CrossMark

Spatial network analysis of surface soil pollution from heavy metals and some other elements: a case study of the Baotou region of China

Xiaodan Han [1,2] • Huajiao Li [1,2] (iD) • Meixia Su [3] • Pengli An [1,2]

Abstract

Purpose With the development of heavy industry, urban soil suffers serious pollution, which threatens the sustainable development of cities. Understanding the spatial distribution characteristics of surface soil pollution aids in pollution prevention and control and promotes sustainable development.

Materials and methods We use China's Baotou as an example. Based on the data of 2820 sampling points in main urban areas and some suburban areas of Baotou, we constructed a relationship network model for sampling points in surface soil by using the complex network method. We combined the network method with spatial geographic information to analyze the spatial agglomeration characteristics of the surface soil pollution in Baotou China.

Results and discussion Sampling points at Dalahai Village (including 506D, 538B, and 538D) and Hayenaobao Village (including 509C, 541A, and 541C), Puerhantu Town within the Kundulun District have the most serious pollution problems, and they are all concentrated in the tailings dam. Sampling points 328D and 544A are scattered in the Leng Community, Kunhe Town, Kundulun District and Changhan Village, Haringer Township, Jiuyuan District, but they have a close co-anomaly relationship with the tailings dam. We suggest that these areas should be unified to give priority to pollution control. There is an obvious difference for Al_2O_3, B, Hg, and U, which are abnormal in the power plant ash storage pools, but normal at the tailings dam. Consequently, pollution control for power plant ash storage pools needs to be different from pollution control at the tailings dam. Sampling points at the Fengying Community (including 580A and 580B), Kunhe Town, Kundulun District, and Gaoyoufang Village (579D and 643B), Rare Earth Road, Qingshan District as well as other sampling points upstream of the Kundulun River have a close co-anomaly relationship with the tailings dam. It is necessary to strengthen the purification treatment of sewage upstream of the Kundulun River to reduce the spread of pollutants.

Conclusions These results provide a theoretical basis for the government to formulate specific cross-regional collaborative governance measures.

✉ Huajiao Li
babyproud@126.com

1. China University of Geosciences, Beijing 100083, China

2. Key Laboratory of Carrying Capacity Assessment for Resource and Environment, Ministry of Land and Resources, Beijing 100083, China

3. Geological Survey Institute of Inner Mongolia Autonomous Region, Hohhot 010020, Inner Mongolia, China

1 Introduction

As part of the Earth's pedosphere, urban soil has achieved certain ecological, environmental, and economic functions, working as both the source and sink of urban pollutants (Rodriguez-Seijo et al. 2017; Yang et al. 2017). Therefore, the pedosphere is related to the quality of urban ecological environments and human health (Simon et al. 2013; Zhang et al. 2015). With the rapid development of industrialization and urbanization, soil pollution caused by metals and metalloids is a serious environmental problem especially heavy

metal pollution in the surface soil (Wei and Yang 2010; Yaylali-Abanuz 2011; Hu et al. 2013; Rodriguez-Seijo et al. 2017). Metal and metalloid contamination in urban local areas spreads through water and air, and eventually precipitates in other soil regions (Chen et al. 2005; Ajmone-Marsan and Biasioli 2010; Bai et al. 2016; Zohar et al. 2017). Therefore, it is necessary to analyze the spatial distribution of soil pollution. Moreover, it is critical to identify soil regions with the same or similar pollution problems for collaborative governance of trans-regional soil pollution.

Currently, studies on the spatial distribution of heavy metals in urban soil focus on the analysis of heavy metal content in soil samples from different land use patterns (Fu and Wei 2013; Tume et al. 2014; Zhang et al. 2015; Perez-Sirvent et al. 2016; Zhao and Hazelton 2016; Bourliva et al. 2017; Zong et al. 2017). However, these studies lack detailed research on the spatial agglomeration characteristics of urban soil pollution. Literature on trans-regional pollution involving collaborative governance is primarily explanatory. In addition, these studies focus on researching large regional backgrounds from a macro perspective. These studies mainly research the importance of cross-regional collaborative governance (Bodin 2017), theoretical benefits and constraints (Bodin et al. 2016), and the importance of leadership (Emerson et al. 2012). However, few studies are based on a regional-scale background. Furthermore, there is a lack of research on identifying specific pollution areas that need to be coordinately controlled.

The analysis of geochemical elements in urban soil contains abundant geological information. Studying the abnormal content of chemical elements in soil is an important basis for establishing environmental quality standards as well as pollution remediation targets. The relationship between soil sampling points and geochemical elements is a complex network. Based on geochemical element content data in urban soil sampling points, this paper uses the complex network method to construct relationship networks among sampling points. The complex network method is an effective tool for describing the relationship between entities and entity status (Fan et al. 2016; Wang et al. 2016), and it can also analyze the relationship among soil sampling points. Specifically, the community division algorithm in complex networks can divide networks into different communities. The links between nodes within one community are relatively tight, and the connections between different communities are relatively sparse (Barigozzi et al. 2011; Zhong et al. 2014). Thus, we analyzed the spatial agglomeration characteristics of soil pollution by using the community division method. On this basis, we summarized the spatial distribution characteristics of soil pollution and identified the soil regions with same or similar pollution problems.

Baotou City, located in the western part of the Inner Mongolia Autonomous Region of China, is one of the earliest industrial bases in ethnic minority areas of China. With the development of industrialization and urbanization, Baotou has become an important industrial city with metallurgical, mechanical, and electric power; meanwhile, the urban soil has become seriously polluted. The object of this paper is to study the surface soil in the main urban areas of Baotou and some suburban areas. This paper combines basic geochemical analysis with the complex network method. Furthermore, we analyze the spatial agglomeration characteristics of Baotou surface soil pollution. Finally, we identify soil regions with the same or similar pollution problems and summarized the spatial distribution characteristics of soil pollution.

2 Materials and methods

2.1 Study area, data, and preparation

The scope of the sampling area is focused on the main urban zone and some suburban areas in and around Baotou. The latitude and longitude for the sample collection area are as follows: north latitude 40° 30′–40° 45′, east longitude 109° 34′–110° 12′. The 1:50000 soil bulk density measurement area is 705 km^2. The sampling density is 4 points/km^2. The number of samples is 2820. The sample collection area primarily covers a steel and iron industrial park, the Qingshan District, the Kundulun District, the Donghe District, the Kundulun River, the Sidaosha River, and some suburban areas. The steel and iron industrial park, located west of the Kundulun District in Baotou, covers an area of 63.94 km^2. On the west side of the Kundulun River is a steel factory, and a residential area is located on the east bank of the Kundulun River. Southwest of the steel and iron industrial park is a tailings dam, which covers an area of 11 km^2. The tailings dam is a "rare earth lake," containing approximately 1.35 t of tailings and approximately 15 million cubic meters of water. The Kundulun River is the main industrial sewage river, and the Sidaosha River is the main living sewage river in Baotou. A description of the soil sample collection method is as follows: the sampling depth is 20 cm, the original weight of each sample is approximately 500 g, and every 4 samples were collected at one field number and marked as filed number-A/B/C/D.

The data are derived from a subsidiary of the Geological Survey of China. Data include map numbers, field sampling point coordinates, field numbers, and 26 analysis indexes. The analytical indicators include pH and 25 geochemical elements. Advanced instrument and a reasonable supporting analysis method are used in soil sample multi-element analysis and testing. The main analytical instruments are X-ray fluorescence (XRF), inductively coupled plasma mass spectrometry (ICP-MS), and inductively coupled plasma optical emission spectrometry (ICP-OES). The most important elemental analysis methods are shown in Table 1. We strictly monitor the quality of various samples by standard reference materials, internal laboratory inspections, password checks, and external inspections.

Table 1 Measurement instruments and methods for determining soil geochemistry (Liao et al. 2012)

Element	Instrument	Methods
As, (Ce), (Cr), (Cu), (La), (Mn), Nb, Ni, S, P, (Pb), (Y), Al$_2$O$_3$, Fe$_2$O$_3$, K$_2$O	XRF	Use powder samples to prepare tablet preparation and measure samples directly.
Cd, (Ce), (Co), (La), (Pb), Th, U, (Y)	ICP-MS	Decompose samples with hydrofluoric acid, nitric acid, and perchloric acid. Dissolve samples in aqua regia and transfer it to a polyethylene tube. A portion of the clear solution was diluted with nitric acid. Then, use ICP-MS to test element content.
(Co), (Cr), (Cu), (Mn), Zn	ICP-OES	Decompose samples with nitric acid, hydrofluoric acid, and perchloric acid. Dissolve samples with aqua regia, and transfer it to a 25-ml plastic tube. Then, use ICP-OES to test element content.

The bracketed elements are compared using two analytical methods to verify and ensure the accuracy of the analytical data

We randomly selected 5% of the samples for repeated analysis. The pass rate is 100%. External inspection results showed that the relative error of most elements is less than 20%.

To make the data processing more accurate and convenient, we performed a pretreatment on the original data, as is explained below.

The pH index data is not complete. We chose 25 indicators, excluding pH, for analysis. These 25 indicators include 17 metal elements (Al$_2$O$_3$, Cd, Ce, Co, Cr, Cu, Fe$_2$O$_3$, K$_2$O, La, Mn, Mo, Ni, Pb, Th, U, Y, Zn) and 8 metalloid elements (As, B, C, F, Hg, P, S, Se). Units for Fe$_2$O$_3$, Al$_2$O$_3$, and K$_2$O are 10^{-2}, and units for the others are 10^{-6}. The original data $A(a_{ij})$:

$$A = \begin{bmatrix} a_{1,1} & \dots & a_{1,25} \\ \vdots & a_{ij} & \vdots \\ a_{2820,1} & \dots & a_{2820,25} \end{bmatrix},$$

where a_{ij} represents the content of element j at sampling point i.

Currently, there are three different evaluation guidelines for soil pollution (Chen et al. 2011; Xia and Lou 2006). The first one is the lower limit for environmental soil anomalies. The second one is the environmental soil quality standard (GB15618-1995). The third one is the threshold value for soil pollution. The GB15618-1995 soil environmental quality standard only contains eight types of heavy metals and two types of organics. In addition, because of the great variety of soil in China, it is difficult to realistically reflect the local soil contamination by a unified national standard value. The lower limit value for environmental soil anomalies is critical in revealing whether there is pollution in soil. If the chemical substance content in soil is higher than this limit, we must be alert (Xia and Lou 2006). The threshold value for soil pollution, which can be determined through a risk assessment of polluted soil, is critical for determining whether or not the soil is contaminated. This paper primarily studies the abnormal situation of the surface soil in Baotou. Therefore, we chose the lower limit for the environmental soil anomalies to preliminarily analyze the soil pollution situation.

Baotou is one of the cities in the Hetao area of Inner Mongolia. The soil types, pH levels, and the geographical environment of Baotou are similar to the entire Hetao area, and climate conditions are the same. Currently, there is no unified geochemical background value of surface soil in Baotou. Therefore, this paper uses the anomaly lower limit value calculated from the surface soil characteristics of the Hetao area in Inner Mongolia to evaluate the soil pollution situation in Baotou. In this paper, we adopted the traditional method for calculating the lower limit anomaly (Galuszka 2007; Matschullat et al. 2000). We used the average (X) and standard deviation (S) of element contents in the surface soil of the Hetao area to calculate $X \pm 2 \times S$. We used these two values as upper and lower iterations to eliminate data until outliers can be eliminated. Furthermore, we used remaining data as background data to calculate the mean and standard deviation. The formula for calculating the lower limit anomaly is as follows (Matschullat et al. 2000; Galuszka, 2007):

$$M \text{ (the lower limit of anomaly)} = X + 2 \times S. \quad (1)$$

2.2 Spatial analytical methods

2.2.1 Construction of co-anomaly and threshold networks

Co-anomaly network: We selected the lower limit of geochemical anomalies in the Hetao, Inner Mongolia area as the standard value (as is shown in Table 2). Based on these standard values, elemental content from surface soil sampling points above the lower limit anomaly was screened out. Normal is represented by "0." Element content higher than the lower limit anomaly is represented by "1". "Anomaly" is referred to as "1". If there is no special explanation, anomaly in this paper means that the element content is higher than the lower limit anomaly. If the content of one or more elements is anomaly at one sampling point, we call it abnormal sampling point. The original data

Table 2 The lower limit of geochemical anomalies in the Hetao, Inner Mongolia area

Element	As	Al$_2$O$_3$	B	C	Cd	Ce	Co	Cr	Cu	F	Fe$_2$O$_3$	Hg	K$_2$O
M	16.48	13.54	71.72	2.87	0.21	96.80	16.15	81.36	33.19	854.42	5.54	0.07	2.65
Element	La	Mn	Mo	Ni	P	Pb	S	Se	Th	U	Y	Zn	
M	50.18	772.98	1.00	40.11	984.38	25.95	1114.68	0.30	14.35	3.60	28.32	91.54	

Units for Fe$_2$O$_3$, Al$_2$O$_3$, and K$_2$O are 10^{-2}; the others are 10^{-6}

matrix A is transformed into a 0–1 matrix $B(b_{ij})$. The matrix B represents the elemental anomaly situation.

$$b_{ij} = \begin{cases} 0, & \text{content of element } j \text{ is normal at coordinate } i \\ 1, & \text{content of element } j \text{ is abnormal at coordinate } i \end{cases} \quad (2)$$

$$c_{ij} = \begin{cases} 0, & \text{there is no common abnormal element between sampling points } i \text{ and } j \\ x, & \text{there are } x \text{ common abnormal elements between sampling points } i \text{ and } j \end{cases} \quad (4)$$

where B' is the transposed matrix of B, matrix C is a symmetric matrix, and the data of the diagonal represent the number of abnormal elements at one coordinate point. Because this paper studies the relationship between different abnormal sampling points, we only retained the data above the diagonal. Thus, we obtained the final co-anomaly matrix $C(c_{ij})$. Sampling points are as nodes. If there are common abnormal elements between sampling points, we define that there are co-anomaly relationships between these sampling points. Co-anomaly relationships for sampling points are as edges. In addition, the number of common abnormal elements between sampling points is as edge weight. Thus, the undirected weighted co-anomaly network was constructed.

Threshold network We defined the number of common abnormal elements between abnormal sampling points as the intensity of the co-anomaly relationship between sampling points. The more common abnormal elements there are between two sampling points, the stronger the co-anomaly relationship. In contrast, the fewer common abnormal elements, the weaker the co-anomaly relationship. To observe the characteristics of co-anomaly networks more clearly, we used the minimum number of common abnormal elements as the threshold. We started the network at the 0 node and gradually added edges in order of edge weight (from large to small). We screened co-anomaly networks with thresholds of 18, 17…1 individually. The threshold of $\tau = i$ denotes edge weight greater than or equal to i.

2.2.2 Analysis of network topology characteristics

Average weighted degree In weighted networks, the weighted degree is used to measure the correlation strength between one node and other nodes. In this paper, the average weighted degree $R_{avg}(w)$ indicates how many common abnormal

By applying the element anomaly situation matrix B to formula (3) (Li et al. 2014), we obtained the co-anomaly matrix $C(c_{ij})$.

$$C = BB' \quad (3)$$

elements exist between one sampling point and other sampling points in the Baotou surface soil, on average. The formula is as follows:

$$R_{avg}(w) = \frac{1}{n} \sum_{j=1}^{n} e_{ij} w_{ij} \quad (5)$$

The variable e_{ij} is the connection property between nodes i and j. The variable w_{ij} is the edge weight between nodes i and j.

Average weighted clustering coefficient: In this paper, the co-anomaly network is an undirected weighted network. The average weighted clustering coefficient represents the close degree of co-anomaly relationships among surface soil sampling points. The greater the weighted clustering coefficient of one node, the closer the co-anomaly relationship is among other sampling points connected with that sampling point. The formula is as follows (Barrat et al. 2004; Gao et al. 2011):

$$C_i = \frac{1}{s_i(r_i-1)} \sum_{j,k} \frac{w_{ij} + w_{ik}}{2} e_{ij} e_{ik} e_{jk} \quad (6)$$

$$C_i' = \frac{1}{n} \sum_i C_i \quad (7)$$

C_i is the weighted clustering coefficient of node i. The variable C_i' is the average weighted clustering coefficient for co-anomaly networks. The variable r_i is the degree of node i. The variable s_i is the point intensity value of node i in co-anomaly networks. Finally, $e_{ij} e_{ik} e_{jk}$ represents whether nodes i, j, and k can constitute a triangle. If the value is 1, it means that the three nodes are connected to each other, and they can form a triangle. If the value is 0, it means that the three nodes cannot form a triangle.

Community analysis: Community division is used to divide the nodes into groups. Nodes in a group are densely

connected, and links between groups are sparse. We chose a heuristic method to divide the network into communities. The algorithm, based on the modular variables, can measure density within the community and between communities. Q represents the degree of modularity, and its value is between -1 and 1. The equation is as follows:

$$Q = \frac{1}{2m} \sum_{i,j} \left[w_{ij} - \frac{A_i A_j}{2m} \right] \delta(C_i, C_j),$$ (8)

where $A_i = \sum_i w_{ij}$ is the sum of the weights for all edges connected to node i. If $C_i = C_j$, $\delta(C_i, C_j) = 1$; otherwise, $\delta(C_i, C_j) = 0$. The sum of weights for all connections in the network is represented by $m = \sum_{i,j} w_{ij}$.

In this paper, community division of co-anomaly networks is divided into two processes. First, we assumed each node i is a community. For every node i, when that node moved to its adjacent community j, we calculated the increment of module degree with the ΔQ algorithm (Blondel et al. 2008). If ΔQ is negative, node i remains in the initial community. If ΔQ is positive, node i moves to adjacent community j. In addition, the maximum value of ΔQ was calculated. The process was repeated until there was no further improvement for all nodes. In the second phase, we constructed a new network whose nodes are the communities found in the first stage. The weight between two different communities is the sum of the number of links between nodes from those two communities. In the new network, edges between nodes from the same community are considered self-loops. After the second stage, the first phase was re-applied to the new network. The two phases iterate until there are no new changes. Thus, the modularity is maximized.

$$\Delta Q = \left[\frac{\sum c_{in} + A_{i,in}}{2m} - \left(\frac{\sum tot + A_i}{2m} \right)^2 \right] - \left[\frac{\sum in}{2m} - \left(\frac{\sum tot}{2m} \right)^2 - \left(\frac{A_i}{2m} \right)^2 \right]$$ (9)

The sum of weights for all links inside community C is $\sum c_{in}$. $\sum tot$ is the sum of weights for links adjacent to all nodes in community C. $A_{i, in}$ is the sum of the weights for links from i to all nodes in community C.

In this paper, we used the community method to analyze spatial agglomeration characteristics of co-anomaly relationships among surface soil sampling points in Baotou. Furthermore, we identified areas with similar pollution problems in the surface soil of Baotou.

2.2.3 Spatial analysis of co-anomaly networks with different thresholds

The intensity of the co-anomaly relationship among sampling points in Baotou may be related to the distribution area and distance between sampling points. We first analyzed the relationship between anomaly intensity and distance of co-anomaly sampling points. According to the distance formula

(10), we calculated the distance between sampling points and constructed the space distance matrix $D(d_{ij})$.

$$d = \sqrt{\left(x_i - x_j \right)^2 + \left(y_i - y_j \right)^2}$$ (10)

$$D = \begin{bmatrix} d_{1,1} & \cdots & d_{1,2820} \\ \vdots & d_{i,j} & \vdots \\ d_{2820,1} & \cdots & d_{2820,2820} \end{bmatrix}$$

(x_i, y_i) and (x_j, y_j) are coordinates of sampling points i and j. Variable $d_{i, j}$ is the distance between sampling points i and j. We analyzed the relationship between the number of common abnormal elements and the spatial distance of co-anomaly points.

Because there are many abnormal sampling points in surface soil, the abnormal intensity of sampling points is different. It is difficult to clearly identify the spatial distribution characteristics of the co-anomaly relationship among abnormal sampling points by analyzing the entire co-anomaly network. Therefore, we used the Spearman correlation coefficient formula (11) to calculate correlation coefficients between the number of co-anomaly elements and the spatial distance of abnormal sampling points. The Spearman rank correlation method is used to analyze whether there is correlation between levels x_i and y_i. We used the Spearman correlation coefficients to determine whether there is a correlation between two variables by examining whether the two variables (X and Y) are synchronous. Variables X and Y were sorted (ascending or descending order, respectively) to get two sets of elements x and y. The elements in set x subtracted the elements in set y correspondingly. In this way, we generated a list of differences (set d, where $d_i = x_i - y_i$). The Spearman correlation coefficient formula is shown below (Gauthier 2001):

$$\gamma = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N(N^2 - 1)}$$ (11)

Threshold networks with large correlation coefficients and strong variation can show clear spatial distribution of anomalous relationships among abnormal sampling points. Therefore, on the basis of the value and changing situation of correlation coefficients, we selected threshold networks with clear spatial distributions for further research.

3 Results and discussion

3.1 Overall analysis of surface soil pollution in Baotou

The co-anomaly networks constructed in this paper represent the co-anomaly relationship between abnormal sampling points. There are 2820 sampling points. As is shown in Table 3, there are 1953 sampling points with co-anomaly

Table 3 The co-anomaly network index statistics

Network	The number of nodes	The number of edges	Average degree	Average weighted degree	Average weighted clustering coefficient
The threshold 1	1953	1,145,486	1173.053	2371.564	0.894
The threshold 11	176	4515	51.307	636.045	0.837

relations in the threshold 1 network, which make up approximately 69.25% of points. According to the average degree of the threshold 1 network, each surface soil sampling point in Baotou has co-anomaly relationships with, on average, 1173 sampling points. The above indicates that there are common abnormal elements among most sampling points in the Baotou surface soil, and the sampling points with co-anomaly relationships are relatively common. However, according to the average weighted degree of the threshold 1 network, we know a sampling point, on average, has 2371 abnormal elements in common with other sampling points. In other words, on average, one sampling point has 2 abnormal elements in common with another sampling point. This implies that co-anomaly relationships among sampling points has a weak intensity. In contrast, as is shown in Table 3, Fig. 1, and Fig. 2, there are 4515 edges in the threshold 11 network, accounting for only 0.39% of edges in the threshold 1 network. That is, co-anomaly relationships between sampling points with a co-anomaly intensity greater than 11 only account for 0.39% of all co-anomaly relationships in the Baotou surface soil. The above signifies that although the sampling points with co-anomaly relationships are relatively common in Baotou surface soil, the sampling points with high-intensity anomalies are only in some local areas. The results are in agreement with Liao and Xu as the overall pollution of surface soil in Baotou
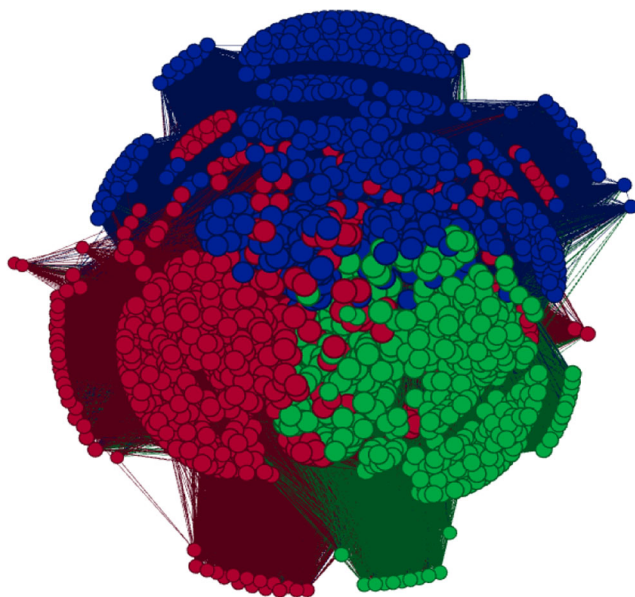
Fig. 1 The threshold 1 co-anomaly network

is not extensive. Serious soil pollution is primarily distributed in some local areas (Xu et al. 2011; Liao et al. 2012).

The weighted clustering coefficient of a node reflects the closeness degree among neighborhoods for this node. The closer the neighbor nodes are, the higher the clustering coefficient is for the node (Gao et al. 2011; Li et al. 2014). As is shown in Fig. 3, the average weighted clustering coefficient in co-anomaly network in Baotou is relatively high. The average weighted clustering coefficient of the threshold 18 network is 0.643, and the average weighted clustering coefficients for other threshold networks are all between 0.75 and 0.90. This reveals that there are close links among abnormal sampling points in the surface soil in Baotou. Once pollution spreads from an abnormal sampling point to its connected sampling points, then pollution may generally spread among those neighbor sampling points, resulting in a wider range of soil contamination.

On the basis of understanding the overall situation of the surface soil in Baotou, we identified key nodes in co-anomaly networks to further analyze soil pollution. In co-anomaly networks, the sampling points with large weighted degrees have important information that indicates soil pollution. The greater the weighted degree of a node is, the greater the pollution influence of a sampling point. If some pollutants are detected in some sampling points with large weighted degrees, then areas with co-anomaly links with those sampling points are likely to be contaminated. In co-anomaly networks, nodes with larger weighted degrees are mainly distributed in steel and iron industrial parks, as well as power plant ash storage pools. Liao et al. studied the spatial distribution characteristics of heavy metals such as Cd and Hg in the Baotou surface soil. Their results confirm that the local area mainly dominated by smelting is significantly polluted, and the environmental quality in the rest areas is good in general (Liao et al. 2012). Therefore, identifying sampling points with a greater impact can provide reference to a priority choice area for soil pollution prevention and control.

3.2 Identification of typical networks of surface soil pollution in Baotou

Figure 4 is a scatter plot showing the number of common abnormal elements and the spatial distance between co-anomaly sampling points. The following conclusions can be drawn from the figure: (1) the distances between abnormal sampling points with large-intensity co-anomaly relations are small. This indicates that surface soil areas with serious

Fig. 2 The threshold 11 co-anomaly network
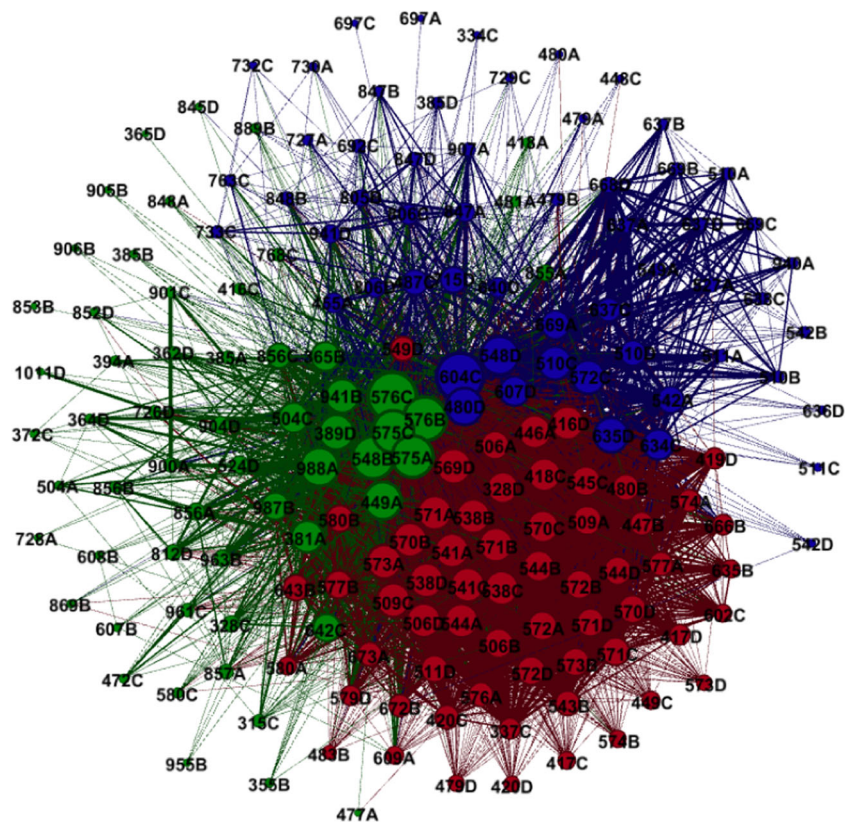
and similar pollution problems are relatively close in distance. Identifying these areas can help us focus on serious soil pollution in Baotou. (2) There are also co-anomaly relationships between the sampling points that are far away from each other. This indicates that there may be the same or similar pollution in different areas, although the distance between these areas is far. Urbanization rarely occurs homogenously across an entire watershed, resulting in spatially variable of runoff and differing contributions of contaminating metals (Tang et al. 2005; Zohar et al. 2017). Therefore, identifying these areas will help

us control soil uniformly. (3) In the case of low thresholds, the relationship between the number of common abnormal elements and the spatial distance of co-anomaly sampling points is almost identical. However, there is a clear break at threshold 14 in the scatter plot. This implies that the co-anomaly relationship between sampling points is very common in surface soil in Baotou, but this general co-anomaly relationship is only a low intensity co-anomaly. The high-intensity co-anomaly relationship is not universal in Baotou surface soil. The high-intensity co-anomaly networks are unique, and the

Fig. 3 The variation curve of average weighted clustering coefficients in co-anomaly networks with different thresholds
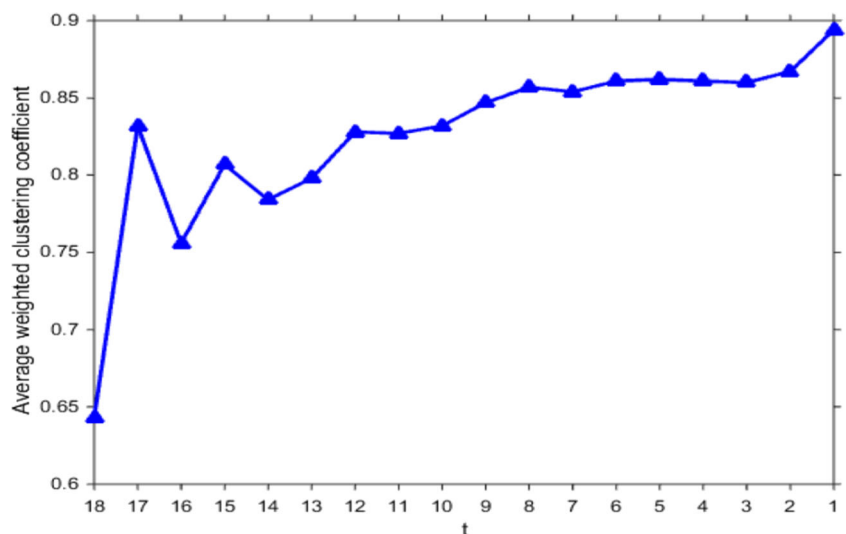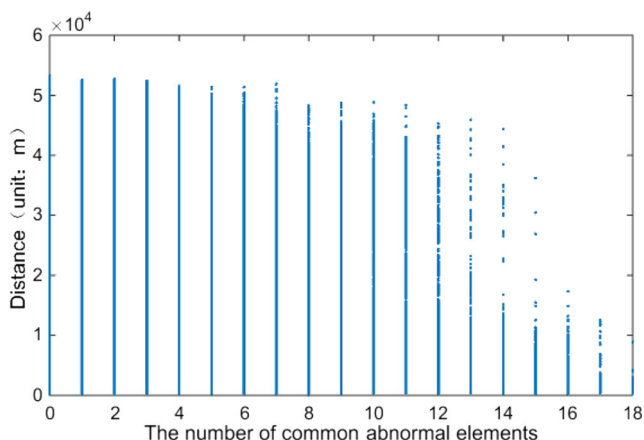
Fig. 4 Scatter plot showing the relationship between the number of common abnormal elements and the spatial distance

threshold networks. There is a negative correlation between the number of common abnormal elements and the spatial distance. In general, the farther the distance between sampling points, the fewer common abnormal elements. Conversely, the closer the distance, the higher number of common abnormal elements. Figure 5 shows that the absolute values of negative correlation coefficients for networks with threshold 17, 16, 15… 11 are larger than those for the other threshold networks. This indicates that the relationships between co-anomaly relationships and the spatial distances are relatively clear in threshold 17, 16, 15… 11 networks. The absolute values of the negative correlation coefficients for networks with a threshold of 10, 9… 1 are relatively small and stable. The above is consistent with the clear break phenomenon in Fig. 4. In addition, threshold 18, 17, 16… 11 networks represent strong co-anomaly relationships. As a result, we selected threshold 18, 17, 16… 11 networks to analyze the spatial agglomeration characteristics of surface soil pollution in Baotou.

regional distribution characteristics of high-intensity co-anomaly networks are more obvious. Consequently, we focused on the analysis of high-intensity co-anomaly networks.

We used the correlation coefficients between the number of common abnormal elements and the spatial distance to select unique threshold networks to do further research. By increasing the correlation coefficients, some connections would be filtered out and the number of connections among the points would be reduced (Tang et al. 2014). Threshold networks with large correlation coefficients and large variation may represent networks with obvious spatial distribution. We took the minimum number of common abnormal elements as the threshold. For different threshold networks, we calculated the correlation coefficients between the number of common abnormal elements and the distance. In addition, we further analyzed the variation of correlation coefficients for different threshold networks.

Figure 5 shows the relationships between the number of common abnormal elements and spatial distances in different

3.3 Spatial agglomeration characteristics of surface soil pollution in Baotou

Recently, many methods such as environmental magnetic methods and geostatistical analysis have been successfully applied as a proxy indicator to determine spatial distribution of soil pollution (Blundell et al. 2009; Zhang et al. 2015). In this paper, the co-anomaly relationship between sampling points signifies soil sampling points with the same or similar pollution problems. Analyzing the co-anomaly relationship between soil sampling points is helpful for us to identify areas with similar pollution problems. We analyzed the spatial agglomeration characteristics of co-anomaly networks with different thresholds, which allows us to explore the spatial distribution of co-anomaly relationships of sampling points in

Fig. 5 Variation curve of correlation coefficients between co-anomaly intensity and spatial distances in different threshold networks
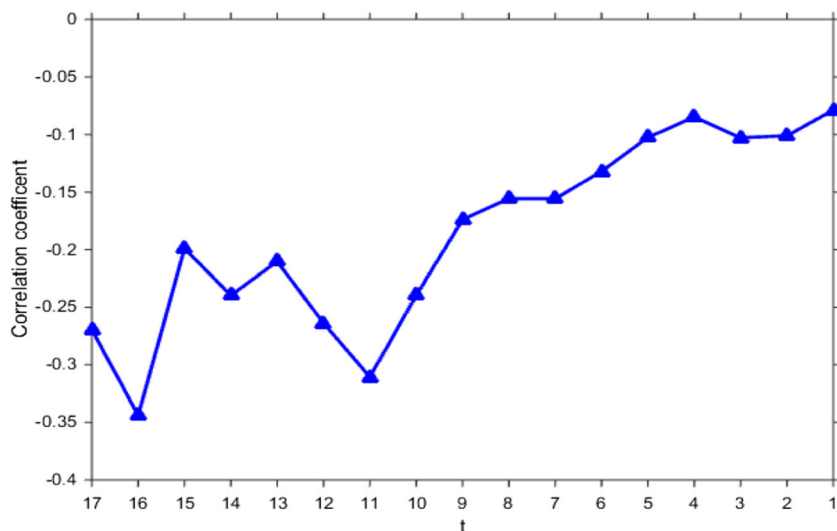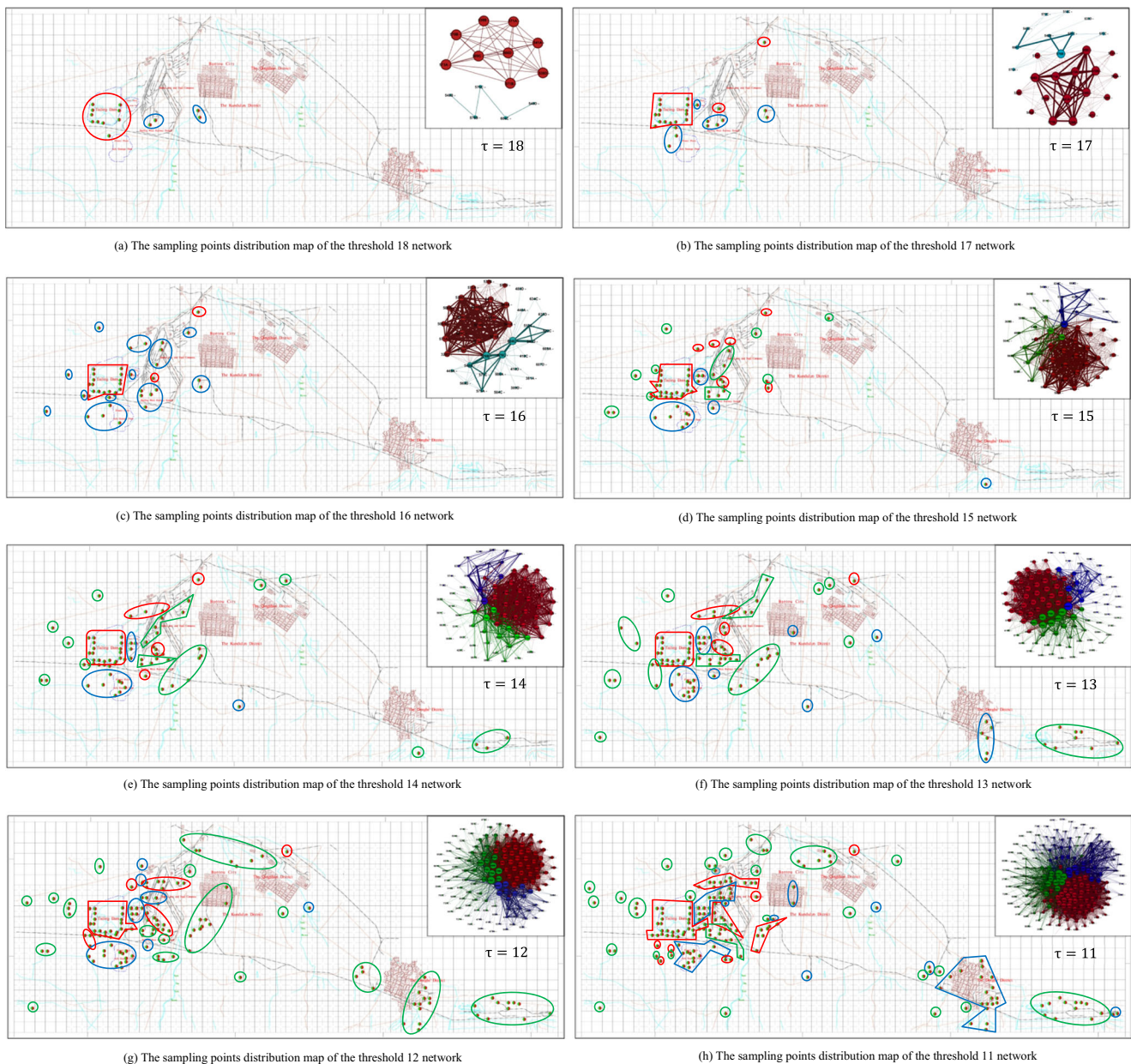
surface soil from different magnification angles. Thus, the co-anomaly network analysis could present an alternative way to extract information from large raw environmental datasets (Fan et al. 2016). On this basis, we identified areas with similar pollution problems in the surface soil of Baotou. Then, we took corresponding pollution control measures for areas with different degrees of pollution. Figure 6 shows the threshold 18, 17, 16… 11 networks and their spatial distribution maps.

The threshold 18, 17, and 16 networks are divided into two communities. The red nodes belong to community 1, and the blue nodes belong to community 2. In the threshold 18 network, nodes from community 1 are concentrated around the tailings dam. All nodes from community 1 have the same 18

abnormal elements: Cd, As, Mo, F, Se, S, P, C, La, Ce, Mn, Fe_2O_3, Ni, Cu, Zn, Pb, Th, and Y. This indicates that 506D, 509C, 538B, 538D, 541A, and 541C and the remaining sampling points around the tailings dam have the most serious and similar pollution problems. The tailings dam occupies a large area of land used to pile up waste. Some tailings are exposed directly to the air, resulting in a series of environmental problems (Li et al. 2011; Xu et al. 2011; Liao et al. 2012). We should focus on these areas to prevent the spread of pollution into other areas. In the threshold 17 and 16 networks, nodes from community 1 are not only concentrated in the tailings dam but also scattered in regions far away from the tailings dam, such as 328D and 544A. Although 328D and 544A are

(a) The sampling points distribution map of the threshold 18 network

(b) The sampling points distribution map of the threshold 17 network

(c) The sampling points distribution map of the threshold 16 network

(d) The sampling points distribution map of the threshold 15 network

(e) The sampling points distribution map of the threshold 14 network

(f) The sampling points distribution map of the threshold 13 network

(g) The sampling points distribution map of the threshold 12 network

(h) The sampling points distribution map of the threshold 11 network

Fig. 6 Sampling points distribution maps of different threshold networks (maps (a) to (h) are threshold 18 to 11 networks' spatial distribution maps)

far away from the tailings dam, these two nodes have 16 or 17 common abnormal elements with the nodes around the tailings dam. This indicates that the two sampling points have similar pollution problems to the tailings dam. Therefore, we should take these two sampling points into account during the treatment of seriously polluted areas around the tailings dam. In general, the tailings dam and its surrounding soil are enrichment areas of most elements, such as Pb, Zn, Cu, Ni, Cd, As, F, etc. (Xu et al. 2011; Liao et al. 2012). The tailings dam has serious, as well as similar, pollution problems, and should be given priority for treatment. In addition, to prevent the spread of pollutants to other areas, we should consider the two sampling points of 328D and 544A. It is inefficient to uniformly control polluted areas where sampling points are loosely distributed and sparsely connected to each other. Consequently, we propose a focus on the governance of core sampling points. For example, the weighted degrees of 576C, 576B, 604C, and 548B sampling points are relatively large and they have great pollution intensity. Moreover, those points are closely linked to the sampling points around the tailings dam, which suggests that we should first control these contaminated areas with limited funding.

The threshold 15, 14, 13, 12, and 11 networks are divided into three communities. The red nodes belong to community 1, the blue nodes to community 2, and the green nodes to community 3. Starting from the threshold 15 and 14 networks, nodes from community 2 initially show concentrated distribution areas, which are mainly concentrated in the two power plant ash storage pools. The remaining residue from coal burning in the power plant caused the enrichment of many elements. Previous studies have confirmed that waste from coal-fired power plants causes radioactive contamination of the surrounding soil (Mishra 2004). Common abnormal elements of sampling points in community 2 are primarily as follows: Al_2O_3, B, C, Cu, Ce, Cd, Hg, La, Mo, Pb, S, Se, Th, U, Y, and Zn. The distance between the ash storage pools and the tailings dam is small. The ash storage pools have many of the same abnormal elements as the tailings dam. Liao et al. confirmed that the contents of F, Zn, Ni, and Cu are all significant positive anomaly (Liao et al. 2012). Xu et al. concluded that due to thermal power generation in the power plants, S and Se elements agglomerated intensively in the surface soil as the atmosphere settles, resulting in a large area of soil pollution (Xu et al. 2011). However, sampling points at the ash storage pools are obviously different from sampling points at the tailings dam. Al_2O_3, B, Hg, and U are abnormal in the ash storage pools, but normal in the tailings dam. Therefore, pollution control for power plant ash storage pools needs to be different from pollution control at the tailings dam. In the threshold 15 and 14 networks, nodes from community 1 are not only concentrated around the tailings dam, but some nodes do not have a concentrated distribution area, such as 328D, 416B, 418C, 446A, 544A, 544B, 544D, and 607D. Those

sampling points are in close contact with the tailings dam. Therefore, we should consider those sampling points for the unified governance of the tailings dam. In the threshold 15 and 14 networks, nodes in community 3 have no concentrated distribution area, but sampling points 576C and 576B are closely related to sampling points in the tailings dam and the ash storage pools. These sampling points have high pollution intensity, and we should give priority to the treatment of these contaminated areas with limited funding. Starting from the threshold 13, 12, and 11 networks, the nodes from community 1 are not only distributed around the tailings dam but also distributed upstream of the Kundulun River. Sampling points 579D, 580A, 580B, 643B, and the remaining sampling points have many abnormal elements in common with the sampling points around the tailings dam. Additionally, the common abnormal elements are primarily As, Cd, Cu, F, Mo, Se, Mn, Pb, and Zn. The enrichment of these elements in surrounding surface soil is related to the sewage from industry (Liao et al. 2012; Xu et al. 2008). Therefore, when we control soil pollution in the tailings dam, we consider the sampling points upstream of the Kundulun River. In the threshold 13, 12, and 11 networks, the nodes from community 2 are concentrated not only in the power plant ash storage pools but also in some areas of the Donghe District (e.g., 732C, 733C, 805B, 806C, 806D). Soil pollution in the Donghe District is primarily caused by an aluminum plant and a brick factory (Sun et al. 2016). Traffic emission and industrial plants are the two main sources of heavy metal emission in urbanized areas (Chen et al. 2010; Dayani and Mohammadi 2010). It has been proved that the main enrichment elements are Cu, Pb, Zn, Cd, Hg, and F in the Baotou residential area (Liao et al. 2012). The major common abnormal elements of sampling points in the Donghe District are Cd, Hg, and Pb, and the enrichment of these elements is primarily related to coal combustion, automobile exhaust emissions, and other human activities (Liao et al. 2012). Finally, the remainder of the sampling points, which are primarily the nodes from community 3, are scattered in a large range, including 576C, 576B, 548B, 575C, 988A, 504C, 449A, 389D, 381A, and 941B. The weighted degrees of these sampling points are large, which implies that these sampling points have a strong influence and need to be controlled and monitored vigorously.

4 Conclusions

The main purpose of this paper is to identify soil regions with common pollution problems. We analyzed co-anomaly relationships between surface soil sampling points in main urban areas and some suburbs of Baotou. On this basis, we constructed a co-anomaly network. We took the minimum number of common abnormal elements as a threshold and screened co-anomaly networks at different thresholds.

Combining the network community method with spatial geographic information, we analyzed the spatial distribution characteristics of co-anomaly networks with different thresholds. Furthermore, we analyzed the spatial agglomeration characteristics of surface soil pollution in Baotou from different threshold angles. Finally, we summarized the distribution area of sampling points that have serious and common pollution problems. The main research results are as follows:

1. In general, the overall amount of surface soil pollution in Baotou is not extensive. Serious soil pollution is primarily concentrated in the tailings dam, the power plant ash storage pools, the Kundulun River, the Donghe District, and the southeastern suburban area.

2. In this paper, we concluded that the sampling points from surface soil in Baotou, which have the most serious and common pollution problems, are mainly distributed in two states. Sampling points around the tailings dam, such as 506D, 509C, 538B, 538D, 541A, and 541C, have the most serious and similar pollution problems. The common abnormal elements in those sampling points are some heavy metals and some radioactive elements. With the discharge of industrial sewage, many abnormal elements and other pollutants will spread to other areas, which will seriously harm human health. Therefore, it is suggested that a unified priority governance should be carried out for these above sampling points that have common pollution problems and are concentrated in the tailings dam. We should investigate the tailings dam in detail and strengthen the environmental quality monitoring around the pollution source to fundamentally prevent further deterioration of the environment. In addition, we should particularly consider sampling points 328D and 544A, which are closely related to the sampling points of the tailings dam, to prevent the spread of pollution to other areas. In terms of the other scattered areas with common and serious pollution problems, we suggest selecting core sampling points for key governance. For example, the weighted degrees of 576C, 576B, 604C, and 548B sampling points are large. These sampling points have high pollution intensity. We should give priority of the limited funding to the treatment of these contaminated areas.

3. There are also some highly polluted areas in Baotou, such as the power plant ash storage pools, the Donghe District, and the Kundulun River. The distance between the ash storage pools and the tailings dam is small, and the ash storage pools have many of the same abnormal elements as in the tailings dam. However, sampling points at the ash storage pools are obviously different from sampling points at the tailings dam. Al_2O_3, B, Hg and U are abnormal in the power plant ash storage pools but normal in the tailings dam. Therefore, pollution control of the power plant ash storage pools needs to be different from the tailings dam. Sampling points 579D, 580A, 580B, 643B, and others distributed upstream of the Kundulun River have a close co-anomaly relationship with the sampling points at the tailings dam. Consequently, we should adopt a collaborative governance for the tailings dam and the Kundulun River. We should control the pollution source at the tailings dam, while strengthening the purification treatment of the sewage in the Kundulun River. In addition, soil pollution in the Donghe District is mainly caused by human activities such as automobile exhaust emissions and coal burning. Therefore, we propose to strengthen collaborative governance on automobile exhaust emissions and coal burning in the Donghe District.

In this paper, we performed research on soil contamination. We only analyzed the enrichment of elements in surface soil and did not study the dilution of elements. In the future research, we will consider a study on element dilution for a more comprehensive analysis.

References

Ajmone-Marsan F, Biasioli M (2010) Trace elements in soils of urban areas. Water Air Soil Pollut 213(1–4):121–143

Bai Y, Wang M, Peng C, Alatalo JM (2016) Impacts of urbanization on the distribution of heavy metals in soils along the Huangpu River, the drinking water source for Shanghai. Environ Sci Pollut Res 23(6):5222–5231

Barigozzi M, Fagiolo G, Mangioni G (2011) Identifying the community structure of the international-trade multi-network. Physica A 390(11):2051–2066

Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proc Natl Acad Sci U S A 101(11):3747–3752

Blundell A, Hannam JA, Dearing JA, Boyle JF (2009) Detecting atmospheric pollution in surface soils using magnetic measurements: a reappraisal using an England and Wales database. Environ Pollut 157(10):2878–2890

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech: Theory Exp. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bodin O (2017) Collaborative environmental governance: achieving collective action in social-ecological systems. Science 357(6352):659

Bodin O, Robins G, McAllister RRJ, Guerrero AM, Crona B, Tengo M, Lubell M (2016) Theorizing benefits and constraints in collaborative environmental governance: a transdisciplinary social-ecological network approach for empirical investigations. Ecol Soc. https://doi.org/10.5751/ES-08368-210140

Bourliva A, Papadopoulou L, Aidona E, Giouri K (2017) Magnetic signature, geochemistry, and oral bioaccessibility of "technogenic" metals in contaminated industrial soils from Sindos Industrial Area, Northern Greece. Environ Sci Pollut Res 24(20):17041–17055

Chen GG, Liang XH, Zhou GH, Zhang M, Lin CH (2011) Grade division method for soil geochemical contamination and its application. Geol China 38(6):1631–1639 (in Chinese)

Chen TB, Zheng YM, Lei M, Huang ZC, Wu HT, Chen H, Fan KK, Yu K, Wu X, Tian QZ (2005) Assessment of heavy metal pollution in surface soils of urban parks in Beijing, China. Chemosphere 60(4):542–551

Chen X, Xia X, Zhao Y, Zhang P (2010) Heavy metal concentrations in roadside soils and correlation with urban traffic in Beijing, China. J Hazard Mater 181(1–3):640–646

Dayani M, Mohammadi J (2010) Geostatistical assessment of Pb, Zn and Cd contamination in near-surface soils of the urban-mining transitional region of Isfahan, Iran. Pedosphere 20(5):568–577

Emerson K, Nabatchi T, Balogh S (2012) An integrative framework for collaborative governance. J Public Adm Res Theory 22(1):1–29

Fan XH, Wang L, Xu HH (2016) Characterizing air quality data from complex network perspective. Environ Sci Pollut Res 23(4):3621–3631. https://doi.org/10.1007/s11356-015-5596-y

Fu S, Wei CY (2013) Multivariate and spatial analysis of heavy metal sources and variations in a large old antimony mine, China. J Soils Sediments 13(1):106–116

Galuszka A (2007) A review of geochemical background concepts and an example using data from Poland. Environ Geol 52(5):861–870

Gao XY, An HZ, Liu HH, Ding YH (2011) Analysis on the topological properties of the linkage complex network between crude oil future price and spot price. Acta Phys Sin 60(6):068902 (in Chinese)

Gauthier TD (2001) Detecting trends using Spearman's rank correlation coefficient. Environ Forensic 2(4):359–362

Hu YN, Liu XP, Bai JM, Shih KM, Zeng EY, Cheng HF (2013) Assessing heavy metal pollution in the surface soils of a region that had undergone three decades of intense industrialization and urbanization. Environ Sci Pollut Res 20(9):6150–6159

Li HJ, Fang W, An HZ, Yan LL (2014) The shareholding similarity of the shareholders of the worldwide listed energy companies based on a two-mode primitive network and a one-mode derivative holding-based network. Physica A 415:525–532

Li YJ, Wang JY, Zheng CL, Shui LY, Yi M, Cai L (2011) Characteristics of heavy metal combined pollution in Baotou Tailing Dam and its neighboring regions. Metal Mine 419:137–148 (in Chinese)

Liao L, Liu HL, Su MX, Duan HL, Wang JF, Zhao LJ (2012) Geochemical characteristics of the soil from Baotou City, Inner Mongolia and its environmental assessment. Geol Explor 48(4):799–806 (in Chinese)

Matschullat J, Ottenstein R, Reimann C (2000) Geochemical background—can we calculate it. Environ Geol 39(9):990–1000

Mishra UC (2004) Environmental impact of coal industry and thermal power plants in India. J Environ Radioact 72(1–2):35–40

Perez-Sirvent C, Hernandez-Perez C, Martinez-Sanchez MJ, Garcia-Lorenzo ML, Bech J (2016) Geochemical characterisation of surface waters, topsoils and efflorescences in a historic metal-mining area in Spain. J Soils Sediments 16(4):1238–1252

Rodriguez-Seijo A, Andrade ML, Vega FA (2017) Origin and spatial distribution of metals in urban soils. J Soils Sediments 17(5):1514–1526

Simon E, Vidic A, Braun M, Fabian I, Tothmeresz B (2013) Trace element concentrations in soils along urbanization gradients in the city of Wien, Austria. Environ Sci Pollut Res 20(2):917–924

Sun P, Li YW, Zhang LK, Li YM, Wang WD, Yu WJ (2016) Heavy metal pollution in topsoil from the Baotou industry area and its potential ecological risk evaluation. Rock and Mineral Analysis 35(4):433–439 (in Chinese)

Tang JJ, Wang YH, Wang H, Zhang S, Liu F (2014) Dynamic analysis of traffic time series at different temporal scales: a complex networks approach. Physica A 405:303–315

Tang Z, Engel BA, Pijanowski BC, Lim KJ (2005) Forecasting land use change and its environmental impact at a watershed scale. J Environ Manag 76(1):35–45

Tume P, King R, Gonzalez E, Bustamante G, Reverter F, Roca N, Bech J (2014) Trace element concentrations in schoolyard soils from the Talcahuano, Chile. J Geochem Explor 147:229–236

Wang XB, Ge JP, Wei WD, Li HS, Wu C, Zhu G (2016) Spatial dynamics of the communities and the role of major countries in the international rare earths trade: a complex network analysis. PLoS One 11(5):e0154575. https://doi.org/10.1371/journal.pone.0154575

Wei BG, Yang LS (2010) A review of heavy metal contaminations in urban soils, urban road dusts and agricultural soils from China. Microchem J 94(2):99–107

Xia JQ, Lou YM (2006) Definition and three evaluation guidelines of soil contamination. J Ecol Rural Environ 22(1):87–90 (in Chinese)

Xu Q, Liu XD, Tang QF, Liu JC, Zhang LL (2011) A multi-element survey of surface soil and pollution estimate in Baotou city. Arid Land Geogr 34(1):91–99 (in Chinese)

Xu Q, Zhang LX, Liu SH, Liu XD, Zhang LJ (2008) Heavy metal pollution of surface soil and its evaluation of potential ecological risk: a case study of different functional areas in Baotou City. J Nat Dis 17(6):6–12 (in Chinese)

Yang JY, Yu F, Yu YC, Zhang JY, Wang RH, Srinivasulu M, Vasenev VI (2017) Characterization, source apportionment, and risk assessment of polycyclic aromatic hydrocarbons in urban soil of Nanjing, China. J Soils Sediments 17(4):1116–1125

Yaylali-Abanuz G (2011) Heavy metal contamination of surface soil around Gebze industrial area, Turkey. Microchem J 99(1):82–92

Zhang JJ, Wang Y, Liu JS, Liu Q, Zhou QH (2015) Multivariate and geostatistical analyses of the sources and spatial distribution of heavy metals in agricultural soil in Gongzhuling, Northeast China. J Soils Sediments 16(2):634–644

Zhao Z, Hazelton P (2016) Evaluation of accumulation and concentration of heavy metals in different urban roadside soil types in Miranda Park, Sydney. J Soils Sediments 16(11):2548–2556

Zhong WQ, An HZ, Gao XY, Sun XQ (2014) The evolution of communities in the international oil trade network. Physica A 413:42–52

Zohar I, Teutsch N, Levin N, Mackin G, de Stigter H, Bookman R (2017) Urbanization effects on sediment and trace metals distribution in an urban winter pond (Netanya, Israel). J Soils Sediments 17(8):2165–2176

Zong YT, Xiao Q, Lu SG (2017) Magnetic signature and source identification of heavy metal contamination in urban soils of steel industrial city, Northeast China. J Soils Sediments 17(1):190–203