMC for the various provinces have been extensively applied in the study of soil heavy metals at different spatial scales in different regions of China. Thus, it is reasonable to apply the background values of Hubei Province to the present study.

The spatial distributions of soil heavy metal concentrations were generated by ordinary Kriging (OK) interpolation. Two accuracy indicators, the Pearson correlation coefficient (r) and the mean absolute error (MAE), were computed to assess the spatial interpolation accuracy based on the standard sample-based “leave one-out” cross-validation technique. In addition, an error rate (ER), defined as the ratio of the MAE over the mean value of soil heavy metal concentrations, was computed to remove any quantitative differences among the various heavy metals in soils. Then, the spatial of accumulations of soil heavy metals were obtained using the difference between the concentrations value and the corresponding background value.

2.3. Principal component analysis (PCA) and absolute principal component scores (APCS)

The PCA/APCS is a common multivariate factor analysis method used in the field of pollution sources identification (e.g., it has been used to estimate sources of particulate matter in Boston by Thurston and Spengler, 1985). The PCA/APCS receptor model can determine quantitatively not only the pollutant load attributed to each source but also the average contribution of each source to each pollutant and to the pollutant in each sample (Jain et al., 2017).

PCA is often used in data reduction to identify a small number of factors that explain most of the variance observed in a much larger number of manifest variables (Guo et al., 2004). In this study, PCA focused on the accumulations of soil heavy metals with the following steps (Song et al., 2006). First, the accumulations of each soil heavy metal are transformed into a dimensionless standardized form,

$$Z_i = (A_{ik} - \bar{A}_k)/\sigma_i,$$

(1)

where \(A_{ik}\) is the accumulation value of heavy metal \(i\) at location \(k\), \(\bar{A}_k\) and \(\sigma_i\) are the mean and standard deviation, respectively, of accumulations of heavy metal \(i\) in the study area. Second, the PCA model is expressed as

$$Z_{ik} = \sum_{j=1}^{p} g_{ij} h_{ij},$$

(2)

where \(k = 1,\ldots, p\) denote sources, and \(g_{ij}\) and \(h_{ij}\) are the factor loadings and factor scores, respectively. This equation is solved by eigenvector decomposition. Varimax rotation is often used to redistribute the variance and provide a more interpretable structure on the factors. It is worth noting that, in the present work, the PCA was applied on the spatial soil heavy metal accumulation distributions considered at dates other than the sampling date for source identification purposes using factor extraction with eigenvalues > 1 after varimax rotation.

Then, APCS were used to estimate the source contributions to each heavy metal based on the following steps:

Step 1: All heavy metal accumulation values were normalized using Eq. (1);

Step 2: By introducing an artificial sample with zero accumulations for heavy metal \(i\) at an extra location (\(A_{i0} = 0\), the true zero for each factor score is calculated as:
Step 3: The factor score of each heavy metal $i$ at location $k$ is obtained by the PCA method, whereas the APCS of each heavy metal at every location is derived by subtracting the factor score of $Z_0$ (the artificial sample) from the factor score of $Z_k$.

Step 4: The contribution of source $p$ to the accumulation of heavy metal $A_i$, denoted as $SC_{p}^{i}$ ($i = 1, ..., n$) is derived by means of the regression equation:

$$b_0 + \sum_{p=1}^{w} SC_{p}^{i} = A_i,$$

$$SC_{p}^{i} = b_p \cdot APCS_p,$$

where $APCS_p$ is the scaled APCS value of the rotated factor $p$ for the considered location, $b_0$ is the constant term of multiple regression for heavy metal $i$, and $b_p$ is the coefficient of source $p$ for heavy metal $i$. The numerical average of $SC_{p}^{i}$ over all locations represents the average contribution of the source $p$.

The factor scores obtained by the PCA results of the estimation approach of APCS were further used to calculated the PC score (PCS) by

$$PCS_p = FS_{kp} \cdot \sqrt{Var_p},$$

where $PCS_p$ and $FS_{kp}$ are the scores of the principal component $PC_p$ and the factor $p$ at the location $k$, respectively, and $Var_p$ denotes the variance of the factor $p$. In this study, $PCS_p$ is considered as the pollution intensity of the source $p$ at location $k$.

2.4. The geographical detector

Based on the PCA/APCS results, we can only roughly detect the source types contributing to the accumulation of soil heavy metals. So, it is necessary to obtain the relationship between landscape and PCs in order to analyze the driving factors of heavy metal accumulation in soils. For this purpose we used the geographical detector (GeoDetector, Wang et al., 2010), which is a spatial variance analysis model that can identify the explanatory variables that influence significantly the dependent variable.

Specifically, in this study the dependent variable is the score of one PC obtained from APCS/PCA, whereas the land cover ($X_1$, Fig. 1), distance to a mine or smelter ($X_2$, left of Fig. S1), distance to main road ($X_3$, middle of Fig. S1) and topographic elevation ($X_4$, right of Fig. S1) are considered as explanatory variables. In addition, because the explanatory variables used in the GeoDetector are of the categorical type,
the continuous variables (such as $X_k$, $X_i$ and $X_j$) were discretized into categorical variables. For $X_i$ and $X_j$, a series of circular buffers at 50 m (equal to the spatial resolution of spatial distributions of PC scores in this study) intervals around each object (mine, smelter, or road) were assumed, and the overlapping parts of the buffers at the same distance from the object were merged. Based on the quantiles, the elevation $X_i$ was divided into 10 categories (Fig. S1). The spatial distributions of $X_k$, $X_i$ and $X_j$ are also shown in Fig. S1.

For each explanatory variable, the study area $D$ was divided into $m$ sub-areas $D_i$ ($i = 1, 2, \ldots, m$; $m$ is the count of categories for the explanatory variable). Let $n_i$ be the number of grids in the sub-area $D_i$, and let $n$ be the total number of grids over the entire study area $D$. Then, the power of the explanatory variable to the $PCS$ of soil heavy metals accumulations is given by

$$q_{DP} = 1 - \frac{1}{nDp} \sum_{i=1}^{m} n_{D_i} \sigma_{PC_{p,i}}^2,$$  

(6)

where $\sigma_{PC_{p,i}}^2$ and $\sigma_{PC_{p,0}}^2$ are the variances of the $PC_p$ scores within the $D$ area and the $D_i$ sub-area, respectively. The range of $q$ is between 0 and 1, and the greater $q$, the stronger the power of the determinant. In addition, to find out whether two explanatory variables, when taken together, weaken or enhance each other, or they are mutually independent in developing soil heavy metal accumulation, the interaction detector model is computed using the following rules (Wang and Hu, 2012):

Nonlinear-weaken: $q(X_i \cap X_j) < \min(q(X_i), q(X_j)).$

Uni-weaken: $\min(q(X_i), q(X_j)) < q(X_i \cap X_j) < \max(q(X_i), q(X_j)).$

Bi-enhance: $q(X_i \cap X_j) > \max(q(X_i), q(X_j)).$

Independent: $q(X_i \cap X_j) = q(X_i) + q(X_j).$

Nonlinear-enhance: $q(X_i \cap X_j) > q(X_i) + q(X_j).$

In the above rules the symbol “∩” means that the $X_i$ and $X_j$ are taken together (or, conjunction of $X_i$ and $X_j$). The $q(X_i)$ and $q(X_j)$ denote the effect of the explanatory variables $X_i$ and $X_j$, respectively, on soil heavy metal accumulation, and the $q(X_i \cap X_j)$ denotes the $q$-statistic of the joint (interactive) effect of $X_i$ and $X_j$. Specifically, the characterization “nonlinear-weaken” indicates a small interactive effect of $X_i$ and $X_j$ compared to their effects considered separately; the term “uni-weaken” suggests a mild interactive effect that lies between the separate effects of $X_i$ and $X_j$; the term “bi-enhance” indicates that the interactive effect of $X_i$ and $X_j$ is bigger than each one of the separate effects of $X_i$ and $X_j$; the characterization “independent” means that the interactive effect is equal to the sum of the separate effects of $X_i$ and $X_j$; and the term “nonlinear-enhance” indicates that the interactive effect of $X_i$ and $X_j$ is stronger than the sum of their separate effects.

2.5 Identification of pollutant land covers and calculation of the contribution of pollution source using multiple linear regression (MLR) with decay function

Initially, the land-covers that could cause soil heavy metal pollution can be easily identified based on the joint implementation of the PCA/APCS and the GeoDetector approach. To further determine whether the initially identified land-covers are actually pollution sources: (i) for each identified land-cover, a series of circular buffers associated with each pollutant source were considered at 50 m intervals; (ii) the average PCS of each buffer ring for each identified land cover was calculated; and (iii) a land cover was determined to be the pollution source only if there is a significant negative correlation between the average PCS of the buffer ring around the land cover and the distance between the ring and the land cover.

Subsequently, the regional area was divided into three categories: pollutant source, pollutant sink, and the area not contaminated by identified pollutant sources termed as pollutant-free area. Notice that the pollutant sources are the land covers determined by the approach (i)-(iii) introduced above. Excluding the areas already identified as pollution sources, the pollutant sink was defined as the area with $PCS > 0$. Correspondingly, a pollutant-free area was defined as the land cover with no $PCS > 0$.

Finally, in order to calculate and compare the contributions of the different pollution sources to the accumulations of soil heavy metals, the source contribution was calculated using MLR with heavy metal accumulation and distance from the pollution sources:

$$Z_{ik} = \sum_{j=1}^{p} b_{jk} D_{ij} = \sum_{j=1}^{p} b_{jk} \exp \left( \frac{D_{kj}}{a_k} \right),$$  

(7)

where $Z_{ik}$ is the accumulation of heavy metal $k$ in the site $i$, $p$ is the number of pollution sources, $b_{jk}$ is the regression coefficient of the distance to sources, and $D_{ij}$ is the decay function of distance from site $i$ to the $j$th pollution source, because the spread character of the heavy metals from emission sources to soil decreases with distance (Huang et al. 2018). In this study, an exponential decay function, as shown in Eq. (7), was used to quantify the decreasing influence from pollution source to soil with increasing distance, $D_{ij}$ is the spatial distance from pollution source $j$ to site $i$, and $a_k$ is the maximum influence distance parameter. Because when $D_{ij} = 3a_k$, the result of $\exp \left( -\frac{D_{kj}}{a_k} \right)$ will be at 5%, close to 0, in this study, $3a_k$ will be considered as the maximum influence distance of pollution $j$ on heavy metal $k$.

3. Results and discussion

3.1 Descriptive statistics and spatial distributions of soil heavy metal accumulations

The histograms of the soil heavy metal concentrations are shown in Fig. S2. And, the summaries of the descriptive statistics of the soil heavy metal concentrations and the corresponding accumulations in the study area are listed in Table S1. The theoretical variogram models of heavy metals in the study area are shown in Table S2. Then, the spatial distribution of the soil heavy metals accumulations in the study area are shown in Fig. S3, where the green, yellow and red colors indicate that the accumulations of heavy metals are lower than, closer to, or higher than 0, respectively.

As shown in Table S1, the mean accumulations of Cd, Cu, Pb and Zn were greater than 0, and more than 50% of sampling points those concentrations of Cd, Cu, Pb and Zn exceeds the corresponding background value, indicating significant Cd, Cu, Pb and Zn accumulation in the Daye region. Meanwhile, as shown in Fig. S3 (a, d, g, h) the spatial distribution of Cd, Cu, Pb and Zn accumulations are almost entirely dominated by yellow and red colors, implying that almost the entire study area was polluted due to the accumulations of the four heavy metals. As shown in Table S1, the percentages of sampling points beyond the corresponding background values (PBV) of Co and Mn are 30.52% and 33.8%. Moreover, nearly half of the study area was covered by chromium resulted from Figs S3c and f, indicating that nearly the entire Daye region was not polluted by Cr and Ni.

The cross-validation results of the spatial distributions of the eight heavy metals accumulations generated by OK spatial interpolation are shown in Table S3. The values of $r$ range from 0.748 to 0.987 ($p < 0.01$), and the ER ranges from 5% to 18%, indicating the high accuracy of spatial interpolation for each heavy metal accumulation due to high $r$ and low ER values. Thus, it is a reliable approach to perform the following analyses using the results of spatial OK interpolation.
3.2. Source identification and apportionment by PCA/APCS

Because the accumulations of Cr and Ni in the largest part of the Daye region were lower than 0 (see, Fig. S3c and f), these two heavy metals will not be considered further in the following analysis. Table S4 shows the PCA results of accumulations of six soil heavy metals after varimax rotation. Two PCs, explaining a total of 86% of the data variation, were extracted because their eigenvalues greater than 1. The first principal component (PC1), accounting for 52.9% of the data variance, is dominated mainly by Cd, Mn, Pb, and Zn. As is shown in Fig. S3a, d, g and h, the high accumulations of Cd, Mn, Pb and Zn were all distributed in the southern and southeastern part of the Daye city, where there are many mines and smelters, related to gold, silver, cooper, iron, molybdenum, sulphur and so on. Thus, PC1, in this study, is interpreted as a source of Exploitation and Smelting involving a variety of Minerals (ES_M). The second principal component (PC2), explaining for 33.1% of the data variance, mainly loads Co, Cu. Meanwhile, as shown in Fig. S2b and d, the high accumulations of Co and Cu were all distributed in the center of the study area, where there is the oldest and still active Copper mine in China (Tonglu Mountain Cooper Mine) with the related tailings and smelters. And, Co often occurs as an associated mineral of cooper deposits. Thus, PC2 can be interpreted as a source of Exploitation and Smelting of Copper ore (ES_C).

Then, the APCS model, was performed on the two extracted principal components (PC1 and PC2) to quantify the two sources to each spatial position. The results of APCS are listed in Table 1 (left). According to the results of Table 1, in the study area, the ES_M accounted, on average, for 64%, 81%, 93% and 78% of the Cd, Mn, Pb and Zn accumulations, respectively. And, ES_C accounted for an average of 86% and 80% of the Co and Cu accumulations, respectively. Overall, the ES_M and ES_C sources contributed 61% and 39% of the total soil heavy metal accumulations in the study area. Meanwhile, the spatial distributions of PC1 scores and PC2 scores are shown in Fig. 3a and b, respectively. The high PC1 scores were mainly distributed in the western and southeastern parts of the study area, where there are many mines and smelters related to various kinds of nonferrous metals. And the high PC2 scores were mainly distributed in the central part of the Daye region, similar to the spatial distribution patterns of Cu and Co. This again confirms that the PC2 was related to the exploitation and smelting of Copper ore.

3.3. Relationship between landscape and PC of soil heavy metals by the Geographical Detector

In this study, the GeoDetector model was used to assess the effect intensities and the interactive effects of the four explanatory variables (X1, X2, X3, and X4; Figs. 1 and S1) on the PC1 and PC2 of soil heavy metal accumulations. I.e., the dependent variables include the scores of PC1 and PC2 shown in Fig. 3a and b, respectively. The GeoDetector software was downloaded from the website (http://www.geodetector.org/), and the results are shown in Table 2, where the values in red and bold font denote the impact of a single explanatory variable on the corresponding PC, and the other values are the impacts of pairs of explanatory variables on the PC. The p values for all q statistics are 0, indicating that all the explanatory variables are found to have significant effects on soil heavy metal accumulations. Especially, the X1 has the greatest q values for PC1 (0.2) and PC2 (0.21), indicating that the distance to mine or smelter was the most important driving factor. Thus, mines and smelters are likely to be the source landscape of heavy metals accumulated in soils. In addition to identifying the separate effects of individual explanatory variable on the accumulation of heavy metals in soils, the interaction relationships between these explanatory variables were further examined. The values in black font in Table 2 show the q-statistic of the interactive effects among the explanatory variables. The interaction results indicate that the interactions between all explanatory variables exhibited nonlinear-enhanced effects on soil heavy metal accumulations, implying that q(X1 ∩ X2) > q(X1) + q(X2) for all explanatory variables. Thus, although the weaker relationships were found to be between the X1, X3, X4 and soil heavy metal accumulations, their interactions with X2 still affected these accumulations. Specifically, the q values of X1 and X2 ∩ X1 for PC1 are 0.15 and 0.59, respectively, indicating that roads also have an important effect on soil heavy metal accumulations. This may be due to the influence of the road network on the distribution of soil heavy metal accumulations during the transportation activities of ore and tailing mines.

Overall, the GeoDetector results show that the selected explanatory variables all have significant effects on soil heavy metal accumulations in Daye city. However, all the q-statistics values are closer to 0 rather than 1, indicating that the explanatory power of the four explanatory variables for PC1 and PC2 is limited. Although, according to the GeoDetector and PCA/APCS results, the mines and smelters are identified as the source landscape of soil heavy metal accumulations, the lower values of the q-statistics suggested that X1 cannot fully explain the spatial distribution of soil heavy metal accumulations in the entire study area. This may be due to the following reasons: (1) it is not necessarily true that all the land covers associated with the identified source landscapes are actually emitting heavy metals into soil; and (2) the influence distance of each real pollutant source is limited, that is, the pollutant sources cannot affect the entire study area. Thus, it is necessary to further identify those land covers that are actually emitting heavy metals to soil from the identified source landscapes. This is the objective of the next section.

3.4. Determination of the land covers of pollutant sources

The spatial distributions of the PC1 and PC2 scores are shown in Fig. 3. Locations where the PC1 and PC2 scores were less than or equal to 0 were determined to be non-polluted land (covering about 43.5% of the study region, see the light grey colored part of Fig. 3c). According to the land cover types shown in Fig. 1 and the identified source types obtained from PCA, the black-colored parts in Fig. 3c were considered to be the suspected land covers of pollutant sources for PC1 or PC2.

According to the methods described in Section 2.5, four mining and metallurgy lands or land groups (marked as 1, 2, 3, and 4 in Fig. 3c) were determined to be the sources of soil heavy metal pollution (black-colored parts of Fig. 3c), because significant negative correlations between the PCS average of the buffer ring around those land covers and the distance between them and the buffer ring were found from the origin to different distance (see Fig. 4). Among them, source #1 is a quarry located under the HuangJin mountain (a forest park); Source #2 involves the famous old Tonglu mountain cooper mine, and some related mineral processing and smelting enterprises; Source #3 is a well-known mining and mineral processing base at the Lion mountain (rich in gold, copper, tungsten, molybdenum, manganese, wollastonite, calcite, dolomite, granite and limestone); And, source #4 is the large mine

Table 1

<table>
<thead>
<tr>
<th>Heavy metal</th>
<th>PCA/APCS</th>
<th>MLR with decay function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ES_M</td>
<td>ES_C</td>
</tr>
<tr>
<td>Cd</td>
<td>64%</td>
<td>36%</td>
</tr>
<tr>
<td>Co</td>
<td>14%</td>
<td>86%</td>
</tr>
<tr>
<td>Cu</td>
<td>20%</td>
<td>80%</td>
</tr>
<tr>
<td>Mn</td>
<td>81%</td>
<td>19%</td>
</tr>
<tr>
<td>Pb</td>
<td>93%</td>
<td>7%</td>
</tr>
<tr>
<td>Zn</td>
<td>78%</td>
<td>22%</td>
</tr>
<tr>
<td>Total</td>
<td>61%</td>
<td>39%</td>
</tr>
</tbody>
</table>

Note: ES_M: the source of Exploitation and Smelting involving a variety of Minerals; ES_C: the source of Exploitation and Smelting of Copper ore.
of the Oujia mountain, which produces many kinds of ores (including copper, gold, silver, molybdenum, tungsten, copper, wollastonite, limestone, calcite and black marble). In addition, about 50.9% of the study area was polluted by the identified pollutant sources (namely, pollutant sink, dark grey part of Fig. 3c).

3.5. Calculation of the contribution of pollution source using MLR with decay function

According to the formula (7), the fitting models for each heavy metal, were developed as follows:

\[
Pb = 28.4 \times e^{-0.73a} + 14.3 \times e^{-0.73b} + 71.3 \times e^{-0.73c} + 167.4 \times e^{-0.73d} \quad R^2 = 0.65
\]

\[
Zn = 239 \times e^{-0.75a} + 131 \times e^{-0.75b} + 368 \times e^{-0.75c} \quad R^2 = 0.66
\]

where the subscript 1, 2, 3, 4 represents the serial number of identified source landscape introduced in the last paragraph. The determinant coefficients \((R^2)\) of the fitting models for Cd, Co, Cu, Mn, Pb, and Zn are 0.7, 0.54, 0.75, 0.51, 0.65, and 0.66, respectively, indicating that the results of models fitting are qualified. It is noticed that only the accumulation of Pb received contributions from all four source landscapes. That is, not all identified pollution source landscapes contributed to the accumulation of each heavy metal, which is reflected in the model by a close to 0 value of the parameter \(b\) (in Eq. (7)). Therefore, the source landscapes that contributed little to the accumulation of some heavy metals were neglected in the fitting models of each heavy metal. The spatial distributions of pollution sources for each heavy metal are shown in Fig. 5, and the contributions of heavy metals by each pollution
source are shown in Table 1 (right). According to the results of Fig. 5 and Table 1, source #1 contributed 5.1% Cd accumulation, 4.1% Mn accumulation, 2.2% Pb accumulation, and 7.7% Zn accumulation, and had influence on Cd, Mn, Pb, and Zn for approximately 3 km, 0.75 km, 0.6 km, and 2.7 km, respectively, from the center this source. Source #2 is responsible for 97.6% Co and 72.4% Cu accumulation in the study area, along with 13.9% Cd, 5.9% Pb, and 16.1% Zn accumulations. The influence distances of source #2 for Co and Cu are approximately 5 km and 6 km, respectively, from the center this source. Source #3 is responsible for 97.6% Co and 72.4% Cu accumulation in the study area, along with 13.9% Cd, 5.9% Pb, and 16.1% Zn accumulations. The influence distances of source #2 for Co and Cu are approximately 5 km and 6 km, respectively, from the center this source. Source #3 is responsible for the accumulation of 41.6% Mn (2.2 km) and 7.9% Pb (1.4 km) accumulation. Source #4 contributed 77.3% Cd, 10.5% Cu, 33.2% Mn, 81.8% Pb, and 61% Zn accumulation. And the influence distances of source #4 for Cd, Cu, Mn, Pb, and Zn are approximately 7.5 km, 3 km, 1.5 km, 3.6 km, and 8.7 km, respectively. Overall, source #1, source #2, source #3, and source #4 contributed 3.2%, 34.3%, 8.3%, and 44% of the total soil heavy metal accumulations in the study area, indicating source #4 is the biggest pollution source for soil heavy metals in the study area. It is noticed that the four pollution sources did not account for the total accumulations for each heavy metal. The rest of accumulations for each heavy metal in soils might come from other small pollution sources, such as roads in the study area.

However, the proposed model of MLR with decay function has potential disadvantage. The model assumes that the influence of pollution source on the surrounding soil in all directions is the same. However, some natural factors, such as topography, runoff, main wind direction, and soil properties, will affect the accumulation of heavy metals in soil. Then, the accumulation of heavy metals in soils by pollution source has directional heterogeneity. Thus, the model needs to consider more environmental or natural factors in the further study.

4. Conclusions

In this work, a synthetic threefold method combining PCA/APCS, GeoDetector and GIS analysis was proposed to identify and quantify the pollution source landscapes that emitted heavy metals into soils. The proposed method was applied to Daye city, China, in which case four pollution source landscapes were identified, and the spatial distribution maps of the accumulation contribution of each source landscape to each soil heavy metal were obtained. Moreover, the total contribution of each pollution source landscape to soil heavy metals was summarized. In particular, the method proposed in this study extends the existing quantitative source apportionment methods (e.g. PCA/MLR, PCA/APCS, PMF) of soil heavy metals to the landscape level. The results obtained by the proposed method are easier to understand and physically more meaningful, because the specific landscapes that cause the accumulation of soil heavy metals, rather than the type of pollution source, are directly identified. At the same time, by comparing the contribution of different pollution source landscapes, we can obtain more specific evidence for the purpose of source assessment and control of heavy metals in regional soils.

Based on the results presented in this paper, more work can further improve the proposed method, as follows: (1) In this study only an

---

Table 2

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Factors</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1 score</td>
<td>$X_1$</td>
<td>0.059</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_2$</td>
<td>0.28</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_3$</td>
<td>0.27</td>
<td>0.59</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_4$</td>
<td>0.18</td>
<td>0.28</td>
<td>0.33</td>
<td>0.062</td>
</tr>
</tbody>
</table>

| PC2 score           | $X_1$   |       |       |       | 0.094 |
|                     | $X_2$   | 0.33  | 0.21  |       |       |
|                     | $X_3$   | 0.27  | 0.31  | 0.022 |       |
|                     | $X_4$   | 0.27  | 0.34  | 0.097 | 0.043 |

Note: $X_1$: Land cover; $X_2$: Distance to mine or smelter; $X_3$: Distance to road; $X_4$: Elevation; all values of $p$ for $q$ statistics are 0, indicates significance at the 0.05 level.
Fig. 5. Spatial distribution of pollution sources for Cd, Co, Cu, Mn, Pb, and Zn, and their contribution to the accumulations of those heavy metals.
exponential model was used to simulate the relationship between the attenuation of soil heavy metals accumulations and the distances from pollution source landscapes. Hence, future research seeking to involve other models (such as the Gaussian and the linear models) in the simulation of this kind of relationship could lead to more accurate results; (2) more landscape factors (such as the main wind direction, runoff and elevation difference) could be taken into account in the above model, which will bring the simulation results closer to the reality of heavy metals migration in soils; (3) in this work a mining and metallurgical city was considered a research area, thus, the identified pollution source landscapes were all related to mining and metallurgical activities. In future studies the proposed method could be applied to additional types of regions to evaluate the applicability of the proposed method.

Acknowledgments

This research was supported by the National Key R&D Program of China (Grant No. 2018YFC1800104), the National Natural Science Foundation of China (Grant Nos. 41671217, 11501339 and 41671399), and the Fundamental Research Funds for the Central Universities, China (Grant No. 2662017PY038).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.caten.2019.104297.

References