Spatial interpolation of soil nutrients using algebra hyper-curve neural network

Chen Liping[1], Zhao Chunjiang[1], Huang Wenqian[1], Chen Tian'en[1], Wang Jihua[1], Liu Zhenyan[2], Hu Jing[2]

(1. *National Engineering Research Center for Information Technology in Agriculture*, *Beijing* 100097, *China*; 2. *Department of Software Engineering*, *College of Software*, *Beijing Institute of Technology*, *Beijing* 100081, *China*)

Abstract: Spatial distribution of and interpolation methods for soil nutrients are the basis of soil nutrient management in precision agriculture. For study of application potential and characteristics of algebra hyper-curve neural network(AHCNN) in delineating spatial variability and interpolation of soil properties, 956 soil samples were taken from a 50 hectare field with 20 m interval for alkaline hydrolytic nitrogen measurement. The test data set consisted of 100 random samples extracted from the 956 samples, and the training data set extracted from the remaining samples using 20, 40, 60, 80, 100 and 120 m grid intervals. Using the AHCNN model, three training plans were designed, including plan AHC1, using spatial coordinates as the only network input, plan AHC2, adding information of four neighboring points as network input, and plan AHC3, adding information of six neighboring points as network input. The interpolation precision of AHCNN method was compared with that of Kriging method. When the number of training samples was big, interpolation precisions of Kriging and AHCNN were similar. When the number of training samples was small, the precisions of both methods deteriorated. Since AHCNN method has no request on data distribution and it is non-linearization of neutron input variables, it is suitable for delineation of spatial distribution of nonlinear soil properties. In addition, AHCNN has an advantage of adaptive self-adjustment of model parameters, which makes it proper for soil nutrient spatial interpolation. After comparison of mean absolute error $\bar{d}$, root mean squared error RMSE, and mean relative error $\bar{d}\%$, and the spatial distribution maps generated from different methods, it can be concluded that using spatial coordinates as the only network input cannot simulate the characteristics of soil nutrient spatial variability well, and the simulation results can be improved greatly after adding neighboring sample points' information and the distance effect as network input. When the number of samples was small, interpolation precision can be improved after properly increasing the number of neighboring sample points. It was also showed that evaluation of interpolation precision using conventional error statistic indexes was defective, and the spatial distribution map should be used as an important evaluation factor.

Keywords: algebra hyper-curve neural network (AHCNN), spatial interpolation, soil nutrients, spatial variability, Kriging interpolation

DOI: 10.3965/j.issn.1934-6344.2008.01.051-056

Citation: Chen Liping, Zhao Chunjiang, Huang Wenqian, Chen Tian'en, Wang Jihua, Liu Zhenyan, et al. Spatial interpolation of soil nutrients using algebra hyper-curve neural network. Int J Agric & Biol Eng. 2008; 1(1): 51－56.

1 Introduction

The obvious variability of soil properties within the same type of soil is an important feature of soil in a field. Even in an area in which texture of soil can be regarded as uniform, soil properties in different locations can be different. This is known as spatial variability of soil nutrients[1,2]. In the past, this variability was analyzed using traditional statistical method founded by Fisher. The traditional method considers samples to be entirely independent and obeys certain known probability distribution, and soil property spatial variability can be delineated through calculating mean, standard deviation, variance, and coefficient of variation of samples and conducting significance tests, with some successful applications in soil research. However, recent research found that soil properties were not entirely spatial-independent, but correlated spatially to some extent[3]. Obviously, spatial variability of soil properties for estimation of unsampled points is the basis of soil

Received date: 2008-06-25 Accepted date: 2008-08-02

Authors: Chen Liping(1973－), PhD, Associate Researcher, majored in Intelligent Decision Making in Precision Agriculture, Head of Department of Precision Agriculture, National Engineering Research Center for Information Technology in Agriculture,. Beijing, China. Email: chenlp@nercita.org.cn.

Huang Wenqian, Research Assistant; Chen Tian'en, PhD, Research Assistant; Wang Jihua, PhD, Professor; Liu Zhenyan, Lecturer; Hu Jing, graduate student.

Corresponding Author: Zhao Chunjiang(1964－), PhD, Professor, Majored in Application of Information Technology in Agriculture, Director of National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China. Voice: +86-10-51503411, Post Box2449-26.
Email: zhaocj@nercita.org.cn.

nutrient management and scientific fertilization in precision agriculture.

In recent years, geostatistics has been widely used in quantitative analysis of spatial variability of soil nutrients. However, there are several important preconditions in application of Kriging interpolation method of geostatistics[4]. If these preconditions are not satisfied, geosatistics cannot be used reliably for spatial variability research[5].

Artificial neural network technology, which has been used in research of soil property spatial variability and some ideal results have been achieved, is a powerful tool to deal with nonlinear system[6-10]. But these researches were based on back propagation (BP) neural network, which was an algorithm developed from establishing configuration network of basic perceptron model. BP neural network, a multilayer feed-forward network with hidden-layers, cannot be used to determine the number of hidden-layers in multi-layer network and the number of nodes in a certain layer. Usually the number is given based on experience, leading to sub-optimal design of network. Moreover, there are still some problems with BP neural network, for example, low learning and convergence speed, and the convergence cannot ensure reaching global minimum. Algebra hyper-curve neural network (AHCNN) develops basic perceptron model in terms of perceptron integrative function which adds some auxiliary units to basic perceptron model. The outputs of these auxiliary units are quadratic functions of each unit x_i (i=1, 2, ···, n) in input vector of basic perceptron model, i.e., AHCNN is non-linearization of neuron input variables. Therefore, AHCNN is a more ideal nonlinear computation tool suitable to delineate spatial distribution of soil properties which are also nonlinear. Furthermore, AHCNN has an advantage of model parameters adaptive self-adjustment, which, in some sense, overcomes the disadvantages of low learning speed and adjustment of many parameters in BP neural network[11,12].

AHCNN model was used to analyze precision of soil nutrient spatial interpolation under conditions of different sampling densities and adding neighboring sampling points' information as network input, and AHCNN model with spatial coordinates as the only network input was compared with ordinary Kriging method. Application potential and characteristics as well as problems of using AHCNN in delineating spatial variability and interpolation of soil properties were also studied.

2 Materials and methods

2.1 Materials

In this study, 956 soil samples were collected from 0～20 cm soil layer, using 20 m×20 m as grid interval with Differential Global Positioning System (DGPS) at No.1 50 hectares field at National Precision Agriculture Demonstration Station in Xiaotangshan Town of Beijing in September, 2007. Each soil sample was a mixture of five sub-samples, with four sub-sample points distributed on ten meter in diameter concentric circle and one sample point soil in the circle center. These samples were air dried within 24 hours after sampling. Then, nutrients of soil samples were measured after sieving.

The alkaline hydrolytic nitrogen contents of soils were used as experimental data. In a diffusion container, as soil was hydrolyzed in alkaline condition, hydrolyzable nitrogen was transformed easily into NH_3 and was absorbed by H_3BO_3 after diffusion. Then, NH_3 in absorbing solution H_3BO_3 was titrated using standard acid, and content of alkaline hydrolytic nitrogen can be calculated.

In order to analyze precision of soil nutrient data interpolation of different interpolation methods at different sampling precisions, the 956 sample points were divided into two groups, which include independent training data sets with 856 sample points and test data set with 100 sample points. Field distribution consisted of entire training data set was Plan a, and sampling points were extracted at 20 m sampling interval based on original field distribution. Plan b was to contain 217 sample points through interlacing extraction in horizontal and vertical directions and sampling interval was 40 m. Plan c was to extract 107 sample points every two lines in horizontal and vertical directions and sampling interval was 60 m. Plan d was to extract 57 sample points every three lines in horizontal and vertical directions and sampling interval was 80 m. Plan e was to extract 44 sample points every four lines in horizontal and vertical directions and sampling interval was 100 m. Plan f was to

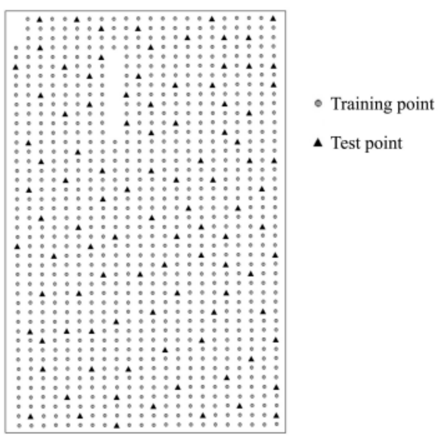


Figure 1 Distribution of the sample points of Plan a

extract 28 sample points every five lines in horizontal and vertical directions and sampling interval was 120 m. The distribution of sample points in Plan a was shown in Figure 1.

2.2 Methods

2.2.1 Kriging interpolation

First, skewness and kurtosis tests of training sample data were conducted. If it did not obey normal distribution, logarithmic transformation would be used to make it obey lognormal distribution. Then, with semi-variance analysis, basic parameters for Kriging interpolation can be calculated by using theoretical semi-variance model with relative high fitting degree to fit semi-variance function. Finally, ordinary Kriging interpolation was conducted with ArcGIS 9.0.

2.2.2 Algebra hype-curve neural network model

Figure 2 showed two-dimensional AHCNN model. On the basis of two nodes x_1 and x_2 in input vector, three auxiliary input nodes f_1, f_2, and f_3 were added. The order of the two-dimensional AHCNN model was 2, and the degree of polynomial was also 2. Given $f_1 = x_1^{\ 2}$, $f_2 = x_1 \cdot x_2$, and $f_3 = x_2^{\ 2}$, the output of perception can be calculated using the following formula:

$$Z = f\left(\sum_{i=0}^{2} w_i x_i + w_3 f_1 + w_4 f_2 + w_5 f_3\right)$$
$$= f\left(-w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^{\ 2} + w_4 x_1 x_2 + w_5 x_2^{\ 2}\right) \tag{1}$$

Hyperbolic tangent sigmoid function was used as transformation function at hidden-layer and linear function at output layer. Before training neural network with these training data, the [−1, 1] normalization processing of training data was conducted for fast convergence. The δ learning rule was used, which is a continuous perceptron learning rule.

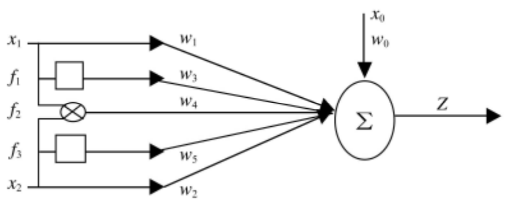


Figure 2 Two-dimensional AHCNN model

Method 1 (AHC1): With spatial coordinates' value as network input, the relation was established only between spatial coordinates and soil property information values which can be described using the following formula:

$$Z = f(X,Y) \tag{2}$$

In formula (2), X and Y are spatial coordinates of sample point or predicted point. AHCNN used X and Y as network inputs and corresponding soil property information values as network outputs, that is, two nodes in input layer and one node in output layer .

Method 2 (AHC2): With spatial autocorrelation as theoretical basis[13], the function of soil nutrient spatial distribution can be expressed as following:

$$Z = f\left(X, Y, A_1, A_2, \cdots, A_n\right) \tag{3}$$

Where Z is the value of soil property; X and Y are spatial coordinates; A_1, A_2, ···, and A_n are n values of soil property of sample points closest to interpolation point. The spatial coordinates X and Y, and the n sample points closest to the interpolation point were used as network inputs, so the number of nodes in network input was 2+n in which the first and second neurons are input coordinates X and Y , respectively.

According to the spatial distance decay law, the effect of sample point on interpolation point declines as the distance between them increases. Considering the effect of distance, in algorithm program the third neuron in network input layer was set as values of soil property of the sample point closest to the interpolation point, and the fourth neuron in network input layer was set as values of soil property of the sample point second closest to the interpolation point and so on. In this study, n was four and total nodes of network input layer were six.

Method 3 (AHC3): In order to compare effects of different numbers of neighboring sample points on results of neural network interpolation, the interpolation results were calculated using six neighboring sample point data.

2.2.3 Evaluation of the interpolation precision

In this study, mean absolute error $\bar{d}$, root mean squared error RMSE, and mean relative error $\bar{d}\%$ were used to evaluate the interpolation precision. The three indexes can be calculated using the following formulas:

$$\bar{d} = \frac{1}{n}\sum \left|\left(\hat{Z}_i - Z_i\right)\right| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Z}_i - Z_i\right)^2} \tag{5}$$

$$\bar{d}\% = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{Z}_i - Z_i\right|}{Z_i} \times 100\% \tag{6}$$

Where $\hat{Z}_i$ is estimated value; Z_i is measured value; n is the number of training samples or test samples. Obviously, the smaller the value of $\bar{d}$, RMSE, and $\bar{d}\%$, the smaller the error, and the better the interpolation precision.

3 Results and discussion

Comparison of interpolation precision of test data set using Kriging and AHCNN based on test data set of different plans with different sampling intervals was shown in Table 1.

As shown in Table 1, when the number of training samples was big, interpolation precisions of Kriging and AHCNN were similar. When the number was small, the precisions of Plan d, e, and f declined.

The skewness and kurtosis of the dataset for this study are −0.284 and 3.241, respectively, indicating normal distribution of the dataset. So the interpolation precision of Kriging is good, and the interpolation precision of AHCNN is not better than that of Kriging. Under the condition that the data set does not obey normal distribution, the advantages of AHCNN will be fully demonstrated.

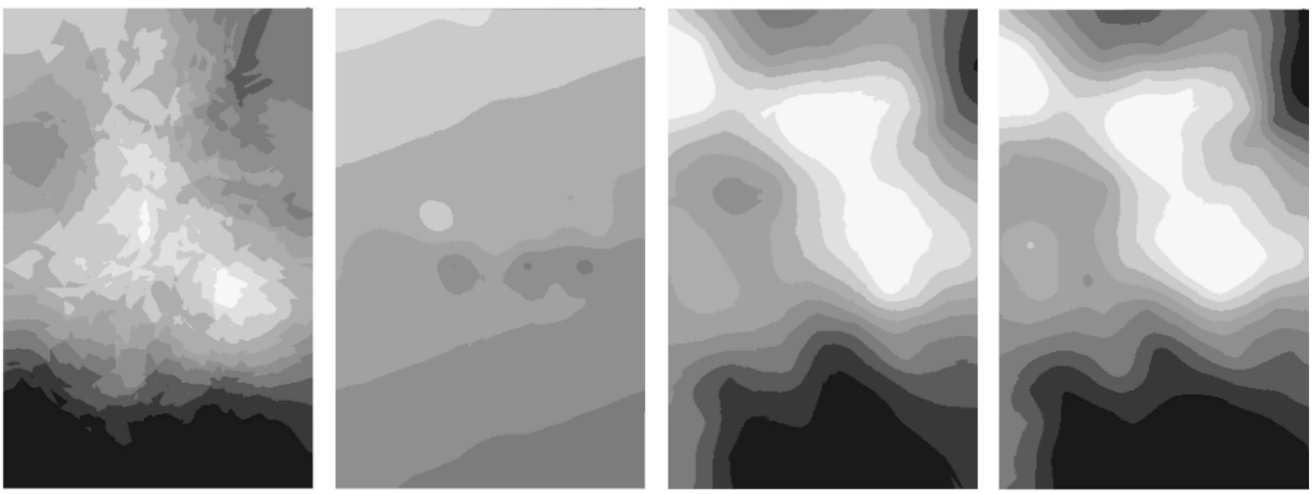
Based on the precision indexes listed in Table 1, several conclusions can be deduced. With neighboring points as neural network input, interpolation precisions can be improved using both AHC2 and AHC3 compared with using ACH1 in Plan a, b, and c when the number of training samples was big. However, when the number of training samples was small, interpolation precisions declined using both AHC2 and AHC3 compared with using ACH1 in Plan d, e, and f. Meanwhile, the number of neighboring points such as 4 or 6 had little effect on interpolation precisions.

Table 1 Comparison of interpolation precisions of test dataset using Kriging and AHCNN
(soil alkaline hydrolytic nitrogen, mg/kg)

Plan	Number of samples	Kriging			AHC1			AHC2			AHC3		
		$\bar{d}$	RMSE	$\bar{d}$ %	$\bar{d}$	RMSE	$\bar{d}$ %	$\bar{d}$	RMSE	$\bar{d}$ %	$\bar{d}$	RMSE	$\bar{d}$ %
a	856	6.37	8.14	8.52	8.52	10.30	11.05	7.66	9.43	9.86	7.14	9.04	9.30
b	217	6.92	8.84	9.33	8.38	10.27	11.27	7.21	9.40	9.78	7.45	9.69	10.10
c	107	7.21	9.18	9.55	8.49	10.13	11.04	7.93	9.93	10.44	8.02	10.04	10.67
d	57	7.89	10.01	10.68	8.47	10.48	11.59	10.49	12.50	13.95	10.45	13.31	14.16
e	44	7.10	8.94	9.60	8.68	10.95	12.06	10.75	13.94	14.50	10.45	13.46	14.28
f	28	7.62	9.49	10.24	9.12	10.85	12.09	8.73	11.25	11.73	9.65	11.78	12.39

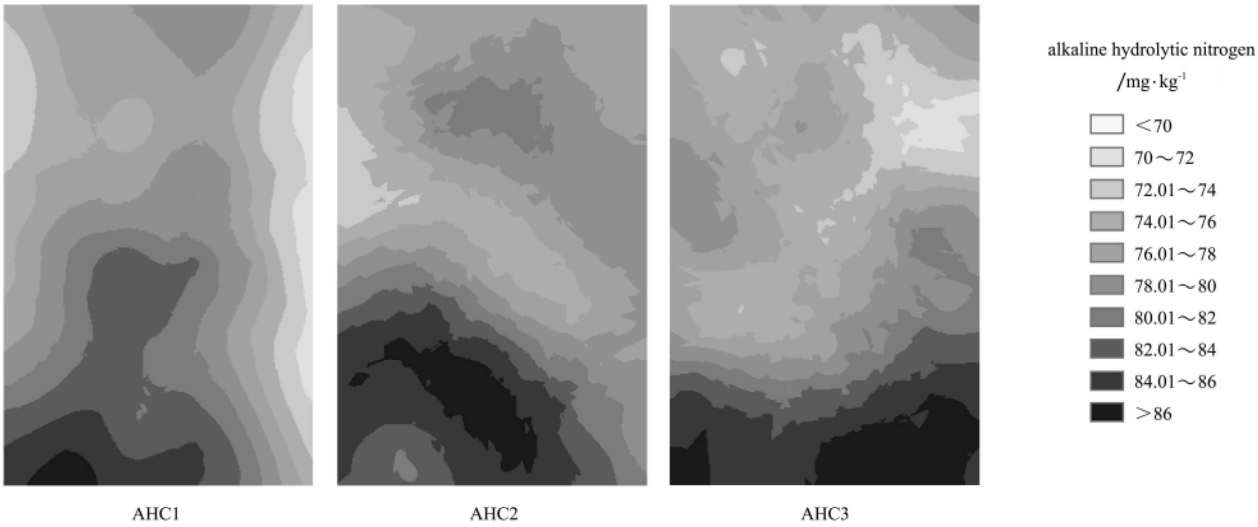
Kriging interpolation for estimated values of alkaline hydrolytic nitrogen of 100 test sample points using different methods was finished and fitting results of alkaline hydrolytic nitrogen spatial distribution in soil were shown in interpolation maps. Comparison of interpolation results using AHCNN and actual measurement data of Plan c and f was shown in Figure 3.

The simulation of spatial distribution of alkaline hydrolytic nitrogen using AHC1 was fairly bad, as AHC1 was based only on relationship between spatial coordinates and value of alkaline hydrolytic nitrogen in soil. The fitting degree of interpolation of spatial distribution was improved as adding the number of neighboring points, which was more obvious in plans with smaller number of training samples such as Plan f.



Spatial distribution of alkaline hydrolytic nitrogen by Kriging AHC1 AHC2 AHC3

a. Interpolation results of Plan c



b. Interpolation results of Plan f

Figure 3 Comparison of spatial distribution of alkaline hydrolytic nitrogen and interpolation results using AHC1, AHC2 and AHC3 of Plan c and f

Precise management of field nutrients is very important in precision agriculture, which is based on soil nutrient spatial distribution. So, high fitting precision of soil nutrient spatial distribution is the aim of research on scientific sampling and interpolation method. At this point, evaluation on interpolation using AHC1, which is based on relationship between spatial coordinates and value of soil properties, is feasible according to conventional statistics; however, it is little helpful for precision agriculture.

4 Conclusions

1) Application of AHCNN in research and interpolation of soil nutrient spatial variability is feasible, and it has no special request on data distribution. AHCNN model is non-linearization of neutron input variables suitable for prediction of nonlinear system such as soil nutrients. In fact, the integrative function of AHCNN model is polynomial. The degree of this polynomial and the coefficient of each term can be obtained automatically through learning. Moreover, if network modeling is unsuccessful in lower degree, the degree of this polynomial can be raised using adaptive degree raising which is useful for fast determination of the order of AHCNN model and higher efficiency of learning and modeling.

2) The precision of soil nutrient interpolation model is determined by the demonstration degree of the model for soil nutrient spatial variability and spatial correlation, which means that the selected regression elements and interpolation elements should be highly correlated[14,15]. The selection of input parameters is important in application of artificial neural network for soil nutrient spatial interpolation. Since there is no correlation between the value of spatial coordinates and the value of soil nutrients, the error would be great if using spatial coordinates of soil nutrients as the only input. Using ordinary error evaluation method, the error is relatively small. But in terms of spatial interpolation map, the simulation of soil nutrient spatial distribution is fairly bad. So, conventional error statistic index such as RMSE cannot be used as the only index for comparison of interpolation results. Compared with using spatial coordinates as the only network input, the simulation degree of soil nutrient spatial distribution was greatly improved using AHCNN with auxiliary information of neighboring sample points as network input considering distance effect. These conclusions agree with the results of Li Qiquan et al[16], who concluded that the error of interpolation of three soil properties in an experiment field was greatly reduced when the information of neighboring sample points was added, the distance effect was considered, as radial basis function neural network (RBFNN) input.

3) Certain data noise is allowable in algebra hyper-curve algorithm (i.e., not all samples in training sample set were used successfully in modeling), which can reduce the time of learning on one hand. On the other hand, modeling can be successful in lower degree and need no degree raising of polynomial, which reduces effect on the accuracy of model. More research work on anti-noise performance of algebra hyper-curve algorithm needs to be done in the future. In terms of multi-source information fusion, nonlinear mapping relationship

between soil properties and neighboring sample point information, and soil forming factors such as parent material, terrain, and climate can be established to delineate spatial distribution of soil properties using nonlinear computation ability and expansiveness of artificial neural network. This provides a solution for large scale study and accuracy improvement of research results.

Acknowledgements

This research was financially supported by the National Natural Science Foundation of China (30600375) and the National High Technology Research and Development Program of China (2006AA10A306, 2006AA10Z271). We would like to thank the National Experimental Station of Precision Agriculture of China.

[References]

[1] Xu Jiyan, Webster R. Optimal estimation of soil survey data by geostatistical method -- Semi-variogram and block Kriging estimation of topsoil nitrogen of Zhangwu County. Acta Pedologica Sinica. 1983; 20(4): 419－430.

[2] Huang Shaowen, Jin Jiyun. Advance in study on spatial variability of soil properties. Soils and Fertilizers. 2002; (1): 8－14. (in Chinese)

[3] Burrough P A. Soil variability: A late 20th century view. Soils and Fertilizers. 1993; 56(5):529－562.

[4] Hou Jingru, Guo Guangyu. The Theory and Practice of Statistical Prediction and Geostatistics for Mineral Deposits. , Beijing: Metellurgical Industry Press. 1993, P325－327. (in Chinese)

[5] Yue Tianxiang, Liu Jiyuan. Multi Sources Information Fusion Data Model, World Science-Technology R&D. 2001; 23(5):1－4. (in Chinese)

[6] Shen Zhangquan, Zhou Bin, Kong Fansheng. Study on spatial variety of soil properties by means of generalized regression neural network. Acta Pedologica Sinica; 2004; 41(3):471－475. (in Chinese)

[7] Shen Zhangquan, Shi Jiebin, Wang Ke. Spatial variety of soil properties by BP neural network ensemble. Transactions of the Chinese Society of Agricultural Engineering. 2004; 20 (3): 35－39. (in Chinese with English Abstract)

[8] He Yong, Zhang Shujuan, Fang Hui. Interpolation method of field information based on the artificial neural network. Transactions of the Chinese Society of Agricultural Engineering. 2004; 20(3): 120－123. (in Chinese with English Abstract)

[9] You Shucheng, Yan Tailai. A study on artificial neural network based surface interpolation. Acta Geodaetica et Cartographica Sinica. 2000; 29(1): 30－34. (in Chinese)

[10] Robinson T P, Metternicht G. Testing the performance of spatial interpolation techniques for mapping soil properties. Computers and Electronics in Agriculture. 2006; 50(2): 97－108.

[11] Liu Zhenyan, Zhang Zhimiao, Wang Wansen. Research of nonlinear neural network. Computer Engineering and Applications. 2004; 40(3): 42－44. (in Chinese)

[12] Qing He, Zhenyan Liu, Zhongzhi Shi. An algebraic multi-class classification method. Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004: 3307－3312.

[13] Zhu Huiyi, Liu Shulin, Jia Shaofeng. Problems of the spatial interpolation of physical geographical elements. Geographical Research. 2004; 23(4): 425－432. (in Chinese)

[14] Mac Eachren A M, Davidson J V. Sampling and isometric mapping of continuous geographic surfaces. The American Cartographer. 1987;14(4): 299－320

[15] Fred C Collins Jr, Paul V Bolstad. A comparison of spatial interpolation techniques in temperature estimation.1996.. http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/collins_fred/collins.html.

[16] Li Qiquan, Wang Changquan, Yue Tianxiang, et al. Error analysis of soil property spatial interpolation with RBF artificial neural network with different input methods. Acta Pedologica Sinica. 2008; 45 (2): 360－365. (in Chinese).