Modeling of Soil Cation Exchange Capacity Based on Fuzzy Table Look-up Scheme and Artificial Neural Network Approach

Ali Keshavarzi (Corresponding author)

Department of Soil Science Engineering, University of Tehran

P.O.Box: 4111, Karaj 31587-77871, Iran

Tel: 98-261-223-1787   E-mail: aliagric@gmail.com, alikeshavarzi@ut.ac.ir

Fereydoon Sarmadian

Department of Soil Science Engineering, University of Tehran

P.O.Box: 4111, Karaj 31587-77871, Iran

Tel: 98-261-223-1787   E-mail: fsarmad@ut.ac.ir

Reza Labbafi

Department of Agricultural Machinery Engineering, University of Tehran

P.O.Box: 4111, Karaj 31587-77871, Iran

Tel: 98-261-280-8138   E-mail: Rezalabbafi@ut.ac.ir

Majid Rajabi Vandechali

Department of Mechanics, Islamic Azad University

Jouybar Branch, Jouybar, Iran

E-mail: majid_r_v@yahoo.com

Abstract

In this study, a new approach is proposed as a modification to a standard fuzzy modeling method based on the table look-up scheme. 70 soil samples were collected from different horizons of 15 soil profiles located in the Ziaran region, Qazvin province, Iran. Then, neural network model (feed-forward back propagation network) and fuzzy table look-up scheme were employed to develop a pedotransfer function for predicting soil CEC using easily measurable characteristics of clay and organic carbon. In order to evaluate the models, root mean square error (RMSE) and R^2 were used. The value of RMSE and R^2 derived by ANN model for CEC were 0.47 and 0.94 respectively, while these parameters for fuzzy table look-up scheme were 0.33 and 0.98 respectively. Results showed that fuzzy table look-up scheme had better performance in predicting and modeling of soil cation exchange capacity than artificial neural network.

Keywords: Fuzzy table look-up scheme, Artificial neural network, Cation exchange capacity, Pedotransfer function, Ziaran

1. Introduction

Cation exchange capacity (CEC) is among the most important soil properties that is required in soil databases (Manrique et al., 1991), and is used as an input in soil and environmental models (Keller et al., 2001). CEC is the amount of negative charge in soil that is available to bind positively charged ions (cations). CEC is used as a measure of fertility, nutrient retention capacity and the capacity to protect groundwater from cation contamination (Akbarzadeh et al., 2009). CEC buffers fluctuations in nutrient availability and soil pH. Soil components known to contribute to CEC are clay and organic matter and to a lesser extent, silt (Seybold et al., 2005). Although CEC can be measured directly, its measurement is especially difficult and expensive in the Aridisols of Iran because of the large amounts of calcium carbonate (Carpena et al., 1972) and gypsum (Fernando et al., 1977).

1.1 Pedotransfer function

The term of pedotransfer function (PTF) was coined by Bouma (1989) as translating data we have into what we need. The most readily available data come from soil survey, such as field morphology, texture, structure and pH.

Pedotransfer functions add value to this basic information by translating them into estimates of other more laborious and expensively determined soil properties. These functions fill the gap between the available soil data and the properties which are more useful or required for a particular model or quality assessment. Various PTFs have been developed to estimate CEC from basic physical and chemical soil properties (Breeuwsma et al., 1986; Manrique et al., 1991; Bell & Van Keulen, 1995; McBratney et al., 2002). In most of these models, CEC is assumed to be a linear function of soil organic matter and clay content (Breeuwsma et al., 1986; McBratney et al., 2002). Results show that greater than 50% of the variation in CEC could be explained by the variation in clay and organic carbon content 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). In B horizons of a toposequence, the amount of fine clay was shown to explain a larger percent of the variation in CEC than the total clay content (Wilding and Rutledge, 1996). Vos et al. (2005) used 12 PTFs and Brazilian's database for prediction of bulk density. Their results showed that the separation of subsoil data from topsoil data did not increase the accuracy of prediction. Similarly, Heusher et al. (2005) and Kaur et al. (2002) reported that the soil texture and organic matter content were the main parameters for estimating of bulk density. Najafi and Givi (2006) used the ANNs and PTFs methods for prediction of soil bulk density. They pointed out that the ANNs are able to predict the soil bulk density better than the PTFs. Amini et al. (2005) estimated the cation exchange capacity in the central of Iran using soil organic matter and clay contents. They used the ANN and five experimental models that were on the basis of regression methods for their predictions. They showed that a neural network PTF with eight hidden neurons was able to predict CEC better than the regression PTFs. Also the ANN model significantly improved the accuracy of the prediction by up to 25%. They concluded that network models are in general more suitable for capturing the non-linearity of the relationship between variables. Jain and Kumar (2006) indicated that the ANN technique can be successfully employed for the purpose of calibration of infiltration equations. They had also found that the ANNs are capable of performing very well in situations of limited data availability.

1.2 Fuzzy inference systems

Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned. Fuzzy inference systems have been successfully applied in fields such as automatic control, data classification, decision analysis, expert systems and computer vision (Sun, 2009). Because of its multidisciplinary nature, fuzzy inference systems are associated with a number of names, such as fuzzy rule-based systems, fuzzy expert systems, fuzzy modeling, fuzzy associative memory, fuzzy logic controllers and simply (and ambiguously) fuzzy systems. It is well known that many elements of land properties have uncertainties. Uncertainty is inherent in decision-making processes, which involve data and model uncertainty. These range from measurement errors, to inherent variability, to instability, to conceptual ambiguity or to simple ignorance of important factors (Keshavarzi et al., 2010). Fuzzy set theory has been widely used in soil science for soil classification and mapping, land evaluation, fuzzy soil geostatistics, soil quality indices (Chang and Burrough, 1987; Burrough, 1989; Zhu et al., 1996; McBratney and Odeh, 1997; McBratney et al., 2003; Zhang et al., 2004; Lagacherie, 2005). McBratney and Odeh (1997) showed the potential of fuzzy set theory in soil science, such as mapping and numeric classification, land use evaluation, modeling and simulation of physical processes. Enea and Salemi (2001) and Klingseisen et al. (2007) used fuzzy logic for evaluating environmental impacts. Metternicht and Gonzalez (2004) presented a fuzzy exploratory model for the prediction of soil erosion hazards. Sadiq and Rodriguez (2004) evaluated and predicted the performance of slow sand filters using fuzzy rule-based modeling. Tran et al. (2002) developed a fuzzy rule-based model to improve the performance of the revised universal soil loss equation (RUSLE). Their approach consisted of two approaches: (1) Multi-objective fuzzy regression (MOFR); and (2) Fuzzy rule-based modeling (FRBM). Tayfur et al. (2003) studied a fuzzy logic algorithm to estimate sediment loads from bare soil surfaces. Predicting the mean sediment loads from experimental runs, the performance of the fuzzy model was compared with that of the artificial neural networks (ANNs) and the physics-based models. The results showed that the fuzzy model performed better under very high rainfall intensities over different slopes and over very steep slopes under different rainfall intensities. Zhu et al. (2010) presented a method to construct fuzzy membership functions using descriptive knowledge. Construction of fuzzy membership functions is accomplished based on two types of knowledge: 1) knowledge on typical environmental conditions of each soil type and 2) knowledge on how each soil type corresponds to changes in environmental conditions. In this study, a new approach is proposed as a modification to a standard fuzzy modeling method. This new method takes randomness into account by considering the statistical properties of training dataset. The method discussed here is called table look-up scheme. The idea is based on all available input-output data pairs (Jang et al., 1997; Liu et

al., 2003), a rule-base will be build. Then the unknown system between the input-output can be approximated using this rule-base.

Hence, the present study was carried out with objective to compare of ANNs model and fuzzy table look-up scheme for estimating and modeling of soil cation exchange capacity using some easily measurable soil parameters in Ziaran region.

2. Materials and methods

2.1 Site description

The study was carried out in Ziaran region, Qazvin province in Iran. The research commenced in 2008 and ended in 2009. The land investigated in the research is located between latitudes of 35°58´ and 36°4´ N and between longitudes of 50°24´ and 50°27´ E which has the area about 5121 hectares. The average, minimum and maximum heights points of Ziaran district are 1204, 1139 and 1269 meters above sea level, respectively (Fig1). The soil moisture and temperature regimes of the region by means of Newhall software are Weak Aridic and Thermic, respectively. Based on soil taxonomy (USDA, 2010), this region has soils in Entisols and Aridisols orders.

2.2 Data collection and soil sample analysis

After preliminary studies of topographic maps (1:25000), using GPS, studying location was appointed. 70 soil samples were collected from different horizons of 15 soil profiles located in Ziaran region in Qazvin Province. Measured soil parameters included texture (determined using Bouyoucos hydrometer method), Organic Carbon (O.C) was determined using Walkley-Black method (Nelson and Sommers, 1982) and CEC (cation exchange capacity in Cmolc kg^{-1} soil) determined by the method of Bower (Sparks et al., 1996).

2.3 Methods to fit PTFs

2.3.1 Artificial neural network

Neural classifiers can deal with numerous multivariable nonlinear problems, for which an accurate analytical solution is difficult to obtain (Park et al., 2010). An artificial neural network is a highly interconnected network of many simple processing units called neurons, which are analogous to the biological neurons in the human brain. Neurons having similar characteristics in an ANN are arranged in groups called layers. The neurons in one layer are connected to those in the adjacent layers, but not to those in the same layer. The strength of connection between the two neurons in adjacent layers is represented by what is known as a 'connection strength' or 'weight'. An ANN normally consists of three layers, an input layer, a hidden layer, and an output layer. In a feed forward network, the weighted connections feed activations only in the forward direction from an input layer to the output layer. On the other hand, in a recurrent network additional weighted connections are used to feed previous activations back into the network. The structure of a feed-forward ANN is shown in Figure 2. This ANN is a popular neural network which known as the back propagation algorithm introduced by Karaca and Ozkaya (2006). This ANN had *k* input and one output parameters. They used this ANN for accurate modeling of the leachate flow-rate. They also reported that the input parameters, number of neurons at the hidden and output layer should be determined according to currently gathered data. Moreover, an important step in developing an ANN model is the training of its weight matrix. The weights are initialized randomly between suitable ranges, and then updated using certain training mechanism (Pachepsky et al., 1996: Schaap et al., 1998; Minasny et al., 1999).

In the feed-forward networks, error minimization can be obtained by a number of procedures including Gradient Descent (GD), Levenberg–Marquardt (LM) and Conjugate Gradient (CG). BP uses a gradient descent (GD) technique which is very stable when a small learning rate is used, but has slow convergence properties (Omid et al., 2009). Several methods for speeding up BP have been used including adding a momentum term or using a variable learning rate. In this study, LM algorithm in the sense that a momentum term is used to speeding up learning and stabilizing convergence is used.

2.3.2 Membership functions and fuzzy table look-up scheme

A general fuzzy system has the components of fuzzification, fuzzy rule-base, fuzzy output engine and defuzzification. Fuzzification converts each piece of input data to degrees of membership by a look-up in one or more several membership functions. The key idea in fuzzy logic is the allowance of partial belongings of any object to different subsets of a universal set, instead of completely belonging to a single set. Partial belonging to a set can be described numerically by a membership function, which assumes values between 0 and 1 inclusive. Intuitions, inference, rank ordering, angular fuzzy sets, neural networks, genetic algorithms and inductive

reasoning can be among many ways to assign membership values or functions to fuzzy variables. Especially, the intuitive approach is used rather commonly because it is simply derived from the innate intelligence and understanding of human beings. Fuzzy membership functions may take on many forms, but in practical applications simple linear functions such as triangular ones are preferable (Tayfur et al., 2003). MFs for the corresponding inputs are recommended by MATLAB (7.8) as triangular membership function. There are five steps in generating fuzzy rules with fixed membership functions. Consider the design of a fuzzy system with two inputs (x_1, x_2) and one output (y) system. Further, there are n data points in the training set.

Step1: Define the fuzzy partition of the input and output variables:

Six and three fuzzy sets are selected to form the partition of this range, respectively. This means the degree of membership can be evaluated for any input values. The fuzzy partition for the output is assumed to have the five fuzzy sets.

Step2: Generate one fuzzy rule for each of the n input-output pairs:

These results are in the initial fuzzy rule base (Eq.1) (Mendel, 2001; Liu et al., 2003):

$$(x_1^{\ p}, x_2^{\ p}, y^p) \Rightarrow Fuzzy\ Rule \quad p = 1,2,...,n \tag{1}$$

From this input-output pair, one fuzzy rule can be generated. One may be reminded of the facts that the fuzzy sets may overlap. Now the question is how to assign the appropriate membership functions to the variables in each data pair. The common practice is that the fuzzy variable is assigned the membership function that produces the maximum membership value.

Step 3: Calculate the degree for each fuzzy rule resulted from rules:

The number of fuzzy rules generated by the input-output pairs is usually large. Inconsistent and redundant rules are inevitable. One is then confronted with the task of eliminating the inconsistency and redundancy.

Step 4: Create the final fuzzy rule base by removing inconsistent and redundant rules:

In the standard approach, the rule having the largest degree is adopted. As an improvement, a new selection approach is proposed here to remove inconsistency and redundancy. The notion of reliability factor is introduced. Specifically, for each given set of k rules with the same antecedent parts, the reliability factor is defined as (Liu et al., 2003):

$$\text{Reliability Factor}\,(RF) = \frac{K_1}{K} \tag{2}$$

Where:

k_1=Number of redundant rules,

k =Total number of the redundant and inconsistent rules having the same antecedent part.

The reliability factor is then used as a weighting factor for computing the effective degree for each rule degree as follows (Liu et al., 2003):

$$\text{Effective Degree}\,(D_{eff}) = D * RF \tag{3}$$

Table 1 shows the example of reliability factors for the inconsistent and redundant rules itemized. The final fuzzy rule-base can now be compiled by choosing the rules with the largest effective degrees. For the redundant and inconsistent rules in table 1, the effective degree is given by (Liu et al., 2003):

$$D_{eff}(\max) = \max(D_{eff(i)}), \quad i = 1,2,...,n \tag{4}$$

Where: D_{eff}= effective degree, and n is the number of membership function.

Step 5: Determine the overall fuzzy system:

Up to this point, the membership functions are defined in step1 and the fuzzy rule-base is compiled in step 4. In this paper, Mamdani's inference scheme is adopted for its simplicity (Fig 3). In carrying out fuzzy inference (reasoning), mathematical operations on the membership functions are invariably required. Any T-norm or S-norm can be used to define the operations involving membership function. In addition, any defuzzification scheme such as the centroid method can be selected. This essentially completes the design procedure in modeling a fuzzy system. In summary, the modified table-look-up scheme offers an effective method for removing inconsistency and redundancy in the process of assembling fuzzy rules. In this study, MATLAB 7.8 software was used for the design and testing of ANN models and fuzzy table look-up scheme.

2.3.3 Performance criteria

The performance of the models was evaluated by a set of test data using the root mean square error (RMSE) and the coefficient of determination (R^2) between predicted and measured values. The RMSE is a measure of accuracy and reliability for calibration and test data sets (Wösten et al., 1999) and is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(z_o - z_p)^2} \tag{5}$$

Where: Z_o is observed value, Z_p is predicted value, and n is number of samples.

3. Results and discussion

3.1 Data summary statistics

Data summary of train and test are presented in Tables 2 and 3, respectively. Data subdivided into two sets: 20% of the data for testing and the remaining 80% of the data were used for training or calibrating. Some soil parameters including clay and organic carbon were input data for prediction of CEC. However, clay and organic carbon were considered as inputs for prediction of CEC. Amini et al. (2005) stated that CEC has high correlation with these inputs. They found that inputs like sand and silt can not improve accuracy of prediction of CEC. Simple linear correlation coefficients (r) between CEC and independent variables were also calculated (Table 4). As Table 4 illustrates correlations between O.C and CEC and between clay and CEC were positive and highly significant. For example the correlation coefficients between CEC and clay content (r = 0.92[**]) is more than between CEC and O.C content (r = 0.56[*]). Positive correlation between CEC, O.C and clay content is related to existence of negative charges on these properties (Manrique et al., 1991; Bell and Van Keulen., 1995; Noorbakhsh et al., 2005). However with regarding to these correlation coefficients, both of them are suitable for developing PTFs for prediction of CEC in soils of Ziaran region.

3.2 Developing PTFs using Artificial Neural Network

After determining of linear correlation coefficients, performance of artificial neural networks was developed for test data set. In the present study for predicting soil CEC we did not increase the input date for constructing artificial neural network. Because according to findings of Lake et al. (2009) and Amini et al. (2005) increasing the number of inputs will decrease the accuracy of the estimations. For example for predicting a soil characteristics if just one types of the input data have low correlation coefficients with output data, the accuracy of the model will automatically decrease. The input data in this model were consisted of the percentages of clay and organic carbon. After determination the complexes of training and testing data, in the next step the various models of neural network having one hidden layer and 1-10 neurons in this layer were made. Then, the optimum structures of network by means of coefficient of determination and RMSE criteria were determined. The RMSE values for various numbers of neurons related to studied soil parameter are presented in the Figure 4. As shown in this Figure, the minimum level of RMSE for CEC is related to the network having seven neurons in the hidden layer. Also, with regarding to this figure can be realize that with 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. The levels of RMSE and R^2 for CEC were 0.47 and 0.94 respectively. Schaap et al. (1998) confirmed applicability of ANNs and concluded that accuracy of these models depend on number of inputs. One of the advantages of neural networks compared to traditional regression PTFs is that they do not require a priori regression model, which relates input and output data and in general is difficult because these models are not known (Schaap and Leij, 1998). The scatter plot of the measured against predicted CEC for the test data set is given in Figure 5. So that 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.

3.3 Developing PTFs using Fuzzy Table Look-up Scheme

Fuzzy rule-base contains fuzzy rules that include all possible fuzzy relations between inputs and outputs. These rules are expressed in the IF-THEN format. In the fuzzy approach there are no mathematical equations and model parameters, however, all the uncertainties and model complications are included in the descriptive fuzzy inference procedure in the form of IF-THEN statements (Tayfur et al., 2003). In this study, fuzzy rules relating the clay and organic carbon contents to soil CEC were inferred from the training data. The antecedent part of the rule (the part starting with IF, up to THEN) included a statement on the clay and organic carbon contents while the consequent part (the part starting with THEN, up to the end) included a statement on soil CEC. For example 'IF the (Clay is Low) and the (O.C is Medium), THEN the (CEC is High)'. Table 5 summarizes the fuzzy rules

constructed in this study. Fuzzy inference engine takes into account all the possible fuzzy rules in the fuzzy rule-base and learns how to transform a set of inputs to corresponding outputs. A general structure of fuzzy system is demonstrated in fig 6. In the main, each fuzzy system consists of three main sections, fuzzifier, fuzzy data base and defuzzifier. At first, input information is made as fuzzy data after bypassing the fuzzifier sections, in which the precise amount value becomes as fuzzy value by membership functions (Fig 7). Defuzzification converts the resulting fuzzy outputs from the fuzzy inference engine to a number. There are several defuzzification methods, such as the weighted average, maximum membership, average maximum membership, and center of gravity, etc. In this study, the centroid method is employed. Later, fuzzy parameters are entered to the fuzzy data base. Fuzzy data base includes two main sections, fuzzy rule-base and inference engine. In fuzzy rule-base, rules related to fuzzy propositions are described. Thereafter, analysis operation is applied by fuzzy inference engine. There are two main fuzzy inference engine-Sugeno and Mamdani- for this purpose. In this paper, Mamdani's inference scheme is adopted for its simplicity (Fig 3) and used for predicting mentioned parameter. Then, the optimum structures of fuzzy table look-up scheme by means of coefficient of determination and RMSE criteria were determined.   The levels of RMSE and R^2 for CEC were 0.33 and 0.98 respectively. In addition, the levels of coefficient of determination and RMSE derived by fuzzy inference system for studied soil parameter had higher values than those derived by artificial neural network (Table 6) which is in line with the work done by Akbarzadeh et al. (2009). The fuzzy inference system for CEC was more suitable for capturing the non-linearity of the relationship between variables. The scatter plot of the measured against predicted CEC for the test data set in fuzzy table look-up scheme is given in figure 8. So that according to this diagram, the best fitted line has the angle of near to 45° that shows higher accuracy of estimation by fuzzy table look-up scheme than neural network model. Liu et al. (2003) found that the modified table look-up scheme can predict the time series more accurately when noise was added to the time series. Akbarzadeh et al. (2009) in their study showed that a hybrid method (ANN and Fuzzy model) predicted soil CEC with very high accuracy. Burrough et al. (1992) demonstrated that fuzzy classification produced a superior number of available areas for agriculture compared to conventional Boolean classification. Zorluer et al. (2010) investigated the application of a fuzzy rule-based method for determination of clay dispersibility. In this study, a fuzzy logic approximation method was developed to combine the different results of the double hydrometer, pinhole, Na (%)-TDS and ESP-CEC methods into a single value. This new method was applied to the dispersibility test results of 29 samples, and it gave more reliable and objective results for identifying the dispersibility of the clay soil. Fernández et al. (2009) worked with fuzzy rules-based on classification systems using a preprocessing step to deal with class imbalance. Their aim was to analyze the behavior of fuzzy rule-based classification systems in the framework of imbalanced datasets through the application of an adaptive inference system with parametric conjunction operators. The empirical results showed that the use of these parametric conjunction operators resulted in a higher performance for all datasets with different imbalanced ratios.

4. Conclusion

In this study, artificial neural network model (feed-forward back-propagation network) and fuzzy table look-up scheme were employed to develop a pedotransfer function for predicting soil cation exchange capacity by using available soil properties. This network was consisted of one hidden layer, a sigmoid activation function in hidden layer, and a linear activation function in output layer and Levenberg-Marquardt training algorithm used due to efficiency, simplicity and high speed. Fuzzy inference system is a rule-based system consists of three conceptual components. There are: a rule-base, contains fuzzy IF-THEN rules, a database, defines the membership function and an inference system, combines the fuzzy rules and produces the system results. First phase of fuzzy logic modeling is the determination of membership functions of input-output variables, second is the construction of fuzzy rules and the last is the determination of output characteristics, output membership function and system results. For predicting the soil property by means of PTFs, the input data were consisted of the percentages of clay and organic carbon for CEC. The performance of the neural network model and fuzzy table look-up was evaluated using a test data set. Results showed that fuzzy table look-up scheme had better performance in predicting soil CEC than neural network model. The fuzzy table look-up scheme for this parameter was more suitable for capturing the non-linearity of the relationship between variables. With regarding to the evaluation criteria, the results of this study revealed that the fuzzy table look-up scheme had superiority to the artificial neural networks for prediction of mentioned soil parameter. This is a crucial result because, since ANN- PTFs formed from local data produce more accurate predictions than those built from data spread from a wider area, the concept of data conservation becomes a critical factor in ANN-PTF construction. However, due to difficulties of direct measurement of soil parameters, we recommend using of neuro-fuzzy models in the future studies for obtaining the logical equations of other soil parameters, especially soil hydraulic properties, in each area and also we recommended testing mentioned formula for CEC in other regions.

References

Akbarzadeh, A., Taghizadeh Mehrjardi, R., Rahimi Lake, H., & Ramezanpour, H. (2009). Application of Artificial Intelligence in Modeling of Soil Properties (Case Study: Roodbar Region, North of Iran). *Environ Research J.*, 3(2), 19-24.

Amini, M., Abbaspour, K. C., Khademi, H., Fathianpour, N., Afyuni, M., & Schulin, R. (2005). Neural network models to predict cation exchange capacity in arid regions of Iran. *Eur. J. Soil Sci.,* 53, 748-757.

Bell, M. A., & Van keulen, J. (1995). Soil pedotransfer functions for four Mexican soils. *Soil Sci Soc. Am. J.,* 59, 865-871.

Bouma, J. (1989). Using soil survey data for quantitative land evaluation. *Advances in Soil Science,* 9, 177-213.

Breeuwsma, A., Wosten, J. H. M., Vleeshouwer, J. J., Van slobbe, A. M., & Bouma, J. (1986). Derivation of land qualities to assess environmental problems from soil surveys. *Soil Sci Soc. Am. J.,* 50, 186-190.

Burrough, P. A. (1989). Fuzzy mathematics methods for soil survey and land evaluation. *Journal of Soil Science*, 40, 447-492.

Burrough, P. A., MacMillian, R. A., & Van Deusen, W. (1992). Fuzzy classification methods for determining land suitability from soil profile observations and topography. *Journal of Soil Science*, 43, 193-210.

Carpena, O., Lux, A., & Vahtras, K. (1972). Determination of exchangeable calcareous soils. *Soil Sci.,* 33, 194-199.

Chang, L., & Burrough, P. A. (1987). Fuzzy reasoning: a new quantitative aid for land evaluation. *Soil Survey and Land Evaluation*, 7, 69-80.

Drake, E. H., & Motto, H. L. (1982). An analysis of the effect of clay and organic matter content on the cation exchange capacity of New Jersey soils. *Soil. Sci.,* 133, 281-288.

Enea, M., & Salemi, G. (2001). Fuzzy approach to the environmental impact evaluation. *Ecological Modelling*, 136, 131-147.

Fernández, A., Jesus, M. J., & Herrera, F. (2009). On the influence of an adaptive inference system in fuzzy rule based classification systems for imbalanced data-sets. *Expert Systems with Applications*, 36, 9805-9812.

Fernando, M. J., Burau, R. G. & Arulanandam, K. (1977). A new approach to determination of cation exchange capacity. *Soil Sci Soc. Am. J.,* 41, 818-820.

Heusher, S. A., Brandt, C. C. & Jardin, P. M. (2005). Using soil physical and chemical properties to estimate bulk density. *Soil Sci. Soc. Am. J.,* 69, 51-56.

Jain, A., & Kumar, A. (2006). An evaluation of artificial neural network technique for the determination of infiltration model parameters. *Appl. Soft Comput.,* 6, 272–282.

Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Karaca, F., & Ozkaya, B. (2006). NN-LEAP: A neural network-based model for controlling leachate flow-rate in a municipal solid waste landfill site. Environ. *Modell. Software.,* 21, 1190-1197.

Kaur, R., Kumar, S., & Gurung, H. P. (2002). A pedotransfer function soil data and its comparison with existing PTFs. *Aust. J. Soil Res.,* 40, 847- 857.

Keller, A., Von Steiger, B., Vander Zee, S. T. & Schulin, R. (2001). A stochastic empirical model for regional heavy metal balances in agroecosystems. *Journal of Environmental Quality.,* 30, 1976-1989.

Keshavarzi, A., Sarmadian, F., Heidari, A., & Omid, M. (2010). Land suitability evaluation using fuzzy continuous classification (a case study: Ziaran region). *Modern Applied Science*, 4(7), 72-81.

Klingseisen, B., Metternicht, G., & Paulus, G. (2007). Geomorphometric landscape analysis using a semi-automated GIS-approach. *Environmental Modelling and Software*, In Press.

Lagacherie, P. (2005). An algorithm for fuzzy pattern matching to allocate soil individuals to pre-existing soil classes. *Geoderma,* 128, 274-288.

Lake, H. R., Akbarzadeh, A., & Taghizadeh Mehrjardi, R. (2009). Development of pedotransfer functions (PTFs) to predict soil physico-chemical and hydrological characteristics in southern coastal zones of the Caspian Sea. *Journal of Ecology and the Natural Environment.,* 1(7), 160-172.

Liu, X., Kwan, B. W., & Foo, S. Y. (2003). Time Series Prediction Based on Fuzzy Principles. *Proc. 'Huntsville Simulation' Conf*, October 8, Huntsville, AL.

Manrique, L. A., Jones, C. A., & Dyke, P. T. (1991). Predicting cation exchange capacity from soil physical and chemical properties. *Soil Sci. Soc. Am. J.,* 50, 787-794.

McBratney, A. B., Mendonca, M. L., & Minansy, B. (2003). On digital soil mapping. *Geoderma*, 117, 3-52.

McBratney, A. B., Minasny, B., Cattle, S. R., & Vervoort, R. W. (2002). From pedotransfer function tosoil inference system. *Geoderma,* 109, 41-73.

McBratney, A. B., & Odeh, I. O. A. (1997). Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma*, 77, 85-113.

Mendel, J. M. (2001). *Rule-Based Fuzzy Logic Systems*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Metternicht, G., & Gonzalez, S. (2004). FUERO: foundations of a fuzzy exploratory model for soil erosion hazard prediction. *Environmental Modelling and Software*, 20 (6), 715-728.

Minasny, B., McBratney, A. B., & Bristow, K. L. (1999). Comparison of different approaches to the development of pedotransfer functions for water retention curves. *Geoderma*, 93, 225- 253.

Najafi, M., & Givi, J. (2006). Evaluation of prediction of bulk density by artificial neural network and PTFs. 10th Iranian Soil Science Congress. Karaj., pp: 680-681. (in Persian)

Nelson, D. W. & Sommers, L. E. (1982). Total carbon, organic carbon, and organic matter. In: Page, AL, Miller, RH, Keeney DR (Eds.), *Methods of Soil Analysis*. Part II, 2nd ed. American Society of Agronomy, Madison, WI, USA. pp. 539-580.

Noorbakhsh, F., Jalalian, A., & Shariatmadari, H. (2005). Prediction of cation exchange capacity with using some soil properties. *Iranian Journal of Science and Technology of Agriculture and Natural Resources*, 3, 107-117 (In Persian).

Omid, M., Baharlooei, A., & Ahmadi, H. (2009). Modeling drying kinetics of pistachio nuts with multilayer feed-forward neural network. *Drying Tech*., 27, 1069-1077.

Pachepsky, Y. A., Timlin, D., & Varallyay, G. (1996). Artificial neural networks to estimate soil water retention from easily measurable data. *Soil Sci. Soc. Am. J.,* 60, 727-733.

Park, B. J., Pedrycz, W. & Oh, S. K. (2010). Polynomial-based radial basis function neural networks (P-RBFNNs) and their application to pattern classification. *Applied Intelligence.,* 32, 27-46.

Sun, R. (2009). Data Mining Based on Fuzzy Rough Set Theory and Its Application in the Glass Identification. *Modern Applied Science*, 3(8), 100-105.

Sadiq, R., & Rodriguez, M. J. (2004). Fuzzy synthetic evaluation of disinfection by-products: a risk-based indexing system. *J Environ Manage,* 73, 1-13.

Sahrawat, K. L. (1983). An analysis of the contribution of organic matter and clay to cation exchange capacity of some Philippine soils. *Commun. Soil Sci. Plant Anal.*, 14, 803-809.

Schaap, M. G., & Leij, F. J. (1998). Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil Till. Res.,* 47, 37-42.

Schaap, M. G., Leij, F. J., & Van Genuchten, M. Th. (1998). Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Sci. Soc. Am. J.,* 62, 847-855.

Seybold, C. A., Grossman, R. B., & Reinsch, T. G. (2005). Predicting Cation Exchange Capacity for soil survey using linear models. *Soil Sci. Soc. Am. J.,* 69, 856-863.

Sparks, D. L., Page, A. L., Helmke, P. A., Leoppert, R. H., Soltanpour, P. N., Tabatabai, M. A., Johnston, G. T., & Summer, M. E. (1996). *Methods of soil analysis*. Soil Sci. Soc. of Am. Madison, Wisconsin.

Tayfur, G., Ozdemir, S., & Singh, V. P. (2003). Fuzzy logic algorithm for runoff-induced sediment transport from bare soil surfaces. *Advances in Water Resources*, 26, 1249-1256.

Tran, L.T., Ridgley, M. A., Duckstein, L., & Sutherland, R. (2002). Application of a fuzzy logic-based modeling to improve the performance of the revised universal soil loss equation. *Catena*, 47(3), 203-226.

USDA. (2010). Soil Survey Staff. *Keys to Soil Taxonomy*. 11 th edition.

Vos, B. D., Meirvenne, M. Quataert, V., Deckers, P. J., & Muys, B. (2005). Predictive quality of pedotransfer

functions for estimating bulk density of forest soils. *Soil Sci. Soc. Am. J.,* 69, 500-510.

Wilding, L. P., & Rutledge, E. M. (1966). Cation exchange capacity as a function of organic matter, total clay, and various clay fractions in a soil toposequence. *Soil Sci. Soc. Am. Proc.,* 30, 782-785.

Wösten, J. H. M., Lilly, A., Nemes, A., & Le Bas, C. (1999). Development and use of a database of hydraulic properties of European soils. *Geoderma,* 90, 169-185.

Zhang, B., Zhang, Y., Chen, D., White, R. E., & Li, Y. (2004). A quantitative evaluation system of soil productivity for intensive agriculture in China. *Geoderma,* 123, 319-331.

Zhu, A. X., Band, L. E., Dutton, B., & Nimlos, T. (1996). Automated soil inference under fuzzy logic. *Ecological Modelling,* 90, 123-145.

Zhu, A. X., Yang, L., Li, B., Qin, C., Pei, T., & Liu, B. (2010). Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma,* 155, 164-174.

Zorluer, I., Icaga, Y., Yurtcu, S., & Tosun, H. (2010). Application of a fuzzy rule-based method for the determination of clay dispersibility. *Geoderma*, In Press.

Table 1. Example of reliability factors and effective degrees for redundant and inconsistent rules

x_1	x_2	y	Degree	RF	D_{eff}
A	B	C_1	D_1	3/6	$D_1*3/6$
A	B	C_1	D_2	3/6	$D_2*3/6$
A	B	C_1	D_3	3/6	$D_3*3/6$
A	B	C_2	D_4	2/6	$D_4*2/6$
A	B	C_2	D_5	2/6	$D_5*2/6$
A	B	C_3	D_6	1/6	$D_6*1/6$

Table 2. Statistics of training data set for cation exchange capacity

	Soil parameter	Min	Max	Mean	Std
Training set	CEC ($Cmol^cKg^{-1}$)	7.42	20.45	15.87	2.64
	Clay (%)	3.20	56.80	23.85	12.05
	O.C (%)	0.04	1.10	0.36	0.23

Table 3. Statistics of testing data set for cation exchange capacity

	Soil parameter	Min	Max	Mean	Std
Testing set	CEC ($Cmol^cKg^{-1}$)	11.10	19.69	16.02	2.21
	Clay (%)	7.20	48.00	23.79	11.79
	O.C (%)	0.09	0.76	0.32	0.20

Table 4. Simple linear correlation coefficients (r) between CEC and independent variables

	CEC ($Cmol^cKg^{-1}$)	Clay (%)	O.C (%)
CEC ($Cmol^cKg^{-1}$)	1	0.92[**]	0.56[*]
Clay (%)	0.92[**]	1	0.22[*]
O.C (%)	0.56[*]	0.22[*]	1

* Correlation is significant at the 0.05 level

** Correlation is significant at the 0.01 level

Table 5. Fuzzy rules relating the clay and organic carbon contents to soil CEC (L= Low; M= Medium; H= High; V=Very)

Rule No.	IF	Clay (%)	and	O.C (%)	THEN	CEC (Cmol^{c}Kg^{-1})
1	IF	VVL	and	L	THEN	L
2	IF	VL	and	L	THEN	M
3	IF	VL	and	M	THEN	M
4	IF	VL	and	H	THEN	M
5	IF	L	and	L	THEN	H
6	IF	L	and	M	THEN	H
7	IF	L	and	H	THEN	H
8	IF	H	and	L	THEN	VH
9	IF	H	and	M	THEN	VH
10	IF	H	and	H	THEN	H
11	IF	VH	and	L	THEN	VH
12	IF	VH	and	M	THEN	VH
13	IF	VH	and	H	THEN	VH
14	IF	VVH	and	M	THEN	VH

Table 6. Statistical parameters in test stage for different methods based on pedotransfer functions

Statistical parameters	Artificial Neural Network	Fuzzy Table Look-up Scheme
RMSE	0.47	0.33
R^2	0.94	0.98
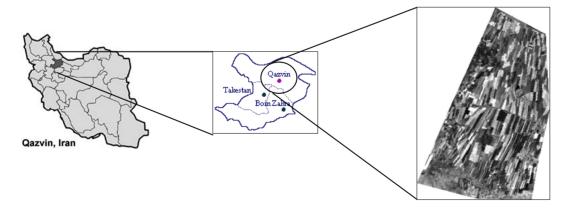
Figure 1. Location of the study area
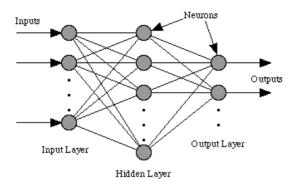
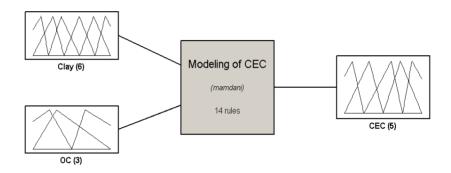Figure 2. Structure of feed-forward ANN

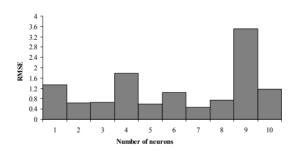Figure 3. Structure of FIS Recommended model

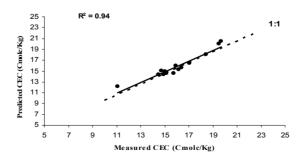Figure 4. RMSE values for 1-10 neurons in hidden layer (cation exchange capacity)

Figure 5. The scatter plot of the measured versus predicted CEC (ANN)
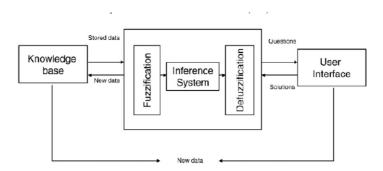
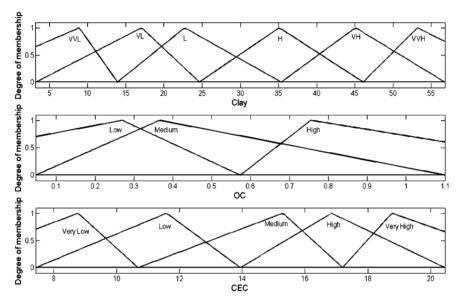Figure 6. The general structure of the fuzzy inference system

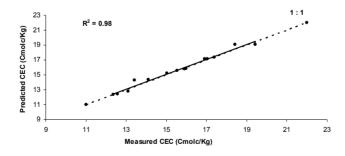Figure 7. Fuzzy membership functions for input-output

Figure 8. The scatter plot of the measured versus predicted CEC (Fuzzy Table Look-up Scheme)