International Journal of GEOMATE, Oct., 2017, Vol.13, Issue 38, pp.23-30
Geotec., Const. Mat. & Env., ISSN:2186-2990, Japan, DOI: http://dx.doi.org/10.21660/2017.38.606721

EVALUATION OF VARIOUS INTERPOLATION TECHNIQUES FOR ESTIMATION OF SELECTED SOIL PROPERTIES

*R. S. Morgan[1], M. Abd El-Hady[2], I. S. Rahim[1], J. Silva[3] and S. Ribeiro[3]

[1]Soils and Water Use Dept., Agricultural and Biological Research Division, National Research Centre, Egypt;[2]Water Relations and Field irrigation Dept., Agricultural and Biological Research Division, National Research Centre, Egypt;[3]NOVA IMS, Universidade Nova de Lisboa, Portugal

*Corresponding Author, Received: 12 Jan. 2017, Revised: 20 Feb. 2017, Accepted: 24 March 2017

ABSTRACT: Efficient soil management practices depend on the spatial distribution of soil properties which varies significantly even within the same field. Considering that it is impossible for any monitoring technique to provide spatially continuous data, spatial interpolation plays an indispensable role in estimating the missing values where no actual value was measured. The objective of this study was to evaluate various interpolation techniques for the estimation of selected soil chemical properties in a study area located about 85km to the north west of Cairo, Egypt. The studied soil properties included soil salinity, available phosphorus and nitrogen. The interpolation techniques included two commonly used techniques namely, Ordinary Kriging (OK) and Inverse Distance Weighting (IDW).The Artificial Neural Networks (ANN) method, which is considered a somewhat new approach was also evaluated. Soil samples were collected at approximately 200×200 m grids at 0-25 cm depth. The cross-validation method was used for evaluating the selected methods utilizing root mean square error (RMSE) and mean relative error (MRE). This study revealed that ANN had the highest accuracy followed by OK then IDW in terms of both RMSE and MRE when interpolating the studied soil properties. Nevertheless, these results are dependent on the accuracy of the designed network which must have an overall accuracy of coefficient of correlation (R) more than 0.80 between the predicted and the actual data. It also revealed that the best IDW with the highest accuracy must have a power of 2 for salinity and nitrogen and a power of 3 for phosphorus.

Keywords: *Soil properties interpolation, Ordinary kriging, Inverse distance weighting, Artificial neural networks*

1. INTRODUCTION

Soil properties which vary spatially and temporally from a field to a larger region scale, are influenced by both intrinsic soil formation factors, such as soil parent materials and extrinsic factors such as soil management practices, fertilization, and crop rotation [1]. Soil properties variations should be monitored and quantified for efficient farming practices. Hence, arise the importance of spatial interpolation methods in providing spatially continuous data that otherwise unfeasible utilizing in-site monitoring techniques.

Most common interpolation techniques calculate the estimates for a property at any given location by a weighted average of nearby data. Weighting is assigned either according to deterministic or statistical criteria[2]. Geostatistics, which is a branch of applied statistics [3], can characterize the regular component of the variation in natural objects, including soils [4]. It can be considered as the tools for studying and predicting the spatial structure of georeferenced variables [5]. The basic tool of geostatistics is known as semivariogram analysis which is used to identify and describe the extent of spatial variability of regionalized variables [2]. Kriging, as a geostatistical method, is based on spatial autocorrelation of the data, which determines the statistical relationship between values where sample observations are available. When the relationship is established, it is used to predict the attribute values at unsampled locations [6].

On the other hand, inverse distance weighting (IDW) is considered among the deterministic interpolation methods. In IDW method, it is assumed that the rate of correlations and similarities between neighbors is proportional to the distance between them and can be defined as a distance reverse function of every point from neighboring points [2]. While kriging requires the preliminary modeling step of a variance-distance relationship, IDW does not require such step and is very simple and quick.

Over the last two decades, conventional interpolation techniques such as OK and IDW have been employed in agricultural practices for predicting spatial variability of soil properties. Nevertheless, various attempts have been introduced to select the most appropriate method. According to [7] OK and IDW methods gave similar root mean square error (RMSE) values when evaluating the studied soil chemical parameters. On the other hand, most studies favored OK over IDW when studying soil chemical properties [2], [8] and [9]. Nevertheless, kriging and the inverse distance weight interpolation methods have various limitations such as strong subjectivity, numerous assumptions, poor adaptive variation, etc. [10].

Artificial neural network (ANN) is an interconnected assembly of simple processing elements, units or nodes (commonly referred to as neurons), whose functionality is loosely based on the animal neuron[11]. ANN models can be used to overcome the non-linearity problem characterizing the soil properties [12]. They also has the advantage over conventional interpolation techniques in defining the functional relationship between the inputs and outputs of a model rather than estimating the output using complex mathematical models [13]. Therefore, recently ANN was introduced as alternative for estimation of soil properties[14]-[18]. Furthermore, when comparing ANN and IDW, ANN gave better RMSE in estimating soil organic matter content [19].

The aim of our work is to compare between recently developed interpolation techniques i.e. ANN, and the conventional interpolation techniques such as OK and IDW in estimating three selected soil chemical properties namely; salinity, phosphorus and nitrogen .

2. MATERIALS AND METHODS

2.1 Study Area and Field Work

The study area is located about 84 km to the north-west of Cairo and covers an area of about 1.5 Km^2. The area is irrigated via drip irrigation system using groundwater as the only source of irrigation water. The study area is considered a relatively new reclaimed area (about twenty years), cultivated with olive trees but with productivity problems that caused changes in land use. The field work was conducted in August 2015. Thirty-two soil samples were collected at approximately 200x200 m grid (Fig. 1) at 0-25 cm depth.

2.2 Laboratory Analysis

The collected soil samples were analyzed for salinity measured as total dissolved solids (TDS) in extract 1:1 (soil: water), using Hanna Instruments HI 2550 Benchtop Meter. Available phosphorus was determined colormetrically using Olsen's sodium bicarbonate extraction method, [20]. Nitrogen was extracted using potassium chloride and determined using Kjeldahl Apparatus according to [21]. All soil properties were expressed in ppm.

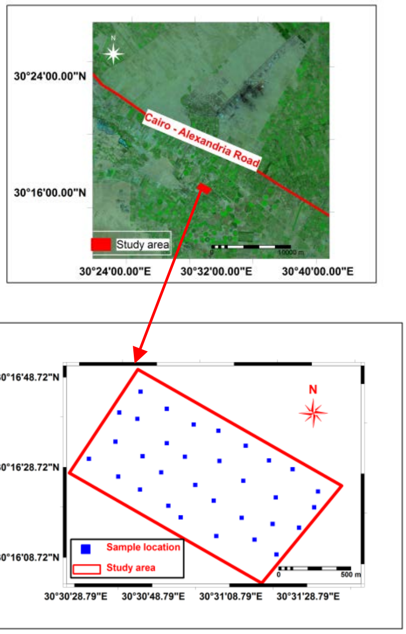

Fig. 1 Study area location and distribution of soil samples

2.3 Data Pre-Processing

Using the ILWIS 3.3 software, a geographic database was designed using the acquired point's locations and the laboratory data and projected into Universal Transverse Mercator (UTM), datum WGS 84 and zone 36N. Furthermore, using Excel software the data were statistically analyzed and the mean, minimum and maximum were computed for each soil property and the correlation between variables was determined.

2.4 Data Processing

Both OK and IDW were performed using ILWIS 3.3 while the ANN was applied using Matlab 12a. The data was directly processed in case of OK and IDW using ILWIS, whereas in case of the ANN the data were imported into Matlab, and processed using the Neural Network toolbox and afterwards exported as tiff into ILWIS software for data visualization and further processing. When designing the neural network, 70% of the samples were used for training, 15% for testing and 15% for validation. The appropriate neural network was selected based on the highest correlation between the actual and predicted soil property expressed as R.

2.4.1 Ordinary Kriging (OK)

Ordinary kriging method was preformed utilizing the semivariogram which measures the strength of the statistical correlation as a function of distance,

[9]. A semivariogram can be calculated as follows [8]:

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

Where $\gamma(h)$ is the semivariance value at distance interval h; $N(h)$ is the number of sample pairs within the distance interval h and $Z(x_i + h)$ and z(xi) are sample values at two points separated by the distance interval h.

When using the semivariogram models, basic parameters including nugget, sill and range must be calculated. The nugget is the variance at zero distance, the range is the distance at which the variables become spatially dependent of one another and the sill is variance at which one point does not influence the neighboring point or at the range distance. In the process of selection of the experimental semivariograms, these parameters were changed until the smallest nugget with the best fitted model was achieved [2].

The semivariograms models that were examined included the Spherical model, Exponential model, Gaussian model, Circular model, the Power model and the best fitted model was selected for each soil property.

2.4.2 Inverse Distance Weighting (IDW)

IDW estimates are made based on nearby known locations. The weights assigned to the interpolating points are the inverse of its distance from the interpolation point. Consequently, the close points are made-up to have more weights than distant points and vice versa [9]. IDW interpolating function is defined as [8]:

$$Z(x) = \sum_{i=1}^{n} W_i Z_i \Big/ \sum_{i=1}^{n} W_i \quad (2)$$

and $W_i = d_i^{-u}$ (3)

where $Z(x)$ is the predicted value at an interpolated point; Z_i is the value at a known point; n is the total number of known points used in interpolation; d_i is the distance between point i and the prediction point; W_i is the weight assigned to point i; and u is the weighting power that decides how the weight decreases as the distance increases.

The main factor affecting the accuracy of the IDW is the value of the power parameter. In this study we compared estimates of IDW using different power parameter from 1 to 4 which are the commonly recommended [7],[8].

2.4.3 Artificial Neural Network

In this study, we used a feedforward perceptron network. In this network the neurons are logically arranged in layers: an input layer, an output layer and one or more hidden layers. The neurons interact with each other via weighted connections and each neuron is connected to all the neurons in the next layer. The input layer is the mean by which data are presented to the network. The output layer holds the response of the network to the input. The hidden layers enable these networks to represent and compute complicated associations between inputs and outputs [14]. The architecture of the three layer feedforward perceptron network used in this study is shown in Fig. 2.

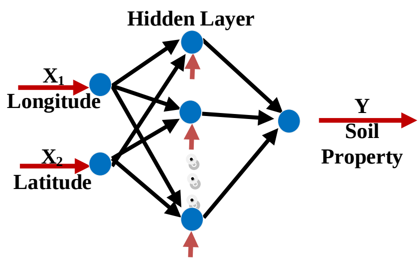

Fig. 2 The neural network architecture (modified after [6])

Currently, there is no analytical way of defining the network structure as a function of the complexity of the problem. The structure must be manually selected using a trial-and-error process [6]. The back-propagation learning algorithm which is the most popular and extensively used neural network algorithm was used. The performance of the ANN model was assessed using the coefficient of determination (R). A well trained model should result in an R value close to 1.

2.5 Evaluation of Interpolation Methods

It is necessary to assess the performance of an interpolation technique to determine if one technique is better than the other. The cross validation is one of the commonly used methods for comparing the interpolation methods [8]. Utilizing this technique, the sample points were arbitrarily divided into two datasets, with one used to train a model and the other used to validate the model [9]. This technique was adopted for evaluating and comparing the performance of the different interpolation methods used in this study. The comparison of performance between interpolation techniques was achieved using the root mean square error (RMSE) and the mean relative error (MRE). MRE is an important measure since RMSE does not provide a relative indication in reference to the actual data [9]. Smaller MRE and RMSE values indicate fewer errors. MRE and RMSE were calculated as follows [8]:

$$MER = \frac{1}{n}\sum_{i-1}^{n}\left|\frac{(Z^*(x_i) - Z(x_i))}{Z(x_i)}\right| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{n}[Z^*(x_i) - Z(x_i)]^2} \quad (5)$$

where $\mathbf{Z}(\mathbf{x_i})$ is the observed value at location i; $\mathbf{Z}^*(\mathbf{x_i})$ is the interpolated value at location i and n is the sample size.

3. RESULTS AND DISCUSSIONS

3.1 Data Visualization and Analyses

Table 1 summarizes the statistical analyses of the studied soil properties, while Table 2 demonstrates the correlation analysis between these properties. Figure 3 shows the spatial distribution of the studied soil properties. The available N ranged from 19.32 - 67.62 ppm while the available P ranged from 2.90 - 8.75 ppm. According to [22] the available P was considered low while available N ranged from low to moderate. This is consistent with the relatively low nutrient condition characterizing the sandy soils alike the soils of the studied area. There was no correlation between the studied soil parameters. The spatial distribution of soil available N and P did not follow any obvious trend. This could be affected by the management practices especially fertilizers applications, taking into account that the study area is divided into small parcels of approximately 200x200 m area with different ownership and consequently different management practices.

Furthermore, data visualization revealed that most of the high soil salinity was clustered in the southern part of the study area. This could be associated with its position on the landscape (flat area with low elevation) (Fig. 4), as the studied area is characterized by similar conditions of climate, parent material, and land use. This area is considered a problematic area where olive orchards were converted to field crops due to low productivity.

Table 1 Statistical summary of the studied soil properties.

Parameter	TDS	N	P
Max	1555.20	67.62	8.75
Min	317.44	19.32	2.90
Mean	609.54	46.94	5.60

Table 2 The correlation analysis between the studied soil properties

Parameter	TDS	N	P
TDS	1.000		
N	0.212	1.000	
P	0.266	0.039	1.000

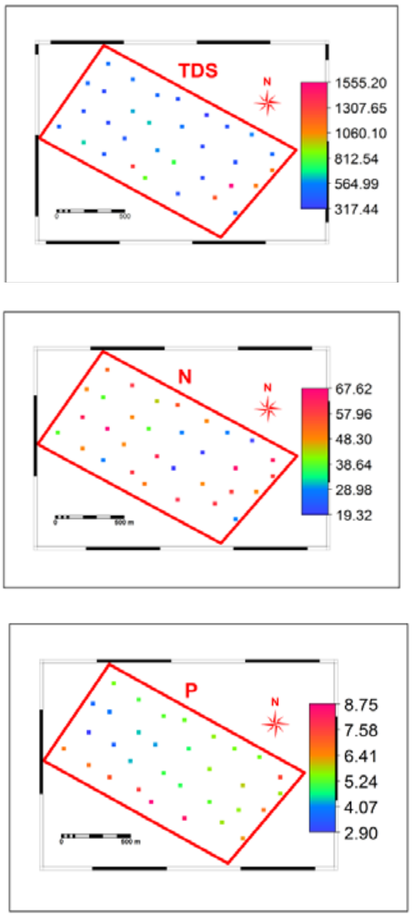

Fig. 3 The spatial distribution of the studied soil properties

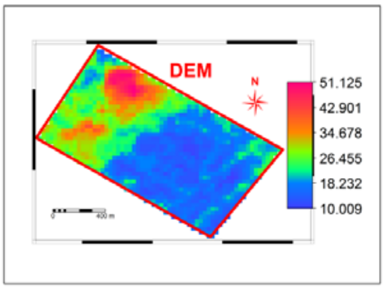

Fig. 4 Digital elevation model of the studied area

3.2 Application of Interpolation Methods

3.2.1 Ordinary Kriging (OK)

The OK was preformed utilizing the geostatistical analysis. Semivariogram for each soil property was calculated and the best model that describe the spatial structure of each property was identified. In this process several models including the Spherical model, Exponential model, Gaussian model, Circular model and the Power model were examined and evaluated in terms of the smallest nugget, largest range and best sill and the best fitted model was selected for each soil property. Spherical, Exponential and Circular models were found to fit well the experimental semivariograms with lag space of 100m for TDS and 75m for N and P. The parameters of the selected semivariogram models are shown in Table 3.

Table 3 Parameters of the semivariogram models for studied soil properties

Variable	Model	Nugget	Sill	Range
TDS	Circular	0.00	75000	180
N	Circular	0.00	180	100
P	Exponential	0.00	1.5	110

As seen in Table 3 the ranges of spatial dependences showed considerable variability among the parameters (from 85 m for P to 180 m for TDS).The range of influence is considered as the distance beyond which observations are not spatially dependent. The range of influence for various soil properties aids in determining where to resample if necessary and in the design of future field experiments to avoid spatial dependency [2]. The difference in ranges of spatial correlation for soil nutrients may be related to the ions mobility in the soil. The interpolation maps for the studied soil properties according to OK are shown in Fig. 5.

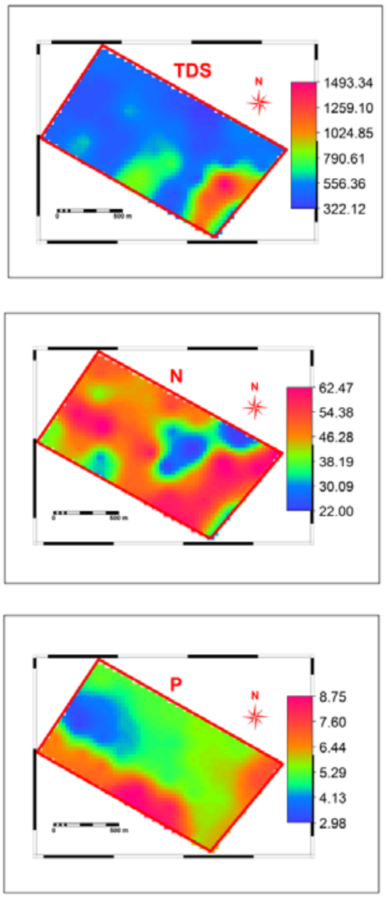

Fig. 5 Interpolation maps of the studied soil properties using OK

3.2.2 Inverse Distance Weighting (IDW)

IDW predictions were performed varying the number of power (from 1-4) and the limiting distances. After examining various distances the limiting distance of 500m gave the best results and was used with all numbers of power. The accuracy of results obtained from the cross-validation procedure for each of the used power value are presented in Table 4.

Table 4 The accuracy of IDW interpolation for the three soil properties using different power values

		Soil properties					
		TDS		N		P	
		RMSE	MRE	RMSE	MRE	RMSE	MRE
Power	1	315.92	21.97	6.95	15.10	0.73	12.62
	2	305.05	21.40	6.62	14.44	0.67	11.28
	3	315.92	21.97	6.76	14.55	0.64	9.84
	4	324.00	24.02	7.00	14.65	0.66	10.14

Both MRE and RMSE were lower for IDW with power of 2 in comparison to that of other powers for all the studied parameters except for P, where the power of 3 gave better results, i.e. lower MRE and RMSE. The interpolated maps of all soil properties using the IDW with the lowest RMSE and MER are shown in Fig. 6.

3.2.3 Artificial Neural Networks (ANN)

Using Matlab a network was designed with the latitude and longitude coordinates as input, one hidden layer and the measured soil parameter as output. The network was trained and the performance of the ANN model expressed as R was evaluated while changing its architect. The results revealed that the optimized network architecture included two input nodes (latitude and longitude coordinates) in the input layer and one node in the output layer (soil parameter). The hidden layer included 20 nodes for TDS, 25 nodes for N, 15 nodes for P. The accuracy of the designed networks for the studied soil parameters expressed as overall R are shown in Fig. 7. The designed ANN showed a very high correlation (R more than 0.9) between the actual and predicted soil salinity and the available phosphorus and the latter had the highest accuracy (R = 0.98). On the other hand, soil available nitrogen attained the lowest correlation between the actual and predicted with R of about 0.72.

After processing the network in Matlab the data were exported into ILWIS for visualization and the predicted soil maps are shown in Fig. 8.

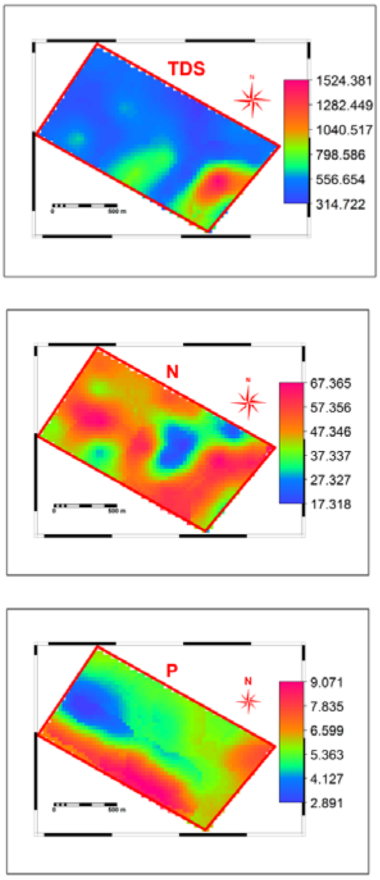

Fig. 6 Interpolation maps of the studied soil properties using IDW

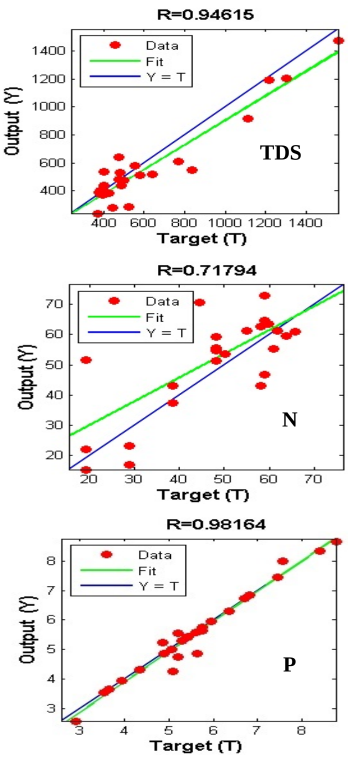

Fig. 7 Accuracy assessment of the designed networks

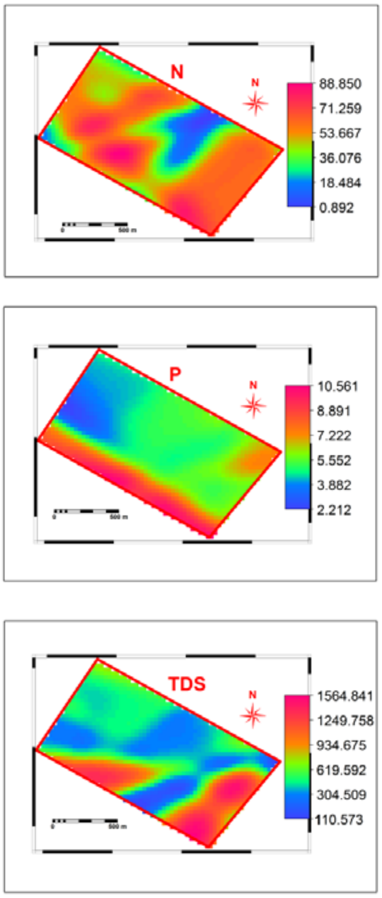

Fig. 8 Interpolation maps of the studied soil properties using ANN

3.3 Evaluation of the Interpolation Methods

The cross validation is applied to evaluate the accuracy of interpolation methods and the result of the cross validation between ANN, OK and the best IDW are shown in Table 5.

Based on these results it was concluded the ANN presented the best accuracy in terms of both RMSE and MRE followed by OK and lastly IDW. But it should be mentioned that this improved accuracy was only achieved when R was more than 0.8 between the actual and predicted values when designing the ANN.

Table 5 Accuracy of the different soil interpolation method

		Interpolation method					
		OK		IDW		ANN	
		RMSE	MRE	RMSE	MRE	RMSE	MRE
Variable	TDS	152.11	13.22	305.05	21.40	124.32	12.03
	N	6.79	13.06	6.82	14.44	6.31	12.41
	P	0.62	7.87	0.64	9.84	0.49	6.93

4. CONCLUSION

The spatial variability of soil properties is the most important component in designing a reliable agriculture management strategy that aim at optimizing crop production and minimizing soil fertility losses and consequently protecting the environment. In this aspect, the best interpolation method for selected soil chemical properties namely; TDS, available P and N which are of main concern in agriculture management was evaluated. The evaluated methods included OK and IDW which are two commonly used techniques for interpolating of soil properties and ANN which is somewhat a recent approach. While IDW is a simpler method to use, our study as well as most previous studies revealed that its accuracy is lower than OK. Contrary, OK which has a higher accuracy has the complications of designing the best fit semivariograms. On the other hand, ANN is considered a promising technique that can offer an alternative method for interpolation of soil properties. In addition to its higher accuracy compared to both OK and IDW, ANN has the advantage of designing simplicity compared to OK and the ability to evaluate the accuracy of the method while designing the network. Nevertheless, geostatistical analysis are still the best way in defining the range of spatial dependence which was found to vary within soil parameters and considered a key factor in designing the sampling strategy.

5. REFERENCES

[1] Sun B, Zhou S and Zhao Q, "Evaluation of Spatial and Temporal Changes of Soil Quality Based on Geostatistical Analysis in the Hill Region of Subtropical China", Geoderma, Vol. 115, 2003, pp. 85- 99.

[2] Yasrebi J, Saffari M, Fathi H, Karimian N, Moazallahi M and Gazni R, "Evaluation and Comparison of Ordinary Kriging and Inverse Distance Weighting Methods for Prediction of Spatial Variability of Some Soil Chemical Parameters", Research Journal of Biological Sciences, Vol. 4, 2009, pp. 93-102.

[3] Bhatti A, "Geostatistical Techniques and Applications for Managing Degraded Soil for Sustainable Production", Quarterly Science Vision, Vol. 8, No. 1, July - September 2002, pp. 27-35.

[4] Kuzyakova, IF, Romanenkov, VA and Kuzyakov YV, "Application of Geostatistics in Processing the Results of Soil and Agrochemical Studies", Eurasian Soil Science, Vol. 34, No. 11, 2001, pp. 1219-1228.

[5] Krasilnikov P, Carré F and Montanarella L, "Soil Geography and Geostatistics Concepts and Applications", Luxembourg: Office for Official Publications of the European Communities, 2008, Introduction.

[6] Gummadi JA, "Comparison of Various Interpolation Techniques for Modeling and Estimation of Radon Concentrations in Ohio", Master of Science Degree in Engineering, The University of Toledo, December 2013.

[7] Robinson TP and Metternicht G, "Testing the Performance of Spatial Interpolation Techniques for Mapping Soil Properties", Computers and Electronics in Agriculture, Vol. 50, 2006, pp. 97-108.

[8] Xie Y, Chen T, Lei M, Yang J, Guo Q, Song B and Zhou X, "Spatial Distribution of Soil Heavy Metal Pollution Estimated by Different Interpolation Methods: Accuracy and Uncertainty Analysis", Chemosphere, Vol. 82, 2011, pp. 468-476.

[9] Bhunia GS, Shit PK and Maiti R, "Comparison of GIS-based Interpolation Methods for Spatial Distribution of Soil Organic Carbon (SOC)", Journal of the Saudi Society of Agricultural Sciences, in press, 2016.

[10] Liu S, Zhang Y, Ma P, Lu B and Su HA, "Novel Spatial Interpolation Method Based on the Integrated RBF Neural Network", Procedia Environmental Sciences, Vol. 10, 2011, pp, 568 – 575.

[11] Gurney K, "An introduction to neural networks", London: UCL Press Limited, 1997, ch. 1.

[12] Zhao Z, Chow TL, Rees HW, Yang Q, Xing Z and Meng F, "Predict Soil Texture Distributions Using an Artificial Neural Network Model", Computers and Electronics in Agriculture, Vol. 65, No. 1, 2009, pp. 36-48.

[13] Kangas LJ, Keller PE, Hashem S, Kouzes RT and Allen PA, "Adaptive Life Simulator: A Novel Approach to Modeling the Cardiovascular System", in Proc. of the 1995 American Control Conference, 1995, pp. 796-800.

[14] Erzin Y, Rao BH, Patel A, Gumaste SD and Singh DN, "Artificial Neural Network Models for Predicting Electrical Resistivity of Soils from Their Thermal Resistivity", International Journal of Thermal Sciences, Vol. 49, 2010, pp. 118-130.

[15] FalamakiA, "Artificial Neural Network Application for Predicting Soil distribution Coefficient of Nickel", Journal of Environmental Radioactivity, Vol. 115, 2013, pp. 6-12.

[16] Aitkenhead MJ, Donnelly D, Sutherland L, Miller DG, Coull MC and Black HIJ, "Predicting Scottish Topsoil Organic Matter Content from Colour and Environmental Factors", European Journal Of Soil Science, Vol. 66, 2015, pp. 112-120.

[17] Keshavarzi A, Fereydoon S, El-Sayed EO and Munawar I,"A Neural Network Model for Estimating Soil Phosphorus Using Terrain Analysis". The Egyptian Journal of Remote Sensing and Space Sciences, Vol. 18, 2015, pp. 127-135.
[18] Aitkenhead MJ and Coull MC, "Mapping Soil Carbon Stocks Across Scotland Using a Neural Network Model", Geoderma, Vol. 262, 2016, pp. 187-198.
[19] Dai F, Zhou Q, Lv Z, Wang X and Liu G, "Spatial Prediction of Soil Organic Matter Content Integrating Artificial Neural Network and Ordinary Kriging in Tibetan Plateau", Ecological Indicators, Vol. 45, 2014, pp. 184-194.
[20] Watanabe FS and Olsen SR, "Test of an Ascorbic Acid Method for Determining Phosphorus in Water and NaHCO3 Extracts from Soil", Soil Science Society of America, Proceedings, Vol. 29, 1965, pp. 677-678.
[21] Cottenie A, Verloo M, Kiekens L, Velghe G and Camerlynck R,"Chemical Analysis of Plants and Soils", Belgium: Laboratory of Analytical and Agrochemistry, State University Ghent, 1982.
[22] Baruah TC and Barthakur HP,"A Textbook of Soil analysis", New Delhi: Viskas Publishing House PVT Ltd., 1997.