

Comparison of Artificial Neural Networks, Geographically Weighted Regression and Cokriging Methods for Predicting the Spatial Distribution of Soil Macronutrients (N, P, and K)

Samad EMAMGHOLIZADEH, Shahin SHAHSAVANI, Mohamad Amin ESLAMI

(Department of Water and Soil Engineering, Shahrood University of Technology, Shahrood 361999-5161, Iran)

Abstract: Soil macronutrients (i.e. nitrogen (N), phosphorus (P), and potassium (K)) are important soils components and knowing the spatial distribution of these parameters are necessary at precision agriculture. The purpose of this study was to evaluate the feasibility of different methods such as artificial neural networks (ANN) and two geostatistical methods (geographically weighted regression (GWR) and cokriging (CK)) to estimate N, P and K contents. For this purpose, soil samples were taken from topsoil (0–30 cm) at 106 points and analyzed for their chemical and physical parameters. These data were divided into calibration ($n = 84$) and validation ($n = 22$). Chemical and physical variables including clay, pH and organic carbon (OC) were used as auxiliary soil variables to estimate the N, P and K contents. Results showed that the ANN model (with coefficient of determination $R^2 = 0.922$ and root mean square error $RMSE = 0.0079\%$) was more accurate compared to the CK model (with $R^2 = 0.612$ and $RMSE = 0.0094\%$), and the GWR model (with $R^2 = 0.872$ and $RMSE = 0.0089\%$) to estimate the N variable. The ANN model estimated the P with the RMSE of 3.630 ppm, which was respectively 28.93% and 20.00% less than the RMSE of 4.680 ppm and 4.357 ppm from the CK and GWR models. The estimated K by CK, GWR and ANN models have the RMSE of 76.794 ppm, 75.790 ppm and 52.484 ppm. Results indicated that the performance of the CK model for estimation of macro nutrients (N, P and K) was slightly lower than the GWR model. Also, the accuracy of the ANN model was higher than CK and GWR models, which proved to be more effective and reliable methods for estimating macro nutrients.

Keywords: precision agriculture; soil characteristics; interpolation; artificial neural networks; geographically weighted regression; Cokriging

Citation: Emamgholizadeh Samad, Shahsavani Shahin, Eslami Mohamad Amin, 2017. Comparison of artificial neural networks, geographically weighted regression and Cokriging methods for predicting the spatial distribution of soil macronutrients (N, P, and K). *Chinese Geographical Science*, 27(5): 747–759. doi: 10.1007/s11769-017-0906-6

1 Introduction

Soil, water and land are vital resources for sustaining good quality of human life and good quality of agricultural production (Das *et al.*, 2009). In recent years with an increase of world population, the need for the betterment of food quality is getting more attention. Sustainable soil and water management is a major challenge for researchers, planner, managers and farmers to insure good quality of food, water and environment. The

balance application of chemical fertilizers becomes very important for efficient agriculture production. For these reasons, the use of modern technologies for solving problems in the agriculture section is essential, especially for precision agriculture. One of the important parameter in precision agriculture is the provision of agriculture land maps, which can predict the nutrient needs for the crop production. The best maps from economic and social points of view are soil fertility maps in agricultural lands due to these reasons: 1) precision fertil-

Received date: 2016-11-16; accepted date: 2017-01-25

Foundation item: Under the auspices of Shahrood University of Technology, Iran (No.348517)

Corresponding author: Samad EMAMGHOLIZADEH. E-mail: s_gholizadeh517@shahroodut.ac.ir

© Science Press, Northeast Institute of Geography and Agroecology, CAS and Springer-Verlag Berlin Heidelberg 2017

izers recommendation on the soil test base; 2) balance plant nutrition; 3) soil organic-C nutrition content and 4) environmental maintenance. The macro nutrients, including nitrogen (N), phosphorus (P) and potassium (K) are necessary for plant growing and they are indicators for soil fertility and also they have a close relationship with land productivity (Reeves, 1997; Malvi, 2011).

Different approaches such as Inverse Distance Weighting (IDW), Radial Basis Functions (RBF) and local/global polynomial interpolation and also geostatistical methods such as Kriging (ordinary, simple, universal) and Cokriging (CK) have been used for spatial interpolation of the soil properties based on soil sampling points (Odeh *et al.*, 1995; McKenzie and Ryan, 1999; Robinson and Metternicht, 2006; Eldeiry and Garcia, 2010; Sun *et al.*, 2012). To achieve interpolation with adequate accuracy by using interpolation method such as Kriging, dense sampling should be used which increases the cost of study (Su *et al.*, 2009). In this study, we tried to predict these parameters with different methods such as ANN, CK and GWR. Recently, CK and GWR approaches got the attention of the soil engineers to prepare maps of soil properties such as soil total nitrogen (TN), soil organic carbon and other parameters (Mishra *et al.*, 2010; Kumar *et al.*, 2012; Rincón-Ruiz *et al.*, 2013). However, the applications of these models to estimate the variation of soil characteristics (N, P and K) are rarely used by some methods such as the ANN model. In recent years, several studies such as Ghorbani *et al.* (2015); Kashi *et al.* (2014) and Emamgholizadeh *et al.* (2016) have reported successful applications of the ANN model for estimating the soil parameters. Also, Li *et al.* (2011) used GIS and back-propagation neural network (BPNN) models to estimate the spatial distribution of soil heavy metals in Huizhou City, and showed that the accuracy of the BPNN model was better than the GIS model for estimating the soil heavy metal content. Liu *et al.* (2013) examined the effectiveness of combining radial basis function (RBF) ANN models with a soil profile depth function to map the three-dimensional distribution of SOM in a subtropical hilly landscape in southern Anhui Province of China.

Based on the literature review studies and the gaps emerged, the objectives of this study are to: 1) develop the ANN, GWR and CK models for soil mapping of nitrogen (N), phosphorus (P) and potassium (K), with limited field data; 2) compare the performance of the

three models; and 3) determine the reasons of variability of soil macronutrients.

2 Materials and Methods

2.1 Study area and data

This study takes advantage of field database provided by Eslami *et al.* (2016). The study area is located near the Aq-Qalla district in the Golestan Province, Iran. The mapping area extends from 54° 28'E to 54° 36'E and 37° 04'N to 37° 07'N (Fig. 1). This study area covers about 3328 ha. A number of 106 composite soil samples were collected from the topsoil (0–30 cm). The locations of soil samples are plotted in Fig. 1. About, 1 km of these samples was prepared and sent to the laboratory for determining physico-chemical properties of the soil. The sampling location was determined with a receiver of the Global Position System (GPS). In the soil samples, the pH, texture (sand, silt, and clay percent) and electric conductivity (EC) were measured with the pH meter, hydrometer technique (Klute and Dirksen, 1986), and with the conductivity meter, respectively. Also, other parameters such as the organic carbon (OC) was measured with the Walkley and Black (1934) method, and the phosphorous, and the potassium were measured with Olsen (1954) and with ammonium acetate, respectively.

Statistical indices (minimum, maximum, mean and std. deviation, SD) of the measured variables (i.e., EC, pH, OC, total N, available P and K, percent of sand, silt and clay) are shown in Table 1. Also for the investigation of the normality of data, the skewness and kurtosis of data were calculated. As indicated in Table 1, the observed soil N, P and K contents in soil samples vary from 0.027% to 0.147%, 1.50 ppm to 31.20 ppm, 106.00 ppm to 560.00 ppm with a mean value of 0.090%, 7.10 ppm and 289.57 ppm, respectively (Table 1 and Fig. 2). Moreover, the pH value of soil samples varies from 7.3 to 8.2, which indicating that all soil samples fall in the alkalinity range. The measured clay and OC values of soil samples show that they vary from 4% to 46% and 0.41% to 1.47 %, respectively. EC of soil samples vary from 1.1 to 31.6. Nearly more than 50% of the study area has EC greater than 8 and therefore its salinity is high. As the soil of the study area contained a low organic matter, most of the pertinent area was concluded to have a low amount of N for that matter.

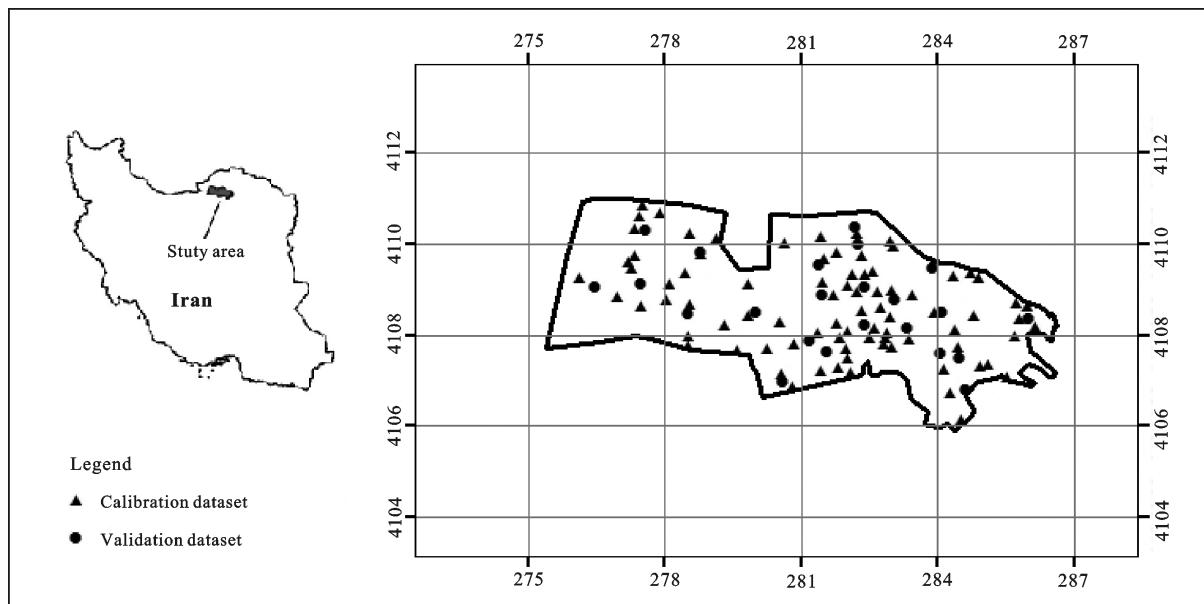


Fig. 1 Study area and locations of the sampling data points. Grid in the figure is kilometer grid (km)

Table 1 Descriptive statistics of soil samples in the study area

	Sand (%)	Silt (%)	Clay (%)	EC mmhos/cm	pH	OC (%)	N (%)	P (ppm)	K (ppm)
Minimum	6.00	44.00	4.00	1.10	7.30	0.41	0.03	1.50	108.02
Maximum	24.00	84.00	46.00	31.60	8.20	1.47	0.15	31.71	560.05
Mean	11.28	60.92	27.84	9.43	7.73	0.94	0.09	7.10	289.57
Median	10.00	60.00	28.00	7.55	7.70	0.94	0.09	6.21	280.02
SD	3.24	7.44	8.90	6.82	0.17	0.24	0.03	4.71	88.81
Skewness	1.08	0.66	-0.46	1.45	0.18	-0.21	-0.19	2.07	0.69
Kurtosis	1.77	1.34	0.09	1.78	-0.021	-0.513	-0.481	6.53	0.13
Kolmogorov-Smirnov Z Asymp. Sig. (2-tailed)	0.000	0.204	0.420	0.002	0.056	0.545	0.228	0.046	0.263

Notes: EC, electric conductivity; OC, organic carbon

2.2 Correlation Analysis

In modeling, choosing the proper input variables (auxiliary or co-variable) is an important step because it affects the accuracy of modeling. For this purpose, six soil variables were preliminarily nominated as candidates of auxiliary variable for the ANN, GWR and CK models to estimate N, P and K. These auxiliary variables include some percents of sand, silt, clay, EC, pH and OC properties. Therefore, a correlation analysis was done to detect the correlation between the N, P and K and the candidates of the auxiliary variable (i.e., clay, sand, silt, pH, EC and OC) (Table 2). Results showed that there is high correlation between N, P and K with OC with correlation (R) of 0.944, 0.306 and 0.534,

respectively. This finding is consistent with the studies of Saikh *et al.* (1998) and Nath (2014). Also, there is a significant correlation between N-clay and N-pH ($R = 0.376$ and -0.324), which is consistent with the findings of Aggelopoulou and Gemtos (2011) and Nath (2014). The correlation between K and clay was 0.357.

In general, nitrogen (N) is correlated with clay, pH, and OC. Also, P and K are correlated with OC, clay, respectively. Therefore, these variables are used as auxiliary variables for the ANN, GWR and CK models. A number of 106 field data sets were divided randomly into two parts: calibration (84 data points) and validation (22 data points). The dividing data in the two sets of calibration and validation were done based on the

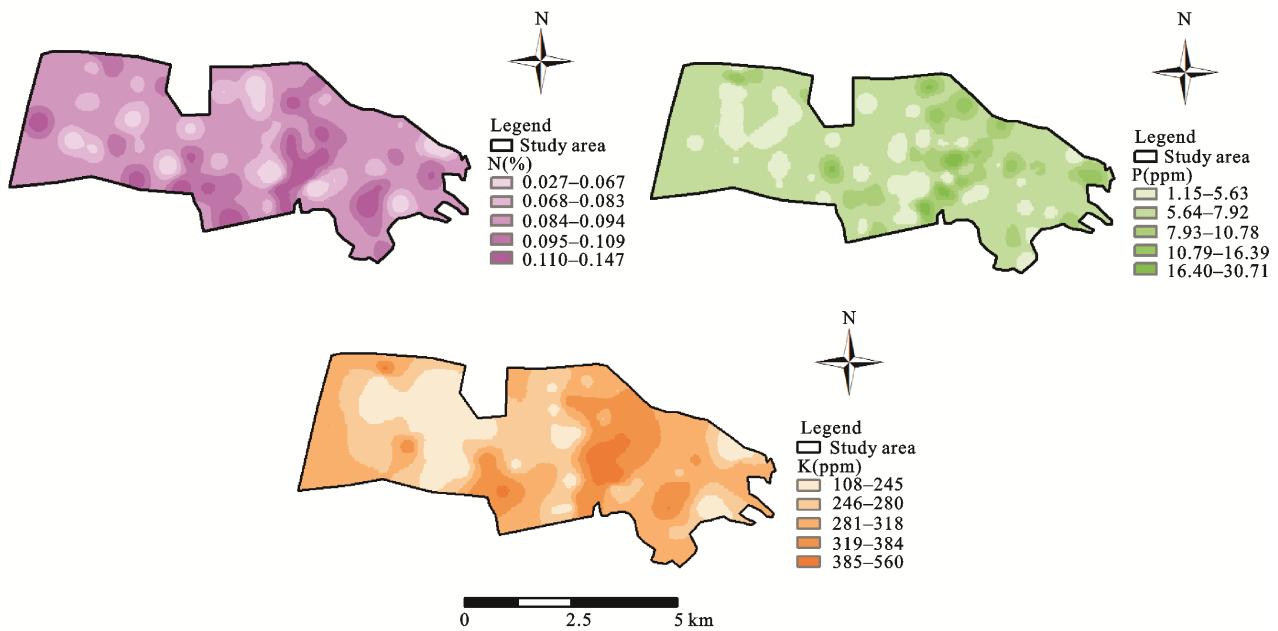


Fig. 2 Spatial distribution maps of N, P and K in study area

Table 2 Pearson correlation between N, P and K and related physical and chemical variables

	Sand	Silt	Clay	EC	pH	OC	N	P	K
Sand	1	0.256*	-0.577**	0.250*	-0.211	0.028	0.057	0.031	-0.101
Silt		1	-0.936**	0.213	-0.001	-0.373**	-0.301**	-0.115	-0.352**
Clay			1	-0.251*	0.073	0.309**	0.376**	0.092	0.357**
EC				1	-0.455**	-0.227*	-0.176	0.006	0.102
pH					1	-0.241*	-0.324**	-0.153	-0.141
OC						1	0.944**	0.306**	0.534**
N							1	0.315**	0.571**
P								1	0.617**
K									1

Notes: ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed)

simple random sampling approach. Figure 3 shows the box plot of data. It graphically shows the distribution of data. Also as the skewness and kurtosis values of data (Table 1) show all soil variables have normal distribution except phosphorus (P), because the acceptable range of skewness is -2 to 2 and kurtosis is -3 to +3 (George, 2011). Also the Kolmogorov-Smirnov test ($P = 0.05$) rejected the null hypothesis of normality for all samples except for phosphorus (P), because the *Sig.* value of the Kolmogorov-Smirnov test for all variable (except phosphorus (P)) is greater than 0.05, so the data have normal distribution. So, before using phosphorus

(P) data, three transformation functions of Box-Cox, logarithmic (log) and square-root were used. The Box-Cox transformation (Box and Cox, 1964) is defined as:

$$y^\lambda = \begin{cases} \frac{y^\lambda - 1}{\lambda} & (\lambda \neq 0), \\ \log y & (\lambda = 0) \end{cases} \quad (1)$$

where y is the response variable (phosphorus) and λ is the transformation parameter. Transforming the data to phosphorus (P) data indicated that the log transformation function for data gives better results; hence, we used this transformation function to make data approximately

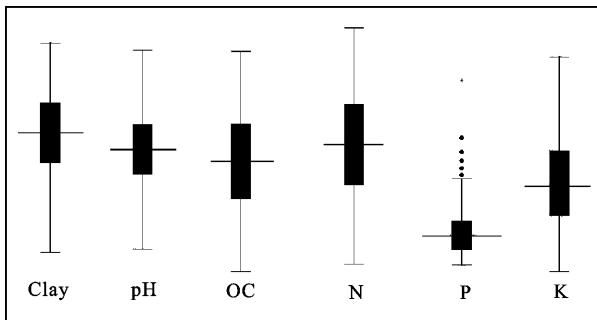


Fig. 3 The box plot of data including clay, pH, OC (organic carbon), N, P and K

follow a normal distribution. Using log transformation function, the Sig. value of the K-S test achieved 0.991 which is more than 0.05, so the distribution of phosphorus (P) changed to normal. Fig. 4 shows the histogram of data values for variable P in two forms before and after using log transformation function. As Fig. 4 shows using transformation function changed the distribution data to a symmetric and normal form.

2.3 Interpolation methods

In the soil studies and also other fields, when sufficient data of the soil parameters such as soil organic carbon, sulfur, N, P and K, and also other spatially-based phenomena are not available in the study area, the interpolation technique can be used to estimate the requested parameters at the unknown location. In the interpolation

methods, the values of the desired variable at un-sampled points were estimated based on known values of surrounding points. For this purpose, several well-known interpolation methods such as Kriging, inverse distance weighting (IDW), Spline, Cokriging (CK), and geographically weighted regression (GWR) can be used.

2.3.1 Cokriging (CK)

CK is the extension of the ordinary Kriging (OK) method. This method is used for the estimation of a poorly sampled variable (target parameter) by using a well-sampled variable (auxiliary or co-variable) to improve the accuracy of the Kriging method (Wackernagel, 2003). Using the co-variables may help to improve estimations (predictions) of the target variable, especially if the desired variable is not well sampled. Usually variable(s) is selected as the candidate auxiliary variable(s) when it well correlates to the target variable and also it is cheaper to measure (Goovaerts, 1998; Wang *et al.*, 2013). In the CK model, it is possible to use auxiliary data to improve estimation. Two methods of Kriging and Cokriging use the semivariogram and cross-semivariogram functions to quantify and describe the spatial patterns of the dependent variable (Western *et al.*, 2004). It should, however, be mentioned that in geostatistical analyses the semivariogram is used to quantify and model the spatial variability degree of the sample data.

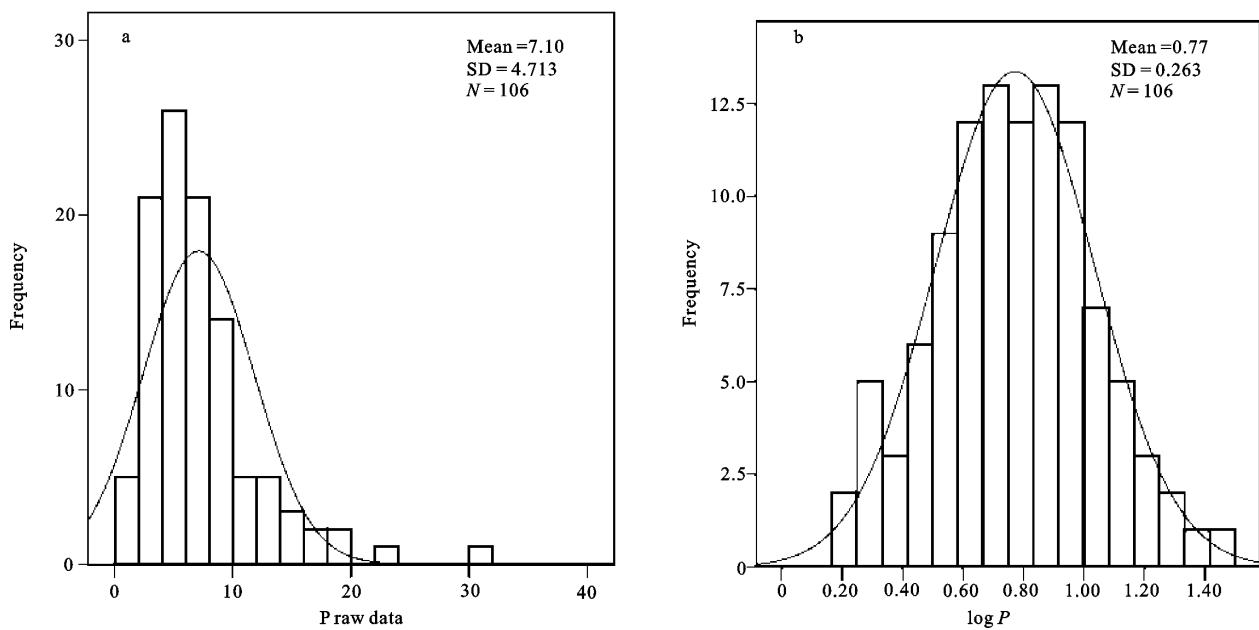


Fig. 4 Histogram of data values for variable P (a) raw data (b) transformed data using log transformed function

For observations $Z_i, i=1, \dots, k$ at locations x_1, \dots, x_k the empirical semivariogram $\gamma(h)$ is defined as (Cressie, 1993; Malinova and Guo, 2004) :

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

where $Z(x_i)$ and $Z(x_i + h)$ are two measured values at locations x_i and $x_i + h$, respectively; $N(h)$ is the number pairs of data locations separated by a distance h .

Cross-variogram $\gamma_{ij}(h)$ is used to determine the spatial interdependence between any two variables calculated by the following equation:

$$\gamma_{ij}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z_i(x_i) - Z_i(x_i + h)][Z_j(x_i) - Z_j(x_i + h)] \quad (3)$$

where $N(h)$ is the number of data pairs of $Z_i(x_i)$ and $Z_j(x_i)$ at a separate distance h .

2.3.2 Geographically weighted regression (GWR)

GWR is an extension of the traditional multiple linear regression (MLR) and it used to model spatially varying relationships between the dependent and explanatory variables. The difference between two methods of MLR and GWR is that, in the model of MLR its parameter is estimated with the assumption of a spatial stationary, so the MLR is known as the global model. Also, in this method a single regression model is fitted to the entire sample data at the study area.

But the spatial data sets are not stationary and they are spatial non-stationary over space, so the assumption of stationary over space will probably be unrealistic (Fotheringham *et al.*, 2003; Murayama, 2012; Wang *et al.*, 2013). Thus, when the spatial data are analyzed, the spatial non-stationary should be taken into account. In this regard, the geographically GWR model overcomes this problem, because it considers the spatial locations of data points and lets the parameter estimates be a function of locations (Zhang *et al.*, 2011). With the GWR the interested soil property was estimated by:

$$y(u) = \beta_0(u) + \sum_{k=1}^p \beta_k(u) x_k(u) + \varepsilon(u) \quad (4)$$

where $y(u)$ and $x_k(u)$ are the value of the dependent and independent variables; y and x_k are at location u ; $\beta_0(u)$ is the intercept; $\beta_k(u)$ is the correlation coefficient for the independent predictor variable x_k and $\varepsilon(u)$ represents the error term. In this method, the regression coefficients of the model for each location are estimated using data within a neighborhood, so the GWR model can measure the spatial variations in the relationships.

2.3.3 Multi-Layer Perception (MLP) Neural Network

The Artificial Neural Network (ANN) is a nonlinear model which comprises many simple computational units and it was first introduced in 1943 by (McCulloch and Pitts, 1943). The ANN tends to mimic the nervous system of a human brain. The advantage of the ANN intelligence model is that it has the ability to recognize the implicit relation which exists in data and employs it to solve problems (Malinova and Guo, 2004). Various types of ANNs (e.g., radial basis function network (RBF), backward propagation (BP), Hopfield networks, multilayer perceptron (MLP) etc.) can be used. The main differences between these methods are related to the network architecture and also the method which they use to determine the weights and functions for inputs and output (Snyder and Redmond, 1995). The MLP network is most commonly used in engineering problems (Azmathullah *et al.*, 2005; Emamgholizadeh *et al.*, 2013; Emamgholizadeh *et al.*, 2015). So in this study, the MLP with a back propagation algorithm was used. Figure 5 shows the architecture of a MLP network, which includes an input layer, a hidden layer and an output layer.

2.4 Models evaluation

The estimated N, P and K with ANN, GWR and CK models were compared with the measured values based

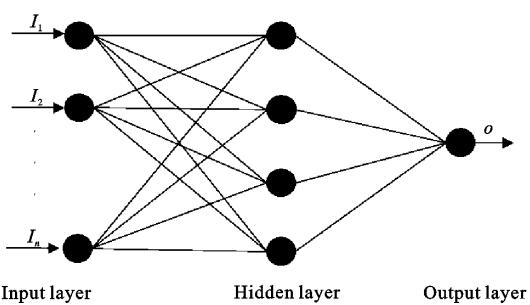


Fig. 5 Structure of Multi-layer perception Neural Network model

on three statistical criteria of coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) as described in Equs. 5, 6 and 7, respectively. They are as follows:

$$R^2 = \frac{\sum_{i=1}^N (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2 \sum_{i=1}^N (P_i - \bar{P})^2}} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{N}} \quad (6)$$

$$MAE = \frac{1}{N} \sum_{i=1}^n |O_i - P_i| \quad (7)$$

where N is the number of data, O_i and P_i are the observed and estimated values, \bar{P} and \bar{O} are the mean values of P_i and O_i , respectively.

3 Results and discussion

3.1 Estimation of macro nutrients (N, P and K) by the GWR and CK models

As mentioned, the GWR and CK are interpolation methods which are used to incorporate the auxiliary variables or information for the interpolation. To examine the spatial correlation and the spatial structure of variables, an empirical semivariogram was fitted on data after normalization. Semivariogram provides the spatial correlation in observations measured at sample locations and it indicates how measurement at one point is correlated to measurements at neighbor points. Different types of semivariogram functions (e.g., Circular, Spherical, Exponential, Gaussian, Stable, etc.), can be used to model the empirical semivariogram. For example (Wang *et al.*, 2013; Ivajnšič *et al.*, 2014) used the Gaussian function in their studies. So, this function was used to model the empirical semivariogram. Fig. 6 displays the semivariogram of the variables of clay, pH, OC, N, P and K. Also, the cross covariogram between macronutrients (N, P and K) and auxiliary variables (clay, pH and OC) are showed in Fig. 7.

The calculation of the key parameters related to the semivariogram and cross-covariogram were given in Table 3. In this table, $C_0/(C_0+C)$ is the ratio of nugget to sill. It shows the spatial dependency of the soil proper-

ties. Ratios from 0–0.25, 0.25– 0.75 and greater than 0.75 show strong, moderate and weak spatial dependence of the variables, respectively (Cambardella *et al.*, 1994). The ratios for N, P, clay, pH and OC are less than 25%, and it shows a strong spatial dependence of variables, respectively. The ratio for K indicates that it has moderate spatial dependencies. The range of semivariogram for N, P, clay, pH and OC were similar and about 503.275–613.400 m (Table 3), and was much greater than that for K with 1201.290 m. In this study, the sampling densities for the six studied variables were 500 m. A comparison of the calculated range (> 500 m) of the variables (N, P, K, clay, pH and OC) and sampling interval indicates a reasonable sampling interval.

Similar to the CK method, the GWR model was applied to model spatially varying relationships between N, P, and K and variables of clay, pH and OC. The Gaussian model with an Adaptive kernel type was chosen for the GWR model. Also the number of neighbors was selected 5 for keeping consistent with that used in the CK method. The calculated values of MAE, RMSE, and R^2 for the estimation of N, P and K from two approaches (CK and GWR) are listed in Tables 4, 5 and 6, respectively.

As the results of these tables show, the CK model makes low accurate estimations of N, P and K, with the RMSE (R^2) of 0.0094% (0.612), 4.68 ppm (0.210) and 76.794 ppm (0.270), in the validation stage. Also the GWR estimated N, P and K, with the RMSE (R^2) of 0.0089% (0.872), 4.357 ppm (0.250) and 75.79 ppm (0.281), respectively. These indicate that the estimation of N, P and K from the GWR model was slightly better than those from the CK model. Compared to the CK model, the GWR model had a lower RMSE by 5.32%, 6.90% and 1.31% for N, P and K, respectively.

3.2 Estimation of macro nutrients (N, P and K) by the MLP-ANN model

As referred in Section 2.3.3, the ANN/MLP model (with a back propagation algorithm) was applied to estimate the N, P and K. For the MLP model, the number of layers and also the number of nodes determines the structure and accuracy of the model. In this study, similar to CK and GWR models, for the estimation of N, five nodes (i.e., X, Y, clay, pH and OC), P, three nodes (i.e., X, Y and OC) and K, four nodes (i.e., X, Y, clay and OC) were used. X and Y indicate the spatial coordinates of

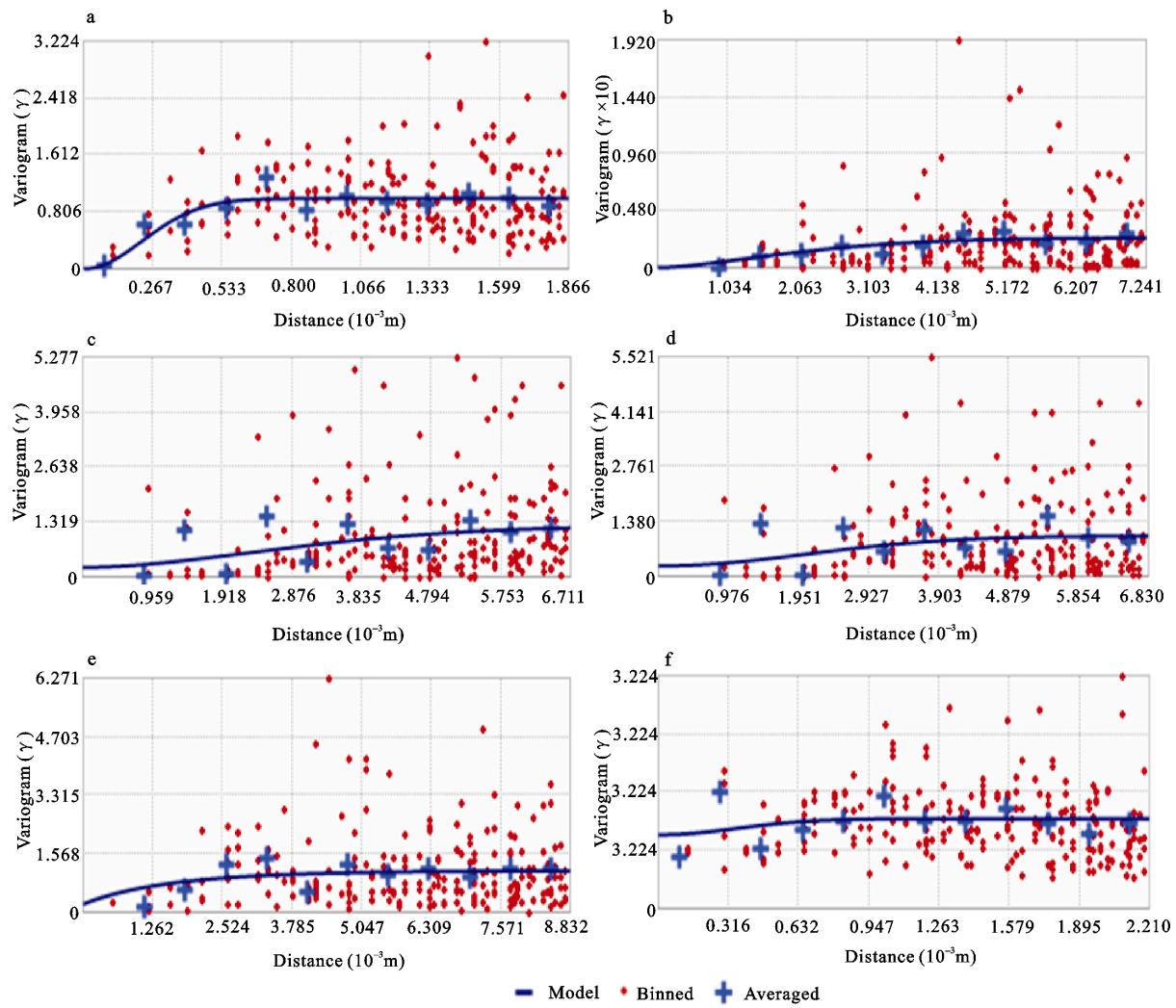


Fig. 6 Auto semivariogram ($\gamma(h)$) of clay (a), pH (b), OC (c), N (d), P (e), and K (f)

Table 3 Parameters of semivariogram models for N, P, K, Clay, pH, OC of the studying area and their cross covariogram models

Variables	C_0	C	Total sill ($C_0 + C$)	Range (m)	$C_0/(C_0 + C)$ (%)	Spatial dependence
N	0.245	0.778	1.023	503.275	23.940	S
P	0.219	0.963	1.182	527.830	18.520	S
K	0.426	0.590	1.016	1201.290	41.930	M
Clay	0.001	0.994	0.995	549.082	0.100	S
pH	0.143	0.743	0.886	503.275	16.090	S
OC	0.251	0.944	1.194	613.400	20.980	S
Cov (N, Clay)	–	0.278	–	576.670	–	–
Cov (N, pH)	–	-0.330	–	503.275	–	–
Cov (N, OC)	–	0.907	–	521.710	–	–
Cov (P, OC)	–	0.421	–	620.570	–	–
Cov (K, Clay)	–	0.474	–	577.450	–	–
Cov (K, OC)	–	0.604	–	527.840	–	–

Notes: Cov is covariogram; M is moderate; S is strong, C_0 is nugget, C is partial sill

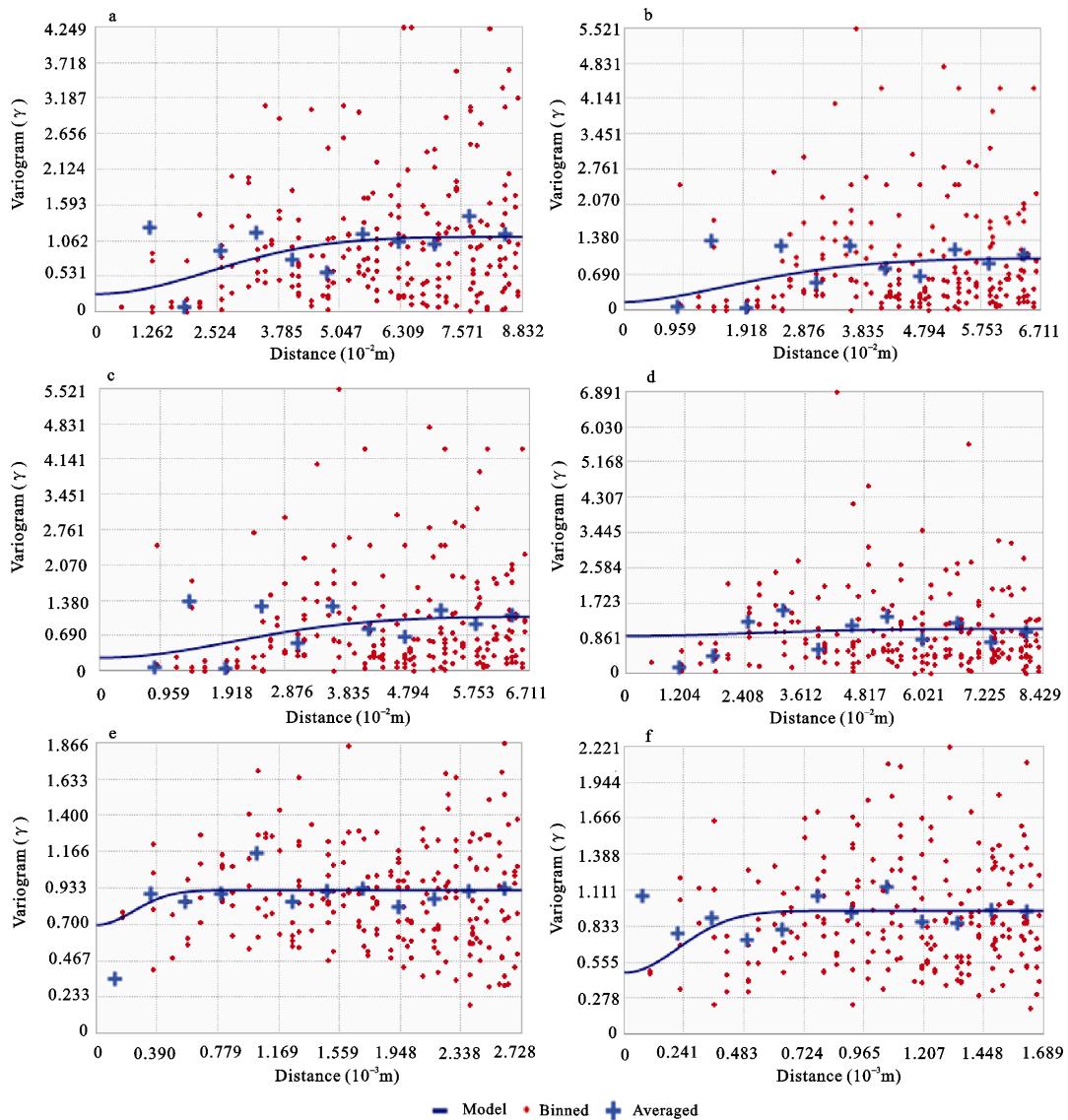


Fig.7 Cross covariogram between N and clay (a), N and pH (b), N and OC(c), P and OC (d), K and clay (e), K and OC (f)

Table 4 Comparing the performance of CK, GWR and ANN models in estimating N using clay, pH and OC as auxiliary data in calibration and validation stages

Model	Calibration stage			Validation stage		
	MAE (%)	RMSE (%)	R ²	MAE (%)	RMSE (%)	R ²
CK	0.0055	0.0091	0.891	0.0084	0.0094	0.612
GWR	0.0053	0.0089	0.892	0.0074	0.0089	0.872
ANN	0.0041	0.0070	0.934	0.0068	0.0079	0.922

Notes: CK, Cokriging; GWR, geographically weighted regression; ANN, artificial neural networks

Table 5 Comparing the performance of CK, GWR and ANN models in estimating P using OC as auxiliary data in calibration and validation stages

Model	Calibration stage			Validation stage		
	MAE (ppm)	RMSE (ppm)	R ²	MAE (ppm)	RMSE (ppm)	R ²
CK	3.677	4.556	0.228	3.653	4.680	0.210
GWR	3.095	4.164	0.267	3.149	4.357	0.250
ANN	1.476	2.114	0.738	2.799	3.630	0.541

Notes: CK, Cokriging; GWR, geographically weighted regression; ANN, artificial neural networks

Table 6 Comparing the performance of CK, GWR and ANN models in estimating K using clay and OC as auxiliary data in calibration and validation stages

Model	Calibration stage			Validation stage		
	MAE (ppm)	RMSE (ppm)	R ²	MAE (ppm)	RMSE (ppm)	R ²
CK	54.659	71.599	0.384	63.769	76.794	0.270
GWR	51.167	70.525	0.388	60.143	75.790	0.281
ANN	22.439	30.842	0.881	47.003	52.484	0.630

Notes: CK, Cokriging; GWR, geographically weighted regression; ANN, artificial neural networks

the soil samples. Furthermore, the selection of the correct number of nodes for the hidden layer is important because few nodes may result in under fitting and too many nodes may result in over fitting, and as a result, the training time of the network increased. Moreover, the transfer function for the hidden or output layers should be selected. In this study, we used four transfer functions, namely Sigmoid ($f(x) = \frac{1}{1 + e^{-x}}$), Gaussian ($f(x) = e^{-x^2}$), Hyperbolic Tangent ($f(x) = \text{Sech}(x)$) and Hyperbolic Secant ($f(x) = \tanh(x)$). Following studies of Bateni *et al.* (2007), Singh *et al.* (2009) and Emamgholizadeh *et al.* (2013) the trial and error method was used to find the best configuration of the ANN model to estimate macro nutrients. Two criteria were used for stopping the training of the MLP model, i.e. the threshold error and iteration number. In other words, when the RMSE between the estimations and observations was less than 0.0001 or when the iterations of the model were reached to maximum of 100 000, the iteration of the model was stopped. Results show that the ANN model with three hidden layers and the Gaussian transfer function gave the best results.

The estimated N, P and K are shown in the Tables 4, 5 and 6 for the calibration (training) and validation (testing) stages. Also Fig. 8 indicates the scatterplot of estimated versus measured N, P and K from the ANN model in two stages of calibration and validation. As shown, the ANN model produced accurate estimation of N, with the RMSE of 0.0070% and 0.0068%, in the calibration and validation stages, respectively. The predicted N indicates that the ANN model has a very small RMSE and MAE and also a high correlation coefficient in both calibration ($R^2 = 0.951$) and validation (validation) ($R^2 = 0.922$).

As the results of this study show, the estimation of N,

P and K from the ANN model was better than GWR and CK models. The main reason is related to the ability of intelligence models, because they can learn the complex relationship between inputs (X, Y, Clay, pH and OC) and outputs (N, P and K), especially when there is a nonlinear relationship between inputs and outputs. So, the ANN model is able to forecast soil macronutrients. Furthermore, in this study we have limited field data (106 soil samples), so the accuracy of GWR and CK models was affected with these limited data. To achieve an interpolation with adequate accuracy with these two models, it is necessary to increase the number of samples and therefore it increases the cost of the study.

4 Conclusions

The present study highlights three techniques applications, namely CK, GWR, and ANN to estimate macronutrients (i.e., N, P and K variables). For this purpose, physical and chemical properties of the soil such as clay, pH and OC were used as the auxiliary soil variables. The results indicated that similar to the estimation of N and P, the performance of the ANN model is better than CK and GWR models to estimate K. A comparison of the results from different methods indicated that the accuracy of the GWR model to estimate macro nutrients (N, P and K) was slightly better than the CK model. This means that the GWR is slightly better than the CK for estimation of macro nutrients based on multiple correlated auxiliary variables. In general, the comparison of statistical metrics (R^2 , RMSE and MAE) for estimation N, P and K revealed that the performance of the ANN model was higher than those of CK and GWR models and it is one of the most effective and reliable methods for estimating macro nutrients.

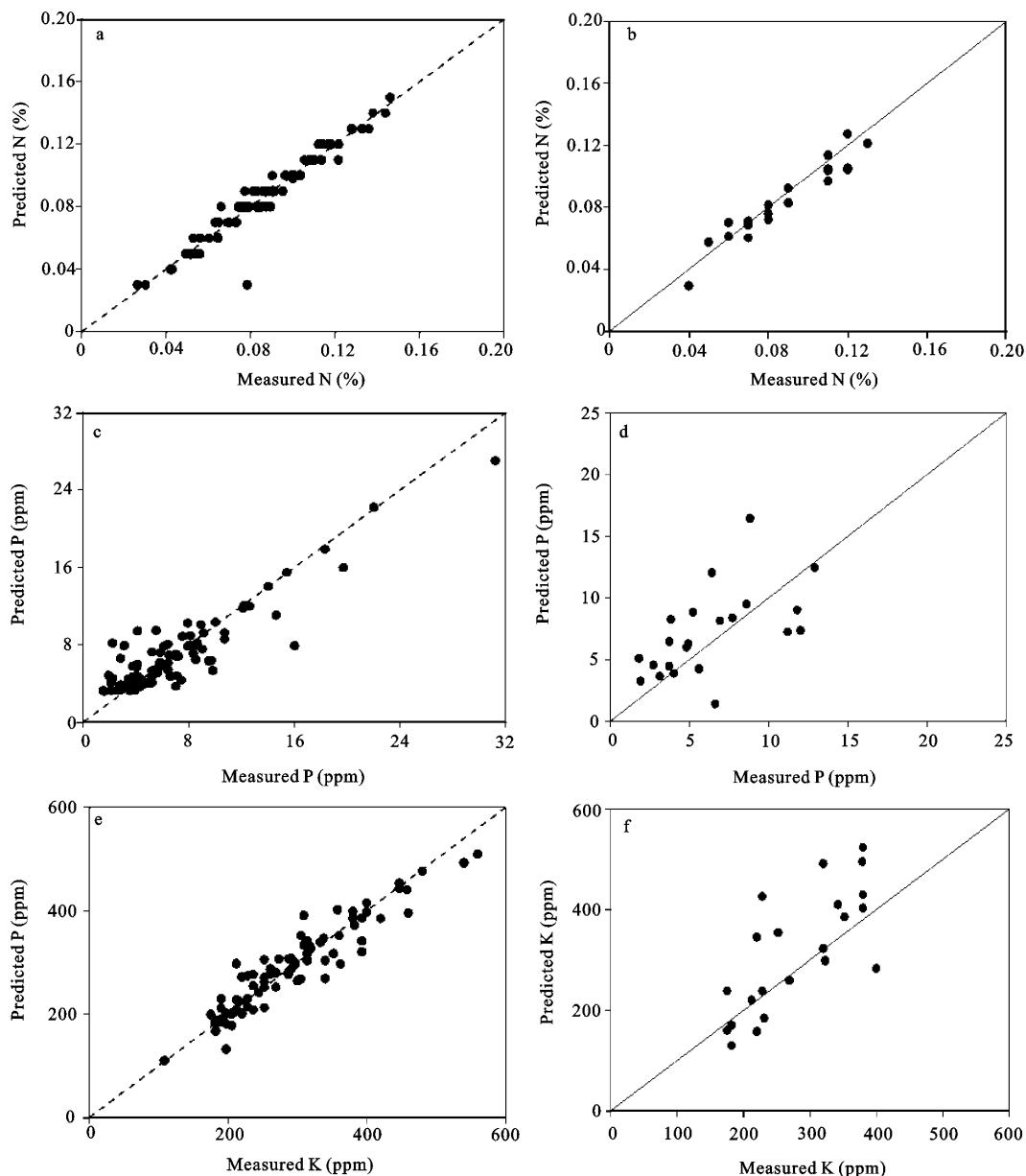


Fig. 8 Scatterplot of estimated versus measured N, P and K from the ANN model (a, c, f) calibration, (b, d, e) validation

References

- Aggelopoulou K, Gemtos T, 2011. Delineation of management zones in an apple orchard: correlations between yield and soil properties. In: Proceedings of the International Conference on Information and Communication Technologies for Sustainable Agri-production and Environment. Skiathos: HAICTA, 443–450.
- Azmathullah H M, Deo M, Deolalikar P B, 2005. Neural networks for estimation of scour downstream of a ski-jump bucket. *Journal of Hydraulic Engineering*, 131(10): 898–908. doi: 10.1061/(ASCE)0733-9429(2005)131:10(898)
- Bateni S M, Borghei S M, Jeng D S, 2007. Neural network and neuro-fuzzy assessments for scour depth around bridge piers. *Engineering Applications of Artificial Intelligence*, 20(3): 401–414. doi: 10.1016/j.engappai.2006.06.012
- Box G E P, Cox D R, 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2): 211–252.
- Camardella C A, Moorman T B, Parkin T B et al., 1994. Field-scale variability of soil properties in central Iowa soils. *Soil Science Society of America Journal*, 58(5): 1501–1511. doi: 10.2136/sssaj1994.03615995005800050033x
- Cressie N, 1993. *Statistics for Spatial Data*. New York: Wiley-Interscience, 15: 105–209.
- Das D K, Bandyopadhyay S, Chakraborty D et al., 2009. Application of modern techniques in characterization and management of soil and water resources. *Journal of the Indian Society*

- of *Soil Science*, 57(4): 445.
- Eldeiry A A, Garcia L A, 2010. Comparison of ordinary kriging, regression kriging, and cokriging techniques to estimate soil salinity using landsat images. *Journal of Irrigation and Drainage Engineering*, 136(6): 355–364. doi: 10.1061/(ASCE)IR.1943-4774.0000208
- Emamgholizadeh S, Bahman K, Bateni S M et al., 2016. Estimation of soil dispersivity using soft computing approaches. *Neural Computing and Applications*, 1–10. doi: 10.1007/s00521-016-2320-x
- Emamgholizadeh S, Bateni S M, Jeng D-S, 2013. Artificial intelligence-based estimation of flushing half-cone geometry. *Engineering Applications of Artificial Intelligence*, 26(10): 2551–2558. doi: 10.1016/j.engappai.2013.05.014
- Emamgholizadeh S, Parsaeian M, Baradaran M, 2015. Seed yield prediction of sesame using artificial neural network. *European Journal of Agronomy*, 68: 89–96. doi: 10.1016/j.eja.2015.04.010
- Eslami M A, Shahsavani S, Roshani G A et al., 2016. Mapping soil fertility of the peyvand cooperative lands by aregis. In: Water and Soil. Shahrood University of Technology, 120.
- Fotheringham A S, Brunsdon C, Charlton M, 2003. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. London: John Wiley & Sons.
- George D, 2011. SPSS for windows step by step: a simple study guide and reference, 17.0 update, 10/e. Pearson Education India.
- Ghorbani H, Kashi H, Hafezi Moghadas N et al., 2015. Estimation of soil cation exchange capacity using multiple regression, artificial neural networks, and adaptive neuro-fuzzy inference system models in golestan province, iran. *Communications in Soil Science and Plant Analysis*, 46(6): 763–780. doi: 10.1080/00103624.2015.1006367
- Goovaerts P, 1998. Ordinary cokriging revisited. *Mathematical Geology*, 30(1): 21–42. doi: 10.1023/A:1021757104135
- Ivajnšić D, Kaligarić M, Žiberna I, 2014. Geographically weighted regression of the urban heat island of a small city. *Applied Geography*, 53: 341–353. doi: 10.1016/j.apgeog.2014.07.001
- Kashi H, Emamgholizadeh S, Ghorbani H, 2014. Estimation of soil infiltration and cation exchange capacity based on multiple regression, ANN (RBF, MLP), and ANFIS models. *Communications in Soil Science and Plant Analysis*, 45(9): 1195–1213. doi: 10.1080/00103624.2013.874029
- Klute A, Dirksen C, 1986. Hydraulic conductivity and diffusivity: laboratory methods. In: Klute A. *Methods of Soil Analysis: Part 1—Physical and Mineralogical Methods*. 2nd ed. Madison: American Society of Agronomy, 687–734.
- Kumar S, Lal R, Liu D S, 2012. A geographically weighted regression kriging approach for mapping soil organic carbon stock. *Geoderma*, 189–190: 627–634. doi: 10.1016/j.geoderma.2012.05.022
- Li Y, Li C K, Tao J-J et al., 2011. Study on spatial distribution of soil heavy metals in huizhou city based on BP-ANN modeling and GIS. *Procedia Environmental Sciences*, 10: 1953–1960. doi: 10.1016/j.proenv.2011.09.306
- Liu F, Zhang G L, Sun Y J et al., 2013. Mapping the three-dimensional distribution of soil organic matter across a subtropical hilly landscape. *Soil Science Society of America Journal*, 77(4): 1241–1253. doi: 10.2136/sssaj2012.0317
- Malinova T, Guo Z X, 2004. Artificial neural network modelling of hydrogen storage properties of Mg-based alloys. *Materials Science and Engineering: A*, 365(1–2): 219–227. doi: 10.1016/j.msea.2003.09.031
- Malvi U R, 2011. Interaction of micronutrients with major nutrients with special reference to potassium. *Karnataka Journal of Agricultural Sciences*, 24(1): 106–109.
- McCulloch W S, Pitts W, 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4): 115–133. doi: 10.1007/BF02478259
- McKenzie N J, Ryan P J, 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89(1–2): 67–94. doi: 10.1016/S0016-7061(98)00137-2
- Mishra U, Lal R, Liu D S et al., 2010. Predicting the spatial variation of the soil organic carbon pool at a regional scale. *Soil Science Society of America Journal*, 74(3): 906–914. doi: 10.2136/sssaj2009.0158
- Murayama Y, 2012. *Progress in Geospatial Analysis*. Japan: Springer Science & Business Media.
- Nath T N, 2014. Status of macronutrients (N, P and K) in some selected tea growing soils of sivasagar district of Assam, India. *International Research Journal of Chemistry*, 7.
- Odeh I O A, McBratney A B, Chittleborough D J, 1995. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. *Geoderma*, 67(3–4): 215–226. doi: 10.1016/0016-7061(95)00007-B
- Olsen S R, 1954. *Estimation of Available Phosphorus in Soils by Extraction with Sodium Bicarbonate*. Washington: United States Department of Agriculture.
- Reeves D W, 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. *Soil and Tillage Research*, 43(1–2): 131–167. doi: 10.1016/S0167-1987(97)00038-X
- Rincón-Ruiz A, Pascual U, Flantua S, 2013. Examining spatially varying relationships between coca crops and associated factors in colombia, using geographically weight regression. *Applied Geography*, 37: 23–33. doi: 10.1016/j.apgeog.2012.10.009
- Robinson T P, Metternicht G, 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture*, 50(2): 97–108. doi: 10.1016/j.compag.2005.07.003
- Saikh H, Varadachari C, Ghosh K, 1998. Changes in carbon, nitrogen and phosphorus levels due to deforestation and cultivation: A case study in simlipal national park, india. *Plant and Soil*, 198(2): 137–145. doi: 10.1023/A:1004391615003
- Singh K P, Basant A, Malik A et al., 2009. Artificial neural network modeling of the river water quality—a case study. *Ecological Modelling*, 220(6): 888–895. doi: 10.1016/j.ecolmodel.2009.01.004

- Snyder P J, Redmond M P, 1995. Understanding neural networks. Computer explorations. Volumes 1 and 2. M. Caudill and C. Butler. Cambridge: MIT Press, 1992. *Brain and Cognition*, 27(1): 128–133. doi: 10.1006/brcg.1995.1012
- Pang S, Li T X, Wang Y D et al., 2009. Spatial interpolation and sample size optimization for soil copper (Cu) investigation in cropland soil at county scale using cokriging. *Agricultural Sciences in China*, 8(11): 1369–1377. doi: 10.1016/S1671-2927(08)60349-1
- Sun W, Minasny B, McBratney A, 2012. Analysis and prediction of soil properties using local regression-kriging. *Geoderma*, 171–172: 16–23. doi: 10.1016/j.geoderma.2011.02.010
- Wackernagel H, 2003. *Multivariate Geostatistics: an Introduction with Applications*. Berlin Heidelberg: Springer Verlag.
- Walkley A, Black I A, 1934. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science*, 37(1): 29–38. doi: 10.1097/00010694-193401000-00003
- Wang K, Zhang C R, Li W D, 2013. Predictive mapping of soil total nitrogen at a regional scale: a comparison between geographically weighted regression and cokriging. *Applied Geography*, 42: 73–85. doi: 10.1016/j.apgeog.2013.04.002
- Western A W, Zhou S-L, Grayson R B et al., 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *Journal of Hydrology*, 286(1–4): 113–134. doi: 10.1016/j.jhydrol.2003.09.014
- Zhang C S, Tang Y, Xu X L et al., 2011. Towards spatial geochemical modelling: use of geographically weighted regression for mapping soil organic carbon contents in ireland. *Applied Geochemistry*, 26(7): 1239–1248. doi: 10.1016/j.apgeochem.2011.04.014