Evaluation of Heavy Metal Pollution in Soil A Mathematical Method for Determining the Source of Pollution

Jiayi Chen[1]

Beijing Forestry University College of science Beijing, China
Chen1012085647@bjfu.edu.cn

Abstract
In recent years, the problem of urban pollution is becoming more and more serious. A large number of toxic substances are discharged, which destroy the living environment of residents and threaten people's health. Therefore, local governments are paying more and more attention to environmental problems, especially heavy metal pollution on the soil surface. The study of pollution in an area is usually divided into two aspects: on the one hand, the evaluation of pollution degree, and on the other hand, the determination of pollution sources. In this paper, the Biharmonic Spline Interpolation method is used for visual analysis of the pollution situation in the research area. And then, using the cumulative index method, the degree of heavy metal pollution in the soil of different regions can be evaluated. Finally, the Gauss diffusion model is optimized according to the actual situation, and the location of the pollution source is obtained by solving the model using the measured data.

Keywords: *Index of Geo-accumulation, Person Correlation Coefficient, Gauss Diffusion Model.*

1. INTRODUCTION

The swift growth of the urban economy and population has accelerated urbanization and industrialization. For better developments, human beings do not hesitate to sacrifice the environment. Nowadays, activities of humankind have a substantial negative impact on urban environmental quality, especially on topsoil. The research on inorganic soil pollution has shown that heavy metal contamination is one of the most significant factors of human survival. To be more specific, the gross area of land with excessive heavy metals has surpassed 50 million acres in China and is still growing [4]. Heavy metal pollution is difficult to control and harmful to human health. Microorganisms can break down these contaminants in soil into methyl compounds. The decomposition products have substantial toxicity that affects cells division. It will be absorbed by crops and then enter food intake. Ultimately, the ingested metallic elements will cause health hazards to people. Accordingly, analyzing the pollution degree of heavy metals and identifying the source of pollution can help determine the reason for metal pollution in different metropolitan areas. Moreover, the survey result will raise residents' attention and provide the basis for the governance scheme.

The soil quantity significantly impacts human health, industrial and agricultural safety. An accurate assessment of the pollution level can provide a rational principle to achieve maximum efficiency in land usage. At present, the single factor index and the Nemerow index methods are commonly calculate environmental pollution indexes and divide contamination levels. The Nemerow index, an extension of the single factor index, could emphasize the importance of extreme value. This method weights multiple environmental impact factors without personal effects. In 2021, this method assessed the comprehensive soil nutrients in eucalyptus plantation of state-owned Bobai Forest Farm in Guangxi [7]. The tree nutrient demand can be divided into several factors. These effect factors are used to evaluate the comprehensive fertility status of forest soil. In addition, this method avoids the cask effect of trees' nutrient demand. The result obtained has become the reference for the fertilizer application in forest production. Utilizing the Nemerow index to evaluate soil pollution will obtain the comprehensive pollution status of heavy metals [3]. Nevertheless, although the Nemerow Index considers the impact of the most severe pollution factors, it overlooks the differences in the impact of different pollutants.

G. Vilas Bhau et al. (Eds.): MSEA 2022, ACSR 101, pp. 406–412, 2023.
https://doi.org/10.2991/978-94-6463-042-8_59

As the toxicity of metals in the soil is also related to the properties and species, the single factor index and the Nemerow index methods do not seem appropriate to access the degree of pollution in different areas. The soil's degree of heavy metal pollution was assessed using the accumulation index method. Muller proposed this method in the 1960s and invented it to quantitatively assess the pollution level of heavy metals in sediments and the specified corresponding classification standard of pollution state [13]. The accumulation index is a proper way to assess the degree of heavy metal pollution in soil and classify it.

There are kinds of ways to determine the source of pollution. Scientists have integrated mathematics and statistics to investigate pollution problems in recent years. The factor analysis method, a very effective way to confirm the source of pollutants, was proposed by Spearman [10]. The factor analysis method, transforming various factors into a few linear independence indicators to simplify data, was proposed to measure the relationship between several variables. However, the linear composite index commonly cannot be precisely observed. The multivariate statistic is the traditional method of determining the source of heavy metal pollution [1]. One of them is Principal component analysis, a data dimension reduction method. In this way, a new composite indicator can be obtained. Using this indicator to score the principal component can be a basis for analyzing the cause of pollution. Nevertheless, principal component analysis is only a statistical method that cannot wholly determine the source of pollutants. Besides, the analyst needs many sample data to obtain sophisticated results, which is unsuitable for a quantitative evaluation of source contribution.

On the problem finding the source of heavy metal pollution, the Gauss diffusion model is suitable for adoption. The model has several applicable conditions. Firstly, the wind speed in the studied area is uniform, and the wind direction is flat. Secondly, the diffusion of pollution obeys the mass conservation. Finally, the source strength of the pollution source is uniform and continuous [9]. According to the propagation characteristics of heavy metal pollution, this method is suitable for determining heavy metal pollution sources.

According to the functional division, urban areas are divided into living areas, industrial areas, mountainous areas, traffic areas and forest areas. Evaluating the heavy metal degree of soil in these five regions and locating the source of pollution, the causes of contamination can be found. In this paper, the Geo-accumulation index is used to determine pollution levels. A rough view of the contamination situation can be obtained through the maps of the pollution distribution about eight major heavy metal elements in each region. Finally, a model is confirmed according to the propagation characteristics of heavy metal pollutants to find out the location of the source of pollution.

2. METHODOLOGY

With the development of the urban economy and the increase in population, human environmental destruction is becoming more intense. The paper explores the geological environment of soil in a town. Subsequently, selecting 319 sampling sites, the concentration of eight metals in the sample sites will be measured. In addition, the samples from natural areas that are far away from people will be evaluated. The measuring results can be viewed as the background values of eight heavy metals in the surface soil of the urban area.

Using mathematical models to evaluate the spatial distribution, the degree of pollution and the source of heavy metals in different areas of the town can be obtained. And then find out the location of the source of pollution.

2.1. Biharmonic Spline Interpolation

The metal spatial distribution is continuous. The concentration of metal is fluid with the elevation of the terrain. However, the measured samples are spatially discrete. Therefore, performing reasonable data interpolation for the position points without data. Biharmonic spline interpolation is a method to interpolate the data. Green function is also called the point source function or influence function [11]. It is used to solve differential equations with boundary conditions. Moreover, Biharmonic spline interpolation is a process of solving Biharmonic Equation based on Green function [5].

A smoothest curve or surface can be found using the cubic spline function by selecting a set of irregular spatial data points. From a mechanical point of view, the process is to force an elastic rod to match each data point. An elastic rod subjected to a vertical force $\delta(X - X_j)$ in point $X_j, j = 1,2, \dots$ is for each point on the elastic rod. The bending deformation degree of elastic thin rod under the action of concentrated force in many different positions is calculated. The displacement function $g(x)$ of the follows harmonic equation as:

$$\nabla^4 g(x - x_j) = 6\delta(x - x_j) \tag{1}$$

The particular solution of equation (1) is:

$$g(x - x_j) = |x - x_j|^3 \tag{2}$$

When interpolating N data points w_j located at x_j, the Biharmonic equation to be solved is:

$$\nabla^4 w(x) = \sum_{j=1}^{N} 6a_j \delta(x - x_j) \tag{3}$$

$$w(x_j) = w_j \tag{4}$$

The particular solution w(x) of equations (3) and (4) can be expressed as a linear combination of green's functions of pressure at each point:

$$w(x)=\sum_{j=1}^{N} a_j g(x-x_j)=\sum_{j=1}^{N} a_j |x-x_j|^3 \quad (5)$$

The weighting coefficient a_j can be obtained by solving the following linear equations:

$$w_i=\sum_{j=1}^{N} a_j |x_i-x_j|^3, i=1,2,\ldots,N \quad (6)$$

After a_j is determined, the value of the interpolation function $w(x)$ at any point can be calculated by using Equation (5).

This method can interpolate discrete sample data and draw topographic maps of the studied areas. Then the spatial distribution characteristics of the eight heavy metals can be obtained.

2.2. Index of Geo-accumulation

The geological accumulation index, based on the study of heavy metal pollution in sediments, was proposed initially by Muller in the 1970s [2]. This index is recorded as I_{geo}, where geo stands for geography, to quantify the pollution degree of heavy metals in sediments or other substances. Subsequently, it was utilized by soil researchers at home and abroad to assess soil heavy metal pollution. This method considers the influence of human activities and natural geological processes on the background value. Moreover, the natural variation of heavy metal distribution and the impact of human activities on the environment can be obtained.

A single element or multi-element comprehensive evaluation is all in the range of application. In the single element evaluation, the pollution degree is measured by the size of I_{geo}, which is a quantitative index to study heavy metal pollution in water environment [8]. It is also commonly used to measure the contamination degree of a single element through calculating the relationship between observed value and its reference value. The index is calculated by MATLAB. The larger I_{geo} is, the more serious soil pollution is. The soil pollution grade is generally determined according to the principle of "from inferior to superior" in the comprehensive evaluation process. The pollution grade corresponding to the most extensive geological accumulation index of each heavy metal element is determined as the comprehensive pollution grade of the sampling site. The calculation formula of geological accumulation index proposed by Muller is:

$$I_{geo}=log_2^{C_i/\alpha B_i} \quad (7)$$

C_i: The measured value of element i in the sample.

B_i: The background value of element i in the sample.

α: An index to eliminate variations in background values that rock differences may cause.

According to the size of index I_{geo}, Table 1 divides the pollution severity into six levels. The higher the level is, the more serious the pollution degree is. The pollution degree of 8 heavy metals in soil was evaluated by Table 1.

According to this table, the heavy metal pollution degree of the five functional areas can be divided, so as to analyze the pollution status and pollution sources of the area. The results can be used as a standard to govern the environment.

Table 1: Land accumulation index and pollution degree classification

I_{geo}	Grade	The degree of pollution
$5<I_{geo}\leqslant 10$	6	Extreme pollution
$4<I_{geo}\leqslant 5$	5	Strong to extreme pollution
$3<I_{geo}\leqslant 4$	4	Strong pollution
$2<I_{geo}\leqslant 3$	3	Moderate to strong pollution
$1<I_{geo}\leqslant 2$	2	Secondary pollution
$0<I_{geo}\leqslant 1$	1	Light to moderate pollution
$I_{geo}\leqslant 0$	0	Pollution-free

Table 2: Index of Geo-accumulation Geo-chemistry background

element	value	element	value
As	3.6	Hg	35.0
Cd	130.0	Ni	12.3
Cr	31.0	Pb	31.0
Cu	13.2	Zn	69.0

Selected the natural area without external pollution as the control group, the soil heavy metal content is measured in this area. The concentrations of different kinds of heavy metals in measurement results can be used as the background value of the index. The specific values are shown in Table 2.

2.3. Person Correlation Coefficient

Pearson, proposed a method to compute linear correlation to measure the correlation between two continuous variables [12]. It can be used to evaluate correlations between different metals. In addition, the output ranges from -1 to 1, with 0 indicating no correlation, negative values indicating negative correlation and positive values indicating positive correlation. In general, the more remarkable the correlation, the more likely they are from the same sources.

The population correlation coefficient ρ is defined as the ratio of the covariance between two variables X and Y and the product of their standard deviations, as follows:

$$\rho_{X,Y}=\frac{cov(X,Y)}{\sigma_X\sigma_Y}=\frac{E[(X-\mu x)(Y-\mu y)]}{\sigma_X\sigma_Y} \quad (8)$$

By estimating the covariance and standard deviation of the sample, the Person correlation coefficient of the sample can be obtained, which is usually represented by *r*:

$$r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \tag{9}$$

The correlation can be obtained by calculating the Person correlation coefficients of 8 heavy metal types in different regions. The more closely related the metals, the more likely they are homologous. Therefore, it is possible to classify homologous metal pollutants into one category to analyze pollution sources.

2.4. *Gauss Diffusion Model*

The Gauss diffusion model is used to calculate the pollution content of spatial point sources [6], and its equation is as follows:

$$C(x,y,z) = \frac{Q}{2\pi\mu\sigma_y\sigma_z} exp[-\frac{1}{2}(\frac{y^2}{\sigma_y{}^2}+\frac{z^2}{\sigma_z{}^2})] \tag{10}$$

In formula (4), C is the pollution content (mg/m^3) in the space point (x,y,z). Q is the pollutant discharge (mg/s) per unit time. μ is the average wind speed (m/s). σ_y , σ_z is the diffusion coefficient of flue gas. These two parameters are related to the stability of the atmosphere and the horizontal distance.

Formula (4) expresses the pollutant concentration in the air, while this article studies the heavy metal content in the soil surface. The formula can be derived as follows:

$$\mathrm{C(x,y,z)} = \frac{\mathrm{A}}{\sigma_\mathrm{y}\sigma_\mathrm{z}} \exp[-\frac{1}{2}(\frac{\mathrm{y}^2}{\sigma_\mathrm{y}{}^2}+\frac{\mathrm{z}^2}{\sigma_\mathrm{z}{}^2})] \tag{11}$$

Factors of atmospheric environment, such as wind speed, are not considered to influence the transmission of pollutant concentration. So, the factors related this in formular (10) are combined into a constant coefficient A.

The Gaussian diffusion equation defaults the source point to the origin. Therefore, assume that the coordinate of pollution source of heavy metal i is (x_{0i}, y_{0i}, z_{0i}). Using coordinate translation, the formula as follows:

$$\begin{aligned} &C(x-x_{0i}, y-y_{0i}, z-z_{0i}) \\ &= \frac{A}{\sigma_y\sigma_z} exp\left[-\frac{1}{2}\left(\frac{(y-y_{0i})^2}{\sigma_y{}^2}+\frac{(z-z_{0i})^2}{\sigma_z{}^2}\right)\right] \end{aligned} \tag{12}$$

where $\sigma_y = 0.2295x^{0.919325}$, $\sigma_z = 0.114682x^{0.941015}$.

3. RESULTS AND DISCUSSION

This article uses biharmonic spline interpolation to interpolate the measured data. Then draw the interpolation data into a three-dimensional topographic map and a sampling point distribution map. The plots intuitively describe the spatial distribution of heavy metals in various city areas. Besides, to determine the degree of heavy metal pollution in soil, the Index of Geo-accumulation can be used. The research obtains the correlation between different metals through Pearson correlation analysis of the data. In this way, metals with the exact source of contamination can be found. Eventually, this article analyses the pollution source's location through the Gauss diffusion model.

3.1. Topography of the area involved and the extend of pollution

Pollutants in soil tend to flow, and their concentration will change with the topography and environmental conditions of the region. Therefore, it is essential to see the topography of the study area to analyze pollutant concentration. The topographic map is shown as follows:

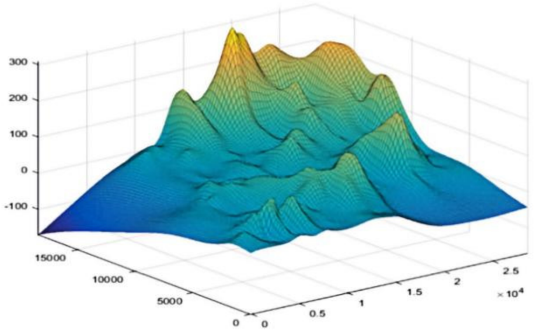

Figure 1. Topographic map

Using the Triangulation method with Linear Interpolation, the sample data can be interpolated. And then, the obtained data will be plotted into maps. It can be seen from Figure 1 that the ground in the northeast corner of the area is higher. Moreover, the ground in the southwest corner is lower. Therefore, the concentrations of metal contamination may spread from higher ground to lower ground. The ground levels of the northwest corner and southeast corner are lower than 0 meters above sea level due to difference interpolation.

The region is divided into five regions (living area, industrial area, mountain area, traffic area and forest area), and samples are taken from these regions. Distribution contour lines of sampling points are shown in the figure below:

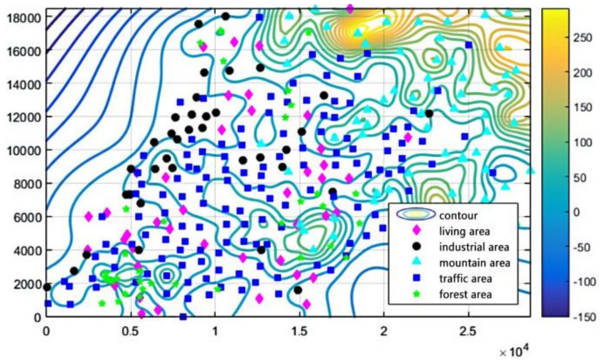

Figure 2. Contour map

It can be seen from Figure 2 that sample points are evenly distributed, indicating that the experimental setting is reasonable. Moreover, the living areas are interspersed between industrial and traffic areas. The mountains and forests are concentrated in the northeast corner and southwest corner, respectively.

The iso-concentration lines of the eight elements are drawn respectively, and the uneven distribution of the eight elements can be seen through the images.

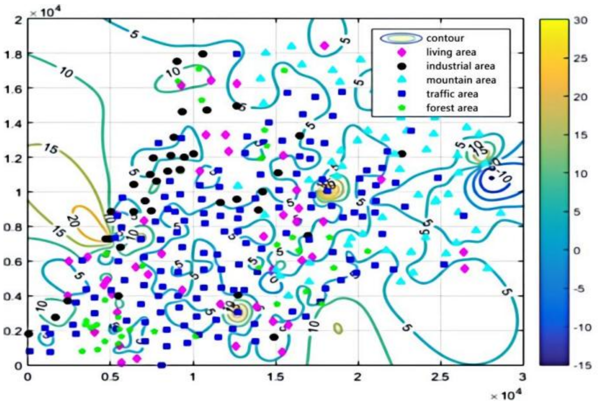

Figure 3. Concentration contour diagram of As

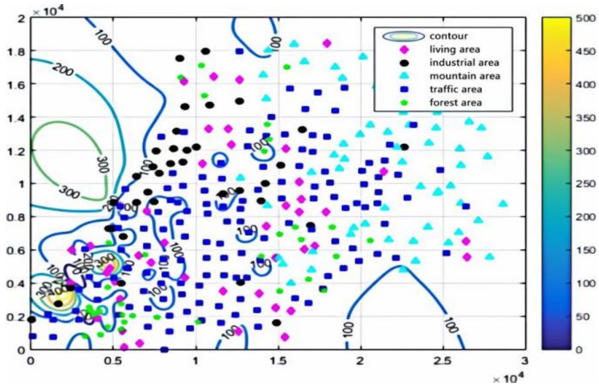

Figure 4. Concentration contour diagram of Pb

By comparing the concentration diagram about Figure 3 and Figure 4, it can be seen that the distribution range of the metal As is the most widely dispersed—moreover, the highest concentration of the metal As are distributed chiefly in traffic areas. Cd is widely distributed in all regions except the northeast. Besides, the distribution range of Cr, Cu and Ni is small, mainly concentrated in the southwest, and the concentration contour maps are similar. Hg, Pb, and Zn are also mainly distributed in the southwest, and Hg and Zn also have aggregation distribution in the middle of the map.

3.2. Degree of pollution determination

Table 3:LAND ACCUMULATION INDEX OF 8 HEAVY METALS

Area	Cu	Pb	Zn	Cr	Ni	Cd	As	Hg
living	0.87	0.23	0.46	0.14	-0.07	0.33	0.12	0.19
industrial	1.37	0.67	0.85	-0.03	-0.03	0.79	0.21	1.31
mountain	-0.39	-0.45	-0.59	-0.44	-0.45	-0.52	-0.53	-0.59
traffic	1.07	0.28	0.65	0.02	-0.18	0.59	-0.05	0.43
forest	0.39	0.12	0.05	-0.16	-0.33	0.18	0.13	0.19

Taking the highest pollution index as the evaluation result of the area, the index of the five regions is 0.876, 1.377, -0.395, 1.074 and 0.395, respectively. According to Table 1, it can be judged that there is no pollution in mountainous areas, little pollution in residential areas and forest areas, and moderate pollution in industrial and traffic areas.

Through the index figure in table 3, Hg, Cu and Zn contribute the most to the pollution in this urban area. Mercury pollution mainly comes from fuel combustion, smelting, and garbage incineration. It is a metal that also enters the soil through fertilization, pesticides, household waste and other ways. The primary pollution sources of copper are produced through the mining and smelting of copper and zinc ore, machinery manufacturing, steel production, etc. Besides, zinc pollution is machine manufacturing and emissions from galvanized and paper making industries. Based on the above analysis, the leading causes of heavy metal pollution in this urban area are industrial and traffic exhaust emissions.

3.3. The Correlation Between Metals

The Person correlation analysis was conducted for eight heavy metal pollutants, and the results are shown as follow:

Table 4:PERSON CORRELATION COEFFICIENTS OF EIGHT HEAVY METALS

Metal	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
As	1.00	0.25	0.18	0.15	0.06	0.31	0.28	0.24
Cd	0.25	1.00	0.35	0.39	0.26	0.32	0.66	0.43
Cr	0.18	0.35	1.00	0.53	0.10	0.71	0.38	0.42
Cu	0.15	0.39	0.53	1.00	0.41	0.49	0.52	0.38
Hg	0.06	0.26	0.10	0.41	1.00	0.10	0.29	0.19
Ni	0.31	0.32	0.71	0.49	0.10	1.00	0.30	0.43
Pb	0.28	0.66	0.38	0.52	0.29	0.30	1.00	0.49
Zn	0.24	0.43	0.42	0.38	0.19	0.43	0.49	1.00

The Person correlation coefficient in Table 4 of Pb and Cd, Ni and Cr is more significant than 0.6. It is more likely that the two groups of metals came from the same source of pollution. Ore mining, non-ferrous metal smelting, coal and fuel burning have been identified as the primary anthropogenic sources of Cd and Pb, most likely in industrial areas. The use of inorganic phosphorus is most likely to lead to the accumulation of Ni and Cr in soil, and the pollution sources are most likely to be located in forest areas.

3.4. Identification of pollution sources

Putting the data into Formula (6), solving the improved Gaussian diffusion equation, the result is as follows:

Table 5: The fitting coefficient of Formula (6)

Heavy metal	K	x_0	y_0	z_0
As	30	-22745	1.4	15
Cr	40	-34434	3.2	45
Cd	66	-54532	5	2
Cu	70	-75687	-7.5	10
Hg	13	-13344	-6	43
Ni	40	-65434	3.3	5
Pb	46	-76765	5.3	23
Zn	78	-24532	1.3	2

The least-square method was used to solve the fitting Equation further, and the pollution sources of the eight metals were obtained. The results are shown in the table below:

Table 6: Sources of contamination for eight metals

Heavy metal	x_0	y_0	z_0
As	12542	10043	80
Cr	24542	11346	6043
Cd	20123	11493	48
Cu	2343	3565	10
Hg	2780	3001	5
Ni	3294	6010	5910
Pb	1742	2973	8
Zn	9342	4322	5024

4.Conclusion

This study divided the region into five functional regions, and sample points were evenly selected within the regions. On this basis, we visualized the sampling points, obtained the rough distribution rules of eight metals, and obtained the rule that the terrain in this region gradually decreases from east to west. In addition, the pollution degree of different functional areas is assessed using the cumulative index method, and it is concluded that the pollution degree of traffic areas and industrial areas located in low-lying areas is the strongest. Finally, the Gaussian diffusion model is optimized, and the least square method is used to solve the location of each metal pollution source. Due to the industrial pollution and precipitation of pollution factors in the atmosphere, most metal pollution sources are concentrated in the low-lying middle region.

References

[1] D. Y. Yu, “Soil heavy metal pollution in China in nearly a decade parsing methods comparison,” Journal of soil bulletin, 2021, 52(4):1000-1008.

[2] G. Müller, “Index of geoaccumulation in sediments of the Rhine River,” Journal of Geology, 1969, 2(3):108-118.

[3] H. Sun, “Assessment of heavy metal pollution in soil surrounding brown land based on improved Nemerow index method,” Environmental protection science, 2018, 44(02):98-102.

[4] J. F. Li, “Status of soil heavy metal pollution and analysis of detection technology,” Communication of Chemical Engineering Design, 2022, 48(02):193-195.

[5] J. Z. Xia, “Application of dual-harmonic spline interpolation method in logging and seismic data integration,” Oil & gas geology, 2010, 31(02):244-249.

[6] J. Q. Zhang, “Application of gaussian diffusion model in coalbed methane leakage discharge,” Petrochemical green and low carbon, 2020, 5(03):59-64.

[7] L. Wu, “Bab forest of eucalyptus plantation soil fertility in index evaluation method analysis,” Guangxi forestry science, 2021, 50(03): 291-296.

[8] L. Li, “Application of ground accumulation index method in evaluation of heavy metal pollution in gold tailings pond,” environmental protection and technology, 2020, 26(02):15-18.

[9] Q. M. Hou, “The Analysis for the Environmental Impact of Natural Gas Leakage and Diffusion into the Atmosphere,” Advanced Materials Research, 2021, 36(01):34-41.

[10] X. J. Li, D. Q. Zhang, J. Qiao and C. H. Qin, “Somewhere based on multivariate statistical method of shallow groundwater pollution source analysis,” Journal of environmental monitoring in China, 2020, 4 (01) : 88-95.

[11] Y. L. Zhu, “Green's function solution of biharmonic equation for plane problems of elasticity,” Journal of Jingzhou Teachers college, 1999(02):12-14.

[12] Y. M. Sun, “Using SPSS software to analyze the correlation between variables,” Journal of Xinjiang Education College, 2007(02):120-123.

[13] Z. B. Jia, “Assessment of heavy metal pollution in sediments of Taizihe River by using geoaccumulation index method,” Journal of Beijing University (Natural Science), 2000(04):525-530.