

Spatial interpolation method based on integrated RBF neural networks for estimating heavy metals in soil of a mountain region

Li Baolei¹ Zhang Yufeng¹ Shi Xinling¹ Zhang Kexin² Zhang Junhua¹

(¹School of Information Science and Engineering, Yunnan University, Kunming 650091, China)

(²Cardiovascular Departments, the Second Affiliated Hospital of Kunming Medical College, Kunming 650031, China)

Abstract: A novel spatial interpolation method based on integrated radial basis function artificial neural networks (IRBFANNs) is proposed to provide accurate and stable predictions of heavy metals concentrations in soil at un-sampled sites in a mountain region. The IRBFANNs hybridize the advantages of the artificial neural networks and the neural networks integration approach. Three experimental projects under different sampling densities are carried out to study the performance of the proposed IRBFANNs-based interpolation method. This novel method is compared with six peer spatial interpolation methods based on the root mean square error and visual evaluation of the distribution maps of Mn elements. The experimental results show that the proposed method performs better in accuracy and stability. Moreover, the proposed method can provide more details in the spatial distribution maps than the compared interpolation methods in the cases of sparse sampling density.

Key words: integrated radial basis function artificial neural networks; spatial interpolation; soil heavy metals; mountain region

doi: 10.3969/j.issn.1003–7985.2015.01.007

Mining and waste water from galvanization factories have aggravated the situation of soil contamination by heavy metals around the world^[1]. Heavy metal contamination in soil can cause serious environmental and health-related problems^[2]. Accumulation of heavy metals in soil constitutes a serious long-term threat to the environment, food quality and human health^[3]. In the geographic information system (GIS), the spatial distribution maps of heavy metals in soil are widely used to visually describe the general features of the spatial patterns of soil heavy metals. Drawing the spatial distribution maps of heavy metals in soil accurately by an interpolation method

according to the measurements at finite sampling sites is a critical fundamental for further spatial evaluation and analysis. However, sampling in a thorough and regular way is limited by expensive costs and some objective reasons such as water areas, cliffs, and mountains. As a result, it is usually difficult to accurately draw samples based on the inadequacy and irregularity of sampling sites.

The traditional linear spatial interpolation methods such as inverse distance weighting (IDW) and ordinary kriging (OK) have been extensively used in soil surveys and pollution distribution mapping^[4]. They can provide adequate and regular raster data through interpolation based on an inadequate and irregular sampled dataset^[5]. However, they suffer from some limitations, such as strong subjectivity, numerous assumptions, poor adaptive variation, etc., which leads to underestimating the local high values or overestimating the local low values^[6].

Recently, artificial neural networks (ANNs) have been introduced as a spatial interpolation method to improve the estimation accuracy of spatial interpolation. The ANNs-based method can automatically develop a nonlinear estimation model by simple self-learning based on a sampled dataset without any assumptions^[7]. Nevertheless, small changes in a sampled dataset or initial conditions may produce distinctly different models^[8], and cause different interpolation performances on un-sampled sites.

Hansen and Salamon^[9] proposed a neural network integration approach and proved that neural network generalization abilities can be improved significantly by simply training multiple neural networks and combining their results. Back propagation ANNs (BPANNs) can adjust network weights and network threshold constantly by reversing transmission errors to ensure a minimum network error quadratic sum. However, their convergence speed is slow, which is unsuitable for the multiple training application in the neural networks integration approach. Unlike BPANNs, the output layer is directly connected to its hidden layer in radial basis function artificial neural networks (RBFANNs). RBFANNs can determine the corresponding network topology structure according to the specific issues as well as self-study, organization and adaptive function^[10]. It has many advantages such as simple parallel distributed computation, distributed storage, and fast

Received 2014-07-19.

Biographies: Li Baolei (1987—), male, graduate; Zhang Yufeng (corresponding author), male, doctor, professor, yfengzhang@yahoo.com.

Foundation items: The National Natural Science Foundation of China (No.61261007, 61062005), the Key Program of Yunnan Natural Science Foundation (No.2013FA008).

Citation: Li Baolei, Zhang Yufeng, Shi Xinling, et al. Spatial interpolation method based on integrated RBF neural networks for estimating heavy metals in soil of a mountain region[J]. Journal of Southeast University (English Edition), 2015, 31(1): 38–45. [doi: 10.3969/j.issn.1003–7985.2015.01.007]

learning^[11]. Therefore, RBFANNs can undertake a wide range of data fusion with a high learning convergence rate.

In this paper, a novel spatial interpolation method based on integrated RBFANNs (IRBFANNs) is proposed for improving the accuracy and stability of the spatial interpolation performance. The advantages of RBFANNs and the neural networks integration approach are combined in IRBFANNs. Concentrations of total manganese (Mn), vanadium (V), nickel (Ni), cuprum (Cu) at 42 soil sample sites in the region of about 10 000 km² in the northern part of Chuxiong and its periphery, Yunnan, China were used to test the proposed method. It was compared with the IDW, OK, BPANNs, RBFANNs, integrated OK (IOK), and integrated BPANNs (IBPANNs) through cross validations by statistic analysis of the root mean square error index and visual evaluation of distribution maps. The results show that the interpolation values based on IRBFANNs are accurate, stable and can provide details for the spatial distribution maps.

1 Spatial Interpolation Methods

1.1 Artificial neural networks-based spatial interpolation algorithm

An ANNs-based spatial interpolation model with input, hidden and output layers is illustrated in Fig. 1. In this scheme, the input vector consists of two variables: longitude x_1 and latitude x_2 . The model should be trained before it is used to interpolate values at un-sampled sites. Training involves feeding the neural networks with training samples and allowing them to learn by adjusting the weight of the neurons. After the training process, the trained ANNs model outputs the interpolation value at a certain site corresponding to the given input data (longitude and latitude of this site).

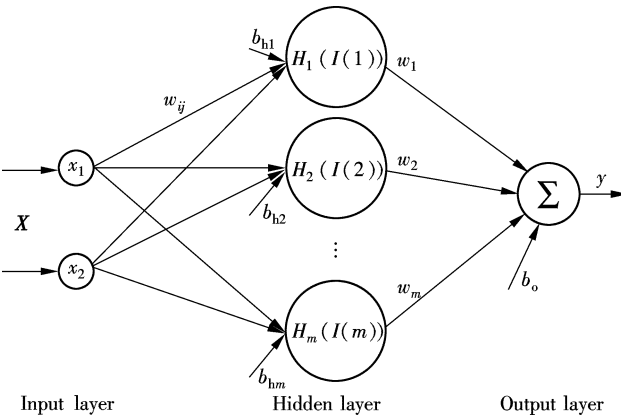


Fig. 1 The architecture of an ANNs-based spatial interpolation model

As shown in Fig. 1, the i -th node in the input layer receives a signal x_i , and then, the j -th node in the hidden layer receives a signal $I(j)$, which is calculated by the

following equation:

$$I(j) = b_{hj} + \sum_{i=1}^2 x_i w_{ij} \quad (1)$$

where w_{ij} is the weight of the connection from the i -th neuron in the input layer to the j -th neuron in the hidden layer; b_{hj} is the bias of the i -th neuron in the hidden layer. The output of the ANNs model is calculated by

$$y = b_o + \sum_{j=1}^m H_j(I(j)) w_j \quad (2)$$

where $H_j(I(j))$ is the output of the j -th hidden unit; w_j is the weight between the j -th hidden layer and the output layer; b_o is the bias of the output neuron; and m is the number of hidden nodes.

In the BPANNs method^[12], the output of the j -th hidden unit is calculated by the bipolar sigmoid activation function:

$$H_j^{bp}(I(j)) = \frac{2}{1 + \exp(-I(j))} - 1 \quad (3)$$

In the RBFANNs method^[13], the hidden layer uses radial basis functions such as Gaussian, multi-quadratic, thin-plate-spline, and cubic approximation^[14] instead of the sigmoid function. The Gaussian function is chosen as the transfer function in this study and the output of the j -th hidden unit is calculated by

$$H_j^{rbf}(X, C_j, \delta_j) = \exp\left(-\frac{\|X - C_j\|^2}{2\delta_j^2}\right) \quad (4)$$

where C_j represents the center of function $H_j^{rbf}(X, C_j, \delta_j)$; δ_j determines the function's width; and X is the input vector.

1.2 Integrated radial basis function artificial neural networks

The IRBFANNs model integrates several individual RBFANNs modules through ensemble techniques to produce dramatic improvements in generalization performance. Bagging is a popular RBFANNs ensemble technique to generate multiple training sets and RBFANNs modules for an ensemble processing^[15].

There are two steps in the creation of the IRBFANNs model, as illustrated in Fig. 2. The first step is to create independent RBFANNs modules based on different training samples, which are derived by using the bagging algorithm. The second step is the appropriate combination of outputs from the RBFANNs modules to produce the ensemble output. Assuming that RBFANNs modules have been trained, then the output of an IRBFANNs model y_{out}^* is calculated by

$$y_{out}^* = \sum_{i=1}^q y_{iout}^* w_{ei} \quad (5)$$

where q is the number of RBFANNs modules in the IRB-FANNs model; w_{ei} denotes the assigned weight of y_{iout}^* , which is the output of the i -th RBFANNs module, and it is calculated by

$$w_{ei} = \frac{1/r_{mse}^i}{\sum_{j=1}^q 1/r_{mse}^j} \quad (6)$$

$$r_{mse}^i = \sqrt{\frac{1}{r} \sum_{k=1}^r (y_{iout}^*(\mathbf{X}_k) - y(\mathbf{X}_k))^2} \quad (7)$$

where r_{mse}^i is the root mean square error of the i -th RBFANNs module, which is trained based on the i -th training dataset in which the number of instances is r . $y_{iout}^*(\mathbf{x}_k)$ is the estimation value of the i -th RBFANNs module and $y(\mathbf{x}_k)$ is the desired value for the input vector \mathbf{x}_k .

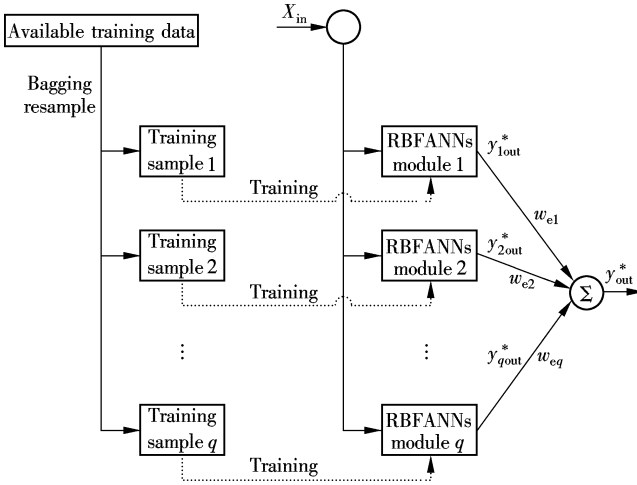


Fig. 2 The IRB-FANNs model

2 Experiments

The northern part of Chuxiong and its periphery, Yunnan, China, is chosen as the research area. The landscape of the study area presents many different types of terrain. Mountainousness is the dominant characteristic of the study area where the elevation ranges from 556 to 3 657 m above sea level. Four kinds of heavy metal concentrations in soil, including Mn, V, Ni and Cu, are considered for the interpolating tests based on 42 collected samples. Prior to the experiments, the collected data were logarithmically transformed to obtain a standard normal distribution whose mean and deviation are 0 and 1, respectively. The location of the investigation area and the spatial pattern of the collected soil samples are shown in Fig. 3.

Based on the 42 collected data, the interpolation results achieved by the IRNFANNs model are compared with other six peer methods. The compared algorithms and executive tool for each method are listed as follows.

- IDW^[16]: The number of sites used in the prediction

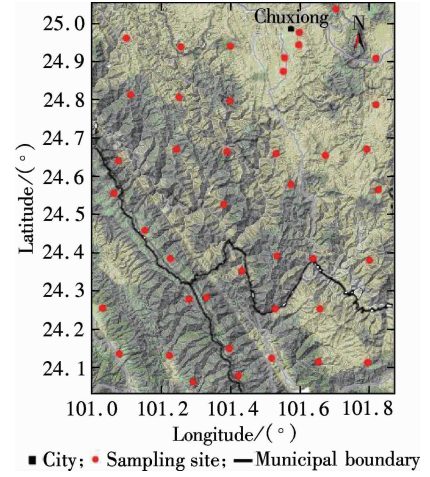


Fig. 3 Location of investigation area

n is 10, and the weighting power μ is 2.

- OK^[17]: The DACE function of the Matlab Kriging toolbox is used for variogram modeling and estimation. The second order polynomial regression function (Regpoly2) and Gaussian correlation function (Corrgauss) are used as the regression model and correlation function, respectively, and the initial correlation function parameter θ is set to be 10.

- BPANNs: The NEWFF function of the Matlab neural networks toolbox is used to generate BPANNs modules.

- RBFANNs: The NEWRB function of the Matlab neural networks toolbox, in which the spread of radial basis functions is set to be 0.05, is used to generate RBFANNs modules. Other parameters of these used functions are optimized automatically by Matlab ToolBox.

- Integrated methods: The results based on the OK, BPANNs and RNFANNs modules are integrated through a bagging ensemble technique to join in the comparison. For all the integrated methods, the collected data are first re-sampled at different densities to generate a training dataset and then 80% of the training dataset are chosen randomly to create an individual module.

In order to make clear the influence of the sampling size reduction on the performance of each interpolation method, three projects for each element are conducted under different sampling densities. In Projects A, B and C, 41, 26 and 16 sampled data are randomly chosen as training datasets, respectively; and the rest are used as testing datasets. Fig. 4 illustrates the steps used for each project. The errors between the prediction values and observed values are used to compute the R_{mse} as follows:

$$R_{mse} = \sqrt{\frac{1}{n} \sum_{i=1}^n (z^*(\mathbf{X}_i) - z(\mathbf{X}_i))^2} \quad (8)$$

where $z^*(\mathbf{X}_i)$ is the prediction value of spatial interpolation methods at location \mathbf{X}_i and $z(\mathbf{X}_i)$ is observed value at the same location; n is the number of testing samples.

The index R_{mse} is used to compare the different methods by comparing how closely the predicted values match the measured values. The method with the smallest R_{mse} is the most accurate interpolation method. In the experiments, 100 tests for each project based on cross-validation are implemented for statistic analysis. Plots of R_{mse} values of four elements using different interpolation methods under different training data sizes are presented for statistic analysis. We also plot the spatial distribution maps of the concentrations for the four soil heavy metals to visually compare the performance of these interpolation methods.

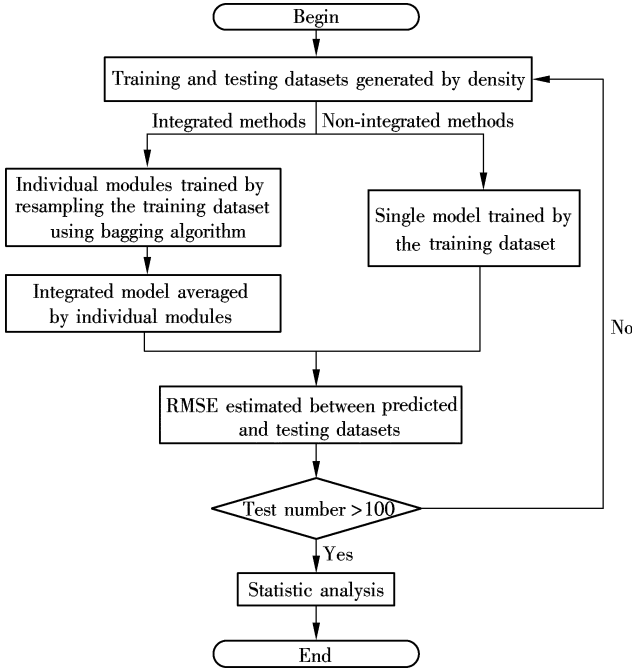


Fig. 4 A flow chart of interpolation methods comparison project

3 Results and Discussion

The R_{mse} values of 100 tests for Mn, V, Ni, Cu achieved by the seven methods under three different sample density projects are used to statistically evaluate the proposed interpolation method performance.

The mean, median, minimum, maximum, upper quintile and lower quintile of the R_{mse} value in a box plot is demonstrated in Fig. 5. The mean and the median of the R_{mse} (M_{mr}) reflect the accuracy of the interpolation method. The smaller the M_{mr} , the more accurate the interpolation method. The difference between the lower quintile and upper quintile of the R_{mse} (D_{qqr}) reflects the stability of the interpolation method. The smaller the D_{qqr} , the more stable the interpolation method. Therefore, the method with small M_{mr} and D_{qqr} is considered to be an accurate and stable interpolation. According to this rule, the prediction results of seven methods with four elements shown in Figs. 6 to 9 are evaluated and compared based on their corresponding box plots of R_{mse} values under different training data sizes.

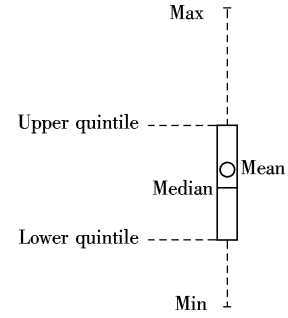


Fig. 5 Description of a box plot

A comparison of the statistic results among Projects A, B and C illustrates that M_{mr} increases along with the decrease of the training data size, which shows that the interpolation accuracy of the seven methods decreases with the decrease of the sampling density. The prediction results of Mn are shown in Fig. 6. Viewing the statistic prediction results of the other elements V, Ni, Cu, we can come to the same conclusion. However, the accuracies of various methods are different. The increase of M_{mr} achieved by the integrated ANNs-based method is smaller than that achieved by the traditional and ANNs-based methods when the training data size decreases. This indicates that the integrated ANNs-based method is less affected by the decrease of the training data size than the traditional and ANNs-based methods. A similar conclusion can be drawn through observing the experimental results of the other three elements.

Although the M_{mr} and D_{qqr} of the seven methods increase with the decrease of the training data, the increase degrees of three statistics achieved by IRBFANNs are smaller than that achieved by other methods. The IRBFANNs method achieves the smallest M_{mr} and D_{qqr} in all the methods when the number of training data is 16. This suggests that the IRBFANNs outperforms other methods in both accuracy and stability in the sparse sampling density situation. The experimental results of other elements also confirm this conclusion.

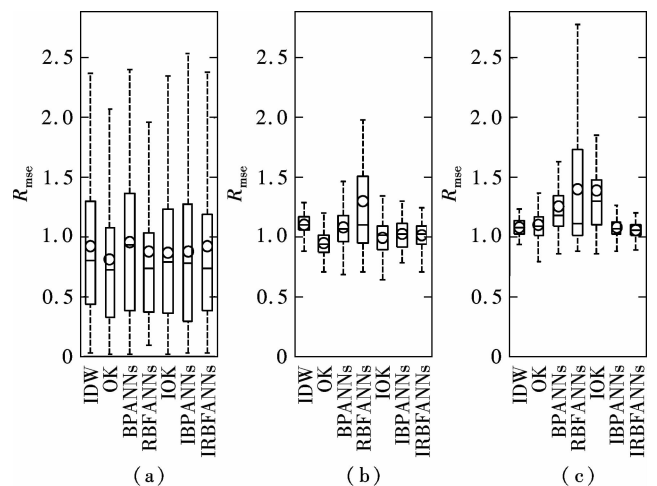


Fig. 6 Box plot of R_{mse} under different training data sizes for Mn. (a) Size of 41; (b) Size of 26; (c) Size of 16

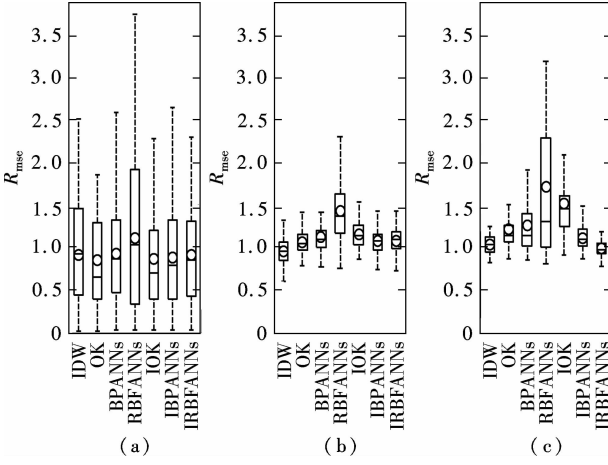


Fig. 7 Box plot of R_{mse} under different training data sizes for V. (a) Size of 41; (b) Size of 26; (c) Size of 16

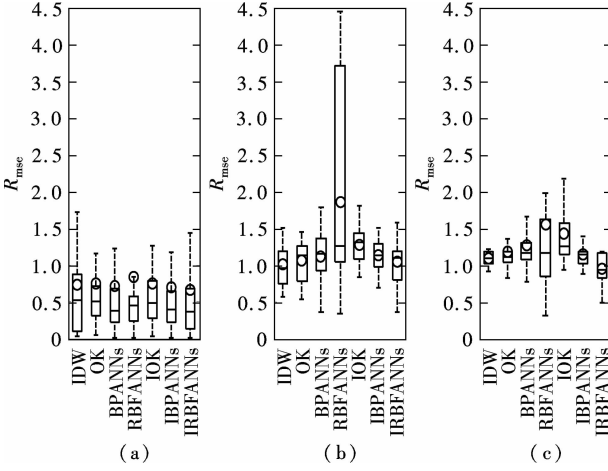


Fig. 8 Box plot of R_{mse} under different training data sizes for Ni. (a) Size of 41; (b) Size of 26; (c) Size of 16

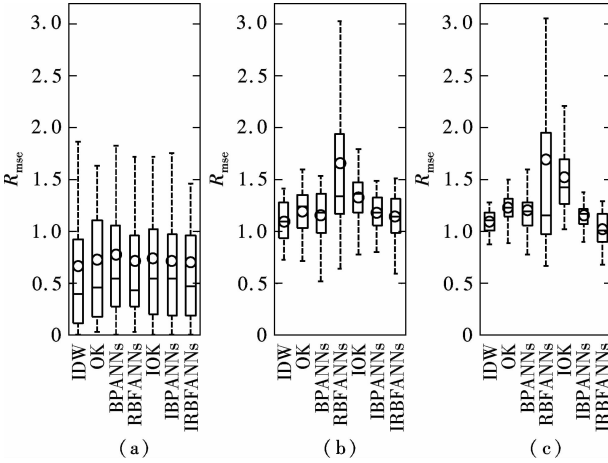


Fig. 9 Box plot of R_{mse} under different training data sizes for Cu. (a) Size of 41; (b) Size of 26; (c) Size of 16

Observing the effects of the sampling density on interpolation accuracy and stability in all the experiments of elements, we can come to consistent conclusions. The interpolation accuracies and stabilities of all the methods de-

crease with the decrease in the sampling density. More specifically, the traditional and ANNs-based methods are less affected by the decrease of the training data size than the integrated ANNs-based methods. The IRBFANNs method outperforms other methods in both accuracy and stability when the training data size is 16.

We compare the performances of seven methods under the same training data size. The statistic results of R_{mse} under the largest training data size of 41 (Project A), the medium training data size of 26 (Project B) and the smallest training data size of 16 (Project C) are discussed as follows:

In the largest training data size case, the statistic results of Mn show that the D_{qqr} values of all the methods are almost the same, and the M_{mr} of OK is a little smaller than that of other methods. Therefore, the geostatistics-based OK method is a little better than other methods in accuracy. The contradistinction of statistic results by the integrated and non-integrated methods shows that the integration of the ANNs modules including BPANNs and RBFANNs helps decrease both the M_{mr} and the D_{qqr} , but the integration of OK modules obtains another reversed result that the M_{mr} and D_{qqr} of IOK are greater than those of OK. These demonstrate that the integration scheme is helpful for improving the interpolation accuracy and stability of the ANNs method but harmful to the OK methods. After observing other elements' experiments in the largest training data size case, we find that the OK method slightly outperforms other methods in the V experiment; the BPANNs method offers the most accurate and stable interpolation results for Ni; the IDW method performs best in the Cu experiment. Thus, it is difficult to find a method which is the best method for all the elements in the largest training data size case because the accuracy and stability of each method vary for different elements. However, the results for all the elements each reach the same conclusion that the integration scheme is helpful for improving the interpolation accuracy and stability of the ANNs-based methods.

In the medium training data size case, the statistic results of Mn show that the M_{mr} of IDW and the D_{qqr} of OK are smaller than those of other methods and the M_{mr} and D_{qqr} of the ANNs-based methods including BPANNs and RBFANNs are greater than those of other methods; which demonstrates that the traditional methods including IDW and OK perform better than the ANNs-based methods in accuracy and stability. Similar to Project A, the integration of the ANNs-based modules can reduce the M_{mr} and D_{qqr} of the ANNs-based methods, which shows that the integration helps in improving the interpolation accuracy and stability of the ANNs-based methods. Observations of all the other elements' experiments in the medium training data size case show that the ANNs-based methods are worse than the traditional methods including IDW and

OK in interpolation accuracy and stability for all the elements except Cu. However, the integration scheme improves the interpolation accuracy and stability of the ANNs-based methods in all the elements' experiments.

When the training sample is reduced to 16 points, the M_{mr} and D_{qqr} of the traditional methods including IDW and OK are smaller than those of the non-integrated ANNs-based methods but greater than those of the integrated ANNs-based methods in the experiments of Mn; which indicates that IBPANNs and IRBFANNs are superior to other methods in the interpolation accuracy and stability. Moreover, the IRBFANNs method outperforms other methods with the most accurate and stable interpolation results. In accordance with Project A and Project B, the M_{mr} and D_{qqr} of the ANNs-based methods are reduced through the integration of the ANNs modules, which shows that the integration scheme can aid in the interpolation accuracy and stability. The results of other elements' experiments in the smallest training data size case are consistent with those of Mn.

From the observations of all the elements' experimental results in the same training data size case, we can draw consistent conclusions that the performance of each method is varied for different elements, and the ANNs-based methods are inferior to the traditional methods but the integration of the ANNs modules improves the performance of the un-integrated ANNs-based methods. Also, the integrated ANNs-based methods outperform other methods in both interpolation accuracy and stability for all the elements in the smallest training data size case.

In order to visually compare the performances of these interpolation methods, the spatial distribution maps of Mn concentration in soil based on the interpolation values of seven interpolation methods are plotted in Fig. 10.

Considering the spatial continuity and spatial variability of soil heavy metal concentrations, the reasonable interpolation spatial distribution map which is helpful for analyzing the spatial distribution trends should be continuous in concentration values, rich in spatial variability and representative in the original data range. In other words, the interpolation data range should not be far beyond the original data range, which are logarithmically transformed to obtain a standard normal distribution with mean 0 and deviation 1 prior to experiments. Therefore, the graphical representations of the interpolation results of four elements using different interpolation methods under different training data sizes as shown in Fig. 10 are discussed in three aspects including spatial continuity, spatial variability and singular interpolation values.

Comparisons among Projects A, B and C show that the spatial distribution map becomes more and more smooth with the decrease of the training data size, which indicates that the decrease of training data size enhances the spatial continuity which is clear in the IDW and OK rows

but reduces the spatial variability which is evident in the RBFANNs and IRBFANNs rows. As shown in IRBFANNs rows, although the interpolation data range as described by the color bar in Project C shrinks, it is more representative in the original data range than that in Project A. The interpolation map of Project A is still better than that of Project C because the interpolation map of Project C offers too little detail in the spatial variation. To sum up, the performances of all the methods become worse when the training data is small, which is consistent with the conclusion achieved by statistic analysis based on box plots of R_{mse} values in 100 tests.

The interpolation maps of seven methods under the largest training data size show that IDW gives an irregular spatial distribution pattern; therefore, the IDW-based result is inappropriate for lacking spatial continuity of interpolation values. The OK-based map which is spotty and less distinct in colors shows rather smooth patterns with fairly low level detail in spatial variability. As shown in the BPANNs and OK rows, the BPANNs method yields more details in the spatial variation than OK. Moreover, the BPANNs-based map illustrates the spatial distribution trends of Mn concentration which are invisible in the OK-based map. Compared with the BPANNs-based map, the surface estimated by RBFANNs provides more details in the spatial variation and spatial distribution trend. But as described by the color bar, the lowest and highest values obtained by RBFANNs are, respectively, much greater or much less than those of the original data, which conforms to the standard normal distribution, so the RBFANNs-based map is not representative of the original data range. In the IOK row, it can be seen that the surface obtained by IOK is almost identical to the surfaces produced by OK except that they are less spotty and the minimum and maximum estimation of Mn concentrations in soil obtained by IOK are shrunken. The surface obtained by IBPANNs is similar to that produced by BPANNs except that the IBPANNs-based map is smoother. As shown in the IRBFANNs row, the IRBFANNs-based map is continuous, rich in spatial variability and representative of the original data range so the surface estimated by IRBFANNs provides a more reasonable spatial representation of Mn concentration in soil than other methods. Comparisons under the same training data size in Project B and Project C also confirm the conclusion, so the IRBFANNs method is more reliable for improving the quality of the interpolation maps of Mn concentrations in soil.

The OK and IOK method interpolation maps have a strong tendency to be smooth, but heavy metal elements in the soil have strong spatial variability and complex spatial distribution. The interpolation maps of IDW, BPANNs, and IBPANNs methods are better than those of OK and IOK in spatial variability; however, the spatial distribution trends in these maps are not clear, especially

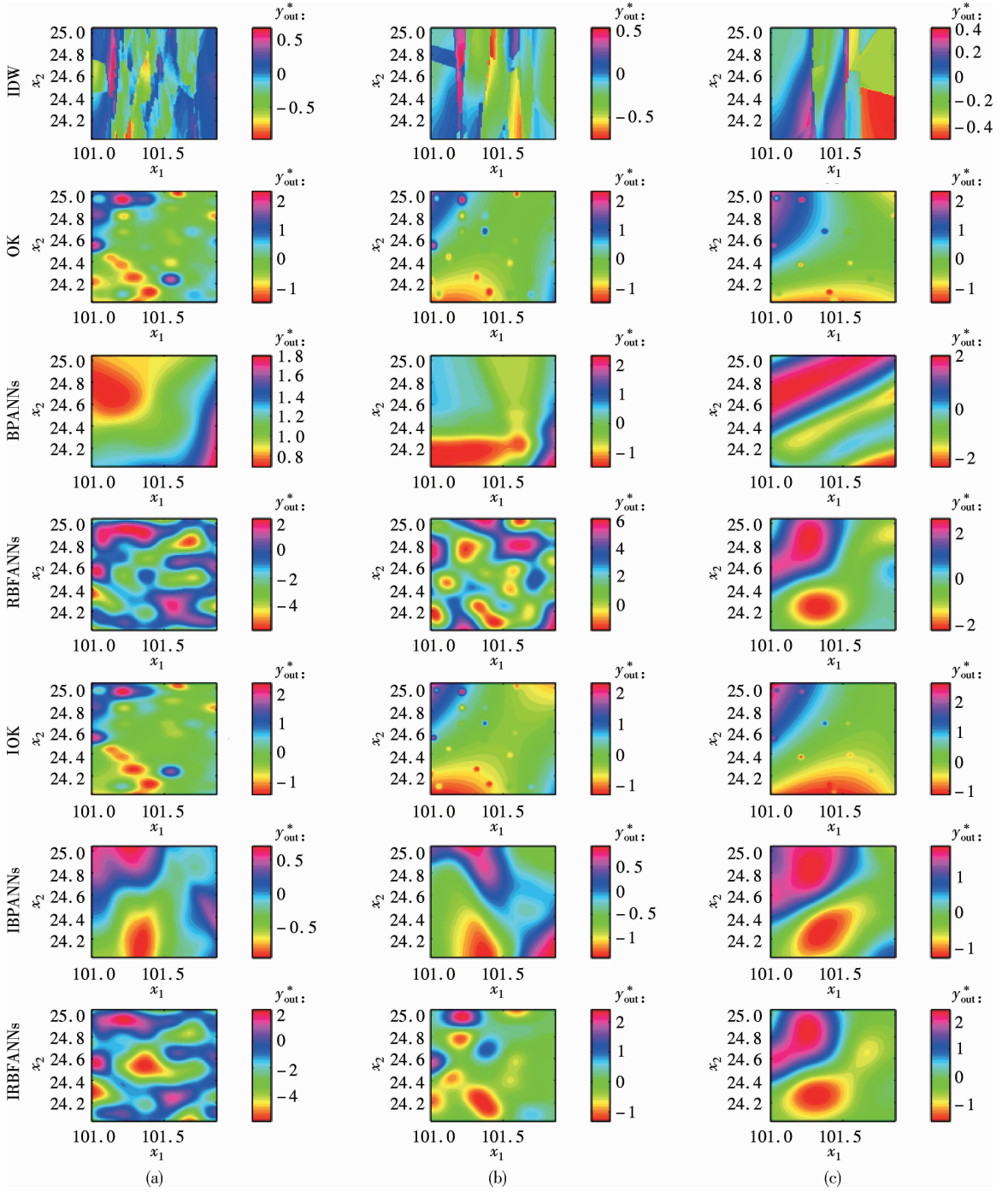


Fig. 10 Plots of interpolated Mn using different interpolation methods under different training data sizes. (a) Size of 41; (b) Size of 26; (c) Size of 16

in the IDW-based maps for lacking of spatial continuity. Although the RBFANNs maps provide the obvious spatial distribution trends of soil heavy metal concentrations, the RBFANNs interpolation values are not the representatives of the original data range. Compared with OK, IOK, IDW, BPANNs, IBPANNs and RBFANNs, the IRBFANNs technique appears stronger in creating continuity, is richer in spatial variation and more representative in the original data range maps with obvious spatial distribution trends especially in the smallest training data size case; which is in accordance with the conclusion obtained by

statistic analysis above.

4 Conclusion

In this paper, a novel spatial interpolation method is proposed to provide an accurate and stable interpolation of soil heavy metal concentrations in a mountain region. Comparative experiments based on statistic analysis and interpolation maps obtain the consistent results that the proposed method obtained the small M_{mr} and D_{qqr} values, meanwhile, offering the high quality interpolation maps with rich spatial variation details and obvious spatial dis-

tribution trends when the training data size decreases. The experimental results show that the IRBFANNs method provides the most accurate and stable interpolation values which can be used to make a high quality interpolation map of heavy metal concentrations in soil among seven considered interpolation techniques in the cases of sparse sampling density. Consequently, IRBFANNs offers an effective interpolation method to improve the accuracy and stability of interpolation values in a mountain region where heavy metal concentrations in soil vary greatly over the land surface and the sampling sites are usually inadequate and irregular, which makes it difficult for traditional interpolation methods to carry out an accurate and stable spatial interpolation.

References

- [1] Qing Chunxiang, Wang Mingming, Xu Yanbo. Current situation of soil contamination by heavy metals and research progress in bio-remediation technique[J]. *Journal of Southeast University: Natural Science Edition*, 2013, **43**(3): 669–674. (in Chinese)
- [2] Nagajyoti P C, Lee K D, Sreekanth T V M. Heavy metals, occurrence and toxicity for plants: a review[J]. *Environmental Chemistry Letters*, 2010, **8**(3): 199–216.
- [3] Beiseyeva G, Abuduwali J. Migration and accumulation of heavy metals in disturbed landscapes in developing ore deposits, East Kazakhstan [J]. *Journal of Arid Land*, 2013, **5**(2): 180–187.
- [4] Li J, Heap A D. A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors[J]. *Ecological Informatics*, 2011, **6**(3): 228–241.
- [5] Stein M L. *Interpolation of spatial data: some theory for kriging* [M]. Berlin: Springer, 1999: 153–154.
- [6] Kishné A S, Bringmark E, Bringmark L, et al. Comparison of ordinary and lognormal kriging on skewed data of total cadmium in forest soils of Sweden[J]. *Environmental Monitoring and Assessment*, 2003, **84**(3): 243–263.
- [7] Philippopoulos K, Deligiorgi D. Application of artificial neural networks for the spatial estimation of wind speed in a coastal region with complex topography[J]. *Renewable Energy*, 2012, **38**(1): 75–82.
- [8] Opitz D W, Shavlik J W. Generating accurate and diverse members of a neural-network ensemble[C]//*Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 1996: 535–541.
- [9] Hansen L K, Salamon P. Neural network ensembles[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, **12**(10): 993–1001.
- [10] Baxter B. The interpolation theory of radial basis functions[EB/OL]. (2010-06) [2014-03-01]. <http://arxiv.org/abs/1006.2443>.
- [11] Zhu Xiaorong, Shen Lianfeng. RBF-based cluster-head selection for wireless sensor networks [J]. *Journal of Southeast University: English Edition*, 2006, **22**(4): 451–455.
- [12] Li Qing, Cheng Jiachang, Hu Yueming. Spatial interpolation of soil nutrients based on BP neural network[J]. *Agricultural Science & Technology*, 2014, **15**(3): 506–511.
- [13] Zou Youlong, Hu Falong, Zhou Cancan, et al. Analysis of radial basis function interpolation approach[J]. *Applied Geophysics*, 2013, **10**(4): 397–410.
- [14] Yee P V, Haykin S. *Regularized radial basis functional networks: theory and applications*[M]. New York: John Wiley & Sons, Inc., 2001: 46–52.
- [15] Carney J, Cunningham P. The neural BAG algorithm: optimizing generalization performance in bagged neural networks[C]//*Proc of European Symposium on Artificial Neural Networks*. Bruges, Belgium, 1999: 135–140.
- [16] Chen Fengwen, Liu Chengwu. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan Province[J]. *Paddy and Water Environment*, 2012, **10**(3): 209–222.
- [17] Oliver M A, Webster R. Kriging: a method of interpolation for geographical information systems [J]. *International Journal of Geographical Information Systems*, 1990, **4**(3): 313–332.

一种用于山区土壤重金属评估的集成 RBF 空间插值算法

李宝磊¹ 张榆锋¹ 施心陵¹ 章克信² 张俊华¹

(¹ 云南大学信息学院, 昆明 650091)

(² 昆明医学院第二附属医院心血管科, 昆明 650031)

摘要:提出了一种在山区能够准确、稳定地预测未采样点土壤重金属浓度的集成径向基函数神经网络空间插值方法(IRBFANNs)。该方法集成径向基函数神经网络和神经网络集成技术的优点。为了研究所提 IRBFANNs 方法的性能,进行了 3 组不同采样密度条件下的实验。通过 Mn 元素插值的均方根误差和分布估计图对 IRBFANNs 和其他 6 个插值方法进行了比较。实验结果表明:IRBFANNs 方法在精确性和稳定性方面优于其他参评方法,且在采样密度稀疏条件下该方法能够提供细节较丰富的分布估计图。

关键词:集成镜像基函数神经网络;空间插值;土壤重金属;山区

中图分类号:TP183