Communications in Soil Science and Plant Analysis
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lcss20

Estimation of Soil Cation Exchange Capacity using Multiple Regression, Artificial Neural Networks, and Adaptive Neuro-fuzzy Inference System Models in Golestan Province, Iran

Hadi Ghorbani[a], Hamed Kashi[a], Naser Hafezi Moghadas[b] & Samad Emamgholizadeh[a]
[a] Department of Water and Soil, Faculty of Agriculture, University of Shahrood, Shahrood, Iran
[b] Department of Earth Science, Faculty of Sciences, Ferdoosi University of Mashhad, Mashhad, Iran
Accepted author version posted online: 16 Mar 2015.Published online: 16 Mar 2015.

Communications in Soil Science and Plant Analysis, 00:1–18, 2015

ISSN: 0010-3624 print / 1532-2416 online
DOI: 10.1080/00103624.2015.1006367

Estimation of Soil Cation Exchange Capacity using Multiple Regression, Artificial Neural Networks, and Adaptive Neuro-fuzzy Inference System Models in Golestan Province, Iran

HADI GHORBANI,[1] HAMED KASHI,[1] NASER HAFEZI MOGHADAS,[2] AND SAMAD EMAMGHOLIZADEH[1]

[1]Department of Water and Soil, Faculty of Agriculture, University of Shahrood, Shahrood, Iran
[2]Department of Earth Science, Faculty of Sciences, Ferdoosi University of Mashhad, Mashhad, Iran

Artificial neural networks (ANN) and adaptive neuro-fuzzy inference systems (ANFIS) provide an alternative by estimating soil parameters from more readily available data. In this article, multilayer perceptron (MLP) and radial basis function (RBF) of ANN and ANFIS models were described to estimate soil cation exchange capacity and compared to traditional multiple regression (MR). Moreover, to test the accuracy of previous functions that estimate cation exchange capacity (CEC), five pedotransfer functions (PTFs) were surveyed. The results showed that the accuracies of ANN and ANFIS models were similar in relation to their statistical parameters. It was also found that ANFIS model exhibited greater performance than RBF, MLP, MR, and PTFs to estimate soil CEC, respectively. Finally, sensitivity analysis was conducted to determine the most and the least influential variables affecting soil CEC. The performance comparisons of used models showed that the soft computing system is a good tool to predict soil characteristics.

Keywords Adaptive neuro-fuzzy inference system, artificial neural networks, cation exchange capacity, multiple regression, pedotransfer function, soil characteristics

Introduction

There is increasing demand for reliable large-scale soil data to meet the requirements of models for planning of land-use systems, characterization of soil pollution, and prediction of land degradation (McBratney et al. 2002). In recent years, the models simulating soil processes have rapidly increased. These models developed to improve the understanding of important soil characteristics. Cation exchange capacity (CEC) is a description of the amount of negative charge in soil that is available to bind positively charged ions (cations). Cation exchange capacity buffers fluctuations in nutrient availability and pH

Received 25 October 2012; accepted 25 March 2014.
Address correspondence to Hadi Ghorbani, Associated Professor in Soil and Environmental Pollution, Department of Water and Soil, Faculty of Agriculture, University of Shahrood, 3619995161 Shahrood, Iran. E-mail: Ghorbani1969@yahoo.com

values in soil. Clay, organic matter, and, to a lesser extent, silt are the main sources of CEC (Seybold, Grossman, and Reinsch 2005). Measuring CEC is both time-consuming and expensive, especially in the Aridisols of Iran, because of the large amounts of calcium carbonate (Carpena, Lax, and Vahtras 1972) and gypsum contents (Fernando, Burau, and Arulanandan 1977).

As a result, the collection of data is the most difficult and expensive step toward modeling the environmental process. Soft computing systems are working by translating this basic information into estimates of other more laborious parameters in soil. The three common methods used to estimate soil characteristics are multiple regressions, artificial neural networks (ANN), and adaptive neuro-fuzzy inference system (ANFIS). Unlike regression, artificial neural networks do not require an a priori regression model, which relates input and output data and in general is difficult because these models are not known (Mermoud and Xu 2006; Schaap, Leij, and van Genuchten 1998). These models can identify and learn correlated patterns between input data sets and corresponding target values. They imitate the learning process of the human brain and can process problems involving nonlinear and complex data even if the data are imprecise and noisy. Thus they are ideally suited for the modeling of agricultural data, which are known to be complex and often nonlinear.

Artificial neural networks have been successfully employed to predict soil properties (Minasny, McBratney, and Bristow 1999; Pachepsky and Rawls 1999; Minasny and McBratney 2002). In most of these models, CEC is assumed to be a linear function of soil organic matter and clay contents (Breeuwsma et al. 1986; McBratney et al. 2002). Some results have shown that greater than 50 percent of the variation in CEC could be explained by the variation in clay and organic carbon contents for several New Jersey soils (Drake and Motto 1982), for some Philippine soils (Sahrawat 1983), and for four soils in Mexico (Bell and van Keulen 1995). Only a small improvement was obtained by adding pH to the model for four Mexican soils (Bell and van Keulen 1995). Amini et al. (2005) tested several published pedotransfer functions and developed two neural network algorithms using multilayer perceptron (MLP) and general regression neural networks based on a set of 170 soil samples to predict soil cation exchange capacity in central Iran. Vos et al. (2005) for prediction of bulk density used twelve pedotransfer functions and the Brazilian database. Their results showed that the separation calibration of topsoil and subsoil layers did not enhance the predictive capacity significantly. Kaur, Kumar, and Gurung (2002) used pedotransfer function for estimating soil bulk density from basic soil data and compared them with existing pedotransfer function. Rasiah (1995) used pedotransfer functions to predict nitrogen-mineralization parameters and get highly accurate predictions. Keshavarzi and Sarmadian (2010) estimated soil cation exchange capacity with ANNs and multiple regression, and their results showed greater accuracy of ANNs than multiple regression. Keshavarzi et al. (2011) showed that ANFIS have better accuracy than ANNs to estimate soil CEC. A comparison of multiple regression, ANN (multilayer perceptron, radial basis function), and ANFIS models for prediction of the swell potential of clayey soils showed that ANN (radial basis function) has the best accuracy (Yilmaz and Kaynar 2011).

The objectives of this study were (i) to develop several ANNs for estimation of soil CEC in the Aridisols of Iran, (ii) to survey the accuracy of some of the most important pedotransfer functions to estimate CEC, and (iii) to evaluate the performance of ANNs compared to ANFIS and multiple regressions.

Materials and Methods

Study Area

The study was carried out in Golestan Province, located in northeastern Iran. The land area is about 20,893 km^2 and located between 36° 44″ and 38° 5″ northern latitudes and 53° 51″ and 56° 14″ eastern longitudes (Figure 1). The average annual temperature is 18.2 °C and the annual rainfall is 556 mm. Geographically, it is divided into two sections: the plains and mountains of the Alborz range. In the eastern Alborz range, the direction of mountains faces northeast and height gradually decreases. The study included four adjacent land parcels under different uses: (1) natural forest, (2) reforested land, (3) cultivated land, and (4) grasslands covered by *Artemissia, Salicornia,* and *Astragallus* species. Large areas of lands are agricultural, with different products. The soil moisture and temperature regimes of the region are Aridic and Thermic, respectively. Based on soil taxonomy (USDA Soil Taxonomy 2010), most soils are Molisols and Aridisols with high values of nonswelling 2:1 clay minerals. The study area was divided into regular, 10-km^2 lattices, and soil samples were collected from each of the lattices (Figure 1).

Data Collection and Soil Sample Analysis

The study location was appointed using GPS, after preliminary studies of the topographic maps (1:25000). Two hundred and twenty soil samples were collected from 0 to 30 cm deep of the study area in Gorgan Province, northeastern Iran. Measured soil parameters including CEC determined by the method of Bower (Sparks et al. 1996). Soil particle-size distribution was determined using Bouyoucos hydrometer method and pH was determined with a pH meter (Gee and Bauder 1986). Organic carbon percentage was determined using the Walkley-Black method (Nelson and Sommers 1982).

Pedotransfer Function Approach

To develop the pedotransfer function approach to estimate soil cation exchange capacity, this parameter was indirectly estimated from the easily measured soil properties such as clay content, organic material, and organic carbon. Therefore, for assessment of the accuracy of some created pedotransfer functions, the best of them have tested. Table 1 shows the selected pedotransfer functions for estimation of soil cation exchange capacity.

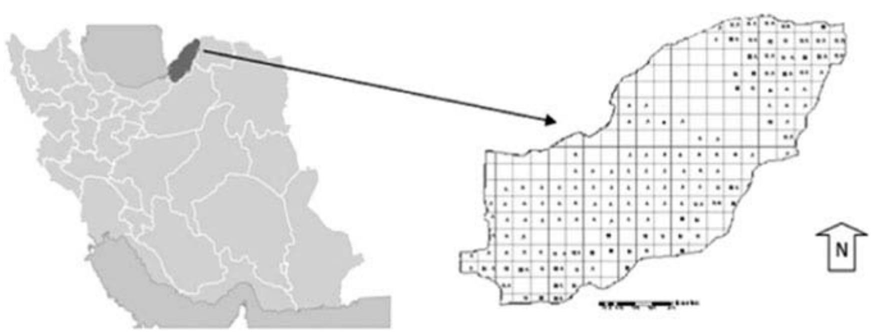

Figure 1. Location of the study area.

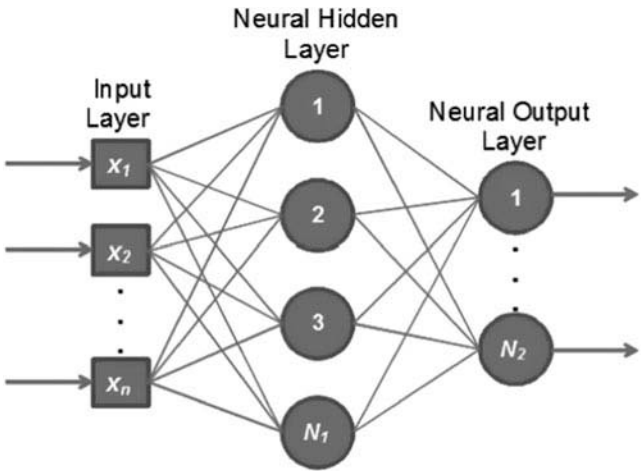

Figure 2. MLP neural network structure.

Table 1
Selected pedotranfer function for estimation of soil cation exchange capacity

Model	Reference	PTF model	Calibration coefficients[a]
F_1	Breeuwsma et al. (1986)	$CEC = a_1 Clay + a_2 OM$	$a_1 = 0.32$ $a_2 = 3.1$
F_2	Manrique, Jones, and Dyke (1991)	$CEC = a_0 + a_1 Clay + a_2 OC^b$	$a_0 = 4.56$ $a_1 = 0.2$ $a_2 = 5.2$
F_3	Bell and van Keulen (1995)	$CEC = a_0 + a_1 Clay + a_2 OM$	$a_1 = 0.2$ $a_2 = 0.3$
F_4	Bell and van Keulen (1995)	$CEC = a_0 + a_1 Clay$	$a_0 = 4.8$ $a_1 = 0.3$
F_5	McBratney et al. (2002)	$CEC = a_0 + a_1 Clay + a_2 Clay^* OC$	$a_0 = 6.9$ $a_1 = 0.1$ $a_2 = 0.16$

[a]CEC is in $cmol_c/kg$, clay and OC are in mass %, and $OM = 1.72^* OC$.

Multiple Regression Model (MR)

Multiple regression is a statistical technique that allows us to predict someone's score on one variable on the basis of their scores on several other variables. It used when it need to exploring linear relationship between the predictor and criterion variables based on statistical technique. The multiple regression equation takes the following form:

$$y = b_1x_1 + b_2x_2 + \ldots. + b_nx_n + c.b_1, b_2, \ldots, b_n$$

where a, b, . . . are the regression coefficients and represent the amount that the dependent variable y changes when the corresponding independent variable changes 1 unit; c is a constant, where the regression line intercepts the y axis, representing the amount of the dependent variable y when all of the independent variables are zero.

Artificial Neural Network (ANN) Models MLP and RBF

Artificial neural networks (ANNs) are nonlinear data-driven self-adaptive approaches as opposed to the traditional model-based methods. They are powerful tools for modeling, especially when the underlaying data relationship is unknown. The ANNs can identify and learn correlated patterns between input data sets and corresponding target values. Neural networks may be used as a direct substitute for autocorrelation, multivariable regression, linear regression, and trigonometric and other statistical analysis and techniques (Singh et al. 2003). A very important feature of these networks is their adaptive nature, where "learning by example" replaces "programming" in solving problems. This feature makes such computational models very appealing in application domains where one has an incomplete understanding of the problem to be solved but where training data are readily available.

The particular network can be defined by three fundamental components: transfer function, network architecture, and learning law (Simpson 1990). It is essential to define these components, to solve the problem satisfactorily. Neural networks consist of a large class of different architectures. Two of the most widely used neural network architecture in literature for classification or regression problems are multilayer perceptron (MLP) and radial basis function (RBF) (Cohen and Intrator 2001, 2003; Kenneth, Wernter, and Macinyre 2001; Lim, Loh, and Shih 2000). Both types of neural network structures are good in pattern classification problems. They are robust classifiers with the ability to generalize for imprecise input data. The MLP are said to distributed-processing networks because the effect of a hidden unit can be distributed over the entire input space; on the other hand, Gaussian RBF networks are said to be local-processing networks because the effects of a hidden unit are usually concentrated in a local area centered at the weight vector.

In this study, two different architectures of ANN (MLP and RBF) are used to estimate of soil cation exchange capacity. All data were first normalized and divided into two data sets, training (70 percent of the whole data) and test (30 percent of the remained data). Table 2 shows the basic statistics of measured soil parameters for training and testing stages. In this study Matlab 7.1 (2010) software was used in neural network analyses.

Multilayer Perceptron (MLP) Model

Multilayer perceptrons (MLPs) represent the most prominent and well-researched class of ANNs in classification, implementing a feed-forward, supervised, and heteroassociative paradigm. In MLP, the weighted sum of the inputs and bias term are passed to activation level through a transfer function to produce the output, and the units are arranged in a layered feed-forward topology called "feed-forward neural network" (Venkatesan and Anitha 2006). A multilayer percepteron has some characteristics: (a) has any number of inputs, (b) has one or more hidden layers with any number of units, (c) uses linear combination

Table 2
Basic statistics of the measured soil parameters

	Training set				Testing set			
Soil parameter	Min	Max	Average	SD	Min	Max	Average	SD
pH	6.5	8.9	7.95	0.3	7.24	8.5	7.96	0.22
Clay (%)	4	64	31.33	12.9	12	62	35.43	10.45
Silt (%)	20	67	44.13	10.5	10	70	39.43	13.01
Sand (%)	10	62	24.54	9.37	10	54	25.13	9.83
OC (%)	0.13	5.73	1.3	0.9	0.34	3.85	1.37	0.70
CEC ($cmol_c/kg$)	4.6	36.95	18.71	7.27	14.39	24.78	20.70	2.64

functions in the input layers, (d) uses generally sigmoid activation functions in the hidden layers, (e) has any number of outputs with any activation function, and (f) has connections between the input layer and the first hidden layer, between the hidden layers, and between the last hidden layer and the output layer. The MLP transforms n inputs to 1 outputs through some nonlinear functions. The output of the network is determined by the activation of the units in the output layer as follows:

$$Xo = f\left(\sum X_h w_{ho}\right)$$

where (f) is activation function, X_h is activation of hidden layer node, and w_{ho} is is the interconnection between hidden layer node and output layer node. The most used activation function is the sigmoid and it is given as follows:

$$sgm\,(x) = \frac{1}{1 + e^{-x}}$$

Radial Basis Function (RBF) Model

Radial basis functions emerged as a variant of artificial neural network in late 1980s. However, their roots are entrenched in much older pattern-recognition techniques. The RBF networks are also good at modeling nonlinear data and can be trained in one stage rather than using an iterative process as in MLP and also learn the given application quickly (Venkatesan and Anitha 2006). The RBF networks are also feed-forward but have only one hidden layer (Figure 3).

The RBF neural networks have some characteristics: (a) has any number of inputs, (b) typically has only one hidden layer with any number of units, (c) use radial combination functions in the hidden layer, based on the squared Euclidean distance between the input vector and the weight vector, (d) typically uses exponential or soft-max activation functions in the hidden layer, in which case the network is a Gaussian RBF network, (e) has any number of outputs with any activation function, and (f) has connection between the input layer and the hidden layer and between the hidden layer and the output layer. Because the hidden units are nonlinear, the outputs of the hidden layer may be combined linearly and so processing is rapid. The output y of an RBF network is computed by the following equation (ASCE 2000):

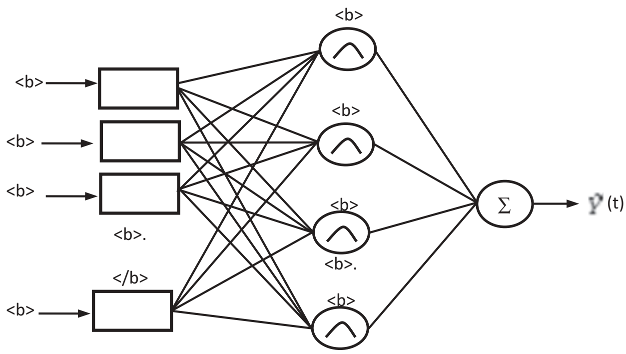

Figure 3. RBF neural network structure.

$$y_k\,(x) = \sum_{J=1}^{M} W_{KJ}\,\varphi_{\mathrm{j}}\,(x) + W_{KO}$$

where M is the number of basic functions, x is the input data vector, w_{kj} represents a weighted connection between the basis function and output layer, and ϕ_j is the nonlinear function of unit j.

Adaptive Neurofuzzy Inference System (ANFIS) Model

The adaptive neuro-fuzzy inference system (ANFIS), first introduced by Jang (1993). The neuro-fuzzy approach combines ANN and fuzzy logic. It effectively integrates the learning capability of neural networks into a fuzzy inference system (FIS). It can be used to approximate any real continuous function on a compact set to any degree of accuracy (Jang, Sun, and Mizutani 1997). Depending on the types of inference operations upon if-then rules, most FISs can be classified into three types: Tsukamoto's system, Mamdani's system, and Sugeno's system (Kişi 2007). In this study, the first-order Sugeno fuzzy model is used because it has been used widely in engineering problems. The ANFIS model is able to use two different optimization methods (hybrid and back propagation) to tune MFs and generate fuzzy rules. The hybrid method is a combination of least squares estimation combined with back propagation method (Matlab User Manual 2010).

Figure 4 shows a typical ANFIS architecture. Every node in layer 1 is an adaptive node with a node function that may be a Gaussian membership function or any membership functions. Every node in layer 2 is a fixed node labeled II, representing the firing strength of each rule. Every node in layer 3 is a fixed node, representing the normalized firing strength of each rule. Every node in layer 4 is an adaptive node with a node function. The single node in layer 5 is a fixed node labeled $\sum$, indicating the overall output (Z) as the summation of all incoming signals (Dastorani et al. 2009). In this study, Gaussian membership function was used for the input variable and Levenberg- Marquaardt algorithm was employed for network training and adjusting its weights.

Modeling Performance Criteria

In this study the performance of all MR, ANN, and ANFIS models was assessed based on four error measures, namely, the correlation coefficient (R^2), which presents the degree

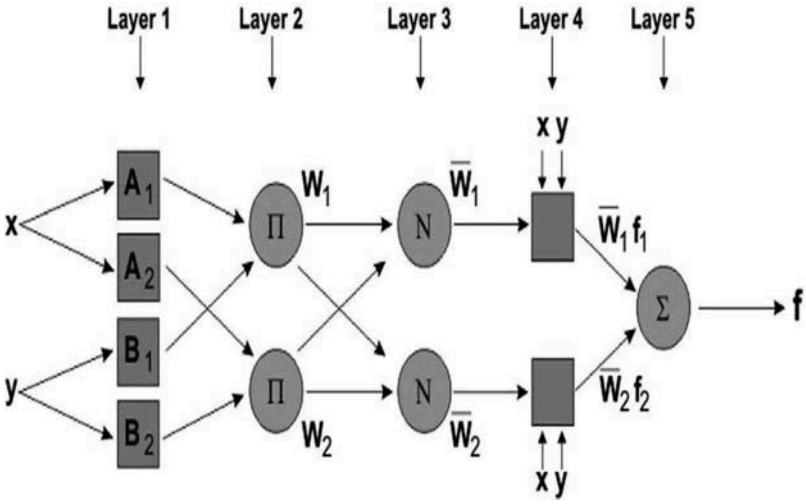

Figure 4. Typical ANFIS structure.

of association between predicted and measured values; root mean square error (RMSE), which is preferred in many iterative prediction and optimization schemes; and mean absolute error (MAE), which is a parameter commonly understood in engineering applications. Expressions for these measures are given as follows:

$$R^2 = \left[\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}\right]^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(Z_O - Z_P)^2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Z_O - Z_P|$$

where Z_O and Z_P are the measured and predicted values, respectively; y and $\bar{y}$ are absolute and average predicted values; and x and $\bar{x}$ are absolute and average of measured values, respectively.

The primary purpose of data transformation is to modify the distribution of input variables so that they can better match outputs. The performance of a neural network is often improved through data transformations (Shi 2000). Therefore in this research to develop ANN and ANFIS models, the raw data of both the independent and the dependent variables were normalized to an interval by transformation means of following equation:

$$X_N = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

where X_N is a normalized value of X and X_{max} and X_{min} represent the maximum and minimum values of each variable of the original data, respectively, which makes data within the range of [0–1].

Results and Discussion

Data Summary Statistics

Data summary of test and train are presented in Table 2. Some soil parameters including sand, silt, clay, pH, and organic carbon were considered as input data for prediction of soil cation exchange capacity. The data sets cover a wide range of soil particle-size distribution. The fraction of clay varies from 4 to 64 percent in the study region. The organic carbon content, ranging from 0.13 to 5.73 and averaging 1.3, is rather small. The simple correlation coefficients (r) between variables are given in Table 3. The results showed that clay and OC percentage have the most correlation with CEC and their correlation is significant at the level of 99 percent. The obtained results were both larger than the respective values reported by Manrique, Jones, and Dyke (1991) for arid soils in the USA ($r = 0.55$ for CEC and clay and $r = 0.23$ for CEC and organic carbon). However, clay and organic carbon have impressive roles, more than other factors, on soil cation exchange capacity. The correlation of sand percentage with CEC is least.

Calibration and Testing of the Published PTFs

From the numerous available pedotransfer functions derived to predict soil CEC, we selected only those regression models that used OC and particle-size distribution as independent variables and had a coefficient of determination, R^2, greater than 0.5. The selected PTFs were calibrated for the study region using a generalized least squares procedure with a subset of the entire data set. The performance indices (RMSE, MAE, and R^2) for training and testing stages are given in Table 4. The accuracy of the F1 function of Breeuwsma et al. (1986) shows a better correlation than other functions ($R^2 = 0.77$). The F4 model of Bell and van Keulen (1995) produced the smallest accuracy ($R^2 = 0.64$).

Table 3

Simple linear correlation coefficients (r) between soil CEC and independent variables

	CEC	pH	OC	Clay	Silt	Sand
CEC	1	−0.43**	0.55**	0.62**	−0.41**	−0.32**
pH		1	−0.33**	−0.14*	0.05	0.13
OC			1	−0.25**	−0.23**	−0.05
Clay				1	−0.69**	−0.49**
Silt					1	−0.30**
Sand						1

**Correlation is significant at 0.01 level.

*Correlation is significant at 0.05 level.

Table 4
Performance indices (RMSE, MAE and R^2) for different models

Estimated soil parameter	Model		RMSE	MAE	R^2
	F_1	Overall	4.65	3.31	0.77
	F_2	Overall	4.40	3.21	0.74
	F_3	Overall	8.46	6.58	0.70
	F_4	Overall	6.49	4.75	0.64
	F_5	Overall	4.78	3.30	0.74
Cation exchange capacity ($Cmol_c$/kg)	MR	Training	3.81	3.02	0.79
		Test	1.83	1.64	0.78
	ANN-MLP	Training	2.56	2.4	0.89
		Test	1.32	1.19	0.80
	ANN-RBF	Training	2.37	2.06	0.90
		Test	1.22	1.11	0.84
	ANFIS	Training	1.97	1.36	0.93
		Test	1.01	0.91	0.88

Multiple Regressions (MR)

Multiple regressions were computed for soil train data set by SPSS software. Multiple regression analysis was carried out to correlate the measured soil CEC. A multiple regression model to predict soil CEC is given below:

$$\mathrm{CEC} = 31.59 - 5.38\,(\mathrm{pH}) - 2.69\,(\mathrm{OC}) + 0.17\,(\mathrm{Sand}) + 0.19\,(\mathrm{Silt}) + 0.44\,(\mathrm{Clay})$$

After determining the equation, performance of multiple regression was developed for the test data set. Correlation coefficient, RMSE, and MAE were obtained as 0.78, 1.83, and 1.64, respectively.

Figures 5 and 6 show the cross-correlation and comparison of predicted and measured values of soil cation exchange capacity for the multiple regression model. Scatter plots between measured and predicted soil cation exchange capacity data serve as a useful visual aid to assess a model's accuracy.

Optimization of Neural Network Models

Two types of ANN (MLP, RBF) are used to find the best models to estimate soil CEC. Therefore, various neurons (1–10) were tested to get better correlation with lower error. The RMSE values for various numbers of neurons related to the studied soil parameter are presented in Figure 7. As shown, the minimum level of RMSE for CEC is related to the network, which has four neurons in the hidden layer. Also, with regard to Figure 7, it can be realize that by increasing the number of neurons, the efficiency of models will decrease, and hence the best efficiency is related to the networks having optimum numbers of neurons. Therefore, the ANN model was run with different hidden nodes and transfer functions including sigmoid ($f(x) = 1/\,(1 + \exp(-x))$, Gaussian ($f(x) = e^{-x.x}$), and hyperbolic tangent ($f(x) = \tanh(x)$). Applying the aforementioned transfer functions and different hidden nodes indicated that the MLP model with the sigmoid transfer function

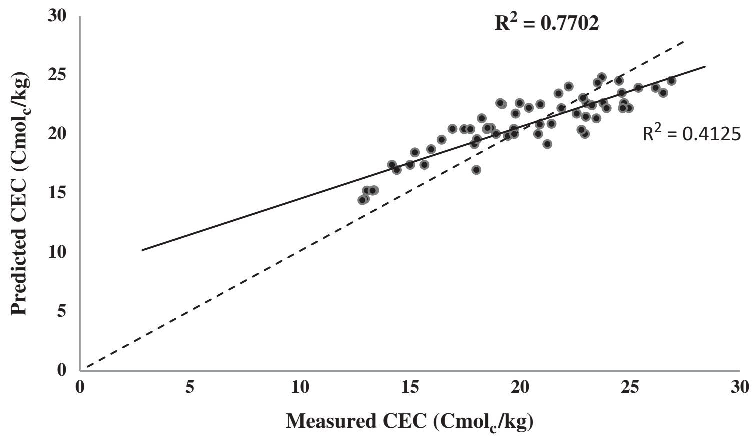

Figure 5. Cross-correlation of predicted and measured values of soil CEC for multiple regression model.

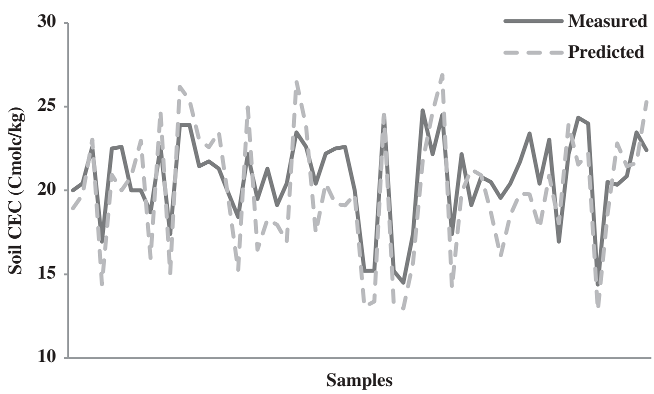

Figure 6. Comparison between predicted and measured soil CEC values using multiple regression model.

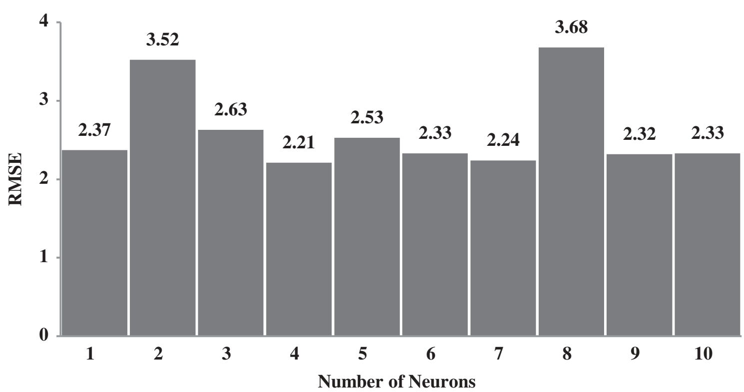

Figure 7. RMSE values for 1–10 neurons in hidden layer for soil CEC.

gives the best results (i.e., the least MAE and RMSE and greatest R^2 values). Also for the RBF model, by trial and error procedure the center selection process found an appropriate tolerance value of 0.005 and the radial basis spread of 25.

The accuracy estimation of soil CEC with MLP and RBF neural networks is shown in Table 4. The results showed that RBF has good accuracy ($R^2 = 0.84$, RMSE = 1.22, and MAE = 1.11) and MLP has less accuracy ($R^2 = 0.80$, RMSE = 1.32, and MAE = 1.19) for estimating soil CEC.

Figures 8–11 show the cross-correlation and comparison of predicted and measured values of soil CEC for MLP and RBF models. The figures show that the RBF models had smaller scatter around the best-fit line than the MLP models for all points. According to this diagram, the best-fit line has the angle near to 45° that shows the high accuracy of estimation by the neural network model.

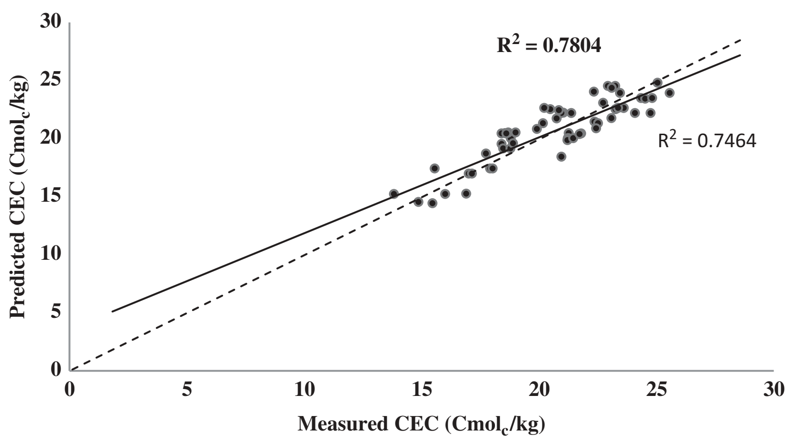

Figure 8. Cross-correlation of predicted and measured values of CEC for MLP model.

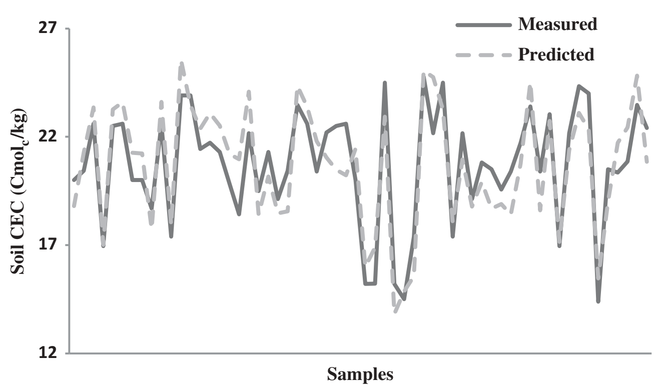

Figure 9. Comparison between predicted and measured soil CEC values using MLP model.

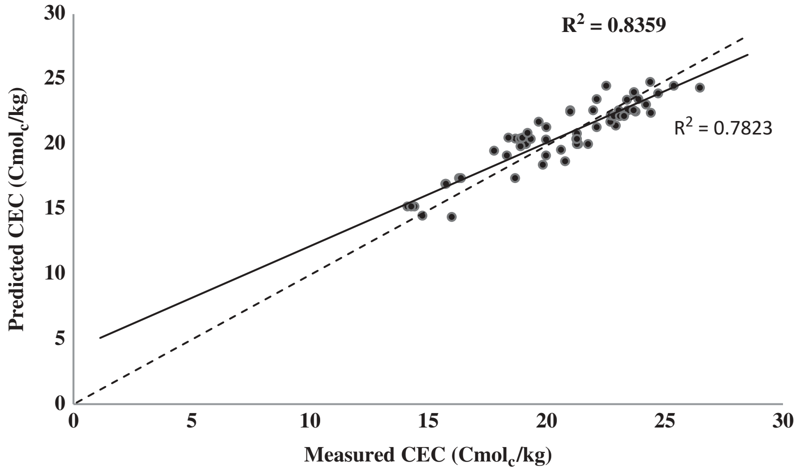

Figure 10. Cross-correlation of predicted and measured values of CEC for RBF model.

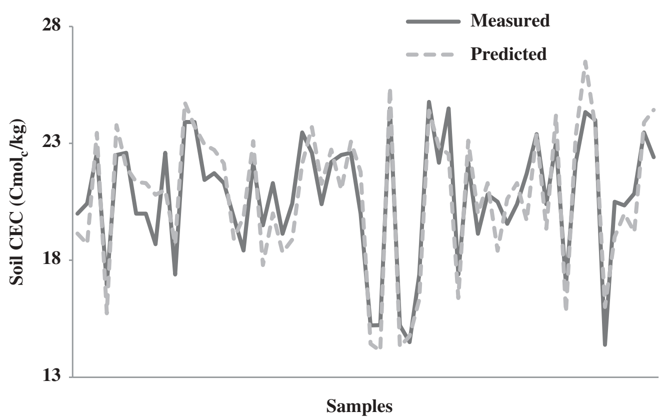

Figure 11. Comparison between predicted and measured soil CEC values using RBF model.

Best Configurations of Neuro-fuzzy Inference System (ANFIS) Models

For the ANFIS model, the Gaussian, triangular, trapezoidal, sigmoid, and bell-shaped input membership functions were used. Because the ANFIS only operates on Sugeno-type systems, two types of constant and linear functions were used for output membership function. The performances of the ANFIS model with different input membership functions, hybrid and back-propagation learning algorithms, and constant and linear output membership functions were investigated. It was found that the best results were achieved when the input membership function is Gaussian, the learning algorithm is hybrid, and output membership function is constant. The ANFIS model has performed for estimation of soil CEC parameter. The results showed that the ANFIS model was the best model to predict soil CEC with high correlation coefficients of 0.93 and 0.88 for training and test, respectively. In other words, its error (RMSE = 1.97, 1.01 and MAE = 1.36, 0.91) was lower and its R^2 was greater than neural networks and linear regression.

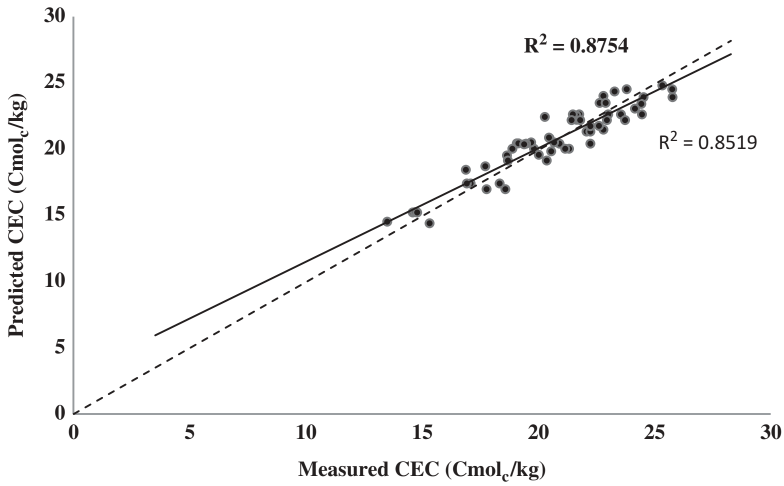

Figure 12. Cross-correlation of predicted and measured values of CEC for ANFIS model.

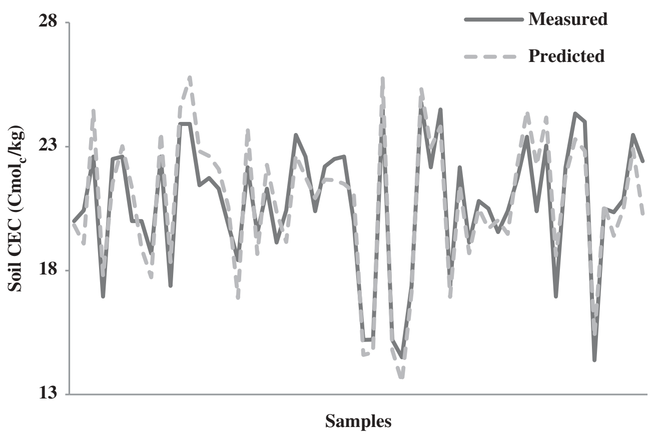

Figure 13. Comparison between predicted and measured soil CEC values using ANFIS model.

Figures 12 and 13 show the cross-correlation and comparison of predicted and measured values of soil CEC for the ANFIS model. Scatter plots between measured and predicted soil cation exchange capacity data serve as a useful visual aid to assess a model's accuracy. These diagrams demonstrate that the predicted values in ANFIS models are close to the experimental values, as many of the data points fall very close to the diagonal line in scatter plot. Clearly the models created by ANFIS models have an agreement with the experimental data. According to this diagram, the best-fitted line has the angle of near to 45° that shows the high accuracy of estimation by the neural network model.

The statistical parameters, i.e., R^2, RMSE, and MAE, for the mentioned models are presented in Table 4. As shown, performance of both artificial neural networks and ANFIS models are quite satisfactory. The accuracy of multiple regressions and selected pedotransfer function was not suitable, because the relations between soil parameters and CEC were not linear.

Table 5
Sensitivity analysis of the input variables on soil cation exchange capacity

	Soil cation exchange capacity		
Method	RMSE	MAE	R^2
Best ANFIS	1.86	1.27	0.92
ANFIS without pH	3.96	2.88	0.80
ANFIS without clay	6.12	4.83	0.68
ANFIS without silt	2.74	1.98	0.85
ANFIS without sand	2.38	1.76	0.88
ANFIS without OC	4.93	3.68	0.75

Sensitivity Analysis

To assign the relative significance of each independent parameter (input variable) on CEC (output), a sensitivity analysis was performed. By comparing the results it was shown that the ANFIS model was better than RBF and MLP models, and therefore the analysis was conducted with this model in the absence of every parameter in the data set. Table 5 gives the sensitivity analysis results for soil CEC. Although the network does not necessarily represent physical meaning through the weights, it suggests that all the input variables have direct relevance and play significant roles in determining the output. However, the results also showed that clay and organic carbon percentages had the best correlation coefficients in CEC and that sand percentage showed the least (Table 5).

The estimation of soil cation exchange capacity with different methods from particle-size distribution and soil organic carbon could not account for all the variation in this study. Moreover, several factors could be implicated for this rather modest accountability. Among them are the type of clay minerals and their morphology, the origin of soil organic matter (Parfitt, Giltrap, and Whitton 1995; Stewart and Hossner 2001), and the large amount of calcium carbonate (Van Bladel, Frankart, and Gheyi 1975) in the Aridisols of the studied region.

Conclusion

This research was designed to evaluate the applicability of new machine learning techniques including artificial neural networks (ANNs) and adaptive neuro-fuzzy inference system (ANFIS) models for estimation of soil CEC in arid rangeland ecosystems. To predict the soil CEC by means of PTFs, the input data consisted of clay, silt, sand, organic carbon, and pH. The performance of the selected pedotransfer function, multiple linear regression, neural network, and ANFIS models were evaluated using a test data set. The results showed that ANFIS model had the best performance to predict soil CEC.

The network model for these parameters was more suitable for capturing the nonlinearity of the relationship between variables. The ANN can model nonlinear functions and has been shown to perform better than linear regression. For accuracy of predicting soil CEC, it seems that the techniques employed in this study can perform quite well, especially in arid rangeland ecosystems, and can be used as powerful tools over existing methods for proper prediction of soil CEC. Because of the poor performance of traditional and statistical equations used in chemical and physical soil properties, interest in applying data-driven models such as ANNs and ANFIS to chemical simulations has to be further accelerated.

The obtained results confirmed the main hypothesis of the research, which was the appropriate performance of ANN and ANFIS models in estimation of soil CEC. It should be mentioned that although large numbers of studies have been carried out on the application of ANNs and ANFIS in soil sciences, more investigations need to be completed on the applications of these techniques in this specific field.

Funding

The authors thank the Department of the Environment of Iran for financial support of the project.

References

Amini, M., K. C. Abbaspour, H. Khademi, N. Fathianpour, M. Afyuni, and R. Schulin. 2005. Neural network models to predict cation exchange capacity in arid regions of Iran. *European Journal of Soil Science* 53: 748–57.

ASCE Task Committee. 2000. Artificial neural networks in hydrology—I: Preliminary concepts. *Journal of Hydrologic Engineering ASCE* 2: 115–23.

Bell, M. A., and H. van Keulen. 1995. Soil pedotransfer functions for four Mexican soils. *Soil Science Society of America Journal* 59: 865–71. doi:10.2136/sssaj1995.03615995005900030034x

Breeuwsma, A., J. H. M. Wösten, J. J. Vleeshouwer, A. M. Van Slobbe, and J. Bouma. 1986. Derivation of land qualities to assess environmental problems from soil surveys. *Soil Science Society of America Journal* 50: 186–90. doi:10.2136/sssaj1986.03615995005000010035x

Carpena, O., A. Lax, and K. Vahtras. 1972. Determination of exchangeable cations in calcareous soils. *Soil Science* 113: 194–99. doi:10.1097/00010694-197203000-00008

Cohen, S., and N. Intrator. 2001. Automatic model selection in a hybrid perceptron/radial network. *Information Fusion: Special Issue on Multiple Experts* 4: 259–66.

Cohen, S., and N. Intrator. 2003. A study of ensemble of hybrid networks with strong regularization. *Multiple Classifier Systems* 5: 227–35.

Dastorani, M. T., A. Moghadamnia, J. Piri, and M. Rico-Ramirez. 2010. Application of ANN and ANFIS models for reconstructing missing flow data. *Environmental Monitoring and Assessment* 166: 421–434. doi:10.1007/s10661-009-1012-8

Drake, E. H., and H. L. Motto. 1982. An analysis of the effect of clay and organic matter content on the cation exchange capacity of New Jersey soils. *Soil Science* 133: 281–88. doi:10.1097/00010694-198205000-00003

Fernando, M. J., R. G. Burau, and K. Arulanandan. 1977. A new approach to determination of cation exchange capacity. *Soil Science Society of America Journal* 41: 818–20. doi:10.2136/sssaj1977.03615995004100040043x

Gee, G. W., and J. W. Bauder. 1986. Particle-size analysis. In *Methods of soil analysis, Part 1*, ed. A. Klute, 383–411. Madison, WI: American Society of Agronomy and Soil Science Society of America.

Jang, J. S. R. 1993. ANFIS: Adaptive network-based fuzzy inference systems. *IEEE Transactions on Systems, Man, and Cybernetics* 23: 665–685.

Jang, J. S. R., C. T. Sun, and E. Mizutani. 1997. *Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence*. Upper Saddle River, NJ: Prentice Hall.

Kaur, R., S. Kumar, and H. P. Gurung. 2002. A pedo-transfer function for estimating soil bulk density from basic soil data and its comparison with existing PTFs. *Australian Journal of Soil Research* 40: 847–57. doi:10.1071/SR01023

Kenneth, J. M., S. Wernter, and J. Macinyre. 2001. Knowledge extraction from radial basis function networks and multi-layer perceptrons. *International Journal of Computational Intelligence and Applications* 3: 369–82.

Keshavarzi, A., and F. Sarmadian. 2010. Comparison of artificial neural network and multivariate regression methods in prediction of soil cation exchange capacity. *International Journal of Environmental and Earth Sciences* 1: 1.

Keshavarzi, A., F. Sarmadian, R. Labbafi, and M. Rajabi Vandekhali. 2011. Modeling of soil cation exchange capacity based on fuzzy table look-up scheme and artificial neural network approach. *Modern Applied Science* 5: 1. doi:10.5539/mas.v5n1p153

Kişi, Ö. 2007. Stream flow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering* 12: 532–53. doi:10.1061/(ASCE)1084-0699(2007)12:5(532)

Lim, T.-S., W.-Y. Loh, and Y.-S. Shih. 2000. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithm. *Machine Learning* 40: 203–28. doi:10.1023/A:1007608224229

Manrique, L. A., C. A. Jones, and P. T. Dyke. 1991. Predicting cation exchange capacity from soil physical and chemical properties. *Soil Science Society of America Journal* 55: 787–94. doi:10.2136/sssaj1991.03615995005500030026x

Matlab User Manual. 2010. *Fuzzy logic toolbox software with Matlab*. Natick, MA: Mathworks.

McBratney, A. B., B. Minasny, S. R. Cattle, and R. W. Vervoort. 2002. From pedotransfer functions to soil inference systems. *Geoderma* 109: 41–73. doi:10.1016/S0016-7061(02)00139-8

Mermoud, A., and D. Xu. 2006. Comparative analysis of three methods to generate soil hydraulic functions. *Soil and Tillage Research* 87: 89–100. doi:10.1016/j.still.2005.02.034

Minasny, B., and A. B. McBratney. 2002. The neuro-m methods for fitting neural network parametric pedotransfer functions. *Soil Science Society of America Journal* 66: 352–61.

Minasny, B., A. B. McBratney, and K. L. Bristow. 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma* 93: 225–53. doi:10.1016/S0016-7061(99)00061-0

Nelson, D. W., and L. E. Sommers. 1982. Total carbon, organic carbon, and organic matter. In *Methods of soil analysis*, *part 2*, 2nd ed., eds. A. L. Page, R. H. Miller, and D. R. Keeney, 539–80. Madison, WI: American Society of Agronomy.

Pachepsky, Y. A., and W. J. Rawls. 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. *Soil Science Society of America Journal* 63: 1748–57. doi:10.2136/sssaj1999.6361748x

Parfitt, R. L., D. J. Giltrap, and J. S. Whitton. 1995. Contribution of organic matter and clay minerals to the cation exchange capacity of soils. *Communications in Soil Science and Plant Analysis* 26: 1343–55. doi:10.1080/00103629509369376

Rasiah, V. 1995. Comparison of pedotransfer functions to predict nitrogen-mineralization parameters of one- and two-pool models. *Communications in Soil Science and Plant Analysis* 26: 1873–84. doi:10.1080/00103629509369414

Sahrawat, K. L. 1983. An analysis of the contribution of organic matter and clay to cation exchange capacity of some Philippine soils. *Communications in Soil Science and Plant Analysis* 14: 803–09. doi:10.1080/00103628309367409

Schaap, M. G., F. J. Leij, and M. T. H. van Genuchten. 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Science Society of America Journal* 62: 847–55. doi:10.2136/sssaj1998.03615995006200040001x

Seybold, C. A., R. B. Grossman, and T. G. Reinsch. 2005. Predicting cation exchange capacity for soil survey using linear models. *Soil Science Society of America Journal* 69: 856–63. doi:10.2136/sssaj2004.0026

Shi, J. J. 2000. Reducing prediction error by transforming input data for neural networks. *Journal of Computing in Civil Engineering* 14: 109–16. doi:10.1061/(ASCE)0887-3801(2000)14:2(109)

Simpson, P. K. 1990. *Artificial neural system: Foundation, paradigm, application, and implementation*. New York: Pergamon Press.

Singh, T. N., R. Kanchan, A. K. Verma, and S. Singh. 2003. An intelligent approach for prediction of triaxial properties using unconfined uniaxial strength. *Mining Engineering Journal* 5: 12–16.

Sparks, D. L., A. L. Page, P. A. Helmke, R. H. Leoppert, P. N. Soltanpour, M. A. Tabatabai, G. T. Johnston, and M. E. Summer. 1996. *Methods of soil analysis*. Madison, WI: Soil Science Society of America.

Stewart, W. M., and L. R. Hossner. 2001. Factors affecting the ratio of cation exchange capacity to clay content in lignite overburden. *Journal of Environment Quality* 30: 1143–49. doi:10.2134/jeq2001.3041143x

USDA Soil Survey Staff. 2010. *Keys to soil taxonomy*, 11th ed. Washington, DC: US Government Printing Office.

Van Bladel, R., R. Frankart, and H. R. Gheyi. 1975. A comparison of three methods of determining the cation exchange capacity of calcareous soils. *Geoderma* 13: 289–98. doi:10.1016/0016-7061(75)90042-7

Venkatesan, P., and S. Anitha. 2006. Application of a radial basis function neural network for diagnosis of diabetes mellitus. *Current Science* 91: 1195–99.

Vos, B. D., M. V. Meirvenne, P. Quataert, J. Deckers, and B. Muys. 2005. Predictive quality of pedotransfer functions for estimating bulk density of forest soils. *Soil Science Society of America Journal* 69: 500–10. doi:10.2136/sssaj2005.0500

Yilmaz, I., and O. Kaynar. 2011. Multiple regression, ANN (RBF, MLP), and ANFIS models for prediction of swell potential of clayey soils. *Expert Systems with Applications* 38: 5958–66. doi:10.1016/j.eswa.2010.11.027