Prediction of CEC using fractal parameters by artificial neural networks

Hossein Bayat[1]*, Naser Davatgar[2], and Mohsen Jalali[1]

[1]Department of Soil Science, Bu Ali Sina University, Hamadan, Iran
[2]Department of Soil Science, Rice Research Institute of Iran, Rasht, Iran

A b s t r a c t. The prediction of cation exchange capacity from readily available soil properties remains a challenge. In this study, firstly, we extended the entire particle size distribution curve from limited soil texture data and, at the second step, calculated the fractal parameters from the particle size distribution curve. Three pedotransfer functions were developed based on soil properties, parameters of particle size distribution curve model and fractal parameters of particle size distribution curve fractal model using the artificial neural networks technique. 1 662 soil samples were collected and separated into eight groups. Particle size distribution curve model parameters were estimated from limited soil texture data by the Skaggs method and fractal parameters were calculated by Bird model. Using particle size distribution curve model parameters and fractal parameters in the pedotransfer functions resulted in improvements of cation exchange capacity predictions. The pedotransfer functions that used fractal parameters as predictors performed better than the those which used particle size distribution curve model parameters. This can be related to the non-linear relationship between cation exchange capacity and fractal parameters. Partitioning the soil samples significantly increased the accuracy and reliability of the pedotransfer functions. Substantial improvement was achieved by utilising fractal parameters in the clusters.

K e y w o r d s: cation exchange capacity, fractal theory, particle size distribution, pedotransfer functions

INTRODUCTION

Measuring cation exchange capacity (CEC) is difficult, costly and time-consuming (Salehi *et al.,* 2008). Moreover, in many cases, CEC values are unavailable in the databases (Seybold *et al*., 2005). Despite the considerable amount of research done to predict CEC from readily available soil properties, improvement of the predictions without additional costs remains a challenging issue. It may be more suitable and economical to develop pedotransfer functions (PTFs) which use some auxiliary variables (the variables which were calculated from soil properties and can be used as input parameters in the PTFs to predict hard-to-measure soil properties) to improve the prediction of CEC. Soil scientists need a reliable estimation for CEC and it would be very important to improve the predictions only by using readily available soil properties. These improvements will benefit all users of soil survey data and will help them in the exact interpretation of their results. Therefore, research in developing methods for the prediction of CEC from readily available soil data is becoming increasingly important (Bishop and McBratney, 2001).

Cation exchange capacity is commonly related to soil texture, mineralogy, and any other information, such as fractal parameters and specific surface area, that depicts their variation. This information can be used to predict CEC (Bishop and McBratney, 2001) leading to successful estimation of CEC from the fractal parameters.

Fractal theory has been employed successfully to explain the soil particle size distribution (PSD) by many authors such as Bird *et al.* (2000). Ersahin *et al.* (2006) related the CEC of the soils to the fractal dimension of PSD using regression method. The estimation of CEC can be improved by using more precise values of fractal dimension that are sensitive to the type of clays in the soils (Ersahin *et al*., 2006). Fractal dimension and CEC increased with increasing finer fractions of the soil and there was a positive correlation between fractal dimension and CEC. The existence of these relationships suggests that fractal parameters should be useful in predicting CEC more accurately.

Calculation of fractal parameters for soils needs the data for the entire PSD curve that is not available in most cases, and some databases. The entire PSD curve can be developed

---

*Corresponding author e-mail: h.bayat@basu.ac.ir

from limited soil texture data (*ie* sand, silt and clay) (Skaggs *et al.*, 2001). Fractal models of PSD can be fitted to the result of PSD curve. The parameters of the developed model and the fractal parameters can be useful to predict CEC.

To our knowledge no one has used fractal parameters, especially fractal parameters and PSD curve model parameters, to predict CEC by artificial neural networks. Thus the objectives of this study were:

– to calculate the fractal parameters and PSD curve parameters using limited soil texture data and to investigate their relationships as new variables with soil CEC, and

– to develop the PTFs by artificial neural networks to evaluate the usefulness for soil CEC prediction of different types of new variables.

MATERIALS AND METHODS

To predict the CEC of the soils 1 662 disturbed soil samples were taken from soil data base of Rice Research Institute of Iran (RRII). Soil samples were passed through a 2 mm sieve. The soil properties that were used in this study included pH in saturated soil paste, soil particle size distribution (sand, silt and clay), organic carbon (OC), and cation

exchange capacity (CEC). Sand, silt and clay mass fractions were measured by hydrometer method (Gee and Or, 2002). Organic carbon was determined by Walky-Black procedure (Nelson and Sommers, 1986). Cation exchange capacity was determined by the ammonium saturation method at pH 7.0 (Soil Survey Division Staff, 1993).

In this study, firstly, we extend the entire PSD curve from limited soil texture data and obtained PSD curve model (the model of Skaggs *et al.* (2001)) parameters. At the second step, the fractal model of Bird *et al.* (2000) was fitted on PSD curve data (that was extended at the first step) and fractal parameters were calculated. At the third step, basic soil properties (including the contents of sand, silt, and clay, OC and soil pH), PSD curve model parameters and fractal parameters were used to predict CEC using three PTFs for all data set. At the fourth step, all data sets were partitioned into eight groups or clusters and basic soil properties and fractal parameters were used to predict CEC using two PTFs for each group. In order to understand the method a flow chart visually shows the procedure (Fig. 1).

Firsts step: Complete PSD for the 20 size classes (0-2, 2-3, 3-5, 5-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-100, 100-150, 150-200, 200-300, 300-400, 400-600, 600-800,
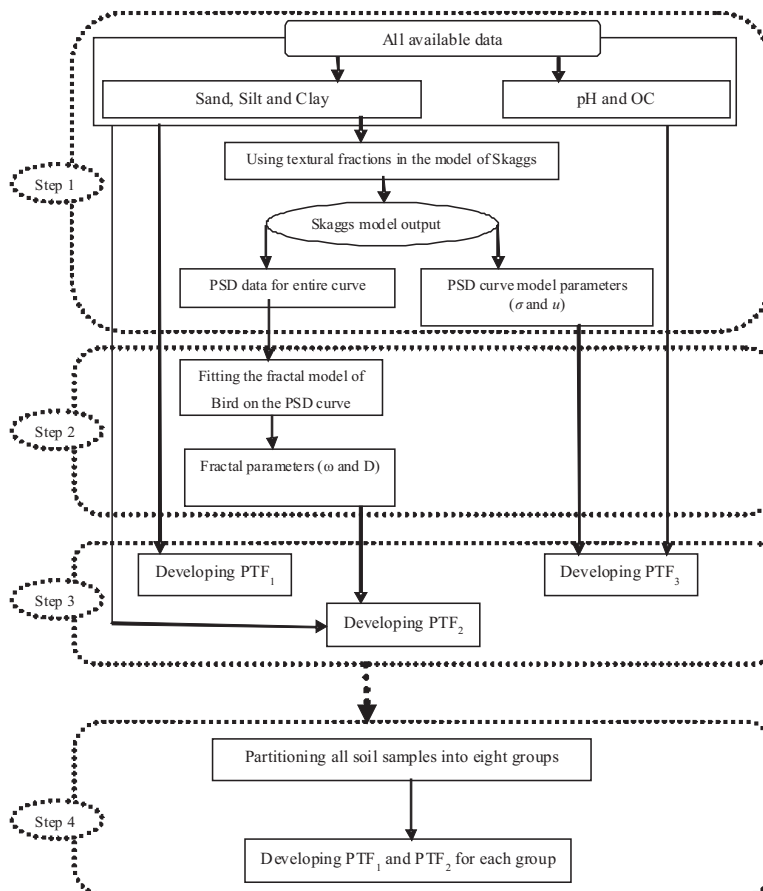
Fig. 1. Flow chart that visually shows the procedure of predicting CEC.

800-1 000, 1 000-1 300, 1 300-1 600 and 1 600-2 000 µm) were estimated from sand, silt and clay, using the Skaggs *et al.* (2001) model:

$$P(r) = \frac{1}{\left(1 + \frac{1}{P(r_0)} - 1\right) \exp\left(-uR_x^{\sigma}\right)}, \tag{1}$$

$$R_x = \frac{r - r_0}{r_0}, \quad r \geq r_0 > 0, \tag{2}$$

where: $P(r)$ is the mass fraction of soil particles with radii less than r, r_0 is the lower limit on radii for which the model applies, and σ and u are model parameters that can be calculated using the following equations:

$$\sigma = \alpha \ln \frac{v}{w}, \quad u = -v^{1-\beta} w^{\beta}, \tag{3}$$

$$v = \ln \frac{\frac{1}{P(r_1)} - 1}{\frac{1}{P(r_0)} - 1}, \quad w = \ln \frac{\frac{1}{P(r_2)} - 1}{\frac{1}{P(r_0)} - 1}, \tag{4}$$

$$\alpha = \frac{1}{\ln \frac{r_1 - r_0}{r_2 - r_0}}, \quad \beta = \alpha \ln \frac{r_1 - r_0}{r_0}, \tag{5}$$

$$1 > P(r_2) > P(r_1) > P(r_0) > 0, \quad r_2 > r_1 > r_0 > 0.$$

To implement the method described in the above section, we must select values for r_0, r_1, and r_2. In this study we used $r_0 = 1$ µm, $r_1 = 25$ µm, and $r_2 = 999$ µm. According to the USDA particle-size classification system, these radii specify that $P(r_0)$ is the clay mass fraction, $P(r_1)$ is the clay plus silt fraction, and $P(r_2)$ is the clay plus silt plus sand mass fraction. Assuming the linear distribution of sand particles, $P(r_2)$ was calculated using the clay plus silt mass fraction and the clay plus silt plus sand mass fraction. However, the assumption may not be correct, but 999 µm is too close to the upper limit on radii of sand fraction (1 000 µm), then this assumption will not introduce much error to the model.

Second step: In this study cumulative solid mass distribution of the Pore-Solid Fractal model was applied to PSD data (Bird *et al.*, 2000):

$$Ms(d \leq d_i) = \omega d_i^{3-D}, \tag{6}$$

where: $Ms(d \leq d_i)$ is the cumulative mass of particles below an upper limit d_i, D is the fractal dimension of the PSD, and ω is the composite scaling constant. The fractal parameters of D and ω were obtained by fitting the above model to the PSD data that was estimated using the Skaggs *et al.* (2001) method. Then the D and ω were used as predictors to estimate CEC. The coefficients of σ and u are the parameters of the PSD curve model as well as D and ω but the first two parameters are related to a non-fractal model and the later two parameters are related to a fractal model.

Third step: The fractal dimension D had non-normal distributions, then, it is normalized by 10^{D-2}. All variables (pH, OC, sand, silt, clay, ω and D, PSD curve model parameters (σ and u) and CEC) were standardized to have a zero mean and unit variance.

At this step, three PTFs were developed as follows. The first pedotransfer function (PTF$_1$) was based on the basic soil properties including the contents of sand, silt, and clay, OC and soil pH. The second pedotransfer function PTF$_2$ included fractal parameters D and ω in the Eq. (6) besides those variables in PTF$_1$ as inputs. In the third pedotransfer function PTF$_3$, PSD curve model parameters σ and u in the Eq. (1) were included besides the basic soil properties.

Fourth step: At the fourth step, all data sets were partitioned into more homogeneous soil groups or clusters to improve the accuracy and reliability of the CEC estimation. The grouping of the data sets was based on soil properties using cluster analysis by SPSS software (SPSS Inc, 1994). Ward clustering method with squared Euclidian distance was used to group all data into eight clusters in order to minimize within-cluster variances and squared Euclidian distances between groups. Then, for every data cluster, PTF$_1$ and PTF$_2$ were developed according to the third step. This is because the main objective of this study was to test the utility of fractal parameters in the prediction of CEC.

The 1 662 data were partitioned into two sets, using a randomised approach, a training set of 1100 data, and a testing set of 562 data and three PTFs (PTF$_1$, PTF$_2$ and PTF$_3$) were developed.

In order to develop the PTFs, the feed-forward multi-layer perceptrons artificial neural networks with one hidden layer were used. Different numbers of neurons in hidden layer, ranging from 3 to 6, were employed. The transfer function in the hidden neurons was tangent hyperbolic. The network with the optimized number of hidden neurons that had the highest accuracy and reliability was selected as the ultimate PTFs. The soil CEC was selected as an output variable.

The data sets in each cluster were split randomly in the training (about two thirds) and testing (about one-third) data sets. Then two PTFs (PTF$_1$ and PTF$_2$) were developed for each data cluster using artificial neural networks method.

The PTFs were developed by the combination of artificial neural networks and the bootstrap method (Efron and Tibshirani, 1993). The input data were selected randomly for 50 different times in order to obtain 50 bootstrap data sets of the same size as the training data set. For each bootstrap data set, a network was trained and CEC was estimated. The mean of all 50 predictions was assumed as the final estimate.

The sensitivity coefficient of the output variable CEC to a given input variable was approximated by allowing changes in the specified input variable within the range of mean

± standard deviation values, while keeping all the other input variables constant, and then dividing the resulting standard deviation of the output variable by the standard deviation of that specified input variable (NeuroSolutions, 2005). The relative importance of input variables was obtained by per- forming sensitivity analyses.

The three criteria to evaluate both the accuracy and reliability of the PTFs were Akaike information criterion (AIC), root mean square error (RMSE), and relative improvement (RI).

In order to examine the differences between various PTFs, the Morgan-Granger-Newbold (MGN) test was conducted (Diebold and Mariano, 2002).
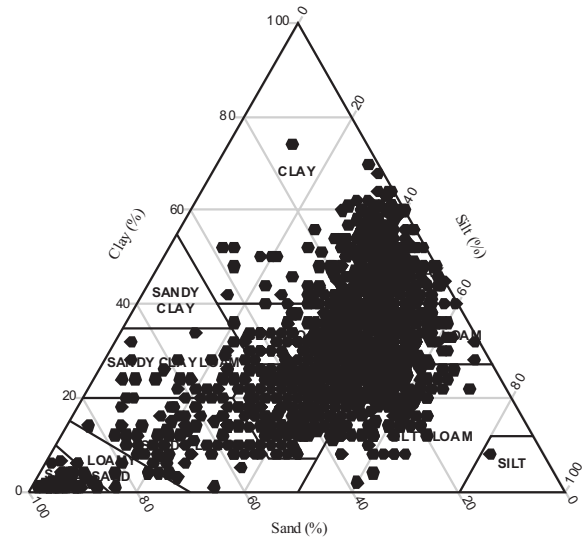
RESULTS AND DISCUSSION

The statistics of the soil properties, fractal parameters and PSD curve model parameters are shown in Table 1. The t-test indicated that the groups were not significantly different between development and validation data sets. The samples used in this study cover a wide range of the physical and chemical soil properties as well as fractal parameters and PSD curve model parameters. The textural range covered by the experimental data set is shown on the USDA texture triangle in Fig. 2. Predominant textural classes were silty clay, silty clay loam, clay loam, and silt loam (Fig. 2). Sand ranged from 0.1 to 98.0%, silt from 1.0 to 71.0% and clay from 1.0 to 74.0%. CEC of the samples showed a wide range from 2.3 to 58.8 cmol$_c$ kg^{-1}. The range of pH was from 3.50 to 8.80. Different values were related to the different land uses. The low pH was characteristic of tea fields and high level of that was related to olive gardens. Standard deviation of OC and sand were high. It seems that OC was affected by field management. Variable managements were applied on these fields. The low OC belonged to the beach sands and olive gardens and high level of that was related to paddy soils that are saturated the most.

Fig. 2. Textural distribution of both training and testing data sets on the USDA soil texture triangle.

Pearson correlations were performed to evaluate the relationships between CEC and input variables (Table 2). Preliminary evaluation of data regarding CEC showed that it was best correlated with fractal parameters ω (R = 0.46, p < 0.001) and D (R = 0.49, p < 0.001). A power function was used to describe the relationship between CEC and fractal parameters. In order to investigate the relation between CEC and fractal parameters different regression equations, including linear, exponential, logarithmic, polynomial and power, were performed (Eq. (13) and (14)). The results indicated that the power regression described the relations between CEC and fractal parameters better than other equations and a strong correlation was found. This is in agreement with Ersahin *et al.* (2006) who reported significant correlation between fractal dimension and CEC:

T a b l e 1. Statistics of the development and validation data sets

Data set	Parameter	pH	OC (%)	Sand (%)	Silt (%)	Clay (%)	ω (g µm^{D-3})	D	σ	u	CEC (cmol$_c$ kg^{-1})
Development set	Mean	6.96	2.13	25.0	42.1	32.9	16.2	2.84	0.409	0.631	26.5
(n = 1100)	SD	0.82	1.51	17.9	11.1	13.9	7.1	0.11	0.078	0.322	9.3
	Min	3.50	0.00	0.3	1.0	1.0	0.1	2.20	0.208	0.004	2.3
	Max	8.70	7.99	98.0	71.0	69.8	34.3	2.97	1.098	3.041	58.8
Validation set	Mean	6.94	2.13	26.1	41.4	32.6	16.0	2.83	0.413	0.625	26.2
(n = 562)	SD	0.82	1.45	19.5	11.8	14.6	7.5	0.11	0.084	0.318	9.1
	Min	3.50	0.00	0.1	3.0	1.0	0.1	2.21	0.194	0.015	5.0
	Max	8.80	8.11	96.0	66.0	74.0	34.6	2.97	0.920	3.513	51.0

OC – organic carbon, ω – constant in the fractal model, D – fractal dimension, σ and u – parameters of PSD curve model.

T a b l e 2. Correlation coefficients (R) between input and output variables

Variable	pH	OC	Sand	Silt	Clay	ω	D	σ	u
OC	0.13***	1.00							
Sand	0.18***	0.03	1.00						
Silt	-0.02	0.14***	-0.64***	1.00					
Clay	-0.23***	-0.15***	-0.79***	0.04	1.00				
ω	-0.22***	-0.14***	-0.90***	0.28***	0.95***	1.00			
D	-0.19***	0.00	-0.93***	0.48***	0.83***	0.86***	1.00		
σ	0.08**	-0.04	0.71***	-0.80***	-0.29***	-0.53***	-0.47***	1.00	
u	-0.07**	-0.03	-0.56***	0.64***	0.23***	0.49***	0.36***	-0.81***	1.00
CEC	-0.09***	0.43***	-0.45***	0.12***	0.48***	0.46***	0.49***	-0.20***	0.11***

Significant correlations at: **p<0.01, ***p<0.001. Other explanations as in Table 1.

T a b l e 3. Development and validation results of the PTFs for all data set

Step		AIC	RMSE	RI	MGN
Development	PTF_1	11 684	6.10		
	PTF_2	11 636	5.97	2.16	4.97
	PTF_3	11 662	6.04	1.02	2.30
Validation	PTF_1	5 540	5.82		
	PTF_2	5 530	5.77	0.86	1.24
	PTF_3	5 531	5.77	0.77	1.13

AIC – Akaike information criterion, RI – relative improvement, RMSE – root mean squared error, MGN – Morgan-Granger-Newbold. MGN value corresponding to significant difference (p < 0.05) for accuracy and reliability is 1.96. In the MGN test, PTF_2 and PTF_3 were compared with the PTF_1.

$$CEC = 0.039\, D^{6.173} \quad R = 0.604, \tag{7}$$

$$CEC = 10.61\, \omega^{0.322} \quad R = 0.604, \tag{8}$$

where: ω is the constant in the fractal model and D is the fractal dimension.

The correlations between CEC and PSD curve model parameters (R = 0.11 for u and R = -0.20 for σ) were lower than the correlation between CEC and fractal parameters but they were significant (p < 0.001).

The correlations between CEC and soil OC and between CEC and clay content were both positive and significant (p < 0.001). The same results have been reported by other authors (Amini *et al*., 2005). The correlation between the sand content and CEC was negative and significant (p < 0.001). This is in agreement with Amini *et al*. (2005). Correlations for silt and pH were weak and lower than 0.2, although highly significant (p < 0.001).

For incorporating fractal parameters and PSD curve model parameters, artificial neural networks were used to develop the PTFs. In fact, this is the third step of the study. According to the method described in the third step, three PTFs (PTF_1, PTF_2 and PTF_3) were developed for all data set and the results for the development and validation are summarized in Table 3. Also, the utility of introducing fractal parameters to the model to predict CEC for all data set is depicted in Fig. 3.

The prediction of soil CEC was improved by the use of fractal parameters and PSD curve model parameters as predictors (Table 3 and Fig. 3). The improvements were significant for the accuracy (development) of the PTFs. However, they were not significant for the reliability (validation) of the PTFs. PTF_2 that utilized fractal parameters produced better results than PTF_3 that utilized PSD curve model parameters. Among the three PTFs that were developed to predict soil CEC, PTF_2 was the most successful. This result confirmed our hypothesis that fractal models simulate the PSD better than the conventional models and CEC may be one of the properties that are controlled by the fractal behaviour of PSD. In fact the fractal theory could develop a more complete description of soil structure and processes that are impossible to describe by the conventional methods based on Euclidean geometry (Sokołowska *et al*., 2001).

Some authors suggested that the fractal dimension of the PSD is useful in quantifying the relationships between soil texture and related soil properties and processes (Hwanga Ii *et al*., 2002). Bayat *et al*. (2011) successfully used fractal parameters to estimate soil water retention curve by artificial neural networks and multi-objective group method of data handling. It may be reasonable to assume that one of the reasons for the insignificant effect of fractal parameters and PSD curve model parameters on the prediction of CEC is that improving PTF reliability may be an issue distinctly different from improving PTF accuracy (Pachepsky and Rawls, 1999). Another reason may be the
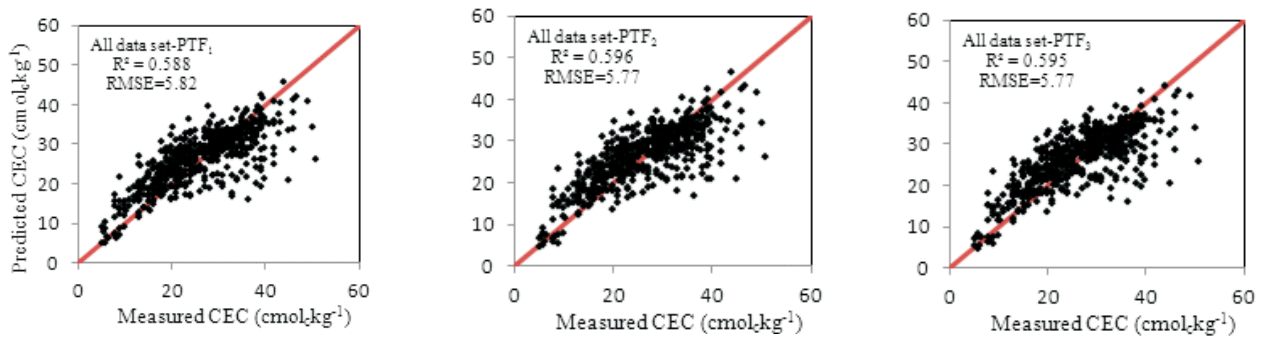
Fig. 3. Distribution of predicted versus measured CEC for PTF$_1$, PTF$_2$ and PTF$_3$ of all data set for the validation step.
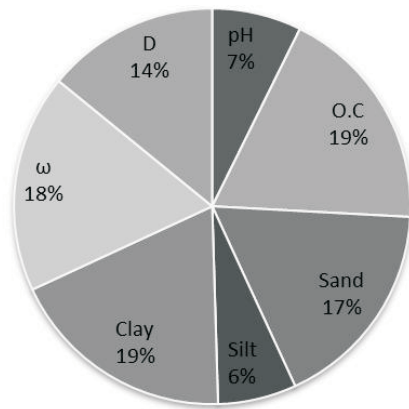
Fig. 4. Sensitivity analysis of output variable (CEC) to the input variables for PTF$_2$.

large and general data set used in this study. Nemes *et al.* (2003) found that PTFs from large and general data sets make large errors. They suggested that having a small set of relevant data is better than using a large but more general data set.

Sensitivity analysis was performed to investigate the contribution of fractal parameters in the prediction of CEC and the result is shown in Fig. 4. The fractal parameters (ω and D) explained 32% of the variation in CEC. The correlation between fractal parameters and soil properties will amplify the contribution of fractal parameters in a model with fractal parameters and soil properties as independent variables when compared with a model with only soil properties as independent variables. This result shows the utility of fractal parameters in describing the relationship between CEC and soil texture. In spite of calculating the fractal parameters from sand, silt and clay, they could significantly improve the CEC prediction. Thus, the results obtained in this study reflect that fractal parameters and PSD curve model parameters can be used as predictors for CEC prediction. Each of OC and clay content explained 19% of the variation in CEC for PTF$_2$. Contribution of organic mat-

ter (Peinemann *et al.*, 2000) and clay (MacDonald, 1998) in the soil CEC has been reported by some authors. However, the relationship between clay and CEC can be highly variable. Different clay minerals have different CECs and the relative proportion of pH-dependant and permanent charge varies among clay minerals. Soil pH and silt content had the least effect on the prediction of CEC. Krogh *et al.* (2000) proposed an equation based on the same inputs (clay, silt, OC and pH) which explained 90% of soil CEC variation.

Stratifying the data based on CEC or textural classes did not lead to any considerable improvement of accuracy or reliability of the PTFs. Data stratification by using cluster analysis with all the variables provided the best-fit models for the soil samples (Table 4).

Stratifying data into more homogeneous groups on the basis of all variables could improve the CEC predictions considerably. When soils with various geneses are included in the PTFs and variables such as mineralogical composition have not been controlled, soil properties become less predictive. In fact partitioning the data according to squared Euclidian distance of the variables between groups increased significantly the accuracy and reliability of the PTFs (Table 4). However, only the accuracy of the PTFs was improved in the study of Pachepsky and Rawls (1999) by grouping the data according to the taxonomic unit, soil moisture regime, soil temperature regime, and soil textural class in the prediction of soil water retention curve. The result of this study may show the superiority of classifying the data according to the squared Euclidian distance to the other classification criteria.

The potential importance of fractal parameters is well illustrated by using artificial neural networks in the prediction of soil CEC for the development and validation data sets in the PTFs after grouping all data (Table 4 and Fig. 5). In all 8 clusters (C1-C8) utilising fractal parameters, CEC predictions improved substantially in comparison to the PTFs without utilizing fractal parameters (Table 4 and Fig. 5). Using fractal parameters as inputs improved the accuracy of CEC predictions (except C5), from 4.39% for C6 to 58.45%

T a b l e 4. Development and validation results of the PTFs for the clusters

| Cluster number	PTF	n	RMSE	RI	MGN	n	RMSE	RI	MGN
			Development				Validation		
C1	PTF$_1$	200	4.95			140	10.80		
	PTF$_2$		4.36	11.95	3.240*		5.25	51.60	13.060*
C2	PTF$_1$	120	4.9			76	8.44		
	PTF$_2$		4.25	13.19	2.739*		5.38	36.20	7.059*
C3	PTF$_1$	90	3.33			34	6.95		
	PTF$_2$		2.62	21.31	3.562*		5.42	22.10	3.099*
C4	PTF$_1$	160	3.1			87	4.05		
	PTF$_2$		2.74	11.57	4.559*		3.75	7.29	2.215*
C5	PTF$_1$	90	1.64			39	4.68		
	PTF$_2$		1.86	-13.1	1.935		3.61	22.80	2.794*
C6	PTF$_1$	200	4.19			103	6.63		
	PTF$_2$		4.01	4.39	2.040*		6.58	0.68	0.920
C7	PTF$_1$	110	11.50			52	14.10		
	PTF$_2$		4.78	58.45	13.790*		6.29	55.50	8.675*
C8	PTF$_1$	110	5.32				7.68		
	PTF$_2$		4.06	23.69	4.985*	51	5.85	23.90	2.818*

n – number of samples in the development and validation steps of each cluster, C1-C8 – clusters 1-8. Other explanations as in Table 3. *Significant differences ($p < 0.05$) between each PTF and the previous one (*ie* PTF$_1$ with PTF$_2$).

for C7. The reliability of CEC predictions improved from 0.68% for C6 to 55.5% for C7. This large improvement in estimates of soil CEC was a benefit arising from the ability of the fractal theory to control the relationship between PSD and CEC.

The utility of introducing fractal parameters to the model to predict CEC for cluster 7 is depicted in Fig. 5. Introducing fractal parameters to the model increased the accuracy and reliability of PTF$_2$ in comparison with PTF$_1$ by decreasing the RMSE from 12.4 to 5.3 and increasing the R^2 from 0.071 to 0.364 and decreasing the distribution of measured versus predicted CEC. In fact the data was distributed away from the 1:1 line in the PTF$_1$, whereas, in PTF$_2$ the data was distributed around the 1:1 line (Fig. 5). Therefore, the fractal approach is a useful tool to describe the processes of porous medium (Hwanga Ii *et al.*, 2002).

The main advantage of using fractal parameters and PSD curve model parameters as predictors was the improvement of the CEC prediction without additional measurements. This result is in agreement with some authors (Giménez *et al.*, 1997) that used fractal dimension in characterizing the porous medium and improving the accuracy and reliability of the models. In the same way, Wu *et al.* (2003) used some easily measurable data to predict soil CEC. Then they utilized the predicted CEC for estimating and mapping of soil copper content by kriging.

In this study the artificial neural networks was run at least 15 times and the PTF with the best performance was selected. The RMSE values obtained for PTF$_2$ of the clusters ranged from 1.86 to 6.58 (cmol$_c$ kg^{-1}) for training and validation. The authors are well aware that the accuracy (based on the RMSE of training) and reliability (based on the RMSE of validation) of the developed PTFs are not very high but they are comparable with the RMSE and coefficient of determination (R^2) values reported in the literature. For example, Manrique *et al.* (1991) used a lot of soil physical and chemical properties such as clay content, organic carbon, sum of exchangeable bases, 1 M KC1 extractable Al, clay plus silt content, pH in soil/water ratio 1:1, dithionite-citrate extractable Al, and Al saturation to predict CEC and reported R^2 values of 0.38-0.93. Using more input variables may increase the accuracy and reliability but their measurement would be more costly and time consuming. In this regard only more readily available data may be used to estimate CEC in our study. Amini *et al.* (2005) tested several published PTFs and developed some PTFs to predict CEC by artificial neural networks with the RMSE values of 2.54-3.97. Ersahin *et al.* (2006) used the textural fractions to predict CEC by the regression method and reported R^2 values of 0.22-0.48 while with employing fractal dimension
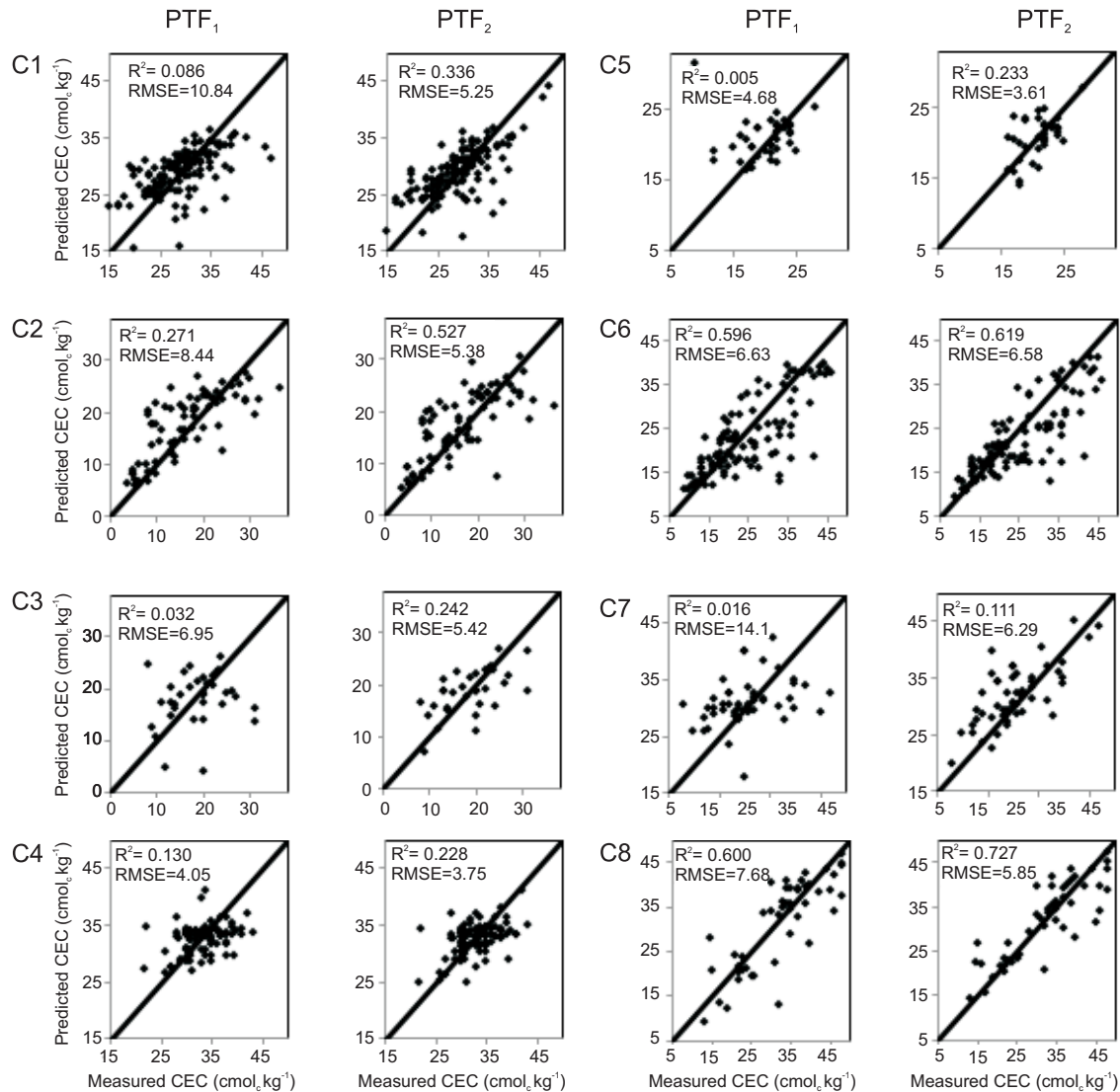
Fig. 5. Distribution of predicted versus measured CEC for PTF_1 and PTF_2 of C1-C8 clusters for the validation step.

R^2 increased to 0.74. Seilsepour and Rashidi (2008), by using the regression method, depending on the variables, reported R^2 values of 0.01, 0.19, 0.21, 0.26 and 0.74. Therefore, the PTF_2 of the clusters developed in the present study may be recommended to be used to predict CEC with good reliability.

On the other hand, according to Davatgar *et al*. (2006), smectite is the dominant clay mineral in most of the soils of the Guilin province, north of Iran. The PTFs developed in this study are also related to smectitic soils that have quite different behaviours compared to soils with other dominant clay minerals, due to large surface area, high cation exchange capacities, chemically active surfaces, reversible interlayer expansibility (Laird *et al*., 1992), high potential of swelling and shrinkage with minimal horizon differentiation due to pedoturbation (Ahmad, 1983), wide and deep cracks

as deep as 50 cm and at least 1 cm wide (Staff, 1975). In spite of numerous PTFs having been offered to predict CEC in different soils, fewer PTFs have been introduced to predict CEC in smectitic soils.

Smectitic soils are very difficult to manage (Millan *et al*., 2002) and the predictability of hard-to-measure soil properties from readily available properties may be less promising. Therefore, developing the PTFs for these soils that could predict CEC only by readily available soil properties that have comparable accuracy with the PTFs reported for the other soils in the literature would be a great step forward.

Moreover, there are two key points in this study. First, the significant improvement of the CEC predictions by PTF_2 (developed simply by using PSD fractal parameters) for most of the clusters in comparison to PTF_1. As a matter of fact, the significant improvement was obtained only by

T a b l e 5. Sensitivity analysis of output variable (CEC) to the input variables for the PTF$_2$ developed for the eight clusters

Parameters	Relative importance							
	Cluster number							
	C1	C2	C3	C4	C5	C6	C7	C8
pH	7.4*	2.3	12.1	24.2	18.4	6.1	6.2	11.5
OC	16.4	6.5	11.2	1.8	8.9	14.2	21.4	22.1
Sand	5.9	17.0	15.0	13.5	9.8	19.9	5.1	13.1
Silt	6.0	13.6	8.8	5.7	2.7	2.8	5.6	4.8
Clay	26.2	5.7	11.7	7.2	13.8	20.6	3.0	18.3
ω	28.8	31.2	22.7	26.4	9.4	23.8	34.9	14.7
10^{D-2}	9.4	23.7	18.5	21.1	36.9	12.7	23.8	15.5

Explanations as in Tables 1 and 4. *Values are in percent.

calculating fractal parameters and using them as predictors, without spending more cost. It clearly demonstrates the utility of fractal theory in describing PSD of Vertisols as has been reported by Millán and Orellana (2001). Second, the effect of clustering in the improvements of the PTFs that was considerable.

Sensitivity analysis was performed to investigate the contribution of fractal parameters in the prediction of CEC in eight clusters (Table 5). In the PTFs (PTF$_{1-2}$) developed for eight clusters (C1-C8), fractal parameters explained between 30.2 and 58.7% of the variation in CEC, with C8 and C7 having the lowest and highest incorporation, respectively. Of the eight PTF$_2$s developed for the clusters, ω had the highest contribution in six clusters (C1, C2, C3, C4, C6 and C7) and in C5 the fractal dimension (D) had the highest incorporation. Only in one cluster (C8), a non-fractal parameter (clay) had the highest contribution in predicting CEC.

Amini e*t al*. (2005) developed several PTFs to estimate CEC by artificial neural networks and multiple regressions. They explained 70% of the variation in CEC using soil clay and organic matter content. They concluded that adding sand and silt did not significantly improve the accuracy of the PTFs. However, there are several factors that could improve the CEC prediction such as the type and the morphology of clay minerals, and the origin of soil organic matter (Stewart and Hossner, 2001) but, they are costly and time consuming. Therefore, finding variables such as fractal parameters that can improve the CEC prediction would be very useful. Our result showed the importance of fractal parameters in the prediction of CEC. Since in this study fractal parameters were calculated only from sand, silt and clay, therefore it could be a great step forward to improve the CEC prediction without additional measurements. However, similar to other authors (Wosten *et al.*, 2001), the PTFs developed in this study can be recommended for application only in the same pedo-environmental conditions within input variable range.

CONCLUSIONS

1. Correlations were found between cation exchange capacity and fractal parameters and particle size distribution curve model parameters. These correlations could be stronger by using non-linear regression equations. As a result, using fractal parameters and particle size distribution curve model parameters in predicting cation exchange capacity by artificial neural networks resulted in improvements, and thus, they successfully could be used as predictors for cation exchange capacity prediction.

2. The artificial neural network models that use fractal parameters as predictors had better performance than the pedotransfer functions that use particle size distribution curve model parameters.

3. Partitioning the data increased significantly the accuracy and reliability of the artificial neural network models. Utilizing fractal parameters in the clusters improved the cation exchange capacity prediction substantially.

4. The results of the study also indicated that fractal parameters are the most important factors which affect cation exchange capacity prediction.

5. Pedotransfer functions that utilize fractal parameters and particle size distribution curve model parameters are interesting and flexible approaches for the prediction of cation exchange capacity by artificial neural networks. Therefore, these artificial neural network models provide an easy, economic and brief methodology to improve the estimation of soil cation exchange capacity.

REFERENCES

Ahmad N., 1983. Vertisols. In: Pedogenesis and Soil Taxonomy. The Soil Orders (Eds L.P. Wilding, N.E. Smeck, G.F. Hall), Elsevier, New York, USA.

Amini M., Abbaspour K.C., Khademi H., Fathianpour N., Afyuni M., and Schulin R., 2005. Neural network models to predict cation exchange capacity in arid regions of Iran. European J. Soil Sci., 56(4), 551-559.

Bayat H., Neyshabouri M.R., Mohammadi K., and Nariman-Zadeh N., 2011. Estimating water retention with pedotransfer functions using multi-objective group method of data handling and ANNs. Pedosphere, 21(1), 107-114.

Bird N.R.A., Perrier E., and Rieu M., 2000. The water retention function for a model of soil structure with pore and solid fractal distributions. European J. Soil Sci., 51(1), 55-63.

Bishop T.F.A. and McBratney A.B., 2001. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma, 103(1-2), 149-160.

Davatgar N., Kavoosi M., Alinia M.H., and Paykan M., 2006. Study of potassiun status and effect of physical and chemical properties of soil on it in paddy soils of guilan province. JWSS – Isfahan Univ. Technol., 9(4), 71-89.

Diebold F.X. and Mariano R.S., 2002. Comparing predictive accuracy. J. Business Economic Statistics, 20(1), 134-144.

Efron B. and Tibshirani R.J., 1993. An introduction to the bootstrap. Monographs on Statistics and Applied Probability, Chapman and Hall Press, London, UK.

Ersahin S., Gunal H., Kutlu T., Yetgin B., and Coban S., 2006. Estimating specific surface area and cation exchange capacity in soils using fractal dimension of particle-size distribution. Geoderma, 136(3-4), 588-597.

Gee G.W. and Or D., 2002. Particle-size analysis. In: Methods of Soil Analysis. Physical methods (Ed. A.D. Warren). SSSA, Madison, WI, USA.

Giménez D., Perfect E., Rawls W.J., and Pachepsky Y., 1997. Fractal models for predicting soil hydraulic properties: A review. Eng. Geology, 48(3-4), 161-183.

Hwanga Ii S., Leeb K.P., Leeb D.S., and Powers S.E., 2002. Models for estimating soil particle-size distributions. Soil Sci. Soc. America J., 66(4), 1143-1150.

Krogh L., Breuning-Madsen H., and Greve M.H., 2000. Cation-exchange capacity pedotransfer functions for Danish soils. Acta Agriculturae Scandinavica, Soil Plant Sci., 50(1), 1-12.

Laird D.A., Barriuso E., Dowdy R.H., and Koskinen W.C., 1992. Adsorption of atrazine on smectites. Soil Sci. Soc. America J., 56(1), 62-67.

MacDonald K., 1998. Development of pedotransfer functions of southern Ontario soils (Eds Y. Pachepsky, W.J. Rawls). Harrow, Ontario, Canada.

Manrique L.A., Jones C.A., and Dyke P.T., 1991. Predicting cation-exchange capacity from soil physical and chemical properties. Soil Sci. Soc. America J., 55(3), 787-794.

Millan H., González-Posada M., and Benito R.M., 2002. Fragmentation fractal dimensions of Vertisol samples: Influence of sieving time and soil pretreatment. Geoderma, 109(1-2), 75-83.

Millán H. and Orellana R., 2001. Mass fractal dimensions of soil aggregates from different depths of a compacted Vertisol. Geoderma, 101(3-4), 65-76.

Nelson D.W. and Sommers L.P., 1986. Total carbon, organic carbon and organic matter. In: Methods of Soil Analysis. Agronomy Handbook American Society of Agronomy and Soil Science Society of America (Ed. A.L. Page). Madison, WI, USA.

Nemes A., Schaap M.G., and Wösten J.H.M., 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. Soil Sci. Soc. America J., 67(4), 1093-1102.

NeuroSolutions, **2005**. Getting Started Manual (www.nd.com). NeuroDimension Press, Gainesville, FL, USA.

Pachepsky Y.A. and Rawls W.J., 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. Soil Sci. Soc. America J., 63(6), 1748-1757.

Peinemann N., Amiotti N.M., Zalba P., and Villamil M.B., 2000. Effect of clay minerals and organic matter on the cation exchange capacity of silt fractions. J. Plant Nutrition Soil Sci., 163(1), 47-52.

Salehi M.H., Mohajer R., and Beigie H., 2008. Developing soil cation exchange capacity pedotransfer functions using regression and neural networks and the effect of soil partitioning on the accuracy and precision of estimation. Int. Meeting Soil Fertility Land Management and Agroclimatology, October 29 – November 1, Kuşadasi, Turkey.

Seilsepour M. and Rashidi M., 2008. Modeling of soil cation exchange capacity based on soil colloidal matrix. American-Eurasian J. Agric. Environ. Sci., 3(3), 365-369.

Seybold C.A., Grossman R.B., and Reinsch T.G., 2005. Predicting cation exchange capacity for soil survey using linear models. Soil Sci. Soc. America J., 69(3), 856-863.

Skaggs T.H., Arya L.M., Shouse P.J., and Mohanty B.P., 2001. Estimating particle-size distribution from limited soil texture data. Soil Sci. Soc. America J., 65(4), 1038-1044.

Soil Survey Division Staff, **1993**. Soil survey manual. Soil Conservation Service, Washington, DC, USA.

Sokołowska Z., Hajnos M., Hoffmann C., Renger M., and Sokołowski S., 2001. Comparison of fractal dimensions of soils estimated from adsorption isotherms, mercury intrusion, and particle size distribution. J. Plant Nutrition Soil Sci., 164(5), 591-599.

SPSS Inc, **1994**. SPSS Professional Statistics. SPSS Press, Chicago, USA.

Staff S.S., 1975. Soil taxonomy. A basic system of soil classification for making and interpreting soil surveys. U.S. Department of Agriculture handbook, Government Printing Office, Washington, DC, USA.

Stewart W.M. and Hossner L.R., 2001. Factors affecting the ratio of cation exchange capacity to clay content in lignite overburden. J. Environ. Quality, 30(4), 1143-1149.

Wösten J.H.M., Pachepsky Y.A., and Rawls W.J., 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol., 251, 123-150.

Wu J., Norvell W.A., Hopkins D.G., Smith D.B., Ulmer M.G., and Welch R.M., 2003. Improved prediction and mapping of soil copper by kriging with auxiliary data for cation-exchange capacity. Soil Sci. Soc. America J., 67(3), 919-927.