SPATIAL ESTIMATION OF SOIL MOISTURE AND SALINITY WITH NEURAL KRIGING

Zhong Zheng [1,2] , Fengrong Zhang [1,*] , Xurong Chai [1] , Zhanqiang Zhu [1] , Fuyu Ma [2]

[1] *College of Resources and Environment, China Agricultural University, Beijing, P. R. China 100094*

[2] *College of Agriculture, Shihezi University, Shihezi, Xinjiang Province, P. R. China 832003*

[*] *Corresponding author, Address: College of Resources and Environment, China Agricultural University, Beijing, P. R. China 100094, Tel: +86-10-62732643, Fax: +86-10-62732643, Email: zhangfr@cau.edu.cn, zhenglxx@gmail.com*

Abstract: The study was carried out with 107 measurements of volumetric soil water content (SWC) and electrical conductivity (EC) for soil profile (0-30 cm) and the estimating accuracy of ordinary kriging (OK) and back-propagation neural network (BPNN) was compared. The results showed that BPNN method predicted a slightly better accurate SWC than that of OK, but differences between both methods were not significant based on the analysis of covariance (ANOVA) test (P >0.05). In addition, BPNN performed much better in EC prediction with higher model efficiency factor (E) and ratio of prediction to deviation (RPD) (E=0.8044 and RPD=3.54) than that of OK (E=0.7793 and RPD=0.39). Moreover, a novel neural kriging (NK) resulting from the integration of neural network (NN) and ordinary kriging (OK) techniques was developed through a geographic information system (GIS) environment for obtaining trend maps of SWC and EC. There was no significance between results of NK and OK through trend maps. Comparing with OK, NK gives better spatial estimations for its great advantage of establishing spatial nonlinear relationships through training directly on the data without building any complicated mathematical models and making assumptions on spatial variations.

Keywords: soil moisture, soil salinity, spatial estimation, ordinary kriging, artificial neural networks

1.   INTRODUCTION

In recent years, agricultural development on the oasis plain of Maigaiti, Northwest China, has being threatened by soil secondary salinization irrigation-induced due to excessive and inefficient water use. Salt accumulation and excessive salt concentration in farmlands has led to land degradation, crop yield decrease, abandoned lands increase, water quality deterioration and environmental degradation (Kitamura et al., 2006). Therefore, consistent and early stage identification of soil salinity as well as assessment of soil moisture is vital for crop production, especially in arid areas where harsh climatic conditions together with rapidly increasing population densities (Farifteh et al., 2007).

In the farmland, it is impractical to sample all the points with the desired temporal frequency in order to research the variability of soil water and salinity content. Optimizing spatial sampling scheme to reduce sampling density and estimation of unsampling values can save time and money (Ferreyra et al., 2002; Li et al., 2007). However, its effectiveness relies on the accuracy of the spatial interpolation used to define the spatial variability. Multivariate techniques such as geostatistics and artificial neural network (ANN) have been widely used as estimation tools. Geostatistics provides descriptive tools such as kriging to directly implements the prediction of an attribute at an unsampled location according to known data points within a local neighborhood surrounding (Emery & Ortiz, 2007). ANN has the ability to model extremely non-linear and complicated relationships between a set of inputs, and are operated by using the available input and output responses without considering inherent system parameters (Sarangi et al., 2006). It can be used as an alternative to predict regionalized variables (RV) which are functions on geographic locations (Huang & Foo, 2002; Farifteh et al., 2007). The performances of different interpolation methods such as ordinary kriging (OK), inverse distance weighting(IDW), splines and so on, have been analyzed in several studies, whereas there have been many conflicting reports concerning the performances of different interpolation methods (Gotway et al., 1996; Patel et al., 2002; Brocca et al., 2007). In addition, many comparisons of various interpolation techniques have been made in respect to different data sets used, different mathematical procedures and different input parameters (Boken et al., 2004; Robinson & Metternicht, 2006). Moreover, very few studies compare the performance of OK and ANN methods simultaneously (Rizzo & Dougherty, 1994).

Hence, the aim of this paper is to: (i) identify the performance of OK and ANN for estimation of soil moisture and salt content in a given area within a threshold of error. (ii) investigate the applicability of neural kriging (NK) resulting from the integration of neural network (NN) and ordinary kriging (OK) techniques; (iii) illustrate trend maps of soil moisture and salinity distribution for study area.

2.   METHODS

2.1   Site description

The study site is located at the west margin of Taklamakan desert in Northwest China. Field research was conducted on a cotton farmland (30°14'29"-39°14'57"N, 78°06'21"-78°07'00" E), which is located in northeast region of Maigaiti County of Xinjiang (Fig. 1). The research field covers 0.54 km^2 (900m x 600m) with a 2-4% slope northwest to southeast. The area is 442 m above sea level and experiences an arid climate with mean annual temperature, precipitation, evaporation and frost-free period of 12.4 °C, 46.5 mm, 2526 mm and 212 d over a 20-year period, respectively. The soil texture is dominant sand soil and its distribution and variability in the topsoil are influenced by Taklamakan desert in arid regions.

The farmland for the present study was reclaimed in 2000 and cotton was planted from 2003. Since the groundwater in the study site has higher mineral degree, soil secondary salinization and drought are the main limiting factors for crop production in arid regions.
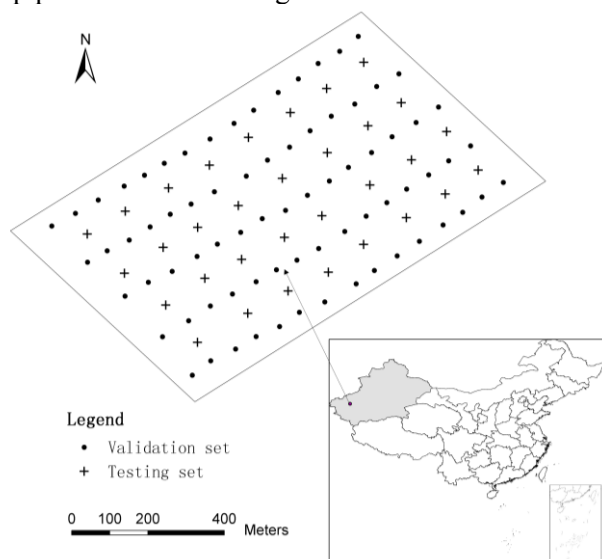

Fig.1: Research area location and map of the distribution of soil samples.

2.2   Data acquisition

A grid sampling scheme (40-60 m sampling space) was imposed on the field with 107 sample measurements of volumetric soil water content (SWC) and electrical conductivity (EC). SWC was measured using a portable Time Domain Reflectometry (Soil Moisture Equipment Corp., TRASE$^®$TDR) and

EC using a portable WET sensor (Delta-T Devices Ltd., Cambridge, UK). On each sampling point, we inserted vertically a triple wire TDR probe to monitor soil moisture and a wire WET sensor probe for soil salinity in the soil profile (0-30 cm). Each EC and SWC measurement was geo-referenced using a Differential Global Positioning System (DGPS). At each sampling grid point, five EC or SWC measurements were made within a 1-m diameter circle. The average reading for each grid point was computed as EC or SWC datum point. Sample measurements were implemented before cotton cultivation at March 12, 2007. Among this set of 107 measurements, a set of 75 data were selected to constitute validation samples, remaining set of 32 for testing samples (Fig. 1).

2.3   Spatial estimation methods

2.3.1   Kriging

Kriging estimate relies on a weighting scheme where closer sample locations have greater impact on final prediction. At an unsampled location and for a given variogram, a kriging estimate is simply an optimally weighted average of the surrounding sampled data (Emery & Ortiz, 2007). In this study, the total numbers in sample site were divided into training sets with 70% of all samples used for developing a geospatial model, and testing sets with the remaining 30% used to test the performance of the models (Fig. 1). According to the integrated performance of different kriging methods, ordinary kriging(OK), for its better performance obtained from a cross-validation procedure in the study, was selected to take part in performance comparison to the following artificial neural network (ANN).

2.3.2   Artificial neural network (ANN)

In general, all estimation techniques require the modeling of the function $Z=f(X, Y)$, where (X, Y) being the station coordinates (latitude–longitude) and Z the regionalized variable (RV). Spatial distribution of soil moisture and salinity is generally related to geographic locations(X coordinate, Y coordinate). Therefore, the network input layer used in this study relates to geographic X coordinate and Y coordinate, while the network output layer relates to soil moisture and soil salinity. The number of neurons in the hidden layer is of great importance, as too many neurons may cause over-fitting problems, and it can be defined using a formula recommended or using a trial-and-error approach (Huang & Foo, 2002). Thus, a 3-layer feed-forward back propagation neural network (BPNN) (topology structure: 2 x 5 x 2) was established (Fig. 2) and used within Neural Network Toolbox of

Matlab 7.0 (The MathWorks Inc. Natick, MA). Tan-sigmoid transfer functions and log-sigmoid transfer functions (non-linear) were selected for the hidden and output layers, respectively. The Levenberg-Marquardt algorithm, which provides a fast optimization, was used for network training. Total of 107 soil samples were divided into two groups as 75 for the development (training and validation) and 32 for the test (Fig.1). Simulated error is 3% and 0.02 mS/cm for SWC and EC in training and testing process, respectively.

However, ANN only allows the RV estimation, but not the predictor variance, which is possible with kriging (Rizzo & Dougherty, 1994; Koike et al., 2001). In some applications ANN is coupled with kriging estimation, which was called neural kriging (NK) by Rizzo and Dougherty (1994). NK is divided into two steps: the first is neural and uses neural network, and the second uses OK. The final estimates are produced as a sum of NN estimates and OK estimates. Hence NK is an integrated interpolation technique. In spatial estimation its utilization is justified by the fact that it extracts its knowledge only from data, which contain information about the spatial distribution of the RV. The present paper follows the NK approach to estimate soil moisture and salinity and compare the obtained estimates with OK technique, and then illustrate trend maps of SWC and EC.
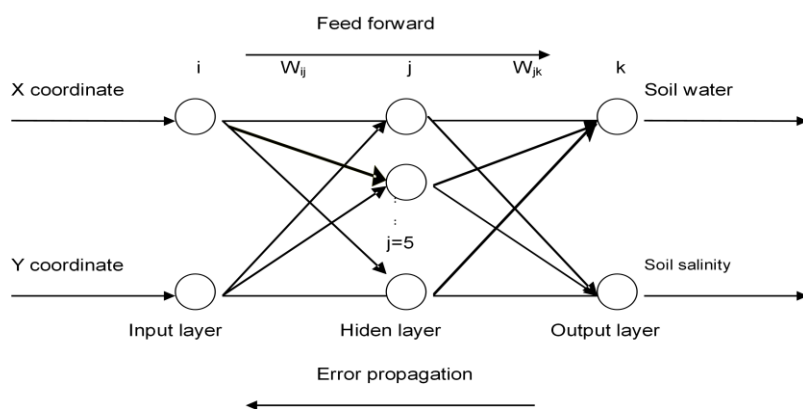

Fig.2: Three-layer feed forward back propagation neural network structure.

2.4 Evaluation criteria

We used cross-validation to validate the accuracy of interpolation algorithms and examine the difference between the measured values and the predicted values using mean absolute errors (MAE) or relative MAE (MAE%) (Eq. (1)), root-mean-square error (RMSE) or relative RMSE (RMSE%) (Eq. (2)), ratio of prediction to deviation (RPD) (Eq. (3)) and model efficiency factor (E)(Eq. (4)). The RPD indicates strength of

statistical correlation between measured and predicted values. MAE, RMSE and E values indicate degree of agreement between measured and predicted values. Detailed descriptions and definitions of these model performance parameters are given by Robinson & Metternicht (2006) and Farifteh et al. (2007).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Z^{*}(x_i) - Z(x_i)\right| \qquad MAE\% = 100\frac{MAE}{\overline{Z}(x_i)} \qquad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Z^{*}(x_i) - Z(x_i)\right)^2} \qquad RMSE\% = 100\frac{RMSE}{\overline{Z}(x_i)} \qquad (2)$$

$$RPD = \frac{\sqrt{\sum_{i=1}^{N}\left(Z^{*}(x_i)\right)^2 - \left[\left([\sum_{i}^{N}Z^{*}(x_i)]^2 / N\right)/(N-1)\right]}}{\sqrt{\sum_{i=1}^{N}\left(Z^{*}(x_i) - Z(x_i)\right)^2 - \left[\left([\sum_{i}^{N}(Z^{*}(x_i) - Z(x_i)]^2 / N\right)/(N-1)\right]}} \qquad (3)$$

$$E = 1 - \frac{\sum_{i=1}^{N}\left(Z^{*}(x_i) - Z(x_i)\right)^2}{\sum_{i=1}^{N}\left(\left|Z^{*}(x_i) - \overline{Z}(x_i)\right| + \left|Z(x_i) - \overline{Z}(x_i)\right|\right)^2} \qquad (4)$$

Where $Z(x_i)$ is the observed (measured) value of z at locations x_i, $Z^{*}(x_i)$ is the predicted value at the same locations, $\overline{Z}(x_i)$, $\overline{Z}^{*}(x_i)$ is the average measured and predicted value, respectively. N is the number of values in the dataset.

Conventional statistical analyses were conducted using the software package SPSS 12.0 for Windows (SPSS Inc., MatLab®, USA). Geostatistical analyses and mapping were performed by using ArcGIS 9.0 software package (Environmental Systems Research Institute, Redlands, CA).

3.   RESULTS AND DISCUSSION

3.1   Descriptive statistics

Based on above estimation methods as discussed, descriptive statistics for SWC and EC used in the development and validation of BPNN and OK are summarized in Table 1. It is observed than both variables are approximately normal distribution avoiding the need for data transformation according to the coefficient of skewness. The soil of study area has wide ranges of SWC and EC with 27.3–48.3% and 0.07–1.90 mS/cm, respectively. The predicted

values of SWC and EC with BPNN method, ranging between 31.4–43.2% and 0.24–1.12 mS/cm respectively, have smaller range and distributions closer to the average measured values than that of OK, which rang between 28.4-47.3% and 0.18-1.47 mS/cm respectively. Besides, there is less standard deviation of BPNN than that of OK. Hence, we primarily conclude that BPNN method could perform somewhat better than OK from the descriptive statistics obtained from cross-validation.

Table 1 Summary statistics of SWC (%) and EC (mS/cm)

Soil property	Mean	Maximum	Minimum	Std.Dev.	Skewness	Kurtosis
SWC (%)						
Measured	36.7	48.3	27.3	5.2098	0.2290	2.1629
OK method	37.3	47.3	28.4	4.4309	-0.0659	2.2501
BPNN method	37.2	43.2	31.4	3.4303	-0.2273	1.7673
EC (mS/cm)						
Measured	0.72	1.90	0.07	0.4208	0.5471	2.6463
OK method	0.66	1.47	0.18	0.3081	-0.0465	1.8989
BPNN method	0.70	1.12	0.24	0.2544	-0.0903	1.8407

3.2   Spatial estimation and its performance

SWC and EC can all be estimated by both methods while the former variable yields very good results (Fig.3a) and the latter shows a bad generalization (Fig.3b). Besides, Fig.4a and Fig.4b also shows good performance of SWC by both methods, respectively. The worse performance of EC can be seen in Fig.4c and Fig.4d, where measured and predicted values are more scattered. Comparing to coefficient of determination (R^2) (Fig.4), R^2 between predictions and observations ranges from 0.5704 to 0.5763 for SWC and from 0.4697 to 0.4948 for EC by both methods. This result reveals that SWC prediction value has better performance than EC. Furthermore, spatial variability of EC is being under-predicted, which is in good agreement with the field finding similar to the study reported by Li et al. (2007). It has to be pointed out that the quality and quantity of the data used for the study were not adequate to the complexity of spatial variability of soil properties. Boken et al. (2004) considered that the accuracy estimate can be improved by enhancing the representation of sampling sites as well as by limiting the estimations to irrigated areas within counties.

Based on the statistical parameters as discussed, both OK and BPNN model are validated using the model efficiency factor E and RPD of predicted values. The E and RPD between reference measurements, i.e. accurate or good prediction if RPD and E values are higher than 2.5 and 0.80 respectively by Farifteh et al. (2007), suggest an accurate to good prediction. It is observed from Table 2 that BP method (with E=0.8578 and RPD=15.06) predicts an accurate SWC nearly similar to that of OK (with E=0.8294 and RPD=16.76), whereas different performances occurred in predicting EC.

BPNN method with much higher E =0.8044 and RPD=3.54 performs well while OK with E=0.7793 and RPD=0.39 performs poorly in EC prediction. In addition to MAE% and RMSE%, it is considered that OK performs slightly better, but differences between BPNN and OK methods in predictions are not significant for both variables based on the analysis of covariance (ANOVA) test (P >0.05).
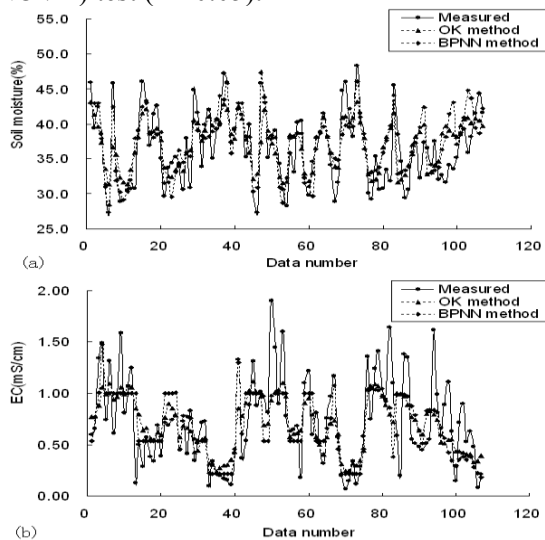

Fig.3: Comparison of the estimated and measured values with both methods for (a) SWC and (b) EC.
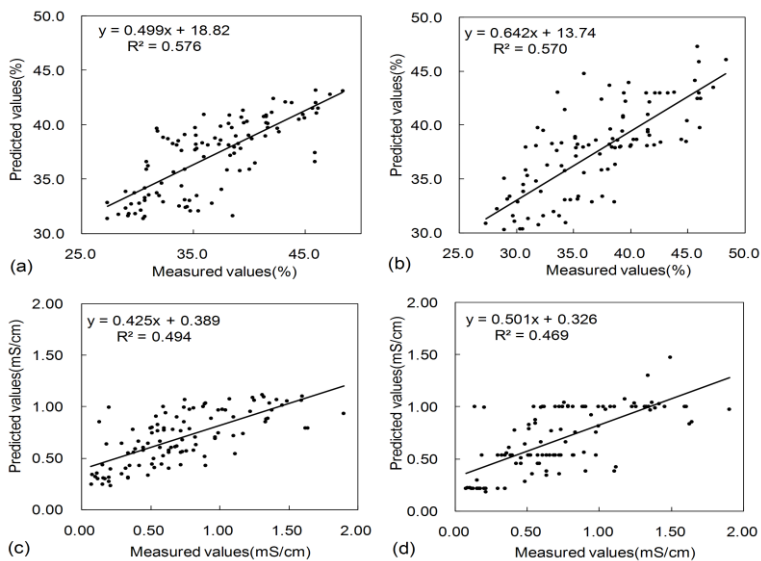

Fig.4: Plots of the measured versus predicted values and the fitted regression line and equation for (a) SWC with OK, (b) SWC with BPNN, (c) EC with OK and (d) EC with BPNN method.

Table 2 Performance of predicted values using OK and BPNN methods

Soil property	MAE	MAE%	RMSE	RMSE%	E	RPD
SWC (%)						
OK method	2.88	7.84	3.45	9.39	0.8294	16.76
BPNN method	2.79	7.63	3.49	9.51	0.8578	15.06
EC (mS/cm)						
OK method	0.24	32.67	0.30	41.9	0.7793	0.39
BPNN method	0.23	32.32	0.31	42.7	0.8044	3.54

3.3   Trend maps of SWC and EC by NK and OK

As explained in section 2.3, the data nonlinear trend can be estimated by neural kriging (NK) better than the ordinary kriging (OK) estimator if neural networks are coupled with kriging. The NK approach presented here for spatial estimation is the result of the integrated of two different techniques: BPNN evaluation and ordinary kriging (OK).
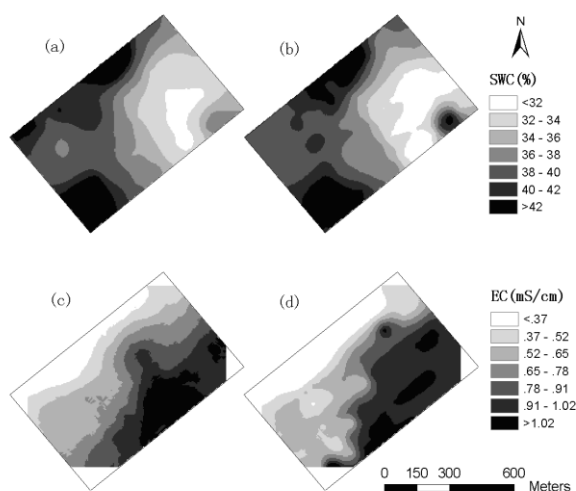

Fig.5: Trend maps obtained from estimated values of (a) SWC by OK, (b) SWC by NK, (c) EC by OK and (d) EC by NK method.

Trend maps of soil moisture and salinity produced by OK and NK are illustrated in Fig.5. Positive values show areas where SWC or EC predictions are higher and negative values represent areas where prediction values are lower. By looking at the maps of Fig.5, differences between results of OK and NK are not significant. In addition, the smoothed contour maps in Fig.5a and Fig. 5b display quite similar patterns with low soil moisture in the eastern section and high in the western and southern parts of the study area, whereas contour maps in Fig.5c and Fig. 5d with high salinity in the eastern section and low in the western sections including northwestern and southwestern parts. Because of the research field with a 2-4% slope

northwest to southeast and soil texture with high sand content and permeability, salt leaching with irrigation and upward transport with evaporation are frequent. This may be result in rapid salt leaching and accumulation in the topsoil in southeastern section of this field. Therefore, water flooding for irrigation should be reduced to a great extent and water-saving irrigation should be promoted in irrigation areas.

4.   CONCLUSIONS

The key aim of the work is to contribute to the problem of spatial estimation of soil properties with a novel solution, through the combined utilization of statistical, geostatistical and artificial neural network (ANN) techniques. The approach of neural kriging(NK), coupled neural network (NN) with ordinary kriging (OK), has a great potential for predicting and mapping soil properties. The procedure of NK requires information on the coordinates (X, Y) of a survey point in the input. After the completion of training, the trained network is tested to estimate soil properties for all sample locations within the area of investigation by producing a corresponding contour map with kriging technique. The NK results compare very well with similar contour maps generated using OK techniques. The main advantage of NK approach is its ability in establishing patterns or nonlinear relationships through training directly on the data without building any complicated mathematical models and making assumptions on spatial variations. It can be seen that this method yields high and significant spatial relations and gives better spatial estimations.

ACKNOWLEDGEMENTS

REFERENCES

Boken, V. K., Hoogenboom G., Hook, J.E., Thomas, D.L., Guerra, L.C., Harrison, K.A.. Agricultural water use estimation using geospatial modeling and a geographic information system, Agric. Water Manage, 2004, 67: 85–199.

Brocca, L., Morbidelli, R., Melone, F., Moramarco, T.. Soil moisture spatial variability in experimental areas of central Italy, Journal of Hydrology, 2007, 333: 356–373.

Emery, X. and Ortiz J. M.. Weighted sample variograms as a tool to better assess the spatial variability of soil properties. Geoderma, 2007, 140: 81–89.

Farifteh, J., Van der Meer, F., Atzberger, C., Carranza, E.J.M.. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). Remote Sensing of Environment, 2007, 110: 59–78.

Ferreyra, R.A., Apezteguia, H.P., Sereno, R, Jones, J.W.. Reduction of soil water spatial sampling density using scaled semivariograms and simulated annealing. Geodenna, 2002, 110: 265–289.

Gotway, C.A., Ferguson, R.B., Hergert,G.W., Peterson, T.A.. Comparison of kriging and inverse-distance methods for mapping soil parameters. Am. J. Soil Sci., 1996, 60: 1237–1247.

Huang, W.R. and Foo, S.. Neural network modeling of salinity variation in Apalachicola River. Water Research, 2002, 36: 356−362.

Kitamura, Y., Yano, T., Honna T., Yamamoto S. and Inosako, K.. Causes of farmland salinization and remedial measures in the Aral Sea basin—Research on water management to prevent secondary salinization in rice-based cropping system in arid land. Agricultural Water Management, 2006, 85: 1–14.

Koike, K., Matsuda, S., Gu, B.. Evaluation of interpolation accuracy of neural kriging with application to temperature-distribution analysis. Mathematical Geology, 2001, 33: 421-448.

Li, Y., Shi,. Z., Wu, C.F., Li, H.X., Li, F.. Improved prediction and reduction of sampling density for soil salinity by different geostatistical methods. Agricultural Sciences in China, 2007, 6: 832–841.

Patel, R.M., Prasher, S.O., Goel, P.K., Bassi, R.. Soil salinity prediction using artificial neural networks. J. Am. Water Resour. Assoc, 2002, 38: 91–100.

Rizzo, D.M., Dougherty, D.E.. Characterization of aquifer properties using artificial neural networks: neural kriging. Water Resources Research, 1994, 30: 483–497.

Robinson, T.P. and Metternicht, G.. Testing the performance of spatial interpolation techniques for mapping soil properties. Computers and Electronics in Agriculture, 2006, 50: 97–108.

Sarangi, A., Singh M., Bhattacharya, A.K., Singh, A.K.. Subsurface drainage performance study using SALTMOD and ANN models. Agricultural Water Management, 2006, 84: 240–248.