Research on Ecological Risk Assessment Model Based on Genetic Algorithm

Shengwei Wang¹, Chang Zhang*, Yue Zhang¹, Tianlong Lou¹, FeiYang Xue¹

¹College of Computer Science &Engineering, Northwest Normal University, Lanzhou, 730070, China
*Corresponding author’s e-mail: dearzhangchang@qq.com

Abstract. In this study, Daxia river, a typical watershed in northwest China, was selected as the research area. The heavy metal content of the soil in the basin is affected by topography, meteorological factors and vegetation. The use of traditional methods to sample heavy metals in the basin soil consumes manpower and material resources, and is affected by geographical environment and weather conditions, resulting in inefficiency. In order to more accurately screen the factors related to soil water content, the characteristics of the global optimization of genetic algorithm are used, and the key factors are screened by controlling the iteration times to accelerate convergence. Furthermore, using the method of multiple regression, using the data from 2015 to 2017, a prediction model for soil heavy metal content was established. The conformity of this model is more than 80%, which can effectively characterize the soil heavy metal content in the study area. The establishment of this model provides a good theoretical and practical basis for the heavy metal content of soil in small and medium-sized watersheds in the future.

1. Introduction
Soil safety is closely related to agricultural production, food safety, human health and ecological health[1]. At the present stage, nearly 20% of China’s land ecological environment is at high risk, with a total area of about 0.11 billion km²[2]. Among them, the effective pollution status of heavy metals in soil environment has received increasing attention[3]. Studies have shown that heavy metals pose a high risk of environmental exposure by spreading and accumulating between plants (main pathways: crops and vegetables) and animals[4]. The heavy metal environmental hazard released by the ecological chain link process has a negative impact on the viability of urban trees[5]. Frequent human activities accelerate the release of heavy metal activity, causing heavy metals in the soil to pose a public health threat[6]. Traditional evaluation methods all require field sampling. However, due to weather, geographical location and other factors, the difficulty of sampling and the time of sample processing have been increased. These factors have caused great obstacles to further understanding the ecological risk of heavy metals in the research area. There are many factors that affect the content of heavy metals in the soil of the watershed, including topography, meteorological factors, vegetation, soil and human factors. In order to better analyze the content of heavy metals in soil, research through the filter associated with the degree of influence factor to establish the model. In order to get the optimal solution, genetic algorithm is introduced. Genetic algorithm can process multiple individuals in a population at the same time [7], and it is easy to achieve parallelization. By using this feature, the crossover operator is used to approach the optimal domain. Then, the local random search function of mutation operator is used to accelerate the convergence to the optimal solution [8]. When the factors
reach the optimal convergence, it indicates the correlation with the content of heavy metals in soil in the study area. Then the pollution degree of different heavy metals in the study area was compared with the soil environmental background value. The research results are expected to provide theoretical references for soil heavy metal ecological risk assessment, prevention and control of heavy metals in farmland and methods of remediation in small and medium-sized watershed.

2. Research area overview

This paper takes the Daxia River flowing through the southwestern part of Gansu Province of China as the research area. The study area is located between 102°02'~103°23' East longitudes and 34°51'~35°48' North latitude. There are two river sources in the Daxia River, namely Xiahe and Luohe. [9]. The Daxia River flows through Gannan and Linxia, which are the main water sources of the two places. The total population of the basin is 788,900[10]. According to the relevant investigation and research on the water conservancy department in Linxia Prefecture, 45.5% of the river water resources in the area are affected by different levels of anthropogenic emissions, of which 11.6% of the river sections are more obvious [11]. In terms of the geographical distribution of water pollution, the quality of the main stream is better than that of the tributary, and the upstream water quality of the main stream is better than the downstream. The water pollution in the industrial and mining area is the most obvious [12]. Water conservation in the Daxia River Basin is an important foundation for the sustainable development of social and economic development in minority areas [13].

3. Data source and processing

The main applications are from the Geospatial Data Cloud (http://www.gscloud.cn/) and the Landsat 8 OLI_TIRS remote sensing satellite data from the LANDSAT series of data provided by the USGS (https://earthexplorer.usgs.gov/). By using ENVI software to read, inlay, trim, and atmospheric correction of remote sensing images with less cloud cover, the unsupervised classification K-Means and the combination of on-site surveys are used to classify the land. The map is processed by Majority/Minority analysis, and the clustering process is carried out to ensure the continuity of the space. The filtering problem is also used to solve the island problem in the picture, so the image is close to the real situation. This study mainly divides land types into waters, grasslands, towns, cultivated land, and forest land, as shown in the following figure 1.

![Figure 1. Classification of land use types of the Daxia River](image)

There are 114 sampling points in this study, including 24 fixed sampling points and 90 random sampling points. When collecting samples, use a hand-held GPS positioner, use a plastic spoon to take the upper layer of 0-20cm soil, and the soil sample is dried and ground, and then passed through a 200 mesh nylon sieve after fine grinding. The sample was then processed by microwave digestion and the
content of heavy metals was determined by inductively coupled plasma optical emission spectrometer (ICP-OES).

Precipitation and temperature are mainly derived from surface meteorological data from the China Meteorological Data Network (http://data.cma.cn/site/index.html). The data for the 2015-2017 study area is mainly used. The temperature data is averaged, the temperature is the monthly average temperature, and the precipitation is the monthly average precipitation.

4. Research methods

4.1. Genetic algorithm

Genetic algorithm[14] is a simulation Darwin the evolution natural selection and genetic mechanism of biological evolution process calculation model, it is a kind of by simulating the natural evolution process to obtain the optimal solution method. The original first put forward by professor Holland in the United States, its advantage is that can avoid the local optimal solution, so as to extract the important impact factors. The basic model of genetic algorithm is as follows:

$$\max \ f(x)$$

$$\text{s.t. } X \in R \ U$$

(1)

Where $x = [x_1, x_2, \ldots, x_n]$ is the decision variable, $U$ is the basic space, $s.t. X$ is the constraint, $R$ is the set of solutions that satisfy all the constraints, and $f(x)$ is the objective function.

In this study, factors related to soil heavy metal content include topography (slope, elevation,), meteorological factors (temperature, humidity, wind speed, etc.), vegetation, soil (pH, precipitation, runoff, etc.), human factors, etc. factor. In order to be able to select the most influence on the heavy metal content of soil from many factors, it is necessary to use the principle of genetic algorithm to screen the optimal solution.

4.2. Multiple linear regression

Environmental factors in a region, such as water, climate, vegetation and topography, affect and restrict each other [15]. The main factors related to the content of heavy metals in soil were selected by genetic algorithm, and the model was built by using the method of multiple linear regression [16]. The general expression of multiple linear regression equation is as follows:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

(2)

Where, $y$ is the observable random variable, that is, the dependent variable; $x_1, x_2, \ldots, x_k$ is the factor affecting $y$, with a total of $k$ factors, and is a non-random variable that can be observed accurately, which is called independent variable or factor. $\beta_0, \beta_1, \cdots, \beta_k$ is $k+1$ unknown parameter, $\epsilon$ is random error.

4.3. Model establishment

4.3.1. Screening of key factors. Genetic algorithm was used to screen the soil heavy metals in Daxia river basin. After iterative calculation of related factors, the impacts are as follows: precipitation > temperature > vegetation index NDVI > altitude > PH value > wind speed. After different iterations, the correlation of each factor is different. The higher the correlation, the greater the influence on the average heavy metal content, and vice versa. The correlation factors screened by the genetic algorithm that were more than 0.8 were precipitation, temperature and NDVI values of vegetation. It is difficult to quantify human factors, so it is temporarily ignored. Therefore, there are three factors involved in modeling, namely precipitation, temperature and NDVI.

4.3.2. Model establishment. Using the heavy metals Pb, Cu, Cr, Cd, and As as the dependent variables, the average precipitation, average temperature and average NDVI value of the study area were used as
the independent linear regression calculation, and the heavy metals and precipitation, temperature and NDVI worthy of expression were obtained. The results are as follows Table 1:

| Soil heavy metal (ug/g) | Multiple linear regression expression | Correlation coefficient |
|-------------------------|---------------------------------------|------------------------|
| Pb                      | $y_{Pb} = -349.983x_1 + 1.308x_2 + 0.02x_3 + 50.168$ | 0.886                  |
| Cu                      | $y_{Cu} = 12.582x_1 - 1.56x_2 + 0.04x_3 + 36.379$ | 0.864                  |
| Cr                      | $y_{Cr} = 95.189x_1 - 2.39x_2 - 0.009x_3 + 52.205$ | 0.832                  |
| Cd                      | $y_{Cd} = -4.140x_1 + 0.124x_2 - 0.001x_3 + 3.199$ | 0.815                  |
| As                      | $y_{As} = 119.965x_1 - 0.639x_2 + 0.11x_3 + 12.763$ | 0.803                  |

In the formula in the above table: $x_1$ represents precipitation, $x_2$ represents temperature, $x_3$ represents NDVI value, $y_{Pb}$ represents the content of heavy metal Pb, $y_{Cu}$ represents the content of heavy metal Cu, $y_{Cr}$ represents the content of heavy metal Cr, $y_{Cd}$ represents the content of heavy metal Cd, and $y_{As}$ represents the content of heavy metal As. The comparison between the calculated results of this formula and the measured soil heavy metal content is shown in FIG. 5. It can be seen that the measured value is not much different from the calculated value, so the formula has a good accuracy and can be used to predict the average heavy metal content.

5. Model application
The average precipitation, average temperature and NDVI value of the study area were substituted into the corresponding expression to calculate the results. The national standard of soil environmental quality standard (GB15618-1995) is often used as the reference standard, as shown in table 2, and the results are obtained as shown in figure 3 below.

| Soil pH | Natural background value | 1 level | 2 level | 3 level | 4 level |
|---------|--------------------------|---------|---------|---------|---------|
| Cd      | 0.2                      | 0.3     | 0.3     | 0.6     | 1       |
| Cr      | 90                       | 250     | 300     | 350     | 400     |
| Cu      | 35                       | 50      | 100     | 100     | 400     |
| Pb      | 35                       | 250     | 300     | 350     | 500     |
| As      | 15                       | 30      | 25      | 20      | 30      |

Figure 2. Comparison of measured and calculated values of heavy metals
As can be seen from the figure of heavy metal pollution level calculated by the expression, the pollution level of heavy metal Pb Cr Cu in the study area is level 1. The pollution situation of heavy metal Cd is as follows: in the study area, the pollution level of Xunhua sala autonomous county and Tongren county in Hezheng county is level 1, the pollution level of cooperative city in Xiahe county, Linxia county, Dongxiang autonomous county is level 2, and the pollution level of Linxia city is level 3. The pollution situation of heavy metal As is as follows: Tongren Xunhua sala autonomous county Hezheng county Hezuo city is level 1, Linxia County, Linxia City, Dongxiang autonomous county is level 2.

According to the classification map of land use types (Figure 3), the causes of heavy metal pollution in soil can be analyzed. Cd pollution mainly comes from cosmetics, lighting lamps, dental materials and coal burning. Heavy metal elements in automobile exhaust and tire additives can also affect the content of Cu in the soil, and the accumulation of these elements is related to the traffic flow, and the pollution degree of Cd in traffic intensive areas is also relatively serious. The high cadmium content in farmland is related to the long-term application of large amount of superphosphate fertilizer with high Cd content impurities. According to the remote sensing images, Linxia has more towns and more human activities, leading to more coal burning for lighting. Therefore, the pollution level of Linxia's Cd is grade 3.

![Figure 3. Heavy metal pollution level map.](image)

6. Conclusions
The overall pollution level of the study area is low, and heavy metals such as Pb, Cr and Cu are primary pollution. Due to the spread of cities and towns and the relationship between human activities, Cd has different levels of pollution. Linxia city has reached level 3, Dongxiang autonomous county, Linxia county, Xiahe county and cooperative city have the pollution level of level 2, and the other areas have the pollution level of level 1. As is mainly generated by coal combustion, so it is mainly concentrated in areas with more urban distribution, namely, Dongxiang autonomous county, Linxia
city, Linxia county and Xiahe county are secondary. Through data verification, it is shown that the model has potential application value and certain execution degree. It laid a foundation for the soil heavy metal pollution in the research area.

Acknowledgement
This work was supported by the National Natural Science Foundation of China under Grant No.61563047.

References
[1] Yu Haiying. Status and Prevention Measures of Soil Heavy Metal Pollution in China[J]. Resources Conservation and Environmental Protection, 2017, 12: 030.
[2] Yang Q, Li Z, Lu X, et al. A review of soil heavy metal pollution from industrial and agricultural regions in China: Pollution and risk assessment[J]. Science of The Total Environment, 2018, 642: 690-700.
[3] Sawut R, Kasim N, Malhemuti B, et al. Pollution characteristics and health risk assessment of heavy metals in the vegetable bases of northwest China [J]. Science of The Total Environment, 2018, 642: 864-878.
[4] Yu K, Van Geel M, Ceulemans T, et al. Vegetation reflectance spectroscopy for biomonitoring of heavy metal pollution in urban soils [J]. Environmental Pollution, 2018.
[5] Doabi SA, Karami M, Afyuni M, et al. Pollution and health risk assessment of heavy metals in agricultural soil, atmospheric dust and major food crops in Kermanshah province, Iran[J]. Ecotoxicology and environmental safety, 2018, 163 : 153-164.
[6] Zhang Yimin, Mu Tingkui, Yan Tingxi, et al. Development and Prospect of Hyperspectral Remote Sensing Technology [J]. Space Return and Remote Sensing, 2018, 6-15: 104-114
[7] Zhou Wei, Ling Xinghong. A Review of Genetic Algorithm Theory and Technology Research[J]. Computer and Information Technology, 2010(04): 37-39+43.
[8] Luo Shuquan. Comparison between traditional optimization algorithm and genetic algorithm[J]. Journal of Hubei University of Technology, 2007(03): 32-35.
[9] Wang Shizhen. Discussion on planning and design of Daxiahe flood control project in Xiahe County, Gannan Prefecture[J]. Journal of Gansu Agriculture, 2016 (20): 39-41.
[10] Wang Hairong. Discussion on the trend and causes of water pollution in Daxia River from 2006 to 2013[J]. Groundwater, 2015 37(2): 83-85.
[11] Chen Nanhai. Discussion on the Causes and Countermeasures of Water Environment Problems in Linxia Prefecture[J]. Agricultural Science & Technology, 2015 (12): 91-92
[12] Luo Ruqi. Discussion on the Causes and Countermeasures of Water Environment Problems in Linxia Prefecture[J]. Agricultural Science & Technology, 2017 (2): 36-37.
[13] Lin Yanhe. Investigation and Analysis of Water Environment Status and Pollution Causes in the Daxia River Basin[J]. People's Yellow River, 2015, 37(1): 74-78.
[14] Mao Haiying, Feng Zhongke, Gong Yuixi, Yu Jingxin. Study on Soil Normalized Water Index of Yongding River by Multispectral Remote Sensing Technology Combined with Genetic Algorithm[J]. Spectroscopy and Spectral Analysis, 2014, 34(06): 1649-1655.
[15] Wang Haiping, Feng Zhongke, Sun He, Li Ying, Cao Menglei. Study on Land Use/Cover Change in Yongding River Basin of Beijing[J]. Forest Investigation Planning, 2011, 36(03):15-20.
[16] Zheng Peichao, Qi Xiang, Wang Jinmei, Yang Rui. Study on liquid cathode glow discharge-atomic emission spectroscopy based on multiple linear regression[J]. Spectroscopy and Spectral Analysis, 2017, 37(10):3209-3213.