Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Environmental Sciences 10 (2011) 1953 – 1960

Procedia
Environmental Sciences

2011 3rd International Conference on Environmental
Science and Information Application Technology (ESIAT 2011)

Study on Spatial Distribution of Soil Heavy Metals in Huizhou City Based on BP--ANN Modeling and GIS

Yin Li, Chao-kui Li*,Jian-junTao,Li-dong Wang

Institute of Geospatial Information Science, Hunan University of Science and Technology, Xiangtan 411201, China
yingxiao_sun@tom.com

Abstract

Using BP neural network model and GIS for heavy metal descript the spatial dynamics of distribution in Huizhou City, Guangdong Province, The Results indicate:(1)BP neural network model can learn the relationship between spatial location of sampling points with content of them and be able to forecast soil heavy metal content. by interpolating more sampling points ,such as Fig.3 and Fig.4, Cd from the original refuses the normal distribution to approximate submit to normal distribution, with the increase in sampling points, the data can be used in geostatistical analysis. the range of Cu was changed from the original 136.24km to 25.20 km, as TABLE 2. more realistic conditions.(2)Import the coordinates of the interpolation data and heavy metal content data into Arcgis, the most significant pollution is Pb, most of the area level over the natural background value .Zn, Cd have a high aggregation in some areas, the majority does not exceed the natural background value. Cu does not exceed the background, as shown in Figure 11 to 14.Therefore, the most noteworthy is the Pb pollution. Use the single factor index method, the four metal did not meet pollution threshold, indicating the study area have not been heavy metal pollution of soil.(3) Compared original data and interpolated data of the background index, as shown in Table 2 and Table 3, the nugget and sill ratio of Cu and Zn decreased, indicating that as interpolation points increase, Cu, Zn by the random factor decreases, increased spatial correlation, Pb nugget and sill ratio increases, the spatial correlation decreases, Cu by the random factors the strongest, spatial correlation the weakest, Zn and Pb is mainly affected by structural factors.

Soil is not a homogeneous body, rather a variant of spatial continuity, this highly complex spatial heterogeneity of soil allows us to study the spatial dynamics of change becomes very difficult, caused not only by nature but also by human activities. These changes are the spatial dynamics of Soil heavy metals(SHM) can become extremely complex and, thus, the description on spatial distribution of SHM and spatial correlation of quantitative is very difficult. In the past there have been many models trying to determine the relationship between them, but these models are more or less the existence of defects[1]. In this paper, an unconventional Modeling Artificial Neural Networks method, combined with GIS technology to the spatial correlation of SHM and its spatial distribution were studied. Using neural network model to determine spatial distribution and pollution of SHM.

Keywords: BP artificial neural network model; GIS technology ;Soil heavy metal; Huizhou city; Spatial distribution; Spatial interpolation; Geostatistics.

doi:10.1016/j.proenv.2011.09.306

Ⅰ. Research Methods and Data Sources

Heavy-metal contents in soil and its spatial position exists between the highly complex nonlinear relationship, should not use the conventional modeling method to solve the problem. The artificial neural network to deal with a "black box" characteristics of the problem [1].Using artificial neural network model to study this relationship, the establishment of various heavy metals and their spatial relationship between the mapping model. Therefore, use this back-propagation neural networks model [2,4] to the relationship between the above studies, using GIS technology to achieve a large area of the spatial distribution of SHM pollution monitoring, mapping and calculating the distribution of heavy metals in some important dynamic information.

Back-Propagation(BP)Neural Network

Artificial Neural Network, is a nonlinear contains many simple computational units, and BP network is one of the most widely used. Commonly BP neural network comprises input, output and hidden layers. Neurons between adjacent layers are interconnected by a weighting factor. BP network layers by learning to modify the connections between neurons weights, so the final error could to minimum.

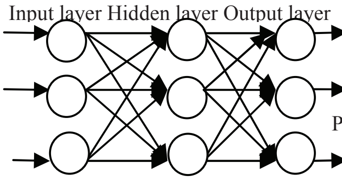


Fig.1 BP network structure

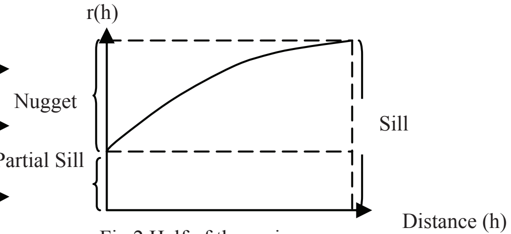


Fig.2 Half of the variograms

Geostatistics Model

Geostatistics, was put forward by a famous French statistician G. Matherom, based on a large number of theoretical studies and the regionalized variable theory format a new statistical statistical branch, the semi-variance function as the main tool to study those Study both random spatial distribution of structural[5]. First, the sample is subject to the normal distribution assumption, if it does not, deal with data transformation, semi-variogram Expression:

$$r(h)= \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[(Z(x_i)-Z(x_i+h)]^2 \qquad (1)$$

Semi-variance function can be fit by spherical model, exponential model and the Gaussian curve fitting models, three basic parameters are:C0,C,and R,as shown in Figure 2,C0 called Nuggets, reflecting Random factors or uncertainties related to the impact of variable space; C called the partial Sill, reflects structural factors or variables to determine factors on the regional impact of spatial autocorrelation; when the distance between sampling points increases h, Semi-variance function r(h) from the initial value of the nugget constant reaches a relatively stable, the value is called the value Sill,the value of C0+ C,said maximum variation properties or regional variables, the greater that ,the overall higher level of spatial heterogeneity[5].Nugget to sill ratio of the value that the variability between samples, the higher the value, indicating more variation between samples is caused by random factors, the weaker space correlation. R is called Range, reflecting the regionalized variables in space relevant range .For a good prediction model should meet Mean Standardized closest to 0,Root-Mean-Square minimum, Average Mean Error closest to the Root-Mean-Square, Root-Mean-Square Standardized closest to 1[6].

Data Sources

The data used in this article from Huizhou City, Guangdong Province .The city a total of 104 sampling points, using GPS for precise positioning,and gives the total Cu in soil, total Zn, Total Ni , total chromium Cr, total Pb, total Cd, total As, total Hg content [7]. we only selected Cu, Zn ,Pb, Cd of four metals for the study. in the use of the data before the data carried out by removing the threshold (the average standard deviation of plus or minus three times the range), not the range after excluding the data point with the maximum and minimum values instead of, the effective point after the number was 102.Cd by the logarithmic (Log) conversion and power (Box-Cox) transformation not follow a normal distribution.

Table 1 Soil environment quality standard (GB15618-1995) (unit: mg/kg)

	Level 1	Level 2			Level 3
	Natural background	*pH*<6.5	*Ph*6.4~7.5	*pH*>7.5	*pH*>6.5
Cd≤	0.2	0.3	0.3	0.6	1.0
Cu farmland≤	35	50	100	100	400
garden≤	——	150	200	200	400
Pb ≤	35	250	300	350	500
Zn dry armland≤	100	200	250	300	500

Data Processing and Analysis

In order to enhance the effect of BP network training and validation of network Generalization, the original data, including the coordinates of sampling points and heavy metal data can not be directly used as network input and output, to them Pretreatment of .104 is equivalent to sampling sites after removing two of the 102 samples were divided by the respective raw data of heavy metals of secondary standard value, as shown in Table 1,to calculate each of the secondary standard index of heavy metals. Statistical analysis of results in Table 2, available, and treatment than before, coefficient of variation of the four metals are larger, indicating that the data were processed larger degree of dispersion .As the degree of dispersion and variability of data is small, the interpolation accuracy will increases, whereas the opposite[1],so from the original value as the nodes directly used for interpolation.

Standardization of Data Coordinate Values and Heavy Metal

In order to improve the training of the network effect, in Matlab7.0 using p (:, i) = (p (:, i)-min (p (:, i)))/(max (p (:, i)) - min (p (:, i))) function to coordinate of all the sample points and content of the heavy metals to [-1, 1],and then complete the restore to the original dimension and quantity . In ordinary Kriging method with the analysis of Cu and Zn with the Log transform, Pb follow a normal distribution, not through the transformation, according to the standard mean square forecast error should be close to a root principle, Cu, Zn with the exponential model, Pb with the spherical model fit. from Table 2 have to deal with before and after Cu of the range was unchanged, Zn (Zn) and Pb in the range was larger, indicating that the correlation between samples larger range. Cu and Pb nugget and sill ratio becomes larger, Zn become smaller, indicating that the Cu and Pb variation between samples more by random factors. As nugget and the sill by its own factors and the impact of a larger unit of measurement can not be used to compare between different variables with the differences in terms of progress, But the ratio of nugget and sill nugget reflects the total t variance of spatial heterogeneity is very significant variability[6],comparing the ratio of nugget of the original data of Cu, Zn, Pb and sill values were 76.93%, 50% and 20.60%,indicating that Cu content of the spatial distribution of the random factors, the greatest impact, and Zn, followed by Pb minimum also reflects the Pb of the spatial correlation of the strongest, Zn next, followed by Cu. Structural factors such as parent material, soil type, climate, soil-forming factors; random factors such as farming, management measures, pollution and other human activities. structural factors weakened the Zn, Pb, particularly Pb of the spatial correlation. in the study area, Pb of the nugget(C0) goes to 115.29,indicating that Pb by random factors and uncertainties that affect the large,Pb is an easy migration of heavy metals, contamination can cause partial large concentration of heavy metals. In addition, research District Cu of the range was large, as 136.24Km,that Cu content of the spatial correlation of the range. and that the change process is 136.24 km and the actual non-compliance, may be too little concerned with the number of sampling points.Zn,Pb of the change process were 13.54Km and 20.60Km, that Zn, Pb content of a range in a smaller memory related.

Table 2 Statistic analysis of secondary standard index of Huizhou soil sample point

Statistics	Cu	Zn	Pb	Cd
Mean	16.12/0.31	55.61/0.28	44.631/0.18	.0961/0.32
Median	12.161/0.21	47.41/0.22	42.40/0.16	0.07/0.23
Mode	5.28/0.11	166.30/0.20	23.41/0.10	0.04/0.13
Standard deviation	1.18/0.27	3.13/0.19	1.75/0.07	0.01/0.48
Coefficient of variation	0.74/0.90	0.57/0.67	0.40/0.42	1.40/1.49
Nugget(C_0)	0.42/0.54	0.15/0.18	$115.29/3.35\times10^{-3}$	——
Sill(C_{0+} C_1)	0.55/0.587	0.3/0.37	$310.44/5.67\times10^{-3}$	——
Range[km]	136.24/25.20	13.54/24.22	20.60 /27.46	——
C_0/C_0+C_1	76.36%/91.99%	50.00%/48.65%	37.14%/59.08%	——
Theory model	Exponential model	Exponential model	Spherecial model	——
Site several	102	102	102	102

Note: The symbol "/" on both sides stand for deal with before and after values.

Spatial Interpolation based on BP network model

Spatial interpolation is a known point by point data to predict the unknown method of data, but also effectively point data through encryption of data changes into the technical side. In setting the appropriate frame deletion unit, the range in the study area estimates within desired results. In this paper, known as the sampling points, will undergo training and testing the BP network model for spatial interpolation, the interpolation function is y = [...]; x = [...]; [b, i , j] = unique (x); x1 = 0:0.025:1; y1 = interp1 (b, y (i), x1, 'linear'), interpolated for each node (interpolation point) after a conversion must correspond to a set of space coordinates in the sequence of coordinates from the 102 into 142, shown in Figure 4.above them as input to the network after training, the use of sim () function to obtain the interpolation point 4 heavy metal content of the forecast results. neural network creation and training function is t = []; p = []; threshold = [0 1; 01]; net = newff (threshold, [206,4], {'tansig', logsig '},' traingdx '); net = init (net); net =

train (net, t, p); y = sim (net, t); y = sim (net, t_test); by interpolating the four coordinates heavy metal content of the standardized value. to determine the spatial dynamics of distribution of various SHM. In order to test the accuracy of the network,the network output and the actual measured value after standardization for regression analysis (Figure 5 ~ 8), where A is the network output, T the actual measured value, R is the output value and the correlation coefficient between the actual value, can be seen from the figure the network after training with very good generalization performance.

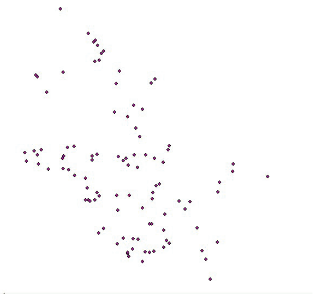


Fig.3 Sample point distribution

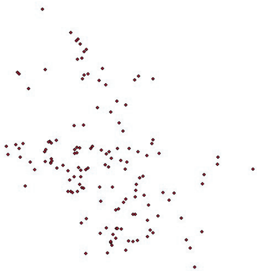
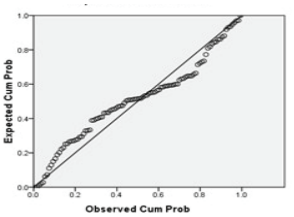
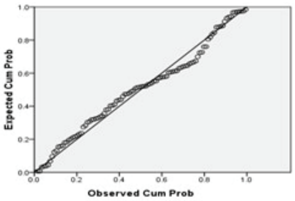


Fig.4 Sample pointdstribution after treatment



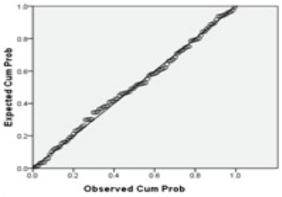
R=0.808 A=1.113T-0.028↵

Fig.5 Cu BP network training and generalization results



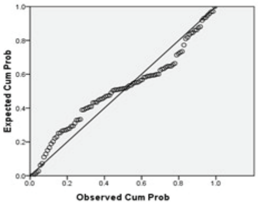
R=0.845 A=1.144T-0.03↵

Fig.6 Zn BP network training and generalization results



R= 0.884 A=1.117T-0.062↵

Fig.7 Pb the training of BP network and generalization results

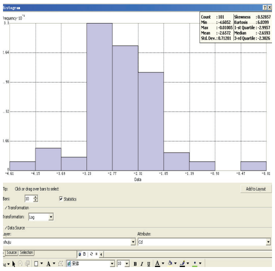


R=0.808 A=1.113T-0.028↵

Fig.8 Cd BP network training and generalization results

Spatial Dynamic Distribution of SHM in Huizhou City

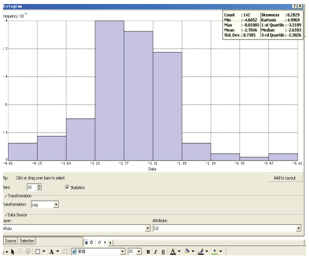
The above process will be in all sites(including known points and interpolation points) coordinate data and heavy metal content data to restore the original dimension and magnitude, import them into GIS, then all sites the value of 4 kinds of heavy metals divided by the natural background values of their respective categories, divided into more or less than two parts of the natural background value.Cu and Zn with the Log transform, order of removal by Second removed, Pb su bject to normal distribution, not through the transformation, Cd transformation does not follow a normal distribution, compared to approximate normal distribution before interpolation, Figure 6 and 7 were interpolated interpolation before and Cd Histogram distribution of samples graph if the interpolation points can be achieved more normal distribution, here with the inverse distance weighting interpolation, Cu , Zn , Pb site statistics interpolation, calculation and analysis using Arcgis obtained spatial dynamics of Nantong City, SHM maps and spatial dynamics of some important info rmation,Figure 8 to Figure 11 for the BP network model based on Spatial Interpolation of Cu , Cd , Zn , Pb spatial distribution, using Arcgis analysis obtained in the SHM Huizhou City is the most significant pollution of Pb ,most of the area level over the natural background value. Zn , Cd in some areas have a high aggregation of more than the natural background value, however, the size of the scope of pollution are relatively small, most no more than the natural background value, Pb, Zn, Cd was the regional distribution of island. Cu does not exceed the background value of the area, north west corner of Cu elements were relatively high, the southeast corner of Cu content of the element in the low state. Therefore, the most noteworthy is Pb contamination. and the environmental quality of the soil standards in the secondary standard (GB15618-1995),as shown in Table 1,use the single factor index method, the area did not meet the four metal pollution threshold, indicating the study area of heavy metals in the soil has not been polluted.



Count : 101
Min : -4.6052
Max : -0.01005
Mean : -2.6572
Std. Dev. : 0.71201
Skewness : 0.52857
Kurtosis : 6.0399
1-st Quartile : -2.9957
Median : -2.6593
3-rd Quartile : -2.3026

Tip: Click or drag over bars to select
Bars: 10 ☑ Statistics
Transformation: Log
Layer: shuju
Attribute: Cd

Fig.9 Former histogram Interpolation for Cd



Count : 142
Min : -4.6052
Max : -0.01005
Mean : -2.7046
Std. Dev. : 0.7385
Skewness : 0.2829
Kurtosis : 4.9969
1-st Quartile : -3.2189
Median : -2.6593
3-rd Quartile : -2.3026

Tip: Click or drag over bars to select
Bars: 10 ☑ Statistics
Transformation: Log
Layer: shuju
Attribute: Cd

Fig.10 After histogram Interpolation for Cd

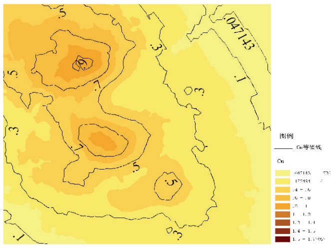


Fig.11 The spatial distribution of soil Cu in Huizhou city

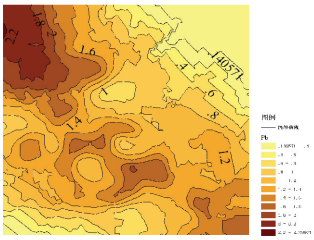


Fig.12 The spatial distribution of soil Pb in Huizhou city

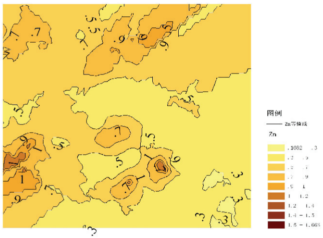


Fig.13 The spatial distribution of soil Zu in Huizhou city

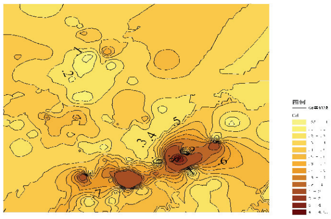


Fig.14 The spatial distribution of soil Cd in Huizhou city

Table 3 Statistic analysis of background index of Huizhou soil sample point

Statistics	Cu	Zn	Pb	Cd
Nugget *(C_0)*	0.28	0.07	0.08	——
Sill *(C_{0+} C_1)*	0.38	0.22	0.21	——
Range*[km]*	25.20	13.44	17.08	——
C_0/C_0+C_1	0.74	0.32	0.38	——
Theory model	Spherecial model	Exponential model	Spherecial model	——
Site several	142	142	142	142

In Table 3 the ratio of nugget and sill and compared to the original data in Table 2, Cu and Zn values become smaller, indicating an increase with the interpolation points ,Cu ,Zn by the random sampling points decreases, increased spatial correlation, Pb nugget and sill ratio increased from the original 37.14% to 38%, the spatial correlation weakening. by Table 1 and Table 2 can be drawn Cu the most affected by random factors, the weakest spatial correlation, random factors such as farming and management practices, cropping systems, pollution and the impact of human activities. Zn and Pb is mainly affected by structural factors, structural factors such as parent material, soil type, climate, soil-forming factors; random factors such as farming and management practices, cropping systems, pollution and other human activities. With the increase of interpolation points, Cu reduced the maximum range was changed to 25.20 km, more realistic conditions than 136.24km,it can be obtained by interpolation more accurate results, Zn ,Pb the maximum range was little change.

Conclusion

(1)Through interpolation more sampling points, Cd from the fully not satisfied with the normal distribution to approximate submit to normal distribution, indicating that the increase with the sampling points, the data follow a normal distribution, can be used in statistical analysis. Cu from the 136.24km variable range into a 25.20 km, more realistic conditions.

(2)Import coordinate data and heavy metals of interpolation points into Arcgis and analysis them, the most significant pollution is Pb, most of the area over the natural background value. Zn, Cd in some areas have a high aggregation of more than the natural background value, but the scope are relatively small, most no more than natural background value. Cu does not exceed the background value. Therefore, the paper argues that in Huizhou City, the most noteworthy is Pb pollution. Use the single factor index method the four metal in the study area did not meet pollution threshold, indicating the study area the soil is not contaminated by heavy metals.

(3)Before processing the data, interpolated data compared to Cu and Zn of the nugget and sill ratio decreased, indicating increased with the interpolation points, Cu, Zn By the random sampling points decreases, increased spatial correlation, Pb nugget and sill ratio increases, the spatial correlation weakening. Cu the most affected by random factors, the space the weakest correlation, random factors such as farming, management practices, cropping systems, pollution and other human activities.Zn and Pb is mainly affected by structural factors, such as parent material, soil type, climate, soil-forming factors; random factors such as farming, management practices, pollution and other human activities.

Acknowledgement

The paper is supported by National Natural Science Foundation of China (41075046)and China Environmental Science Research Institute of Soil Quality Information Management Information System

(D10870).Corresponding Author is Chao-kui Li (1967-),Male, Hanshou, Hunan Province, Professors. Engaged in the acquisition and application of geospatial information, his email is chkl_hn@163.com.

References

[1]Da-wei Hu,Xin-ming Bian,Shu-yu Wang,Wei-guo Fu :Study on spatial distribution of farmland soil heavy metals in Nantong City based on BP-ANN modeling. Journal of Safety and Environment. Vol.7(2007).p. 91~95

[2]Hongxia Xia, Hong Zhou, Luo Zhong, etal: Assessment of efficient lifetime of concrete structure by means of neural network. Journal of Safety and Environment. Vol.4(2004).p.29-31.

[3]KRUSKIN: Theory of artificial neural network(.trans Ping-fan YAN ,Tsinghua University Press
(2002)

[4]Ceng-ren Yuan: Theory of artificial neural networks and its application. Tsinghua University Press(2001)

[5]Zhen-jun Sun, Dong-xing Zhou: Ecology Research Methods. Science Press (2009).

[6]Guo An Tang, Xin Yang. ArcGIS Spatial Analysis Experimental Course, Science Press (2006).

[7]Bo-xi Shen :Soil Environmental Evaluation Model Based on Cokriging interpolation and Spatial Factor Central South University Press(2008).