Scientific and Social Research, 2022, Volume 4, Issue 1
http://journals.innosciencepress.com/index.php/ssr
ISSN Online: 2661-4332

Establishment and Evaluation of Baseline Model of Heavy Metal Pollution in Mine Topsoil

Xiaohong Xu*

Anhui University of Finance and Economics, Bengbu 233000, China

***Corresponding author:** Xiaohong Xu, 1193637814@qq.com

Abstract: The elements polluted by heavy metals in the surface soil of different regions differ from one another from the perspective of the earth environment. The parts of the mine research data set from recent years are used as a reference for the mine topsoil, and the baseline model of heavy metal pollution in the mine surface soil is calculated using scientific and reasonable methods, which is critical for environmental pollution control. In this paper, the numbers of heavy metal elements in mine topsoil are established by using statistical theory and method: No. 1 As (arsenic), No. 2 Cd (cadmium), No. 3 Co (cobalt), No. 4 Cr (chromium), No. 5 Cu (copper), No. 6 Hg (mercury), No. 7 Mn (manganese), No. 8 Ni (nickel), No. 9 Pb (plumbum) and No. 10 Zn (zinc). This paper uses SPSS25.0 software to comprehensively analyze and process data, create a model using linear regression analysis, and thoroughly examine and evaluate overall heavy metal pollution in mine surface soil. The report then assesses the baseline model of heavy metal pollution in mine surface soil, provides reasonable treatment recommendations based on potential ecological hazards, delves into the heavy metal pollution of mine surface soil in depth, and offers some pollution control and treatment recommendations.

Keywords: Mine topsoil; Heavy metal pollution; Factor analysis; SPSS25.0 software

Online publication: January 20, 2022

1. Introduction

Heavy metal pollution in mine soils has become a severe concern in recent years. It is required to construct a set of baseline models to completely evaluate the degree of pollution and the weight of different elements through the study of heavy metal contamination in the soil. This work uses the possible ecological hazard evaluation approach developed by Swedish researcher Hakanson as a source of data, develops a model using linear regression, then analyses and discusses the results.

2. Existing problems and data sources

With the continuous development and mining of mines and the increasing impact of human activities on topsoil and environmental quality, the environmental problems caused by heavy metal pollution in topsoil have attracted more and more attention. We consulted the data, investigated the topsoil geological environment of a mine, and analyzed the collected soil with special instruments to obtain the data of various heavy metal elements contained in each sample.

This paper refers to the elements of the mine literature data set in recent years and the mine environmental geochemical baseline at home and abroad, takes the potential ecological hazard evaluation method established by the Swedish scholar Hakanson [1-2] as an effective method for evaluating heavy metal pollution and ecological hazard, applies the principle of sedimentology [3], and based on the on-site collection, testing and analysis of mine topsoil samples, studies the environmental geochemical characteristics of heavy metal pollution elements As (arsenic), Cd (cadmium), Co (cobalt), Cr (chromium), Cu (copper), Hg (mercury), Mn (manganese), Ni (nickel), Pb (plumbum) and Zn (zinc) in the topsoil in this area. Combined with the reference values, the pollution toxicity coefficients of 10 heavy metal elements are calculated as follows:

As=10, Cd=30, C0=2, Cr=2, Cu=5, Hg=40, Mn=9, Ni=10, Pb=5, Zn=1

Referring to the potential ecological hazard index and classification relationship table of heavy metal pollution, the environmental pollution factors of 10 kinds of heavy metal elements were established by linear analysis method to form the basic model of heavy metal pollution baseline in mine topsoil. According to the mine environmental geochemical baseline and the hazard degree of each heavy metal in each urban area, the pollution intensity of heavy metals in mine topsoil divided into five grades was established and evaluated. Suggestions on pollution control and treatment were put forward.

3. Basic hypothesis

(1) Hypothesis 1: The data of the sampling point fully reflect the basic situation of heavy metal pollution on the topsoil of the mine.
(2) Hypothesis 2: The data sources are true, reliable and statistically significant.
(3) Hypothesis 3: It ignores the effect of abnormal state on the concentration of heavy metals.

4. Linear regression analysis

Table 1. Results of Linear Regression Analysis (n=10)

	Unstandardized Coefficients		Standardized Coefficients	t	p	VIF	R^2	Adjust R^2	F
	B	Standard error	Beta						
Constant	9.988	2.033	-	4.914	0.008**	-	0.913	0.804	F(5,4)=8.393, p=0.030
Pollution level 1	-0.065	0.145	-0.346	-0.451	0.676	27.087			
Pollution level 2	0.510	0.210	2.460	2.424	0.072	47.336			
Pollution level 3	0.577	0.254	3.414	2.274	0.085	103.599			
Pollution level 4	-2.022	0.375	-13.976	-5.399	0.006**	307.999			
Pollution level 5	1.002	0.203	8.135	4.933	0.008**	125.038			

Dependent variable: number; D-W value: 1.285; p<0.05 ** p<0.01

Regression analysis is used to study the influence relationship between X (quantitative or categorical) and Y (quantitative), whether there is an influence relationship, and how about the influence direction and degree. First, analyze the model fitting. Analyze the model fitting through the R-square value and analyze the variance inflation factor (VIF) value to judge whether there is a collinearity problem in the model and solve it by stepwise regression. Second, get the model formula. Third, analyze the significance of X. If it is significant (p value is less than 0.05 or 0.01), it indicates that X has an influence on Y, and then analyze the direction of the influence relationship. Fourth, compare and analyze the influence of X on Y combined with the value of regression coefficient B.

Before regression analysis, box diagram can be used to check whether there is abnormal data, or scatter diagram can be used to visually show the correlation between X and Y. After regression analysis, the normality of the saved residual values can be observed and displayed using the normal diagram. Otherwise, the scatter chart can be used to observe and display the heteroscedasticity of the regression model. If the residual has no relationship with the scatter between X, there is no heteroscedasticity.

It can be seen from **Table 1** that taking pollution level 1, pollution level 2, pollution level 3, pollution level 4 and pollution level 5 as independent variables and number as dependent variable for linear regression analysis, the model formula is number = 9.988-0.065 * pollution level 1 + 0.510 * pollution level 2 + 0.577 * pollution level 3-2.022 * pollution level 4 + 1.002 * pollution level 5, and the R square value of the model is 0.913, which means pollution level 1, pollution level 2, pollution level 3, pollution level 4 and pollution level 5 can explain the reason of 91.3% change of the number. It is found that the model passes the F-test when performing the F-test on the model ($F = 8.393$, $p = 0.030 < 0.05$), which means that at least one of pollution level 1, pollution level 2, pollution level 3, pollution level 4 and pollution level 5 will have an impact on the number. In addition, according to multicollinearity of the model, if the value of VIF in the model is greater than 10, it means that there is a collinearity problem. Ridge regression or stepwise regression can be used to solve the collinearity problem. Meanwhile, it is recommended to check the closely related independent variables, eliminate the closely related independent variables and re analyze them.

The final specific analysis shows that:

(1) The regression coefficient value of pollution level 1 is -0.065 ($t = -0.451$, $p = 0.676 > 0.05$), which means that pollution level 1 will not affect the number.

(2) The regression coefficient value of pollution level 2 is 0.510 ($t = 2.424$, $p = 0.072 > 0.05$), which means that pollution level 2 will not affect the number.

(3) The regression coefficient value of pollution level 3 is 0.577 ($t = 2.274$, $p = 0.085 > 0.05$), which means that pollution level 3 will not affect the number.

(4) The regression coefficient of pollution level 4 is -2.022 ($t = -5.399$, $p = 0.006 < 0.01$), which means that pollution level 4 will have a significant negative impact on the number.

(5) The regression coefficient value of pollution level 5 is 1.002 ($t = 4.933$, $p = 0.008 < 0.01$), which means that pollution level 5 will have a significant positive impact on the number.

According to the summary and analysis, pollution level 5 will have a significant positive impact on the number and pollution level 4 will have a significant negative impact on the number. However, pollution level 1, pollution level 2 and pollution level 3 will not affect the number (view **Figure 1** and **Figure 2** below).

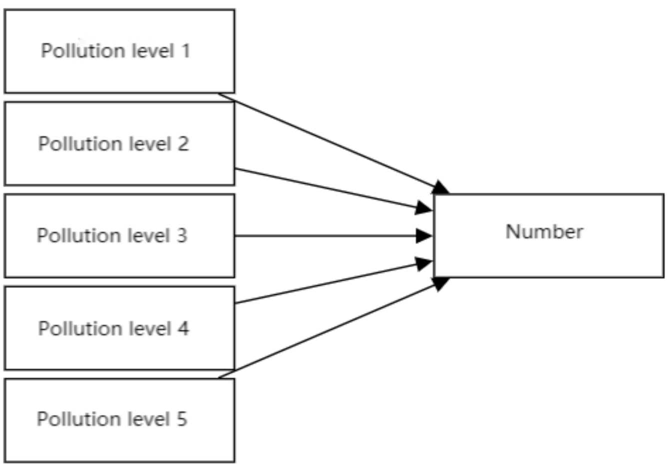

Figure 1. Model

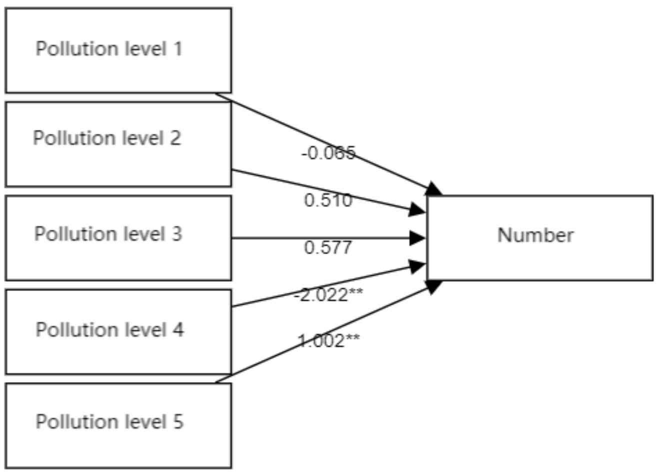

Figure 2. Result of model

Table 2. Results of linear regression analysis - simplified format

	Regression coefficient	95% CI	VIF
Constant	9.988** (4.914)	6.004 ~ 13.972	-
Pollution level 1	-0.065 (-0.451)	-0.349 ~ 0.218	27.087
Pollution level 2	0.510 (2.424)	0.097 ~ 0.922	47.336
Pollution level 3	0.577 (2.274)	0.080 ~ 1.075	103.599
Pollution level 4	-2.022** (-5.399)	-2.756 ~ -1.288	307.999
Pollution level 5	1.002** (4.933)	0.604 ~ 1.400	125.038
sample capacity		10	
R ²		0.913	
Adjust R ²		0.804	
F value		F (5,4)=8.393,p=0.030	

Dependent variable: number; D-W value: 1.285; p<0.05 ** p<0.01 (in brackets is t value)

Table 3. Model summary (intermediate process)

R	R ²	Adjust R ²	RMSE	DW value	AIC value	BIC value
0.955	0.913	0.804	0.847	1.285	37.065	38.881

Regression analysis is used to study the influence relationship between X (quantitative or categorical) and Y (quantitative), whether there is an influence relationship, and how about the influence direction and degree. First, the R-square value of the model is used to analyze the fitting of the model. For example, 0.5 means that the fitting degree of the model is 50%. Second, adjust the R-square value to punish arbitrarily placing too many X, which is usually less used. It can be seen from **Table 3** that taking pollution level 1, pollution level 2, pollution level 3, pollution level 4 and pollution level 5 as independent variables and number as dependent variable for linear regression analysis, the R square value of the model is 0.913, which means that pollution level 1, pollution level 2, pollution level 3, pollution level 4 and pollution level 5 can explain 91.3% of the change reasons of number [4].

Table 4. ANOVA table (intermediate process)

	Square sum	*df*	Mean square	*F*	*p* value
Regression	75.321	5	15.064	8.393	0.030
Residual	7.179	4	1.795		
Total	82.500	9			

F test is used to test whether the regression model is meaningful. First, if the model passes the F test ($p < 0.05$), it shows that the model is meaningful and at least one X will affect Y. Second, if the model fails to pass the F test ($p > 0.05$), it indicates that the model construction is meaningless, and X will not affect Y. It can be seen from **Table 4** that the model passed the F test ($F = 8.393$, $p = 0.030 < 0.05$), which means that the model construction is meaningful.

Table 5. Regression coefficient (intermediate process) (n=10)

	Unstandardized Coefficients		Standardized Coefficients	t	p	95% CI	VIF
	B	Standard error	Beta				
Constant	9.988	2.033	-	4.914	0.008**	6.004 ~ 13.972	-
Pollution level 1	-0.065	0.145	-0.346	-0.451	0.676	-0.349 ~ 0.218	27.087
Pollution level 2	0.510	0.210	2.460	2.424	0.072	0.097 ~ 0.922	47.336
Pollution level 3	0.577	0.254	3.414	2.274	0.085	0.080 ~ 1.075	103.599
Pollution level 4	-2.022	0.375	-13.976	-5.399	0.006**	-2.756 ~ -1.288	307.999
Pollution level 5	1.002	0.203	8.135	4.933	0.008**	0.604 ~ 1.400	125.038

Dependent variable: number; p<0.05 ** p<0.01

5. Evaluation of baseline model of heavy metal pollution in mine topsoil

5.1. Evaluation

From the results of the baseline model of heavy metal pollution in the surface soil of the mine (**Figure 2**), it can be seen that although the respective proportions of polluted heavy metals are different, they have caused varying degrees of pollution to the mine topsoil. From the distribution of level 5 pollution, the greater the level is, the more serious the pollution is. It can be seen that the level 3 and 4 pollution is very serious. At the same time, among As (arsenic), Cd (cadmium) and Co (cobalt), Cr (chromium), Cu (copper), Hg (mercury), Mn (manganese), Ni (nickel), Pb (plumbum) and Zn (zinc), the four, As (arsenic), Cd (cadmium), Hg (mercury) and Ni (nickel), have the highest pollution weight. Obviously, these four elements play an important role in environmental pollution.

5.2. Suggestions on control and treatment of heavy metal pollution in mine topsoil

Hg and Cd are the most harmful heavy metal contaminants on mine soils, and there are regions where human activities are taking place. Additional than mines, there might be other coal-fired power stations, Hg electroplating operations, or non-ferrous metal plants near these sites. We must regulate the heavy metal contamination in the mine's surface soil after becoming familiar with it. The fraction of four heavy metals, namely As (arsenic), Cd (cadmium), Hg (mercury), and Ni (nickel), is now the highest (nickel). As a result, it is vital to determine the source of contamination. If pollution is being generated by relevant factories, it must be addressed through consultation and collaboration with the appropriate department. For the surface soil in the mine area, some chemical methods can be used to prepare a corresponding precipitant for treatment, and real-time monitoring points can be set up on a regular basis to detect heavy metal concentrations, with early warning given to those with large fluctuations in dynamic monitoring [5]. Meanwhile, in the mining field, we should focus on the proper categorization and treatment of residential waste in order to minimize the environmental source of heavy metal pollution produced by trash dumping, and utilize contemporary clean energy as much as possible to reduce hazardous gas emissions such as coal.

Disclosure statement

The author declares no conflict of interest.

References

[1] Hakanson L, 1980, An Ecological Risk Index for Aquatic Pollution Control a Sedinen to Logical

Approach. Water Research, 14(8): 975-1001.
[2] US Environmental Protection Agency, 2019, Exposure Factors Handbook. EPA/600/P-95/002, 104-126.
[3] Tang J, Chen C, Li H, et al., 2019, Potential Ecological Hazards and Health Risk Assessment of Soil Heavy Metals in Built-Up Areas of Daqing City. Geosciences, 3(1).
[4] Zhu W, Chen H, Liu Z, et al., 2018, Study on Surface Water Pollution by Factor Analysis. Journal of Fudan University, 40(6).
[5] Li Q, Huang Y, 2017, Changes and Pollution of Heavy Metals in Vegetable Soil in the Suburbs of Chongqing. 33(2).