

# Analysis of Soil Pollution Degree and Causes Based on Mathematical Model

Zeng Yunhui (Corresponding author)

College of Intelligent Science and Engineering, Jinan University

No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province, China

Tel: 86-136-977-21415 E-mail: yunhui@stu2015.jnu.edu.cn

Zhang Yuanbiao

College of Intelligent Science and Engineering, Jinan University

No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province, China

Tel: 86-137-5008-5338 E-mail: jnu-mcm@163.com

Huang Shusheng

College of Intelligent Science and Engineering, Jinan University

No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province, China

Tel: 86-158-2034-4259 E-mail: 16429633569@qq.com

Gong Hongfei

College of Intelligent Science and Engineering, Jinan University

No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province, China

Tel: 86-130-2570-6555 E-mail: ghf-2005@163.com

Chen Yilin

College of Intelligent Science and Engineering, Jinan University

No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province, China

Tel: 86-131-9220-0223 E-mail: 2283519758@qq.com

Shen Shiyue

College of Intelligent Science and Engineering, Jinan University

No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province, China

Tel: 86-131-7265-7008 E-mail: 1094439399@qq.com

Huang Li

College of Intelligent Science and Engineering, Jinan University

No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province, China

Tel: 86-159-4690-5549 E-mail: 3101191558@qq.com

Hu Wenjuan

College of Intelligent Science and Engineering, Jinan University

No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province, China

Tel: 86-159-1321-7124 E-mail: 15913217124@qq.com

Received: Sep. 15, 2018 Accepted: Sep. 27, 2018 Published: December 24, 2018

doi:10.5296/jee.v9i2.13655 URL: <https://doi.org/10.5296/jee.v9i2.13655>

## Abstract

Taking the heavy metal pollution in the urban surface soil as the object, this paper analyzes the degree of heavy metal pollution in different areas in the city and the main causes of pollution. Firstly, this paper uses Kriging interpolation method to increase the sample data together with Surfer software to draw the spatial distribution map of eight heavy metals, and then compares the single factor index method and the geological accumulation index-Nemero index method to make a comprehensive evaluation of the heavy metal pollution degree in different areas of the city. It is concluded that the pollution level in the area from slight to heavy is: mountain areas, park green areas, living areas, traffic areas, industrial areas. Then, the main comprehensive index of heavy metals is extracted by the principal component analysis, and the spatial distribution map of the main factors is drawn based on it. According to the spatial distribution map, the main cause of heavy metal pollution is the emission of automobile exhaust and industrial waste, which provides a reliable theoretical basis for the prevention and treatment of heavy metal pollution in the urban surface soil.

**Keywords:** heavy metal pollution, Kriging interpolation, Surfer, single factor index method, principal component analysis

## 1. Introduction

Heavy metal pollution in soil refers to excessive heavy metal concentration caused by excessive deposition of microscale metal elements in soil due to human activities. With the rapid development of urban economy and the continuous increase of urban population, the impact of human activities on urban environmental quality has become increasingly prominent. The verification of urban soil geological environment anomalies and how to apply the vast amount of data obtained from the verification to carry out urban environmental quality assessment as well as study the evolution pattern of urban geological environment under the influence of human activities have become the focus of attention. At present, plenty of studies have been carried out on the evaluation of heavy metal pollution levels in urban soil at home and abroad. However, overall the evaluation of academia for the degree of heavy metal pollution in urban soils is mainly through a single method such as the single factor index method, the geological accumulation index method and the Nemero index (Xu, 2008). Meanwhile it is only the surface analysis of the obtained pollution degree results, directly to the cause of pollution, not through the scientific and effective model for analysis and verification. For example, only the geological accumulation index method and Excel tool are used to study the degree of heavy metal pollution in the soil, and then the cause of pollution is directly obtained through the result of pollution degree (Xu et al., 2012), or only the single factor pollution index method is used to analyze the degree of heavy metal pollution in cities (Yue et al., 2012). Although some used single factor index method and Nimero comprehensive pollution index method to evaluate the pollution degree of soil heavy metal elements, the pollution reason is directly obtained according to the analysis result of pollution degree, without considering the correlation between heavy metals and using effective way to verify the conclusion (Gu, et al., 2013; Li, et al., 2014). In fact, the single evaluation method of urban soil metal pollution degree existing today has its own scope and defects. For example, the geological accumulation index can reflect the natural variation characteristics of heavy metal distribution, but only giving the degree of pollution of a single heavy metal in the sampling point; the Nemero index specifically considers the most polluting factor, avoiding the influence of subjective factors in the weight coefficient in the weighted process, but does not consider the natural distribution characteristics of the main metal distribution. Therefore, only using a single evaluation method to evaluate the degree of soil heavy metal pollution will inevitably result in an unobtrusive or inaccurate evaluation result. Besides, the spatial distribution of some heavy metals may have a certain correlation. Metal with greater correlation may have a certain relationship between source and origin, and related heavy metals may interfere with each other during analysis. Therefore, when we analyze the main reason of heavy metal pollution, the relationship between heavy metal pollution must be considered.

This paper, to evaluate the problem of heavy metal pollution in different regions, combined with the single factor index method and the geological accumulation index-Nemero index method, draws a comprehensive conclusion showing the degree of heavy metal pollution in different regions. At the same time, in this paper, the principal component analysis method is used to convert the eight heavy metal elements mentioned in the data into several linear

uncorrelated comprehensive indicators, which help to eliminate mutual interference between metal elements and simplify data processing. Finally, according to the spatial distribution map of the main factors, the main causes of heavy metal pollution are obtained, making up for the limitations of the previous studies on the degree and causes of heavy metal pollution in urban soil, and providing a reliable theoretical basis to prevent and control heavy metal pollution in urban surface soil.

## 2. Method

### 2.1 Data Acquisition and Assumptions

The original data of the research object of this paper is from the 2011 National College Students Mathematical Modeling Competition (China Society for Industrial and Applied Mathematics, 2011). The sampling points in the original data are from the main functional areas of living areas, industrial areas, mountain areas, main road areas and park green areas. The data sheet lists the location of the sampling points, the altitude and information of its functional areas, together with the concentration of the eight major heavy metal elements at the sampling point and the background values of the eight major heavy metal elements. To facilitate analysis and understandings, we make the following assumptions about the model:

- (1) The data given in the topic is true and authentic;
- (2) No consideration of the degradation of heavy metals;
- (3) Assuming that the probability of airflow occurring in all directions is the same, that is, the atmospheric propagation characteristic is considered to be spherical diffusion;
- (4) Ignore the effect of gravity on heavy metal ions.

Include in these subsections the information essential to comprehend and replicate the study. Insufficient detail leaves the reader with questions; too much detail burdens the reader with irrelevant information. Consider using appendices and/or a supplemental website for more detailed information.

### 2.2 Kriging Interpolation

The fact of Kriging interpolation (Zhou,2016) is using the original data of the regionalized variables and the structural characteristics of the variogram and making linear unbiased and optimal estimation of the unknown samples.

Due to trends and expectations of the unknown data, a regionalized variable valuation method of unknown expectations is chosen under ordinary Kriging estimation.

Here, The regionalized variable  $Z(x)$  consists of expectation  $m$  and residual  $e(x)$ :

$$Z(x) = m + e(x) \quad (1)$$

Among them, expectation  $m$  is unknown and the expectation of residual  $e(x)$  is 0.

The second-order stationary of the variable  $Z(x)$  is assumed to be:

$$E[Z(x+h) - Z(h)] = 0 \quad (2)$$

$$\text{Var}[Z(x+h) - Z(h)] = 2\lambda(h) \quad (3)$$

Among them,  $\lambda(h)$  is its variogram.

The estimation formula for the ordinary Kriging estimation method is:

$$Z^*(x_0) = \sum_{i=1}^n l_i Z(x_i) \quad (4)$$

Among them,  $Z^*(x_0)$  is the estimate of estimated position  $x_0$ , and the observation value of known location  $x_i$ ,  $\lambda_i$  is the weight assigned to  $Z(x_i)$ ,  $n$  is the estimated number of observations to use.

The equation for ordinary Kriging estimation is:

$$\begin{cases} \sum_{j=1}^n l_j C(x_i - x_j) - m = C(x_i - x_j), i = 1, 2, \dots, n \\ \sum_{i=1}^n l_i = 1 \end{cases} \quad (5)$$

The estimated error variance is:

$$s_{OK}^2 = u + \sum_{i=1}^n l_i C(x_i - x_0) + C(0) \quad (6)$$

## 2.3 Evaluation Model

### 2.3.1 Single Factor Pollution Index Method

It is a relative dimensionless index used to evaluate soil pollution degree or soil environmental quality levels. The pollution degree of the soil is represented by a single factor, and then the comprehensive pollution index  $P$  is calculated. The comprehensive pollution index  $P$  is a combination of single pollution indexes. Its formula is (Li et al., 2008):

$$P_i = \frac{C_i}{S_i} \quad (7)$$

Among them,  $P_i$  is the pollution degree of heavy metal  $i$ ,  $C_i$  is the actual concentration of pollutants  $i$ , and  $S_i$  is the evaluation criteria for pollutants, generally taking the secondary standard of inorganic pollutants in soil environmental quality standards.

Finally calculate the comprehensive pollution index  $P$ . The formula is as follows:

$$P = \frac{\sum_{i=1}^n P_i}{n} \quad (8)$$

### 2.3.2 Geological Accumulation Index - Nemeru Index Method

#### i. Geological accumulation index

The geological accumulation index is widely used to study the quantitative indicators of heavy metal pollution in sediments and other materials (Yang et al., 2014). It not only considers the influence of natural geological processes such as sedimentary diagenesis on the background values, but also pays attention to the impact of human activities on heavy metal pollution. However, the index can only be used to indicate the pollution degree of a certain heavy metal, not providing comprehensive pollution degree of various heavy metals. Its expression is as follows:

$$I_{ij} = \log_2 \frac{C_{ij}}{1.5 \cdot BE_j} \quad (9)$$

In the formula,  $I_{ij}$  is the geological accumulation index of the  $j_{th}$  heavy metal in the  $i_{th}$  region,  $C_{ij}$  is the actual measured value of the  $j_{th}$  heavy metal in the  $i_{th}$  region, and  $BE_j$  is the background concentration of the  $j_{th}$  heavy metal. Besides, 1.5 is the correction factor and usually used to indicate sedimentary characteristics, rock geology, and other effects.

#### ii. Nemeru Comprehensive Pollution Index

The Nemeru index is a weighted multifactor environmental quality index considering the extreme value or the maximum value. It highlights the impact of the most polluting substances on the degree of pollution. The basic formula is:

$$P_i = \sqrt{\frac{\max_i^2 + \text{ave}_i^2}{2}} \quad (10)$$

In the formula,  $P_i$  is the Nemeru Comprehensive Index in the  $i_{th}$  region,  $\max_i$  is the maximum value of the single factor pollution index of each heavy metal in the  $i_{th}$  region,  $\text{ave}_i$  is the average of the single factor pollution index of each heavy metal in the  $i_{th}$  region.

#### iii. Comprehensive evaluation

Substituting the results obtained in the formulae (9) into (10), the indicators for comprehensive evaluation of the pollution elements in different regions can be obtained.

## 2.4 Pollution Cause Analysis Model

### 2.4.1 Spearman Coefficient

Spearman coefficient is a statistic that measures the degree of correlation between grading and ordering variables. For the original grade data, the one-side opening data, the data whose overall distribution is unknown and the data that does not obey the normal distribution, or meet the correlation coefficient of the product, the Spearman coefficient formula is (Zhang et al.,2016):

$$r = 1 - \frac{6 \sum d_i^2}{n^3 - n} \quad (11)$$

### 2.4.2 Principal Component Analysis

#### i. Standardized processing

Suppose the original data matrix be:

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}$$

and use the following formula:

$$X_{ij}^* = \frac{(X_{ij} - \overline{X_j})}{S_j} \quad (i = 1, 2, \dots, p; j = 1, 2 \dots, p) \quad (12)$$

In the formula,  $S_j$  is the standard deviation,  $\overline{X_j}$  is the mean. The original data is normalized to eliminate the dimension relation between the variables, making the data comparable.

#### ii. Calculating the eigenvalues and eigenvectors of the correlation coefficient matrix

The correlation coefficient is calculated as:

$$r_w = \sum_{k=1}^n X_{kw}^* X_{kj}^* / (n-1) \quad (w = 1, 2, \dots, p; j = 1, 2, \dots, p) \quad (13)$$

In the formulae,  $r_{jw} = r_{wj}$ ,  $r_{jj} = 1$ ,  $r_{wj}$  is the correlation coefficient between the indicator  $w$  and the indicator  $j$ . Its matrix is:

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix}$$

iii. Calculating the eigenvalues and eigenvectors

Supposing that  $\lambda_1, \lambda_2, \dots, \lambda_p (\lambda_1 > \lambda_2 > \dots > \lambda_p)$  are the eigenvalues of R and  $u_1, u_2, \dots, u_p$  are the corresponding eigenvectors, then there is the following linear transformation relationship:

$$\begin{cases} Y_1 = u_{11}X_1 + u_{12}X_2 + \cdots + u_{1p}X_p \\ Y_2 = u_{21}X_1 + u_{22}X_2 + \cdots + u_{2p}X_p \\ \dots\dots\dots \\ Y_p = u_{p1}X_1 + u_{p2}X_2 + \cdots + u_{pp}X_p \end{cases}$$

iv. Contribution rate and cumulative contribution rate

Contribution rate:

$$\eta_t = \lambda_t / \sum_{j=1}^p \lambda_j \quad (t = 1, 2, \dots, p) \tag{14}$$

cumulative contribution rate:

$$\alpha_t = \sum_{j=1}^t \lambda_j / \sum_{j=1}^p \lambda_j \quad (t = 1, 2, \dots, p) \tag{15}$$

v. Calculating principal component load

$$I_{ij} = p(Z_i, X_j) = \sqrt{\lambda_i} e_{ij} \quad (i, j = 1, 2, \dots, p) \tag{16}$$

vi. Calculating each principal component score

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & \Lambda & Z_{1m} \\ Z_{21} & Z_{22} & \Lambda & Z_{2m} \\ M & M & & M \\ Z_{n1} & Z_{n2} & \Lambda & Z_{nm} \end{bmatrix}$$



### 3. Model Establishment and Solution

#### 3.1 Spatial Distribution of Heavy Metals in Soil Based on Kriging Interpolation

Using the Surfer software to the data in the attachment x and y, the concentration of eight heavy metals are filled by Kriging interpolation method, and the three-dimensional spatial distribution map was drawn. Due to the space limitation, only concentration of As, Cd is shown in Figure 1, Figure 2.

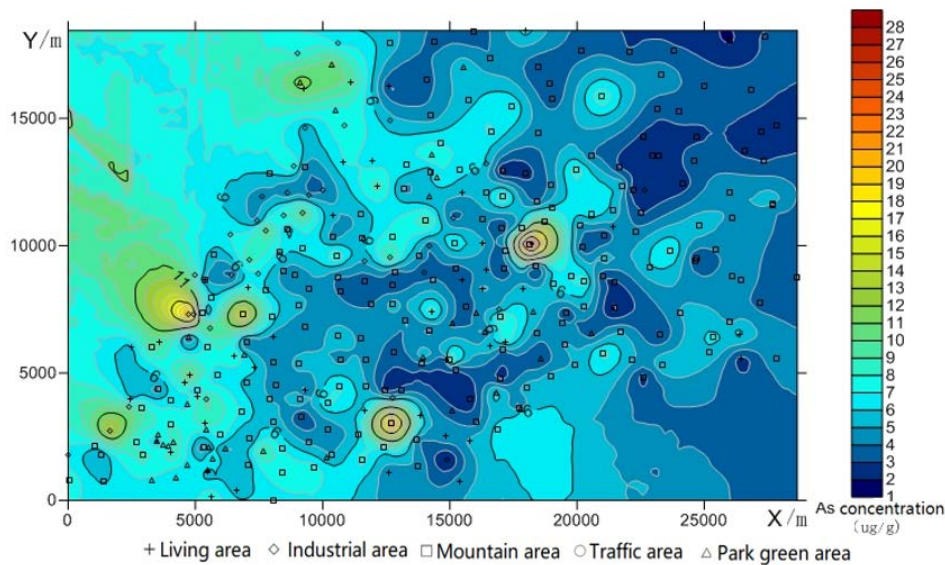


Figure 1. Spatial distribution of as concentration in the urban area

Different colors are used to indicate the concentration of As in the region. The closer to red, the higher the concentration of As in the area. On the contrary, the closer to blue, the lower the concentration of As in the area. It can be seen intuitively from Figure 1 that As is more concentrated in the central and southwestern part of the urban area, that is, the concentration of As in the industrial area and the traffic area is higher. It can be seen from Figure. 2 that Cd is more concentrated in the southwest direction of the urban area, that is, the concentration of Cd in the industrial area is very high.

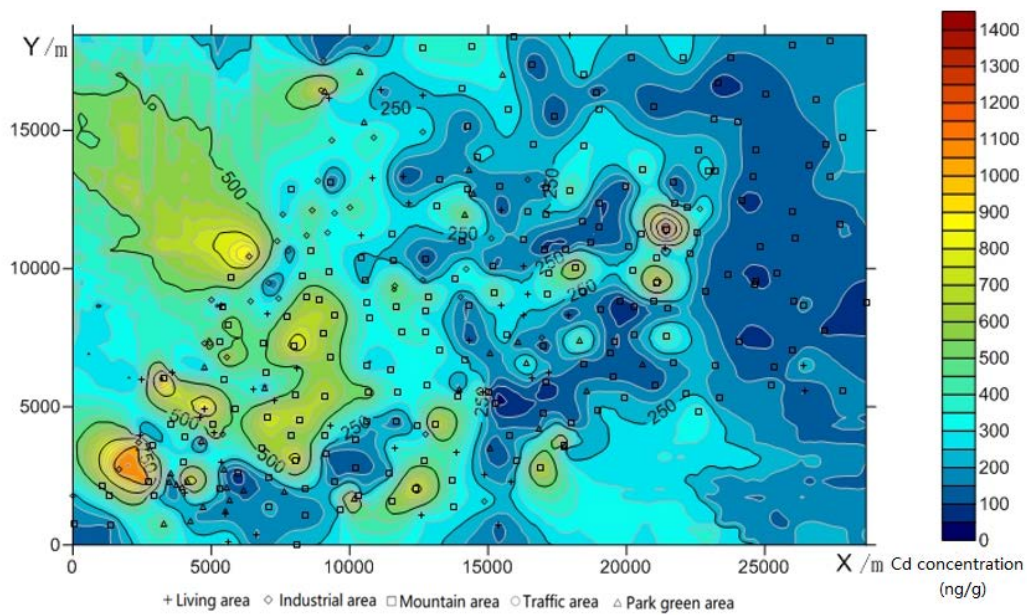


Figure 2. Spatial distribution of Cd concentration in the urban area

### 3.2 Analysis of Pollution Degree of Heavy Metals in Different Regions

#### 3.2.1 Solution of Evaluation Model

i. Solving the single factor index method

a) Constructing a single factor index matrix

Calculate the average concentration of each heavy metal in each region, and then calculate the single factor index of each heavy metal in each region according to formulae (7). The obtained matrix is shown in Table 1:

Table 1. Single factor index matrix

Area	As ( $\mu\text{g/g}$ )	Cd ( $\text{ng/g}$ )	Cr ( $\mu\text{g/g}$ )	Cu ( $\mu\text{g/g}$ )	Hg ( $\text{ng/g}$ )	Ni ( $\mu\text{g/g}$ )	Pb ( $\mu\text{g/g}$ )	Zn ( $\mu\text{g/g}$ )
1	0.25	0.96	0.23	0.49	0.18	0.36	0.23	0.94
2	0.29	1.31	0.17	1.27	1.28	0.39	0.31	1.11
3	0.16	0.50	0.12	0.17	0.08	0.30	0.12	0.29
4	0.22	1.20	0.19	0.62	0.89	0.35	0.21	0.97
5	0.25	0.93	0.14	0.30	0.22	0.30	0.20	0.61

b) Calculating the comprehensive pollution degree index

Calculate the single factor comprehensive pollution degree by using the data in Table 1 by formula (8). The results are shown in Table 2:

Table 2. Single factor comprehensive pollution degree

Area	1	2	3	4	5
P	0.7535	1.0767	0.1670	0.9468	0.7137

Among them, the higher the P value, the more serious the pollution.

c) Grading according to the comprehensive index pollution degree

According to the comprehensive index pollution degree grading standards (Appendix II), the degree of pollution of each heavy metal substance in each region and the comprehensive pollution degree of each region can be evaluated. The results are shown in Table 3:

Table 3. Comprehensive pollution degree

Area	1	2	3	4	5
Grade	2	3	1	2	2
Ranking	3	5	1	4	2

Among them, the grade represents the pollution degree in the region, and the ranking represents the ranking of the pollution degree in the five regions.

By using single factor index method to analyze each heavy metal in each region, it was found that Zn and Cd were the most polluted in living areas and green areas; in addition to high Zn and Cd in the industrial zone, the pollution degree of Cu and Hg is also quite high; the highest degree of pollution in the mountain areas is Cd; the pollution degree in the traffic areas are high in order of Cd, Zn, Cu and Hg. Besides, the comprehensive pollution degree of the five regions from slight to high are ranked from mountain areas to park green areas, living areas, traffic areas and industrial areas.

ii. Heavy metal pollution degree model based on geological accumulation index-Nemero index

a) Constructing a geological accumulation index matrix

Firstly, calculate the average concentration of each heavy metal in each region. Then, according to the geological accumulation index formulae (9), the single factor indexes I of each heavy metal substance in each region is calculated. The results are shown in table 4:

Table 4. Geological accumulation index matrix

Area	As ( $\mu\text{g/g}$ )	Cd ( $\text{ng/g}$ )	Cr ( $\mu\text{g/g}$ )	Cu ( $\mu\text{g/g}$ )	Hg ( $\text{ng/g}$ )	Ni ( $\mu\text{g/g}$ )	Pb ( $\mu\text{g/g}$ )	Zn ( $\mu\text{g/g}$ )
1	0.21	0.57	0.56	1.30	0.82	-0.0086	0.57	1.19
2	0.42	1.01	0.19	2.68	3.61	0.10	1.00	1.42
3	-0.41	-0.35	-0.25	-0.19	-0.35	-0.25	-0.34	-0.49
4	0.08	0.88	0.32	1.65	3.0893	-0.06	0.45	1.23
5	0.21	0.52	-0.09	0.60	1.1311	-0.27	0.38	0.57

## b) Calculating comprehensive pollution indicators using the Nemer index

Calculate the comprehensive pollution degree index  $P$  in each region using all the geological accumulation indexes in each region in table 4 and improve the basic Nemer formulae (10). The following formula is obtained:

$$P_i = \sqrt{\frac{(I_{ij})_{\max}^2 + (I_{ij})_{\text{ave}}^2}{2}} \quad (17)$$

In the formula,  $P_i$  is the comprehensive pollution index for the  $i_{\text{th}}$  region,  $(I_{ij})_{\max}$  is the maximum value of the eight geological accumulation indexes in the region  $i$ , and  $(I_{ij})_{\text{ave}}$  is the average value of the eight geological accumulation indexes in the region  $i$ . Use the formulae (17) to obtain the comprehensive pollution index of the five regions. The results are shown in Table 5:

Table 5. Geological accumulation - Nemer comprehensive pollution degree

Area	1	2	3	4	5
P	0.9781	2.7270	0.2778	2.2937	0.8532
Grade	2	4	1	4	2
Ranking	3	5	1	4	2

By analyzing each heavy metal in each region by geological accumulation index, it was found that the living area was seriously polluted by Zn; the park green area was seriously polluted by Hg; the industrial area was seriously polluted by Cd, Cu, Hg, Pb, Zn; the mountain area was not polluted; the traffic area was seriously polluted by Cu, Hg, Zn. In addition, the comprehensive pollution levels of the five regions from low to high are followed by mountain areas, park green areas, living areas, traffic areas and industrial areas, which are the same as those obtained by the single factor index method, indicating the accuracy of the final conclusions.

### 3.3 Solution of Heavy Metal Pollution Analysis Model

#### 3.3.1 Verification of Normal Distribution

Using SPSS to process data (Liu et al., 2003), and verify the normal analysis, due to space limitations, only the normal distribution map of Hg, Zn is listed here, as shown in Figure 3 and Figure 4 below.

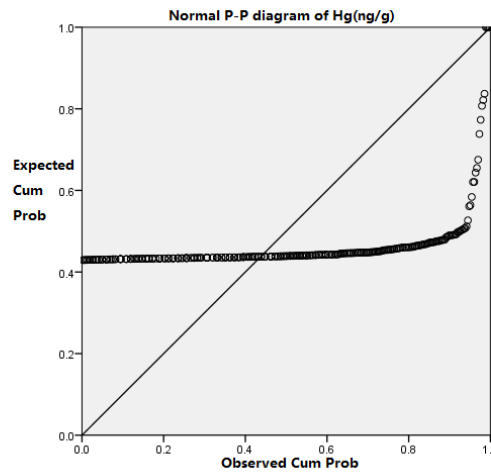


Figure 3. Normal distribution test diagram of Hg element

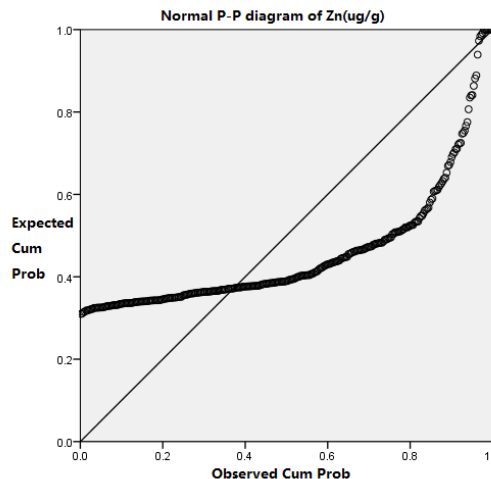


Figure 4. Normal distribution test diagram of Zn element

It can be seen intuitively from the figure that the distribution of heavy metal element concentration does not satisfy the normal distribution, so that the Spearman coefficient can be used.

### 3.3.2 Verification of Correlation Analysis

Verification of correlation analysis is to analyze whether there is a certain relationship between the various heavy metals, that is, whether they can be considered separately. Since the heavy metal pollution indicators do not obey the normal distribution, the SPSS software is used to verify the correlation of the Spearman coefficient. Some of the results are shown in Table 6.

Table 6. Correlation verification of several heavy metal pollution Spearman coefficients

			As(μg/g)	Cd (ng/g)	Cr (μg/g)	Cu (μg/g)	Hg (ng/g)	Ni (μg/g)	Pb (μg/g)
Spearman	As (μg/g)	Correlation coefficient	1.000	0.132*	0.508**	0.410**	0.382**	0.594**	0.215**
	Cd (ng/g)	significance	0.000	0.019	0.000	0.000	0.000	0.000	0.000
	Cr (μg/g)	N	319	319	319	319	319	319	319
	Cu (μg/g)	Correlation coefficient	0.132*	1.000	0.384**	0.705**	0.537**	0.292**	0.811**
		significance	0.019	0.000	0.000	0.000	0.000	0.000	0.000
		N	319	319	319	319	319	319	319
		Correlation coefficient	0.508**	0.384**	1.000	0.573**	0.398**	0.792**	0.406**
		significance	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		N	319	319	319	319	319	319	319
		Correlation coefficient	0.410**	0.705**	0.573**	1.000	0.639**	0.490**	0.764**
		significance	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		N	319	319	319	319	319	319	319

Among them, the \*\* marks the correlation is significant at 0.01 (two-tailed), and the \* marks that the correlation is significant at 0.05.

It can be seen from the table that there is a strong correlation between various heavy metals, such as Cr and Ni, Cd and Pb. Therefore, it is judged from the ingredients that they have strong correlations in the source.

### 3.3.3 Principal Component Analysis

#### a) Principal component rationality analysis

KMO and Bartlett are used to verify whether the original variables are suitable for principal component analysis. The results are shown in Table 7.

Table 7. Verification of KMO and Bartlett

Kaiser-Meyer-Olkin	Measuring sampling legitimacy	.778
Bartlett's spherical check	chi-square	905.711
	Df	28.000
	significance	0.000

From Table 7, the value of KMO is 0.778. The original hypothesis of the Bartlett Sphericity test is that the correlation coefficient matrix is a unit matrix, and the Sig value is 0.000 which is less than the significant level of 0.05. The original hypothesis is rejected, indicating that there is a correlation between the variables and that is suitable for principal component analysis.

## b) Determining the principal component factors

The number of principal component factors was determined by SPSS software. The results are shown in Table 8.

Table 8. The quantitative analysis of principal component factor

Component	Starting eigenvalue			Codified sum of squares load		
	Statistics	Accumulated %	Accumulated %	Statistics	Accumulated %	Accumulated %
1	3.560	44.500	44.500	3.560	44.500	44.500
2	1.150	14.377	58.877	1.150	14.377	58.877
3	0.965	12.063	70.941	0.965	12.063	70.941
4	0.768	9.596	80.537	0.768	9.596	80.537
5	0.578	7.220	87.756	0.578	7.220	87.756
6	0.432	5.399	93.156	0.432	5.399	93.156
7	0.301	3.769	96.924	0	0	0
8	0.246	3.076	100.000	0	0	0

It can be seen from the table that the cumulative variance contribution rate of the first six factors is greater than 90%. Therefore, the first six factors are extracted as the principal factors.

## c) Calculating the factor load value

The principal component load value is acquired by calculating the eigenvectors, cumulative contribution rate and other indicators. And the Kaiser normalized orthogonal rotation method is used to obtain the load values of rotated factors by factor rotation. The values are as shown in Table 9.

Table 9. Factor load value after rotation

element	component					
	1	2	3	4	5	6
Z-score: As (µg/g)	0.135	0.130	0.016	0.974	0.043	0.084
Z-score: Cd (ng/g)	0.223	0.918	0.148	0.086	0.031	0.126
Z-score: Cr (µg/g)	0.859	0.187	-0.016	0.004	0.245	0.141
Z-score: Cu (µg/g)	0.395	0.177	0.273	0.031	0.810	0.110
Z-score: Hg (ng/g)	0.017	0.137	0.967	0.016	0.173	0.069
Z-score: Ni (µg/g)	0.890	0.101	0.051	0.196	0.103	0.170
Z-score: Zn (µg/g)	0.261	0.235	0.082	0.097	0.122	0.916
Z-score: Pb (µg/g)	0.073	0.718	0.061	0.162	0.496	0.275

## 3.3.4 Analysis of the causes of heavy metal pollution

To carry out the analysis better, the SPSS software is used to convert the table data

corresponding to the eight original variables into the data corresponding to the six principal factors. Because there is a certain positive linear relationship between the principal factor variable and the meta variable, it can determine the spatial distribution of the original variables based on the spatial distribution of the principal factors. Due to space limitations, only the spatial distribution of the first two principal factors is shown here in Figure 5 and Figure 6.

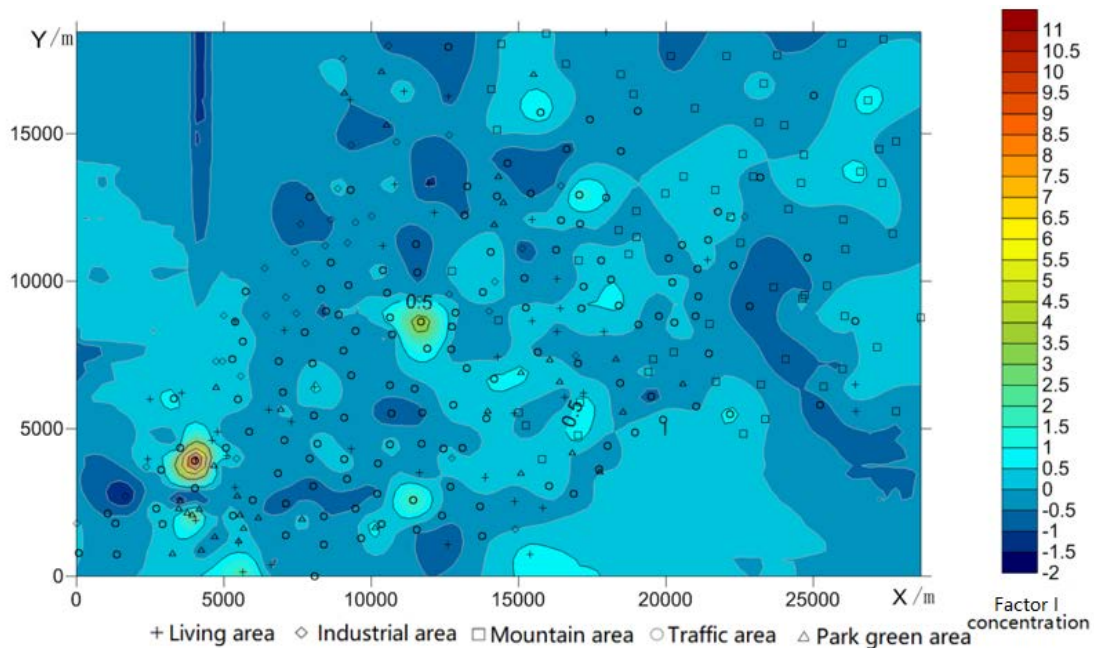


Figure 5. Factor I spatial distribution contour map

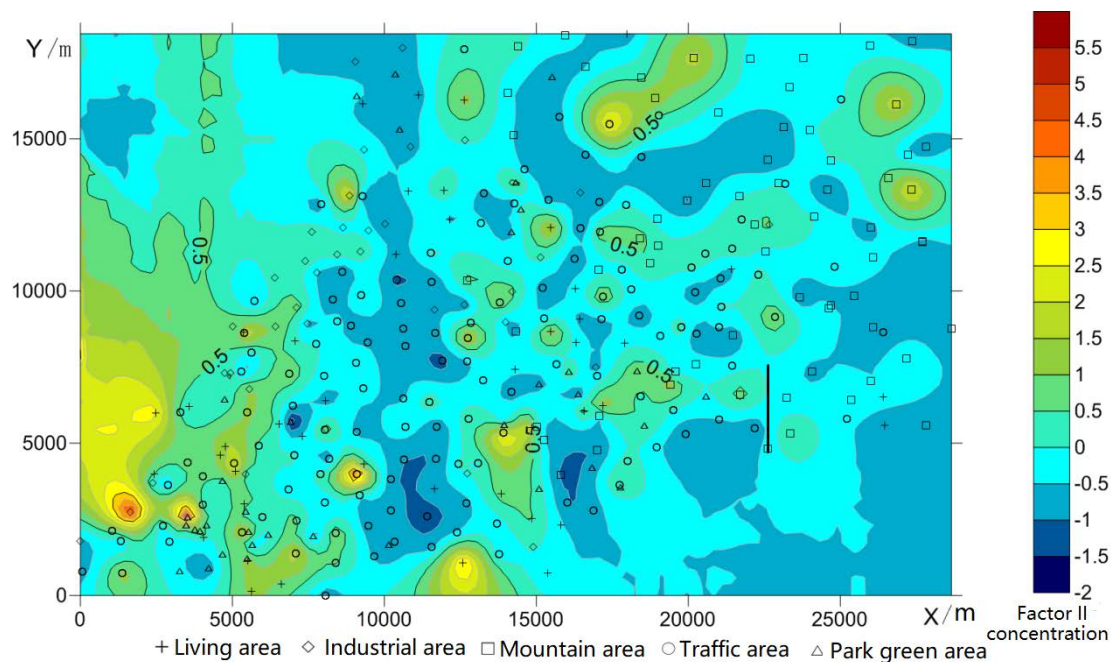


Figure 6. Factor II spatial distribution contour map

According to the spatial distribution contour map of Factor 1, the elements Ni and Cr are



closely related in source, and their concentrations in the living area and traffic area are high. In real life, there are many overlapping parts in the two areas. Therefore, the analysis is that the excessive content of Ni and Cr is caused by the domestic sewage and automobile exhaust emissions.

From the spatial distribution contour map of factor 2, Cd and Pb are closely related in source, and the concentrations are high in the traffic area and industrial area. The analysis is that it mainly caused by the emission of automobile exhaust and industrial waste water.

It can be seen from the spatial distribution contour map of the factor 3 that the element Hg has a higher concentration in the traffic area and the industrial area. However, in the lower left part of the figure, an anomaly that spreads to one side appears. The analysis is that the uneven Hg in the propagation process may be due to the presence of wind or rain erosion in a certain direction.

From the spatial distribution contour map of Factor 4, it can be seen that the concentration of element As in the traffic area and industrial area is high. Combined with the actual situation, the main source of As is the discharge of industrial As wastewater, which is consistent with the actual situation.

According to the spatial distribution contour map of Factor 5, the concentration of element Cu in the industrial area is relatively high. Combined with the actual situation, it is known that Cu mainly comes from the chemical industry and the three wastes in the plastics, printing and dyeing industries, which is consistent with the actual situation.

It can be seen from the spatial distribution contour map of Factor 6 that the concentration of element Zn is high in the living area and industrial area. Combined with the actual situation, it is known that the source of Zn is mainly the discharge of domestic sewage and industrial wastewater, which is consistent with the actual situation

It can be seen from the above results that the main source of heavy metal pollution is consistent with the actual situation according to the spatial distribution map of heavy metal concentration. The content of various heavy metals in the traffic area and industrial area is relatively high, and it is known that the emissions of automobile exhaust and industrial wastes are the major sources of heavy metal pollution.

At the same time, the concentration of Cr and Ni is higher in the traffic area and is also affected by the living area. The analysis is that the pollution is mainly caused by automobile exhaust gas (the domestic sewage also accounts for a part of the proportion). The concentration of Cd and Pb is relatively high in the traffic area and industrial area. And analysis is that it mainly caused by automobile exhaust and industrial waste gas or industrial sewage. The concentration of Hg is relatively high in the traffic area and industrial area. The concentration of As relatively high in the traffic area and industrial area. The concentration of Cu is high in the industrial area and living area, and it is also high in the traffic area mainly due to the industrial printing production. The concentration of Zn is higher in the living area and traffic area.

#### 4. Conclusion

In this paper, at first, the Kriging interpolation method is used to linearly estimate the concentration of heavy metals in the unsampled region, so that it can increase the sample size objectively. Based on this, the spatial distribution maps of eight heavy metals are obtained. At the same time, the single factor index method and the geological accumulation index-Nemero index method are introduced to evaluate the degree of heavy metal pollution in different regions. It is found that their descriptions of the pollution degree are similar, and the comprehensive pollution index is compared to obtain the pollution degree of different areas. The degree from slight to heavy is: mountain areas, park green areas, living areas, traffic areas, industrial areas.

Subsequently, the correlation analysis of the eight heavy metal elements given in the title is carried out. With the premise of ensuring that many variables were fully explained by a few factors, principal component analysis is used to transform them into several linear and irrelevant comprehensive indicators. It is used to eliminate mutual interference between metal elements and simplify data processing. Finally, according to the spatial distribution map of the principal factors, it is known that the main cause of heavy metal pollution is the emission of automobile exhaust and industrial waste, which provides theoretical basis and guiding significance for preventing and treating heavy metal pollution in urban soil.

However, the research in this paper has some shortcomings because the data given by the title is limited, which are embodied in:

(1) The number of sampling points given in the question is small, which causes a large error after interpolation.

(2) The data provided is sampled in a grid of one kilometer without specific geographical distribution. Therefore, when analyzing pollution sources, it has certain ambiguity because it is based on interpolation.

#### Acknowledgments

This work was supported by the National Natural Science Foundation of China (51475095), the Fundamental Research Funds for the Central Universities (21618412), Key Project of Guangdong Natural Science Foundation (2016A030311041), 2015 Guangdong Special Support Scheme (2014TQ01X706), High-level Talent Scheme of Guangdong Education Department (2014-2016), the Guangdong Natural Science Foundation (2017A030313401).

#### References

China Society for Industrial and Applied Mathematics. (2011). Higher Education Club Cup National Contest on Mathematical Modeling for College Students. [Online] Available: [http://www.mcm.edu.cn/html\\_cn/node/a1ffc4c5587c8a6f96eacefb8dbcc34e.html](http://www.mcm.edu.cn/html_cn/node/a1ffc4c5587c8a6f96eacefb8dbcc34e.html) .

Gu, Yi-Qiu., Shen, Jie-Jie., & Yao, Wen-Xu. (2013). Analysis of Heavy Metal Pollution in the Surface Soil of City. *Journal of Xichang University (Natural Science Edition)*, 27(1). Mar. 24-28.

Haibo, L., Zhonghua, T., Qingjun, D., & Jun, W. (2014). The Comprehensive Environmental Quality Evaluation and Pollution Analysis about the north Area of Guangdong. *Applied Mechanics and Materials*, 507, 798. <https://doi.org/10.4028/www.scientific.net/AMM.507.798>.

Liu, R. X., Kuang, J., Gong, Q., & Hou, X. L. (2003). Principal component regression analysis with SPSS. *COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE*, 71(2), 141-147. [https://doi.org/10.1016/S0169-2607\(02\)00058-5](https://doi.org/10.1016/S0169-2607(02)00058-5).

Weixin, Li., Xuxiang, Zh., Bing, W., Shilei, S., Yansong, Ch., Wenyang, P., Dayong, Zh., & Shupe, Ch. (2008). A comparative analysis of environmental quality assessment methods for heavy metal-contaminated soils. *PEDOSPHERE*, 18(3), 344-352. [https://doi.org/10.1016/S1002-0160\(08\)60024-7](https://doi.org/10.1016/S1002-0160(08)60024-7).

Xiaopeng, Y., & Nan, W. (2012). Analysis of heavy metal pollution of urban topsoil. *Technology Innovation and Application*, 33, 116-117. [http://www.wanfangdata.com.cn/details/detail.do?\\_type=perio&id=qgsj201233110](http://www.wanfangdata.com.cn/details/detail.do?_type=perio&id=qgsj201233110).

Xu, J. D. (2008). Research on spatial distribution and pollution evaluation of heavy metals in soil of Jinhua city based on GIS. [Online] Available: <http://cdmd.cnki.com.cn/Article/CDMD-10319-2008152738.htm> (May 2008).

Xu, X. J., Wu, J. L., Yuan, Q. Q., Wang, T. L., & Xu, Y. S. (2012). Analysis of heavy metal pollution of urban topsoil. *Science & Technology Information*, 25, 41-63. [Online] Available: <http://www.cnki.com.cn/Article/CJFDTotl-KJXX201225029.htm>

Yang, G., Chaofeng, Sh., & Meiting, Ju. (2014). Heavy Metal Contamination Assessment and Partition for Industrial and Mining Gathering Areas. *INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH*, 11(7), 7286-7303. <https://doi.org/10.3390/ijerph110707286>

Zhang, W. Y., Wei, Z. W., Wang, B. H., & Han, X. P. (2016). Measuring mixing patterns in complex networks by Spearman rank correlation coefficient. *PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS*, 451, 440-450. <https://doi.org/10.1016/j.physa.2016.01.056>

Zhou, T. P. (2016). The PM2.5 prediction in Kunming base on kriging interpolation method. [Online] Available: <http://xuewen.cnki.net/ArticleCatalog.aspx?filename=1016230443.nh&dbtype=CMFD&dbname=CMFD201701> (June 2016).