Contamination assessment and source apportionment of heavy metals in agricultural soil through the synthesis of PMF and GeogDetector models

Xufeng Fei\textsuperscript{a,b,*}, Zhaohan Lou\textsuperscript{a}, Rui Xiao\textsuperscript{c}, Zhouqiao Ren\textsuperscript{a,b}, Xiaonan Lv\textsuperscript{a,b}

\textsuperscript{a} Zhejiang Academy of Agricultural Sciences, Hangzhou, China
\textsuperscript{b} Key Laboratory of Information Traceability of Agriculture Products, Ministry of Agriculture and Rural Affairs, China
\textsuperscript{c} School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

HIGHLIGHTS

• A novel framework based on spatial analysis for source apportionment is proposed.
• Combined with auxiliary data, the new model provides foundations for source analysis.
• Cr (80.72\%) was derived mainly from natural sources while As and Pb had mix sources.
• Cd (73.68\%) was closely associated with agricultural activities.
• Hg (92.38\%) was mainly attributed to industrial activities.

GRAPHICAL ABSTRACT

Abstract

Heavy metal pollution in soils has attracted great attention worldwide in recent decades. Selecting Hangzhou as a case study location, this research proposed the synthesis application of positive matrix factorization (PMF) and GeogDetector models for quantitative analysis of pollution sources, which is the basis for subsequent soil pollution prevention and remediation. In total, 2150 surface soil samples were collected across the study area. Although the mean concentrations of As, Cd, Cr, Hg, and Pb in the soils were lower than the National Environmental Quality Standards for Soils in China, the mean contents of As and Cd were higher than their corresponding local background values by approximately 1.31 and 1.59 times, respectively, indicating that heavy metals have been enriched in topsoil. Agricultural activities, industrial activities, and soil parent materials were the main sources of heavy metal pollution in the soils, accounting for 63.4\%, 19.8\%, and 16.8\% of the total heavy metal accumulation, respectively. Cr was derived mainly from soil parent materials (80.72\%). Cd was closely associated with agricultural activities (73.68\%), such as sewage irrigation and application of fertilizer. Mercury was mainly attributed to industrial activities (92.38\%), such as coal mining and smelting. As was related to agricultural (57.83\%) and natural (35.56\%) sources, and Pb was associated with industrial (42.42\%) and natural (41.83\%) sources. The new synthesis models are useful for estimating the source apportionment of heavy metals in soils.

1. Introduction

Soil is one of the most important ecosystems for human survival and development, and soil safety is the fundamental guarantee for national food security and human health (Fei et al., 2019; He et al., 2019; Wang et al., 2019). With the rapid urbanization and industrialization of
society, soil pollution, especially soil heavy metal contamination and accumulation, has become a serious problem, attracting great public attention worldwide (Hu et al., 2017a; Huang et al., 2018; Yang et al., 2018). As China is a developing nation that has undergone rapid urbanization in recent decades, soil heavy metal pollution has become a main environmental problem (Niu et al., 2013; Hu et al., 2017b; Liu et al., 2019). According to the Chinese National Soil Contamination Survey Report released by the Ministry of Land and Resources and the Ministry of Environmental Protection of China, approximately 16.1% of soil samples analyzed were contaminated by heavy metals to various degrees (MEP, 2014). Because of the heavy metal characteristics of bioavailability, persistence, and toxicity, heavy metal accumulation in soil can lead to reductions in soil fertility and crop production. Heavy metals can also be transferred easily and accumulate through biomagnification in the food chain, which poses significant risks to food safety and human health, as heavy metals are absorbed by the human body through inhalation, ingestion, and dermal absorption (Xu et al., 2017; Zang et al., 2017; Zhao et al., 2014a). Therefore, it is necessary to quantitatively evaluate the characteristics, contamination, and sources of heavy metals in soil.

Generally, heavy metals in soil come from two major sources: natural sources, driven by weathering and pedogenesis processes, and anthropogenic activities, such as industrial manufacturing, sewage irrigation, vehicle exhaust, agricultural fertilizer, mining, and smelting (Sun et al., 2014, 2019; Yang et al., 2013). Therefore, to effectively reduce the cost and workload of soil remediation, it is important to quantitatively clarify the sources of soil heavy metal pollution (Fei et al., 2018; Huang et al., 2018). Many previous studies have used multivariate qualitative/quantitative statistical methods, such as spatial deviation, correlation analysis, cluster analysis, geographic information systems, and principal component analysis (PCA), often in combination with multiple linear regression, to determine the variability in and possible sources of heavy metals in soils (Davis et al., 2009; Dong et al., 2019; Gu et al., 2012; Lv, 2019; Nanos and Martín, 2012). Among these approaches, the positive matrix factorization (PMF) model recommended by the US EPA, an ideal method for ensuring nonnegative source contributions, has been applied widely and successfully to identify and quantify pollution sources of heavy metals in soils (Dong et al., 2019; Huang et al., 2018; Wang et al., 2019).

However, most of these methods cannot analyze the effect of categorical variables such as soil type and parent materials, which have an important influence on the accumulation of heavy metals. Few studies have taken the spatial information of the sampling points into consideration, and the definition of each principal component obtained from a single PCA/PMF model is based mainly on previous researches and experience, thus precluding the ability to quantitatively define the detailed effects and spatial characteristics of every source on different heavy metals. Moreover, previous methods are unable to identify the importance of specific sources to the principal components based on the spatial analytical data, which is vital for controlling local HM pollution and remediation in soils. To address these gaps, this study proposes the synthesis of PMF and GeogDetector models for soil heavy metal source apportionment and evaluates the approach in a case study using the city of Hangzhou, China. Compared to previous receptor models, by including spatial information and auxiliary data (categorical and numerical variables), such as soil parent material, road networks, industrial production, etc., this approach could provide an in-depth understanding of multiple source contributions and a better definition of principal factors. To our knowledge, this is the first study using GeogDetector to interpret results derived from a PMF receptor model.

The main objectives of this study were to (1) analyze the contents and basic characteristics of heavy metals in soil in the city of Hangzhou, China, (2) identify the potential sources of heavy metals through the PMF model, and (3) develop a new combined spatial method for quantitative source apportionment and definition through various forms of auxiliary data. The results of this study provide useful information for the prevention and control of heavy metal pollution in soil.

2. Materials and methods

2.1. Study area

Hangzhou is located in the eastern coastal area of China (29°11′–30°33′ N, 118°21′–120°30′ E). It is the capital city of Zhejiang Province and one of the most important cities in the Yangtze River Delta urban agglomeration (Fig. 1). The city has a moderate subtropical climate with a hot and humid summer and cold and dry winter. The annual average values of temperature, precipitation, and sunshine hours are 17.8 °C, 1454 mm, and 1765 h, respectively. In total, 13 districts/counties are included in the study area, which has a total area of approximately 16,854 km². According to the population census in 2017, the population of the area is approximately 9.47 million. The northeastern area consists of plains, but the southwestern area is mountainous. The main soil types are red soil and paddy soil derived from eluvium and aluvial soil parent materials, respectively. In the past few decades, this area has experienced rapid industrialization and urbanization, and the consequent increasing accumulation of heavy metals has caused great concern (Chen et al., 2009; Fei et al., 2018; Zhang and Ke, 2004).

2.2. Soil sampling and chemical analysis

According to the agricultural soil vector map, there are 211,931 ha of agricultural soil in the study area. The 1 km*1 km grid layer was overlapped with the distribution of agricultural soil to extract the expected sampling points locations in the study region (Ren et al., 2019). In total, 2150 surface soil samples were collected across the study area (Fig. 1). Five subsamples were collected and mixed thoroughly to obtain a representative sample, and the sampling locations were recorded by a global positioning system (GPS). All soil samples were air-dried at room temperature and then ground to 100-mesh size for chemical analysis. For the determination of Cd, Cr, and Pb contents, the soil samples were digested with HNO₃–HClO₄–HF, and the metals were analyzed by inductively coupled plasma mass spectrometry (ICP-MS). For the determination of As and Hg contents, soil samples were digested with HNO₃–HCl, and atomic fluorescence spectrometry (AFS) was used for the analyses. Blind duplicates and standard reference materials (GSS–3, China National Center for Standard Materials) were used for quality assurance and control. Standard sample recovery ranged between 90 and 110%, and the relative standard deviations of duplicate samples were between 3 and 8%.

2.3. Analysis methods

In this work, the analysis consisted of both a statistical and a spatial analysis part, as follows.

Statistical analysis: (i) Statistical indicators (mean, median, coefficient of variation, etc.) and Spearman correlation analysis were employed to describe the basic characteristics of heavy metals in agricultural soil and their internal correlation, and (ii) a PMF model was used for pollution source apportionment of heavy metals, as well as for calculation of the pollution component scores of each site.

Spatial analysis: (i) Getis–Ord Gi* analysis (Getis and Ord, 1992) was implemented to explore the spatial distribution pattern of the component scores derived from PMF analysis. (ii) Ordinary kriging was used to draw the spatial distribution map of each component score (grid format). (iii) The results of ordinary kriging and the detailed pollution source dataset (soil parent materials, distance to the main road, etc.) were combined, and the GeogDetector model was employed to determine the spatial correlation between each component score and specific
pollution sources, which could provide additional valuable information for quantitative source apportionment.

The overall flow chart of this study is shown in Fig. 2, and the specific description of each method is as follows:

2.4. PMF model

The PMF model, an efficient receptor model for pollution source apportionment, was used in this study to analyze the sources of heavy metals in the soil (Huang et al., 2018; Wang et al., 2019; Zheng et al., 2018). In brief, this model decomposes the matrix of the original dataset $X_{ij}$ into two factor matrices (the source contribution matrix $g_{ik}$ and the source profile matrix $f_{jk}$) and a residual matrix $e_{ij}$. The basic equation is as follows:

$$X_{ij} = \sum_{k=1}^{p} g_{ik} f_{jk} + e_{ij};$$

where $X_{ij}$ is the concentration of the $j$th heavy metal at the $i$th sampling location, $g_{ik}$ is the contribution of the $k$th source to the $i$th sample, $f_{jk}$ is the concentration of the element $j$ from the $k$th source, and $e_{ij}$ is the residual error matrix, which can be calculated through the minimum value of the objective function $Q$. The value of $Q$ is calculated as follows:

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left( \frac{e_{ij}}{u_{ij}} \right)^{2};$$

where $u_{ij}$ is the uncertainty of the $j$th heavy metal in the $i$th sample, which is calculated from the species-specific method detection limit (MDL), the concentration, and the provided error fraction. If the heavy metal concentration is greater than the MDL, the uncertainty can be calculated as follows:

$$u_{ij} = \sqrt{\text{error fraction} \times \text{concentrations}}^{2} + (\text{MDL})^{2}.$$

Otherwise, the uncertainty is estimated through the following equation:

$$u_{ij} = \frac{5}{6} \times \text{MDL}.$$  

All calculations follow the US EPA PMF 5.0 User Guide (U.S. Environmental Protection Agency, 2014).

2.5. Spatial analysis model

After the factor values of each component in the samples were estimated through PMF, Getis–Ord Gi* analysis (Getis and Ord, 1992) was employed to explore the spatial distribution pattern of the component scores. The spatial clustering of the samples’ component scores was determined through the $Z$ values, which were calculated as follows:

$$Z_{j} = \frac{\sum_{j=1}^{n} w_{ij} x_{j} - \bar{x} \sum_{j=1}^{n} w_{ij}}{\sqrt{\frac{1}{n} \sum_{j=1}^{n} x_{j}^{2} - \overline{x}^{2} \sum_{j=1}^{n} w_{ij}}} \times \frac{1}{\sqrt{\frac{n}{n-1} \left[ \sum_{j=1}^{n} w_{ij}^{2} - \left( \sum_{j=1}^{n} w_{ij} \right)^{2} \right]}};$$

where $Z_{j}$ is the Getis–Ord Gi* $Z$ value of sample $j$, $x_{j}$ is the component score of sample $j$, $n$ is the total number of samples, $w_{ij}$ is the weight of the neighbor sample $i$ to $j$, and $\overline{x}$ is the mean value of the component score in all samples. $Z$ values higher than 1.96 indicate significantly high
clustering, and values lower than \(-1.96\) mean significantly low clustering at the 0.05 level (Getis and Ord, 1992; Fei et al., 2016).

In addition, ordinary kriging interpolation (Fei et al., 2019; Gribov and Krivoruchko, 2011) was executed to construct spatial distribution maps of each component score to validate the spatial distribution pattern of the sampling points and provide a foundation for the GeogDetector model.

As one of the most commonly used geostatistical methods, ordinary kriging has been successfully applied in various disciplines, especially in the assessment of soil heavy metal pollution in recent years (Li et al., 2020; Wang et al., 2020; Wu et al., 2020). This tool realizes the optimal linear unbiased estimation of spatial data based on sampled data and is useful for exploring the spatial distribution pattern and variation characteristics of variables (Christakos, 1998). The semivariogram in the ordinary kriging process can be calculated as follows:

\[
\hat{\gamma}_X(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (X(s_i + h) - X(s_i))^2
\]

where \(\hat{\gamma}_X(h)\) is the semivariogram value at the distance of \(h\), \(X(s_i + h)\) and \(X(s_i)\) are the values of studied variables at the locations of \(s_i + h\) and \(s_i\), respectively, and \(n(h)\) is the number of pairs of sample points at a distance of \(h\) (Olea, 2006). Spatial distribution maps of component scores derived from ordinary kriging can be used for overlay analyses with pollution source variables through GeogDetector models. Moreover, semivariogram information could provide fundamental reference for source analysis (Wang et al., 2020; Wu et al., 2020).

2.6. GeogDetector model

The GeogDetector model (Wang and Hu, 2012) was then used to determine the spatial correlation between each component score and the detailed pollution source dataset (soil parent material data obtained from Wu et al. (2013) as a proxy for natural sources; the distance from the sampling points to the main roads calculated in ArcGIS 10.2 as a proxy for traffic, or vehicle emission sources; and industrial and agricultural production values from the Hangzhou Statistics Yearbook (http://tjj.hangzhou.gov.cn) as proxies for industrial and agricultural sources), which provided additional information for source identification of each component obtained from the PMF model.

The detailed theory of this model is available in previous works (Fei et al., 2016; Fei et al., 2018; Li et al., 2013; Wang et al., 2010; Wang and Hu, 2012). Briefly, there are four detectors in this model, i.e., a risk detector, factor detector, ecological detector and interaction detector (Wang et al., 2010; Jiang et al., 2020; Zhang et al., 2020). In this study, the factor detector and interaction detector were employed to quantitatively detect the degree of influence of the pollution sources on the...
component scores derived from the PMF model. The basic assumption of the factor detector model here is that if one component extracted from the PMF can be defined as a specific source, then the spatial distribution of this component should be similar to that of its source. For example, if component 1 (F1) has a spatial distribution pattern similar to that of the soil parent material, then F1 has a great probability of representing a natural source. The strength of this spatial similarity can be calculated quantitatively through the power determination (PD) value, which can be estimated as follows:

\[ PD = 1 - \frac{1}{AV} \sum_{i=1}^{n} A_i V_i, \]  

(7)

where \( A \) is the total area of Hangzhou, \( V \) is the component score variance of all samples, \( A_i \) is the area of subregion \( i \) for the soil parent material, the subregions were defined as every category of parent material; for numerical variables such as distance and production, the subregions were classified through maximizing/minimizing the dispersion variance between within subregions, respectively, and \( V_i \) is the component score variance of samples in subregion \( i \). In this case, the PD value, ranging from 0 to 1, represents the strength of the spatial correlation, from weakest to strongest.

Then the interaction detector was implemented to assess the combined influences of pairs of pollution sources on the component scores (Ren et al., 2019; Jiang et al., 2020). The PD values of each pair of pollution sources were labeled \( \text{PD}(S_1) \) and \( \text{PD}(S_2) \), and their interaction influence was calculated by the interaction detector as \( \text{PD}(S_1 \cap S_2) \). Finally, the interaction of the two sources was assessed by comparing the value of \( \text{PD}(S_1 \cap S_2) \) to the sum of \( \text{PD}(S_1) \) and \( \text{PD}(S_2) \), which resulted in five interaction phenomena (Table S1).

Descriptive statistics and Spearman correlation analysis were conducted using SPSS 19.0 software (IBM Inc., Armonk, NY, USA). The PMF model was applied using US EPA PMF 5.0 (U.S. Environmental Protection Agency, 2014), and Getis–Ord \( G^* \) analysis and ordinary kriging interpolation were conducted using ArcGIS 10.2 (ESRI Inc., USA), GeoDetector software (Wang and Hu, 2012) was employed for the GeoDetector model. \( P \) values at the 0.05 level were considered significant, and all \( P \) values and 95% CIs were two-sided.

3. Results and discussion

3.1. Descriptive statistics of heavy metals in soil

The summary statistics of the heavy metal concentrations in the soil samples are shown in Table 1. The contents of As, Cd, Cr, Hg, and Pb ranged from 2.63 to 43.20 mg/kg, 0.05 to 1.42 mg/kg, 10.50 to 104.00 mg/kg, 0.03 to 0.50 mg/kg, and 17.00 to 62.60 mg/kg, with median values of 7.07, 0.20, 54.40, 0.11, and 30.90 mg/kg, respectively.

<table>
<thead>
<tr>
<th>Heavy Metal</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>CV</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Background of Zhejiang</th>
<th>Average of China</th>
<th>Risk screening values</th>
</tr>
</thead>
<tbody>
<tr>
<td>As</td>
<td>2.63</td>
<td>43.20</td>
<td>10.03</td>
<td>0.03</td>
<td>17</td>
<td>31.66</td>
<td>38.47</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>Cd</td>
<td>0.05</td>
<td>1.42</td>
<td>0.5</td>
<td>0.03</td>
<td>0.38</td>
<td>0.62</td>
<td>0.36</td>
<td>6.88</td>
<td>0.3</td>
</tr>
<tr>
<td>Cr</td>
<td>54.40</td>
<td>104</td>
<td>52.9</td>
<td>0.13</td>
<td>3.09</td>
<td>16.62</td>
<td>46.09</td>
<td>1.07</td>
<td>0.05</td>
</tr>
<tr>
<td>Hg</td>
<td>0.11</td>
<td>0.50</td>
<td>0.11</td>
<td>0.03</td>
<td>0.36</td>
<td>0.62</td>
<td>0.36</td>
<td>6.88</td>
<td>0.3</td>
</tr>
<tr>
<td>Pb</td>
<td>30.90</td>
<td>62.60</td>
<td>31.66</td>
<td>0.17</td>
<td>35.7</td>
<td>17.16</td>
<td>17.16</td>
<td>11</td>
<td>0.05</td>
</tr>
</tbody>
</table>

\[ \text{Risk screening values} = 0.156 \times \text{As} + 0.156 \times \text{Cd} + 0.156 \times \text{Cr} + 0.156 \times \text{Hg} + 0.156 \times \text{Pb} \]

The PMF model was used further to identify the sources and quantify the contributions of heavy metals in the soils. The number of factors for the model was initially set at 2, 3, and 4, the start seed number was randomly obtained, and the number of runs was 20. The most suitable number of factors was determined by assessing the smallest and most stable Q value. Finally, it was determined that three factors resulted in good model fitting, with prediction residuals normally distributed within —3.0 to 3.0 and a prediction \( R^2 \) greater than 0.68. Two error estimation models, classical bootstrap (BS) and displacement of factor elements (DISP), were implemented to assess the bias and uncertainty of the PMF results (Hu et al., 2020). The results showed that approximately 91% of the base factors were reproduced in the BS model, and no factor swaps were observed in the DISP model, indicating that the three-factor PMF solution was stable.

3.2. Source apportionment of heavy metals in PMF

The PMF model was used further to identify the sources and quantify the contributions of heavy metals in the soils. The number of factors for the model was initially set at 2, 3, and 4, the start seed number was randomly obtained, and the number of runs was 20. The most suitable number of factors was determined by assessing the smallest and most stable Q value. Finally, it was determined that three factors resulted in good model fitting, with prediction residuals normally distributed within —3.0 to 3.0 and a prediction \( R^2 \) greater than 0.68. Two error estimation models, classical bootstrap (BS) and displacement of factor elements (DISP), were implemented to assess the bias and uncertainty of the PMF results (Hu et al., 2020). The results showed that approximately 91% of the base factors were reproduced in the BS model, and no factor swaps were observed in the DISP model, indicating that the three-factor PMF solution was stable.

<table>
<thead>
<tr>
<th>Heavy Metal</th>
<th>As</th>
<th>Cd</th>
<th>Cr</th>
<th>Hg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cd</td>
<td>0.361**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cr</td>
<td>0.249**</td>
<td>0.156**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hg</td>
<td>0.124**</td>
<td>0.305**</td>
<td>0.121**</td>
<td></td>
</tr>
<tr>
<td>Pb</td>
<td>0.309**</td>
<td>0.465**</td>
<td>0.266**</td>
<td>0.475**</td>
</tr>
</tbody>
</table>

** Significant at the 0.01 level.
The results of the PMF analysis are shown in Fig. 3. Factor 1 (F1) contributed to 16.8% of the heavy metals in the soils and included a heavy loading for Cr (80.72%) and moderately loadings for Pb (41.83%) and As (35.56%). The mean concentration of Cr was lower than the local background values, and only 6.33% of the samples showed a concentration of Cr exceeding the background value, indicating that there were few external pollution sources of Cr. In addition, Cr showed lower CV values than the other heavy metals. The above statistics imply that Cr mainly comes from natural sources. Previous studies have also reported that Cr in soils originates from parent materials (Fei et al., 2019; Sun et al., 2019; Zhou et al., 2016). Hence, F1 may represent a natural source, such as soil parent materials and pedogenic processes.

Factor 2 (F2), accounting for 19.8% of the total contribution, was weighted primarily on Hg (92.38%) and Pb (42.42%). Mercury had a high CV value, indicating that anthropogenic sources were the main origin of Hg pollution (Cai et al., 2015; Mamut et al., 2017; Baltas et al., 2020). In China, smelting, coal mining, and combustion are regarded as the main sources of Cd. Previous studies also concluded that long-standing farming practices, such as the application of fertilizers and pesticides, can lead to the accumulation of heavy metals such as As in soils (Jiang et al., 2017; Liu et al., 2017; Qiao et al., 2011). Thus, F3 may be associated with agricultural activities.

3.3. Spatial analysis of components

To understand the spatial distributions of the component factors further and provide information for source identification, the spatial pattern of sampling point component scores was detected by Getis–Ord Gi* analysis and the distribution of component factors across the study area was mapped by ordinary kriging (Fig. 4). Samples with low F1 values were clustered in the northern and central-eastern areas, whereas high F1 values were distributed in a wide region including the northern, central, and southwestern areas, where various eluvium and alluvium soil parent materials are distributed. The kriging interpolation map of F1 was consistent with the Getis–Ord Gi* analysis of the samples. It is worth noting that the variogram model of F1 had relatively strong spatial autocorrelation, as indicated by the low ratio (0.31) of nugget/(nugget+sill) and the long model range (34,527 m). Furthermore, there were no obvious point sources of F1 (Lv, 2019; Sun et al., 2019). As discussed above, F1 had a high loading on Cr, which had a concentration lower than the mean of local background value and relatively small CV. Consequently, F1, with strong spatial autocorrelation and a long dependence range in the soils, was probably controlled by parent materials.

There was only one sample with a low F2 value, which was located in the southwestern area, whereas there were abundant samples with high F2 values that were clustered in the northeastern suburban area, where industrial development and urban expansion are intense (Zhang et al., 2013). The kriging interpolation map of F2, with significant point sources in the high-cluster area, was in accordance with the
point results. Moreover, the variogram model for F2 had moderate spatial autocorrelation, indicated by the high ratio (0.64) of nugget/(nugget + sill) and the short dependence range (4136 m). As discussed above, F2 was heavily loaded on Hg, which had a relatively high CV. Therefore, F2, with low spatial autocorrelation and significant point sources in suburban areas, was probably controlled by intense industrial and urbanization activities.

There was no sample with a low F3 value across the study area, but many samples with high F3 values were clustered in the southern and southwestern rural areas. The kriging interpolation map of F3 was also in accordance with the point results. The variogram model for F3 had strong spatial autocorrelation, indicated by the low ratio (0.21) of nugget/(nugget + sill) and the short dependence range (4437 m). As discussed above, F3 was heavily loaded on Cd, which had a relatively high CV and is typically used as an indicator for agricultural activities (Shao et al., 2016; Zhao et al., 2014b). Hence, F3, with strong spatial autocorrelation in rural soils, was probably controlled by agricultural nonpoint source pollution, such as sewage irrigation and fertilizer application.

3.4. Source identification in GeogDetector

In the final step, the GeogDetector model was used to determine the spatial correlation between each component factor and the detailed pollution sources. The PD values of each possible pollution source proxy for the components extracted from the PMF are shown in Table 3. Soil parent material type had a significant spatial correlation with F1 (PD = 0.336, P < 0.01), which indicated that F1 represents natural sources of soil pollution. However, industrial production and agricultural production had significant spatial correlations with F2 (PD = 0.178, P < 0.01) and F3 (PD = 0.143, P < 0.01), indicating that F2 and F3 represent anthropogenic sources of soil pollution (industrial and agricultural activities, respectively).

The interaction influences of pairs of sources on the spatial distribution of the component factors derived from the PMF models are listed in Table 4. The joint effect of soil parent materials and agricultural production had the strongest influence on F1 (0.437), followed by the interaction of soil parent materials and industrial production (0.423). For F2, the joint effect of industrial production and agricultural production had the highest PD value (0.326), which further confirmed that F2 mainly comes from anthropogenic activities. In addition, the interaction effect of soil parent materials and agricultural production had the strongest influence on F3 (0.240). The PD values of the interaction of the two sources were all greater than the sums of the single PD values, which corresponded to previous studies that found that the interaction effects of multiple pollution sources showed nonlinear enhancements, indicating that the sources of heavy metals in soils are complex and diverse (Ren et al., 2019; Jiang et al., 2020).

3.5. Summary of source apportionment

Based on spatial analytical data with auxiliary datasets (both categorical and numerical variables), the proposed comprehensive model (integrated PMF and GeogDetector models, as well as Getis–Ord Gi* analysis and ordinary kriging) in this study could provide an in-depth understanding of multiple source apportionment.

Table 3

<table>
<thead>
<tr>
<th>Component Factors</th>
<th>Soil parent materials</th>
<th>Industry</th>
<th>Agriculture</th>
<th>Distance to road</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>0.336**</td>
<td>0.028</td>
<td>0.004</td>
<td>0.013</td>
</tr>
<tr>
<td>Factor 2</td>
<td>0.073</td>
<td>0.178**</td>
<td>0.014</td>
<td>0.033</td>
</tr>
<tr>
<td>Factor 3</td>
<td>0.086</td>
<td>0.024</td>
<td>0.143**</td>
<td>0.01</td>
</tr>
</tbody>
</table>

** Significant at the 0.01 level.

Table 4

<table>
<thead>
<tr>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPM&amp;IP</td>
<td>0.423NLE</td>
<td>0.282NLE</td>
</tr>
<tr>
<td>SPM&amp;AP</td>
<td>0.417NLE</td>
<td>0.121NLE</td>
</tr>
<tr>
<td>SPM&amp;DR</td>
<td>0.376NLE</td>
<td>0.123NLE</td>
</tr>
<tr>
<td>IP&amp;AP</td>
<td>0.198NLE</td>
<td>0.320NLE</td>
</tr>
<tr>
<td>IP&amp;DR</td>
<td>0.054NLE</td>
<td>0.212NLE</td>
</tr>
<tr>
<td>AP&amp;DR</td>
<td>0.029NLE</td>
<td>0.064NLE</td>
</tr>
</tbody>
</table>

SPM: Soil parent materials.
IP: Industrial production.
AP: Agricultural production.
DR: Distance to the main road.
NLE: Non-linear enhancement.

F1 derived from the PMF model showed low clusters in the northern and central-eastern areas and high clusters in a wide region in the northern, central, and southwestern areas (Fig. 4). Combined with the distribution map of F1 produced by ordinary kriging, the spatial pattern of F1 showed high spatial consistency with the distribution of soil parent materials. Moreover, the variogram model of F1 also revealed strong spatial autocorrelation and a long model range, indicating that F1 was more natural and continuous in space, with less nonstructural heterogeneity caused by human activities. Finally, the soil parent material type had the highest and most significant spatial correlation with F1 in the GeogDetector model, and there was a certain nonlinear enhancement of the interaction between the soil parent material and other pollution proxies. In conclusion, F1 could be defined as natural sources.

F2 obtained from the PMF model showed high clusters in the northeastern suburban area, where industrial development and urban expansion were intense (Fig. 2). It can also be seen from the spatial distribution map of the kriging data that high values of F2 were mainly distributed in the suburbs. In addition, the variogram model for F2 had moderate spatial heterogeneity and a short dependence range, indicating that F2 had low spatial autocorrelation and significant point sources in suburban areas. Finally, industrial production had the highest and most significant spatial correlation with F2 in the GeogDetector model. In terms of interaction effects, the combined influence of industry and agriculture had the highest PD value, which further confirmed that F2 mainly comes from anthropogenic activities. Thus, F2 is mainly controlled by intense industrial and urbanization activities.

F3 derived from PMF exhibited high clusters in the southern and southwestern rural areas. The kriging interpolation map of F3 was also in accordance with the Getis–Ord Gi* analysis results (Fig. 4). In addition, the variogram model for F3 had strong spatial autocorrelation and a short dependence range, indicating nonpoint source pollution. Finally, agricultural production had the highest and most significant spatial correlation with F3 in the GeogDetector model. In terms of interactions effects, the combined influence of soil parent material and agriculture had the highest PD value. Hence, F3 is associated with agricultural activities such as sewage irrigation and fertilizer application.

Regarding the contribution of various pollution sources to heavy metals, 83.2% of the heavy metals in the soil were ascribed to anthropogenic sources, and 16.8% were attributed to natural sources. Among the anthropogenic sources, agriculture accounted for 63.4% of the total pollution, and industry was responsible for 19.8%. The source determination of each heavy metal indicated that Cr was mainly from soil parent materials, Cd was closely associated with agricultural activities such as sewage irrigation and fertilizer application, Hg was mainly attributable to industrial activities such as coal mining and smelting. As was related to natural and agricultural sources, and Pb was related to natural and industrial sources.
4. Conclusions

This study developed a new combined spatial method for quantitative source apportionment and definition of heavy metal pollution in soils through various auxiliary data and performed a case study in the city of Hangzhou, China. Although the mean concentrations of As, Cd, Cr, Hg, and Pb in the soils were lower than the National Environmental Quality Standards for Soils in China (GB15618-2018), the mean contents of As and Cd were higher than their corresponding local background values by approximately 1.31 and 1.59 times, respectively, indicating that heavy metals were enriched in the topsoil through human activities. F1 derived from the PMF model showed strong spatial autocorrelation, a long model range and the highest spatial correlation with soil parent material; thus F1 was defined as natural sources. F2 (high values were clustered in suburban areas where industrial development and urban expansion were intense) obtained from the PMF model showed moderate spatial heterogeneity, a short dependence range and the highest spatial correlation with industrial production; hence, F2 was mainly controlled by intense industrial and urbanization activities. F3 (high values were clustered in the southwestern rural areas) extracted by the PMF model revealed strong spatial autocorrelation, a short dependence range and the highest spatial correlation with agricultural production; thus, F3 could be attributed to agricultural activities such as sewage irrigation and fertilizer application. Agricultural activities, industrial activities, and natural sources accounted for 63.4%, 19.8%, and 16.8% of the total heavy metal accumulation in the soils, respectively. Agricultural activities were the main source of Cd (73.68%) and provided a portion of As (57.83%); industrial activities dominated the contribution of Hg (92.38%) and provided a portion of Pb (42.42%). In addition, Cr (80.72%) and a portion of As (35.56%) and Pb (41.83%) were closely related to the soil parent materials. These results suggest that anthropogenic activities have strong influences (83.2%) on heavy metal accumulation and distribution in the soils of the study area, which calls for the prevention and control of heavy metal pollution in the region.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.141293.

CRediT authorship contribution statement


Declaration of competing interest

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

Acknowledgements

This work was partially supported by the National Key Technology Research & Development Program of China (2018YFD0200500) and the National Natural Science Foundation of China (41801302 and 41671399). The foundation had no involvement in the design of the study, collection of the data, analysis or interpretation of the results, writing of the report, or the decision-making process for submission of the article for publication.

References


Yang, S., Zhou, D., Yu, H., Wei, R., Pan, B., 2013. Distribution and speciation of metals (Cu, Zn, Cd, and Pb) in agricultural and non-agricultural soils near a stream upriver from the Pearl River, China. Environ. Pollut. 177, 64–70.


