

A locating method based on the Anselin elocal spatial autocorrelation model which researches in the heavy metal pollution source

Jinming Fu^{1, a}

¹College Of Science, Huazhong Agricultural University, 430070, Wuhan, China

Yuanfan Li ^{1, a}(Corresponding Author)

² School of Foreign Languages ,Wuhan Textile University ,430073,Wuhan, China

Keywords : Pollution source, Cokriging , Anselin spatial clustering

Abstract. Based on the given data of As content in the soil ,using Log transformation, the Box-Cox transformation to exclude outlier for preprocessing data to meet the normal distribution.Using Anselin elocal spatial autocorrelation model to carry on local spatial clustering in order to cluster the similar attribute values to a class.Preliminary view is that the sources of pollution distribute within the region of tne distribution of these points.The thesis build Cokriging interpolation model to interpolate of the global by the ArcGis software in order to higher concentration of the surface domain.The similar points obtained by spatial clustering properties fall on the higher concentration of the surface domain obtained by Cokriging interpolation is the region for the location of pollution sources.We focus on the analysis of As,finalize two As sources of pollution of surface domain,one of the sources of pollution range is (9277,11121), (16148,16432)]andthe other is [(3573, 4777),(6213,4897)].Then the artical uses the cross validation error analysis methods to test Cokriging interpolation model and found that cross-validation result is very good, model checking has reached a certain accuracy.

Introduction

With the rapid development of urban economy and the constant increase of urban population , the impact of human activities on urban environmental quality is becoming more and more obvious .So the way to find Urban Pollution to eliminate pollution problems from the source, is increasingly becoming the focus of attention.Cokriging model in geostatistics is a good way to analysis that both have both structural and random data.Be able to analyze the interdependence and the regional between multiple soil properties ,the model is widely applied to analysis spatial variability of soil properties[5]、 [6]、 [7]、 [8] .it is also widely applied in the field of hydrology, meteorology, environmental protection, precision agriculture and geological exploration and so on.Spatial clustering analysis is a very important aspect of spatial analysis,it is to describe the spatial variables and spatial characteristics of the object from the perspective of the overall, global. reflect the Information of the location of groups of similar space things with the purpose to analysis the cluster of space objects to reveale a certain geographical mechanisms or as the basis of other analyzes.Now most of the studies is the use of a single method ,the thesis based on the actual issues to combine the two methods reasonably to ensure the result is more accurate. Known by literature knowledge, heavy metal pollutants are broadly spread from the direction of the high concentrations to low concentrations and spread from the high altitude location to the low-lying position,and its propagation speed is slower. So we can get the characteristics of pollution sources is that they compared to the surrounding content of polluting elements with the large difference,and the concentration changes uniformly. Based on this characteristic, this article utilizes Anselin local spatial autocorrelation model to cluster of heavy metal pollutants to find the points which have the similar property values.Then use the the Cokriging model to identify the location of the sources of pollution .

Anselin spatial autocorrelation clustering model

Spatial autocorrelation refers to the same variable in different spatial positions, is a measure of the

spatial unit attribute value of the degree of aggregation. In the spatial autocorrelation theory, the most commonly used indicators are the local Moran index and the G index. The former is proposed by Anselin, it is defined

$$I_i = \frac{x_i - \bar{x}}{S_i^2} \sum_{j=1, j \neq i}^n w_{i,j} (x_j - \bar{x})$$

Among them, x_i is the attribute values of spatial unit i , w is space matrix, $w_{i,j}$ represents the degree of influence of the spatial units i and j ,

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n w_{ij}}{n-1} - \bar{X}^2$$

Positive I_i means the property values of the space unit and adjacent units are similar as well as negative I_i indicates the attribute values of the space unit and adjacent units are not similar.

I_i can be standardized by the following formula,

$$Z(I_i) = \frac{I_i - E(I_i)}{\sqrt{\text{VAR}(I_i)}}$$

and $E(I_i)$ is the theoretical expectations as well as $\text{Var}(I_i)$ is the theoretical variance

Cokriging model

Assuming that there is a group of synergic and regionalized variables which can be characterized by a set of spatially random functions (i.e. $Z_k(x)$). In the hypothesis of second-order stationary, get:

$$E\{Z_k(x)\} = m_k$$

Cross Covariance is:

$$C_{k'k}(h) = E\{Z_{k'}(x+h) \cdot Z_k(x)\} - m_{k'} \cdot m_k$$

Variogram is:

$$\gamma_{k'k}(h) = \frac{1}{2} E\{[Z_{k'}(x+h) - Z_{k'}(x)] \cdot [Z_k(x+h) - Z_k(x)]\}$$

Assuming that there is a main variable K_0 , which is to be measured in the K of regionalized variables. The mean of main variable $Z_{k_0}(x)$ is $Z^*_{v_{k_0}}$ in the domain of V . The equation as follow:

$$Z^*_{v_{k_0}} = \sum_{k=1}^K \sum_{a_k}^{n_k} \lambda_{a_k} Z_{a_k}$$

Where n_k is the amount of valid data in the domain of V . Likewise, we can get the cokriging equations according to the validity and unbiased.

$$\left\{ \begin{array}{l} \sum_{k'=1}^K \sum_{\beta_{k'}}^{n_{k'}} \lambda_{\beta_{k'}} \overline{C_{k'k}}(v_{\beta_{k'}}, v_{a_k}) - u_k = \overline{C_{k_0k}}(v_{k_0}, v_{a_k}) \\ \sum_{a_{k_0}=1}^{n_{k_0}} \lambda_{a_{k_0}} = 1 \\ \sum_{a_k}^{n_k} \lambda_{a_k} = 0 \end{array} \right.$$

Where: $a_k = 1, L, n_k, k = 1, L, K, k \neq k_0,$

Cokriging variance is:

$$\delta_{V_{k_0}}^2 = \overline{C}_{k_0 k_0}(V_{k_0}, V_{k_0}) + u_{k_0} - \sum_{k'=1}^K \sum_{\beta_{k'}}^{n_{k'}} \lambda_{\beta_{k'}} \overline{C}_{k'k}(v_{\beta_{k'}}, v_{a_k})$$

The relationship between predictive value and standard value of pollution source as follow:

$$Z^*_{v_{k_0}} - c_0 \geq 0$$

Where $Z^*_{v_{k_0}}$ is predictive value, c_0 is standard value of pollution source.

The solution and examination

Both Anselin model and Cokriging model have strictly mathematical theories and algorithmic supports. The solution of the models depend on ArcGis10.0 software. According to the principle of geostatistics, we get the steps as follows:

Step1: To assure the initial data comply to the normal distribution of second-order stationary, we applied the QQ test chart or the KS test method. If it goes well, then go to Step2. If not, then take the measures of Log transformation, Box-Cox transformation and removing the abnormal data. To begin with, we utilize the program of SAS to do the test of normality for the data of As, only to find that it cannot comply to the normal distribution. So we take the measures of Log transformation and removing the abnormal data.

Step2: To draw trend analysis figure, if there is the apparent second-order trend, then remove it. There is no existence about second-order trend, so it is no necessary to remove the trend.

Step3: By utilizing the visualized ToolBox Anselin local Moran I of software ArcGis10.0, we get the result of Anselin spatial clustering

Step4: According to given sample data, we take a measure of Cokriging Interpolation and get the level distribution of heavy metal pollutions. Thus the scope of heavy metal pollutions can be located. Furthermore, taking an example of heavy metal As, we implement Anselin spatial clustering synchronized with Cokriging Interpolation.

By utilizing the visualized ToolBox Anselin local Moran I of software ArcGis10.0, we get the result in the figure 1.

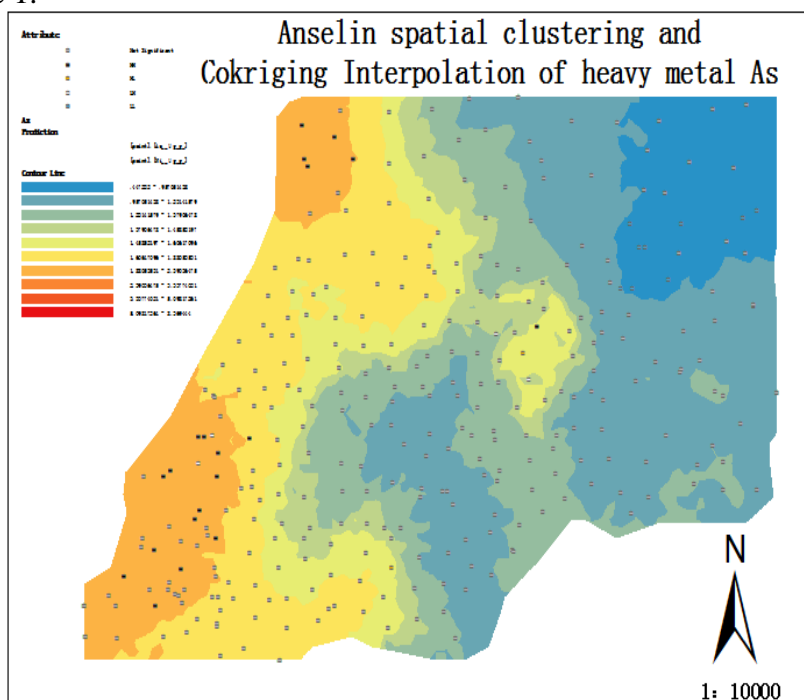


Figure 1: Anselin spatial clustering and Cokriging Interpolation of heavy metal As

Taken the intersection of the results of Anselin spatial clustering and Cokriging interpolation, accorded to the least square principle, selected the points which they property values were closed in the deepest region of the Cokriging interpolation, and ultimately determined that there were basically two As sources of pollution regions, which one ranged from [(9277, 11121),(16148,16432)], another ranged from [(3573, 4777),(6213,4897)].

The analyze the cross-validation error of method was usually taken to Cokriging model checking, that was through the part of known data to predict another part of data, then analyze its error. In this paper, the 318 datum of the heavy metals As were used to predict the 319th data, the prediction error as shown below table:

Prediction Errors	
Samples	319 of 319
Mean	-0.006714204
Root-Mean-Square	0.7830805
Mean Standardized	-0.02974327
Root-Mean-Square Standardized	1.259903
Standard Error of Mean	0.6514207

Table 2: Cross Validation

Cokriging interpolation predicted the results which always were unbiased, that was the standard average of the prediction error should be close to 0. It usually hoped that the uncertainty assessment had effective, that was the normalized root mean square error should be close to 1(Ref. [4]). If the normalized root mean square error is greater than 1, then it indicates estimate of the variability was too low in the forecast; If the normalized root mean square error is less than 1, then it indicates estimate of the variability was too high in the forecast. The figure showed, the standard of the mean almost closed to 0, the normalized root mean square almost closed to 1. So the result of cross-validation was better, the model checking had reached a certain accuracy.

Conclusions and application:

The Anselin local spatial autocorrelation clustering model not only considered the pollution concentration, but also included the spatial location at the same time, which made the results more objective. The model could also be used to solve the population growth rate of spatial association relations and other issues. The Cokriging Model overcame the problems which were difficult to analyze, it could do the best linear unbiased for the spatial distribution. We combined the two models, both full considered the characteristics of the distribution of pollution sources, and more precise positioned the position of the pollution sources.

References

- [1]. Zhang Songlin, Zhang Kun, Contrast Study on the Local Indices of Spatial Autocorrelation, Statistical Research, Vol.24 No7,Jul.2007
- [2]. Zhang Songlin, Zhang Kun, Comparison between Genera I Moran's Index and Getis-O rd General G of Spatial Autocorrelation, ACTA SCIENTIARUM NATURALIUM UNIVERSITATIS SUNYATSENI,Vol.46 No4, Jul.2007
- [3]. Xu jianming, Indices and Assessment of Soil Quality[M], Beijing: Science Press,2010.
- [4]. Wang xuejun, Li bengang, Spatial Analysis of Soil Trace Metal Content[M], Beijing: Science

Press,2005.

- [5]. Webster B,Burgess T M.Optimal interpolation and isarithmic mapping of soil properties III:Changing drift and universal kriging [J]. Soil Sci.,1981,31:505-524
- [6]. Hurvich CM,Simonoff JS,Tsai C-L . Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion[J]. J R Stat Soc 1998;60:271-293
- [7]. Leung Y,Mei CL,Zhang WX. Statistical tests for spatial nonstationarity based on the geographically weighted regression model[J]. Environ Plan 2000,32:9-32
- [8]. Mariusz Szymanowski,Maciej Kryza. Application of geographically weighted regression for modelling the spatial structure of urban heat island in the city of Wroclaw (SW Poland)[J].Mariusz Szymanowski and Maciej Kryza/Procedia Environmental Sciences ,2011,3:87 - 92.