DESERT

Desert
Online at http://desert.ut.ac.ir

Desert 25-2 (2020) 227-238

Determining the features influencing physical quality of calcareous soils in a semiarid region of Iran using a hybrid ACO-ANN algorithm

H. Shekofteh[a*], H. fatehi Marj[b]

[a] *Soil Science Department, University of Jiroft, Jiroft, Kerman, Iran*
[b] *Department of Electrical Engineering, Vali-e-Asr University of Rafsanjan, Kerman, Iran*

Received: 1 January 2020; Received in revised form: 21 May 2020; Accepted: 15 August 2020

Abstract

Soil quality indicators are measurable characteristics of the soil affecting the soil capacity for crop production or environmental performance. Among these indicators, air capacity (AC) and relative field capacity (RFC) are believed to be the most important ones. To select the best combination that affects soil physical quality indicators (AC and RFC), we employed a hybrid algorithm: an ant colony organization (ACO) in combination with an artificial neural network (ANN). Multiple linear regression and support vector regression models were constructed for the comparison of performances. The results obtained from running ACO-ANN to select the best combination revealed that a combination with four input variables, including soil organic matter, clay, carbonate calcium equivalent, and bulk density, had the lowest error. The R^2 values in the ACO-ANN model for the AC and RFC predictions were respectively 0.91 and 0.95 whereas they were 0.75 and 0.53 respectively in support vector regression model, and 0.54 and 0.53 in the multiple linear regression model. Since the results obtained from the ACO-ANN algorithm are acceptable, this algorithm could be applied to other locations of the world in order to tackle environmental problems. The results form sensitivity analysis for the ANN model showed that carbonate calcium equivalent and clay content had the highest and the lowest effects on AC and RFC indicators, respectively.

Keywords: Soil physical quality indicators, Feature selection, Modeling, Artificial neural network, Ant colony optimization

1. Introduction

Researchers insist that the soil physical environment needs to be improved for better plant growth (da Silva and Kay, 2004), chemical (Drury *et al*., 2003) and biological (Allmaras *et al*., 2003) conditions of soil . To evaluate the physical quality of soil, its indicators were introduced in their “optimal” or “ideal” range; in other words, the maxim crop yield and the minim soil degradation were attained (Reynolds *et al*., 2009). Once soil physical indicators (SPQIs) are not in an optimal range, the following symptoms occur in soils: poor aeration, poor water infiltration, surface run-off, hard-setting, crust formation on the surface, and poor rootability. To date, there has not been a unique measure of soil physical quality, according to several researchers (for example Dexter, 2004; Reynolds *et al*., 2009). Hence, the measurement and integration of a range of properties is suggested to effectively combine various information for a multi-objective decision. Two top physical quality indicators for agricultural soils are air capacity (AC) and relative field capacity (RFC). They represent the capacity of soil to store and provide essential water, air, and nutrients for crops (Topp *et al*., 1997; Reynolds *et al*., 2007; Moncada *et al*., 2014). Moreover, they determine soil pore space volume and pore size distribution. As an example, AC, and RFC are directly dependent on soil porosity, water release properties, and soil aeration (Reynolds *et al*., 2002; Dexter, 2004). The mentioned SPQIs indicate the soil dynamic properties that change with time and

* Corresponding author. Tel.: +98 913 3481945
Fax: +98 34 43347065
E-mail address: h.shekofteh@ujiroft.ac.ir

management systems. Therefore, such findings might add to our understanding of the linkage between soil physical quality, crop productivity, environmental impact, and water and solute dynamics within the soil profile. In order to evaluate the temporal changes in soil physical quality, sustainability of agricultural, land management practices, certain methods have been applied. These methods help to measure single indicators and minimum data sets (da Silva *et al*., 1997; Dexter, 2004), and calculate indices among indicator groups as well (Karlen and Stott, 1994; Andrews *et al*., 2004). Real application of these tests beyond research must be cost-effective. Currently, for the assessment of integrated soil quality, there are very few inexpensive experimental methods which could be widely applied by governments, farmers, and consultants. Most of these methods are laborious, time-consuming, and not easy to standardize. The present research was conducted to carefully investigate and predict soil physical quality and its indicators utilizing easily measurable properties sensitive to management practices (Brejda *et al*., 2000). There is a variety of economical approaches or methods that help functions to translate data into SPQIs predictions. In this research, we utilized the feature selection approach. Feature selection (FS) is usually applied to machine learning with high dimensional datasets, for selecting the best subset of features. For this purpose, redundant features, which are significantly correlated and have no predictive information, are removed (Vieira et al., 2010). The large amounts of input data is a challenge to regression and classification analysis methods. As an example, when a large number of input features are used to create a PTF and predict SPQIs, the estimation of a large number of parameters within the regression process and, therefore, the measurement of further data might be required. Ideally, an independent set of information should be added for each feature in the regression process. However, once these features are significantly correlated, there is some redundancy in the available information; this redundancy may reduce the accuracy of the regression (Pal and Foody, 2010). Lately, meta-heuristic algorithms have been inspired by the nature and used for feature selection in soil science; an example could be particle swarm optimization (PSO) which helps to select the most important features in soil quality indices (Shirani *et al*., 2015), and ant colony organization optimization (ACO) to select the effective soil CEC properties (Shekofteh et al., 2017).

The current work aimed to: examine an advanced ACO-ANN combination in order to select the best effective subset on SPQIs, and model this subset using the ANN, SVR-GA (Genetic algorithm) combination and multiple linear regression (MLR) approaches, and ultimately, to compare the three obtained results.

2. Materials and Methods

2.1. Study area

Rabor region (29° 27′ to 38° 54′ N and 56° 45′ E to 57° 16′ E) located in southwestern Kerman province, southeast Iran, was studied as the study site (Fig. 1). Herein, the required data for selecting the best subset affecting SPQIs were collected from typical semiarid farm lands with a cold temperate climate. The region receives an average annual precipitation of 250 mm and the mean annual temperature is 15 °C.

2.2. Soil sampling and measurement

In total, 104 soil samples were collected from the topsoil (0-15 cm depth) representing four different land uses. The samples were brought to the laboratory where they were air-dried and grounded to pass a 2 mm sieve. The content of soil organic matter (SOM) was determined using the Walkley–Black method with dichromate extraction and titrimetric quantization (Nelson and Sommers, 1982). We also measured the weight percentages of clay (>0.002 mm), silt (0.002–0.05 mm), and sand (0.05–2 mm) fractions employing the sieving and sedimentation method of Gee and Bauder (1986). The particle density (PD) and calcium carbonate equivalent (CCE) were determined via Blake and Hartge (1986) and the back-titration methods, respectively (Nelson, 1982). Utilizing a digital pH-meter (Model 691, M0065 trohm AG Herisau, Switzerland), soil pH was measured in a saturated paste (Thomas, 1996). We determined electrical conductivity (ECe) in the same extract with an electrical conductivity meter (Model Ohm-644, Metrohm AG Herisau, Switzerland)(Rhoades, 1996).

2.3. Soil physical quality indicators

Several soil physical quality indicators were used in this study and are described briefly in the following. For further details, with respect to their optimal ranges or critical limits, several measures could be taken (for instance Reynolds *et al*., 2002; Dexter, 2004; Reynolds *et al*., 2008).

2.3.1. Air capacity

Air capacity, AC (m^3m^{-3}), is often defined as (White, 2006):

$$AC = \theta_S(\Psi = 0) - \theta_{FC}(\Psi = -1m); 0 \leq AC \leq \theta_S (1)$$

where θ_S ($m^3\ m^{-3}$) is the saturated soil water content, θ_{FC} ($m^3\ m^{-3}$) is the water content at field capacity defined as the water content at a matric potential (Ψ) of –100 cm.

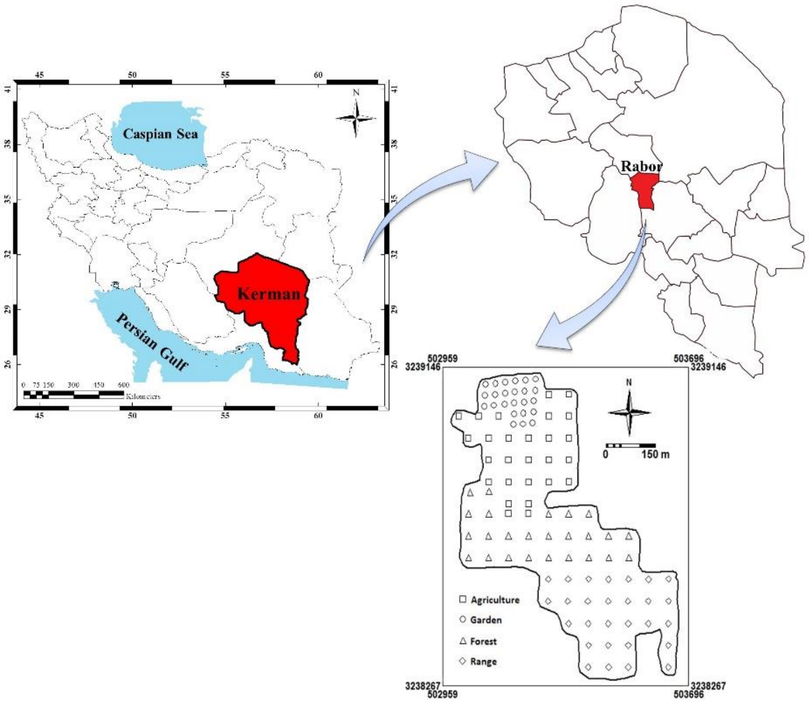


Fig. 1. Location of the study area along with sampling points in different land uses

2.3.2. Relative field capacity

Relative field capacity, RFC (dimensionless), is defined by Reynolds *et al*. (2008) as:

$$RFC = \left(\frac{\theta_{FC}}{\theta_S}\right) = \left[1 - \left(\frac{AC}{\theta_S}\right)\right]; 0 \leq RFC \leq 1 \qquad (2)$$

RFC indicates the ability of soil to store water and air relative to the total pore volume of the soil (θ_S). The optimal balance between root-zone soil water capacity and soil air capacity for a rain-fed agricultural soil occurs for a value of RFC between 0.6 and 0.7.

2.4. Modeling approaches

2.4.1. ACO-ANN

The design of the ACO approach is similar to the one by Shekofteh *et al*. (2017).

2.4.2. Artificial neural network

Artificial neural network (ANN) is a nonlinear regularization technique used for modeling complex relationships between inputs and outputs. More information about ANN can be found in the study by Schaap *et al*. (1998). A Feed forward back-propagation (BP) neural network was utilized in this study. The BP-based ANN structure includes three layers: input, hidden, and output layers. The number of hidden neurons determines the complexity of the network. In this paper, ANN models with various numbers of hidden neurons were tested so that we could find the best value with a perfect efficiency in both calibration and testing phases. Finally, a hidden layer with 10 neurons was selected. The logarithmic sigmoid (log sig) and linear transfer functions (purelin) were referred to the activation function in the hidden and output layers, respectively.

2.4.3. Support vector machine and genetic algorithm

Support vector machine (SVM), initially introduced by Boser *et al*. (1992), is an approach to solving the classification and regression problems. Regarding support vector regression

(SVR), a mapping is at first performed from an input space onto a highly-dimensional variable one. Secondly, a linear regression is implemented by a hyper plane in the latter space by ε-insensitive loss. Applying a kernel function, SVR helps hyper plane surface fit to training data. For more accuracy of SVR prediction, setting of kernel parameters is of paramount importance; see Vapnik (1995) for more details. This study made use of radial basis function (RBF) kernel. Through the ACO-ANN algorithm, input variables were selected and used for the modeling and prediction of AC and RFC indicators.

SVR calculations were done with MATLAB software. SVR parameters known as the penalty parameter C, width parameter γ, for the RBF kernel, and variable ε are important factors in SVR training (Wohlberg *et al*., 2006). These parameters were attained and tuned with genetic algorithms (GA), which works on the basis of direct analogy to Darwinian natural selection and genetics in biological systems. This research applied real-valued GAs (RGAs). For using GA, the initial step is the determination of the objective function whose value for each individual is normally a measure of the individual's fitness. Herein, relative mean absolute percentage error (RMAPE) was considered as the main objective to avoid the variable scale. The objective and fitness functions are defined as below:

Objective function =

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|Y(p_i)-Y(o_i)|}{Y(o_i)}\right)\times 100 \qquad (3)$$

Fitness function=100- MAPE (4)

where *Y (pi)* and *Y(oi)* represent SPQIs observed and estimated values, respectively; and *n* (104) is the number of data.

In a GA, a population of points (solutions) is randomly generated for the first time. Fitness is computed for every chromosome in the population. In this research, we utilized a fitness-proportionate method, known as roulette wheel selection (Goldberg, 1989), to select individuals for reproduction based on their fitness values. Once the parent is selected, genetic operation of crossover, referred to as crossover probability, is implemented on each mated pair with a defined probability. Common crossover operations could be uniform, single-point, two-points, or arithmetic (Michalewicz, 1994). The arithmetic crossover for an RGA is simple and effective. In this study, an arithmetic crossover was selected and designed for the crossover operation. Afterwards, a mutation operation was utilized. A Gaussian mutation was then selected and designed to the latter. After the creation of next generation (offspring), stopping criteria was checked. The algorithm was repeated until a certain termination criterion was met; for example, a limit in the maximum number of generations or no obvious change in the fitness or preset one. Herein, for finding the best parameters, different values were examined for GA.

2.5. Evaluation criteria

The predictive capabilities of the proposed models were evaluated via the following equations: mean absolute percentage error (MAPE), root mean square error (RMSE), and coefficient of determination (R^2) between the observed and estimated values. MAPE, RMSE and R^2 are defined as follows:

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|Y(p_i)-Y(o_i)|}{Y(o_i)}\right)\times 100 \qquad (5)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[Y(pi) - Y(oi)]^2} \qquad (6)$$

$$R^2 = \left\{\frac{\sum_{i=1}^{n}(Y(p_i)-\overline{Yp})(Y(o_i)-\overline{Yo})}{\sqrt{\sum_{i=1}^{n}(Y(p_i)-\overline{Yp})^2(Y(o_i)-\overline{Yo})^2}}\right\}^2 \qquad (7)$$

in which *Y(pi)* and *Y(oi)* represent measured and predicted SPQIs values, respectively; $\overline{Yp}$ and $\overline{Yo}$ are the means for measured and predicted SPQIs values, and *n* is equal to the total number of observations.

3. Results and Discussion

3.1. Statistical analysis of data

Table 1 illustrates some statistic properties of soil variables, which were used to select the best input variables for predicting AC and RFC indicators.

Table 1 gives a summary of descriptive statistics for soil physical and chemical properties that help select the best subset of features in SPQIs. The SOM content varied between 0.23 and 7.93% with an average value of 2.2%. Clay content ranged between 5.5 and 23.5% with an average of 12.61%. In general, USDA soil texture of the studied area was classified as sandy loam. The CCE content varied between 19.99 and 47.28% with 19.1% as the mean value. Totally, under an arid climate, the examined soil was calcareous. Soil pH varied between 6.74 and 7.99 with an average value of 7.75. Among all the measured variables, the lowest coefficient of variation belonged to soil pH due to the buffering capacity related to the high content of calcium carbonate. The maximum, minimum and mean soil bulk densities were1.66, 0.96 and 1.27 g cm^{-3}, respectively. The highest, lowest and mean soil particle densities were respectively 2.74, 1.98 and 2.37 g cm^{-3}. Soil EC ranged from 0.325 to 2.21 dS m^{-1}.

3.2. Determining effective input variables for predicting AC and RFC using ACO-ANN

The results obtained from running ACO-ANN algorithm for selecting effective input variables in AC index demonstrated that the lowest RMSE (0.0195) belonged to a four-variable combination (SOM, clay, BD, and CCE properties) (Fig. 2).

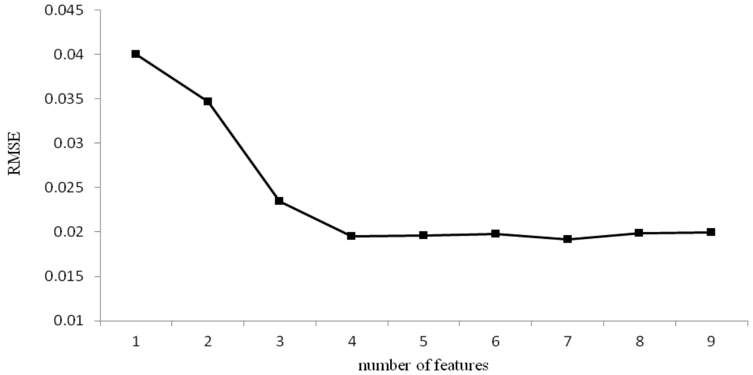


Fig. 2. Obtained RMSE values in selecting feature numbers using ACO-ANN hybrid algorithm for prediction of AC parameter

Table 1. Basic statistics of physiochemical properties of soils under studied area

	Maximum	Minimum	Mean	Median	CV (%)
EC (dS m^{-1})	2.21	0.32	0.69	0.60	44.94
pH	7.99	6.74	7.75	7.80	2.78
Clay (%)	23.50	5.50	12.61	12	31.50
Silt (%)	43.50	9.50	31.47	31.50	15.81
Sand (%)	85.00	36.50	55.91	56.50	13.54
SOM (%)	7.93	0.23	2.20	1.39	82.39
Porosity	0.57	0.32	0.46	0.46	10.77
BD (g cm^{-3})	1.66	0.96	1.27	1.27	10.48
PD (g cm^{-3})	2.74	1.98	2.37	2.38	5.77
CCE (%)	47.28	19.99	31.10	30.41	4.77

EC: Electrical Conductivity, BD: Bulk density, PD: Particle density, CCE: Carbonate Calcium Equivalent; SOM: Soil Organic Matter, CEC: Cation Exchange Capacity, CV: Coefficient of Variation

One of the most influential variables in AC indicator was SOM which enhanced aggregates and pores stability, increased soil porosity, improved water infiltration and soil aeration as well as water holding capacity and oxygen holding capacity. Therefore, SOM can influence AC indicator which depends on water drained from soil due to gravity, followed by air penetration into the soil instead of water depletion. Similar results by Hati *et al.* (2006) confirm SOM effects on water retention and flow in soils.

Among the primary particles, sand fraction was dominant in the region. Clay particles can bind to sand grains and create aggregate and soil structure owing to their strong adhesion and high plasticity. Pores are large between aggregates, yet small within them. Thus, clay particles are capable of changing soil pore size distribution, which in turn influences water flow and air flow in the AC. In addition, clay particles have high surface areas that absorb more water; this has a considerable effect on soil water retention capacity, saturated soil water content, gravity drained water content, and as a consequence, on the AC parameter.

BD is associated with soil porosity and saturation soil water content. As such, soil porosity decreased in line with the increase in BD. Furthermore, such transportations occurring in soil as water flow and air flow are influenced by BD. Consequently, BD reflects soil ability of functioning for structural properties, water and solute movement, and soil aeration.

Another feature influencing AC index was observed to be CCE. With respect to high percentage of CCE in the region (due to its existence in the semi-arid location and non-leaching), this cementing agent causes soil particles flocculation, aggregate formation, and improvement in soil structure. Regarding CCE effect on soil structure, carbonates act as a source of Ca^{+2} and help flocculation of clay particles. The creation of soil structure increases macro-pores and facilitates water flow. In general, water flow thorough soil mostly depends on micropores and air flow mainly occurring in macropores.

Running ACO-ANN algorithm for selecting the best effective combination and predicting RFC indicator led to different results. A combination with nine variables (pH, EC, silt, clay, sand, CCE, BD, PD, and SOM) was selected as the best for predicting RFC indicator (RMSE = 0.053, Fig. 3). However, a combination with the four soil variables of SOM, clay, CCE, and BD had an RMSE value close to the already mentioned nine variables (RMSE = 0.056). Therefore, considering the spent time and expenses and laboratory effort, the four-variable combination was selected for building RFC model.

The soil pH, sand, EC, silt, and PD contents were considered redundant among the input variables for both AC and RFC modeling. For this reason, they were not used in the datasets for modeling. Soil pH was removed from the input variables due to its narrow range in the region (Table 1). Soil pH has a considerable effect on chemical properties. In order to provide a significant effect on soil physical quality, it should be in an extensive range. Accordingly, within its narrow range, its values were not likely sufficient (see Table 1) to have a significant effect on AC and RFC indicators. Another redundant variable was sand fraction in view of its relation with clay particles. Sand particles did not exhibit any significant roles in the formation of soil structure due to low specific surface area and charge. Furthermore, silt content was not distinguished as a significant variable of SPQIs with the ACO-ANN algorithm since its effects can be explained by clay particles. Soil EC had no effects on AC and RFC indicators because of low salinity and range in the region.

PD shows the solid phase of soil, which has an inverse relationship with pore volume. It could be obtained from SOM and soil mineral matter. To reach this correlation, it was omitted and not considered as an influential property in SPQIs.

3.3. SPQIs modeling after selecting proper features

It should be noted that our train and test data were the same for all the models (ANN, SVR and MLR). The MAPE and RMSE amounts between the ANN model and observed AC indicator for the train data were 5.94 % and 0.0107, respectively. Drawing on ANN results for the train data, the R^2 value between the observed and estimated AC was 0.94 (Fig. 4). MAPE and RMSE between the ANN model and the observed data were 8.02 % and 0.013, respectively, for the testing data. According to ANN results for the testing data, R^2 value between the observed and estimated AC was also 0.91 (Fig. 4).

For the training data, MAPE and RMSE between SVR data and the observed AC indicator were 13.15 and 0.021, respectively. For the same data, R^2 between the observed AC indicator and the SVR-estimated data was 0.76 (Fig. 5). For the test data, R^2 (Fig. 5), RMSE, and MAPE between the observed AC indicator and the SVR-estimated data were respectively 0.75, 0.027, and 15.22.

Table 2 represents the PTFs obtained by the regression technique for AC and RFC indicators. For the train data, MAPE and RMSE between MLR model and the observed AC indicator were 12.45 and 0.025%, respectively (Table 2). Additionally, R^2 between MLR and the observed AC data was 0.72 (Fig. 6). Based on the test data, MAPE and RMSE between MLR model and the observed AC were respectively 14.17 and 0.03%. R^2 between the observed and the estimated AC for the test data via the MLR model was also 0.54 (Fig. 6).

The results obtained from the training data indicated that ANN model can understand the relationship between the input variables and AC indicator with more accuracy compared to MLR and SVR. The values of performance criteria for the testing data revealed that ANN is a more accurate method than MRL and SVR in predicting the AC indicator in the region.

According to the training data, MAPE and RMSE between ANN and the observed RFC indicator were 1.78 % and 0.014, respectively. Moreover, R^2 value between the observed and the estimated RFC indicator via ANN was 0.97 (Fig. 7). On the other hand for the testing data, MAPE and RMSE between the ANN model and the observed RFC value were 1.9 and 0.022%, respectively. R^2 value between the observed and estimated RFC indicator via ANN for the test data was also 0.95 (Fig. 7).

Based on the train data, MAPE and RMSE between SVR data and the observed RFC indicator were respectively 8.05 and 0.056. Accordingly, R^2 between the observed RFC indicator and estimated by SVR was 0.54 (Fig. 8). R^2 (Fig. 8), RMSE, and MAPE between the observed RFC indicator and the SVR-estimated data for the test data were 0.53, 0.073 and 11.45, respectively.

MAPE and RMSE values between MLR model and the measured RFC parameter were 8.74 and 0.059%, respectively, according to the train data. For the same data, R^2 between MLR and the measured RFC parameter was 0.49 (Fig. 9). Meanwhile, based on the test data, R^2 (Figure 9), RMSE and MAPE between the observed RFC indicator and the MLR-estimated data were 0.45, 0.079 and 11.83, respectively.

According to the results of the training data, ANN could attain the relationship between the input indicator and AC and RFC indicators with more accuracy than MLR and SVR. Based on the values of performance criteria for the test data, ANN is a more accurate method compared to MRL and SVR in predicting AC and FRC indicators in the region.

According to the evaluation indices, it was observed that the conventional regression model was fairly weak in predicting AC and RFC. Therefore, conventional regression techniques (multiple-linear regression) might fail in consistency for predicting the SPQIs in the region.

3.4. Sensitivity analysis of the ANN model

Fig. 10 displays the results from RMSE sensitivities for the ANN model. For AC prediction, the RMSE sensitivities corresponding to CCE, BD, SOM, and clay removals are 38, 30, 11, and 10 %, respectively. For RFC prediction, the RMSE sensitivities corresponding to CCE, BD, SOM, and clay removals are 30, 25, 8.5, and 10.5 %, in the same order of appearance. This figure indicates that CCE and BD had the highest effect on SPQIs.

Table 2. Goodness-of-fit of the proposed MLR model for the prediction of soil physical quality indicators AC and RFC

Multiple Linear Regression	R^2
$AC = 0.462 - 0.003\,SOM - 0.003\,Clay - 0.482\,P - 0.62\,BD + 0.348\,PD$	0.72
$RFC = 0.576 + 0.005\,SOM + 0.01\,Clay + 1.62\,BD - 1.03\,PD + 2.77\,P$	0.49

AC: air capacity, RFC: relative field capacity, SOM: soil organic matter content, BD: bulk density, PD: particle density, and P: porosity

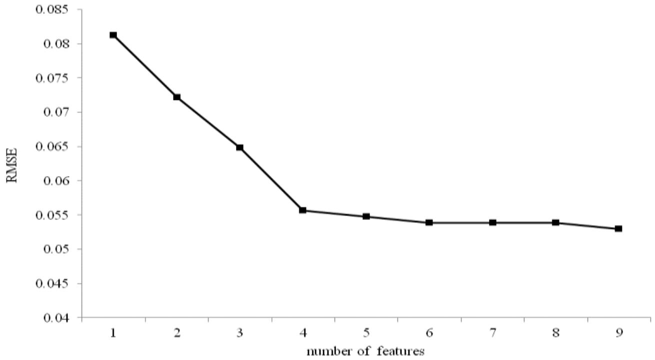


Fig. 3. Obtained RMSE values in selecting feature numbers using ACO-ANN hybrid algorithm for prediction of RFC parameter

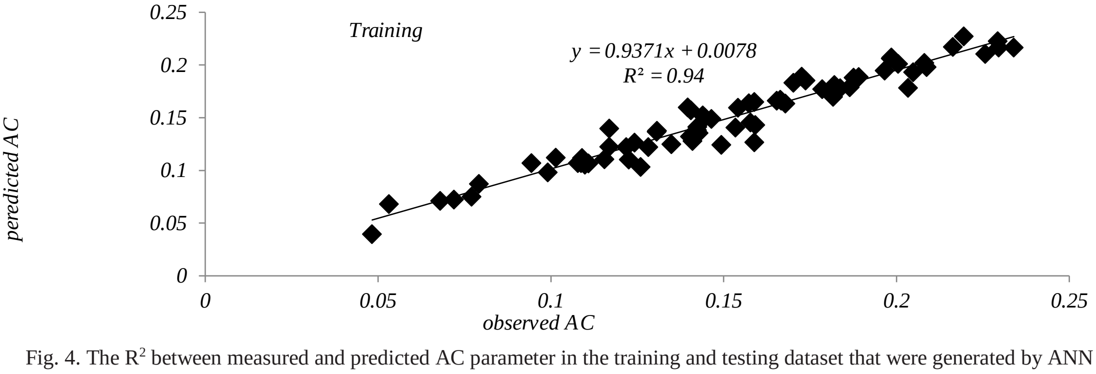
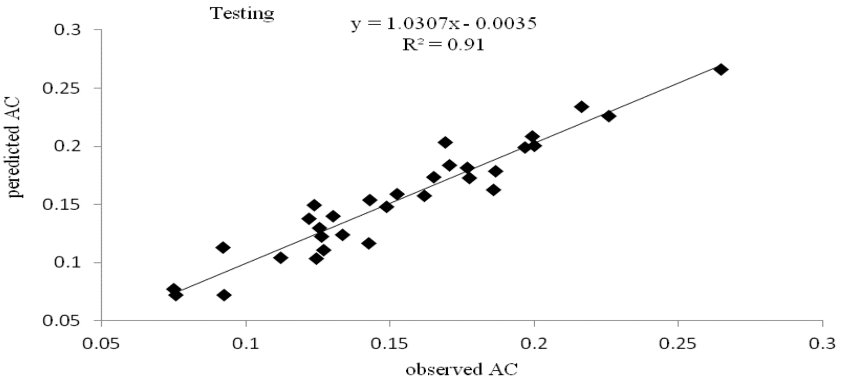


Fig. 4. The R^2 between measured and predicted AC parameter in the training and testing dataset that were generated by ANN model



Continued Fig. 4. The R^2 between measured and predicted AC parameter in the training and testing dataset that were generated by ANN model

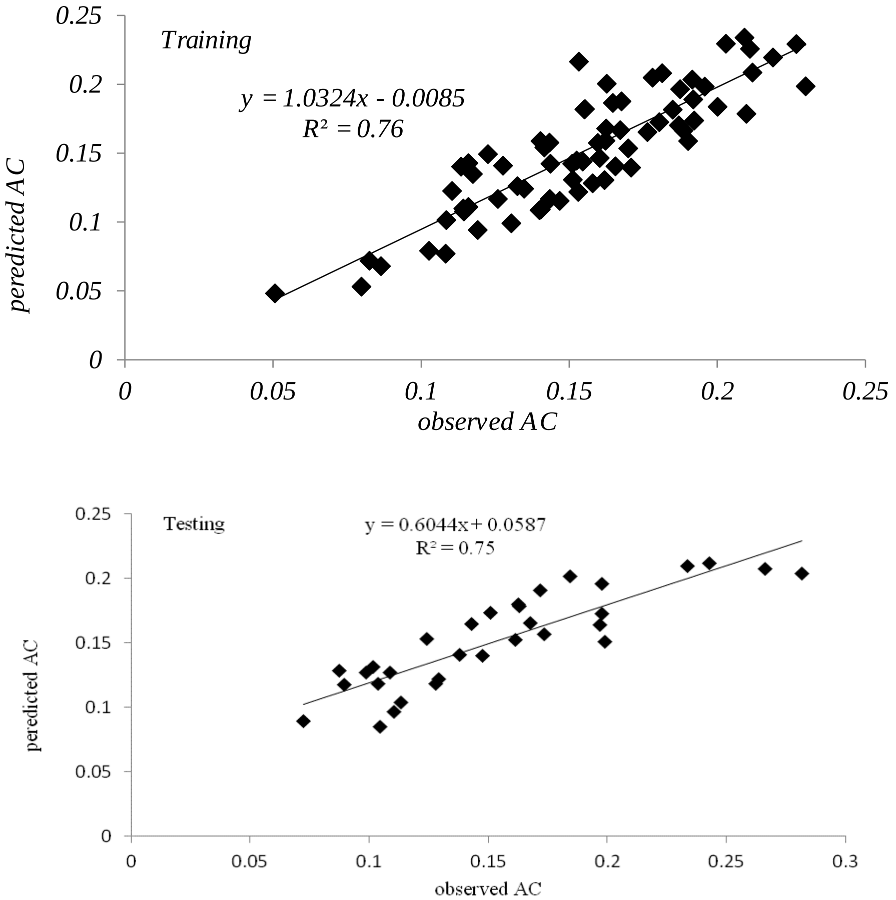


Fig. 5. The R^2 between measured and predicted AC parameter in the training and testing dataset that were generated by SVR model

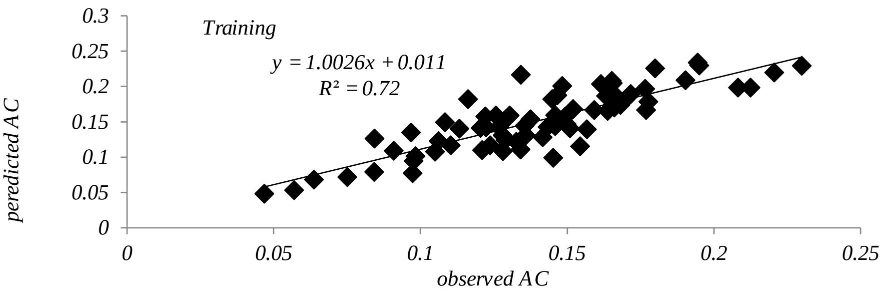
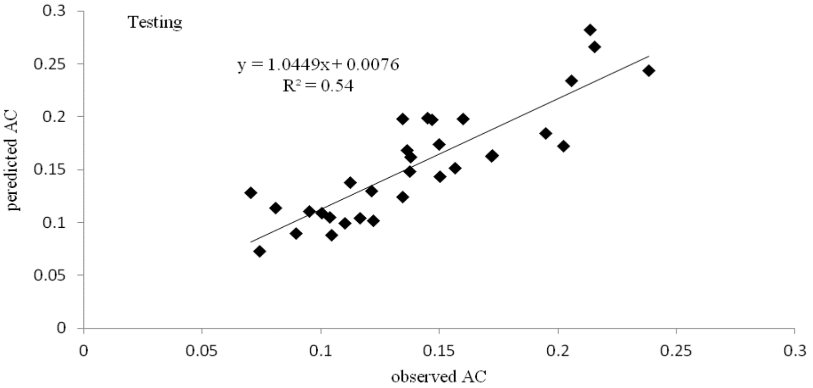


Fig. 6. The R^2 between measured and predicted AC parameter in the training and testing dataset that were generated by MLR model



Continued Fig. 6. The R^2 between measured and predicted AC parameter in the training and testing dataset that were generated by MLR model

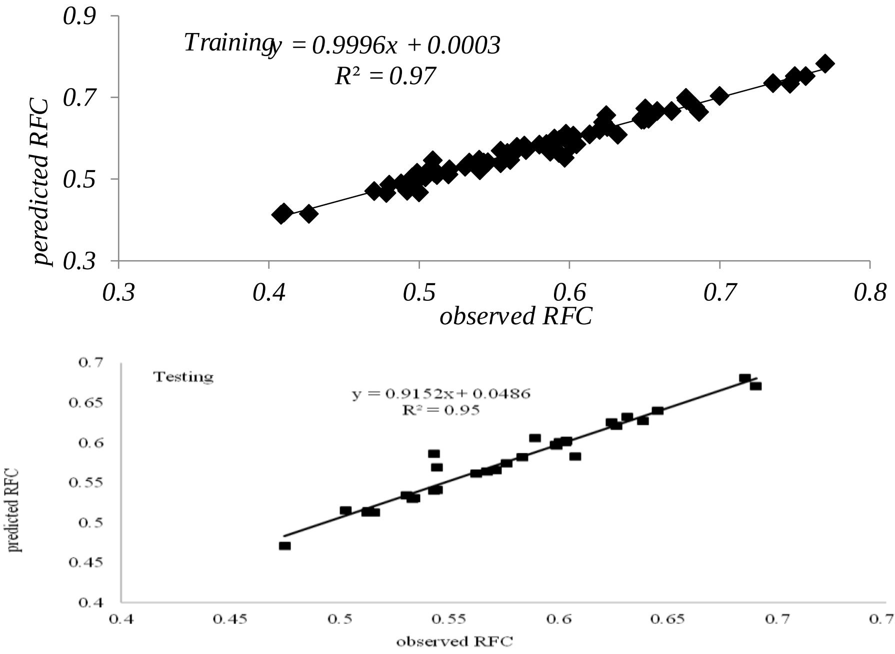


Fig. 7. The R^2 between measured and predicted RFC parameter in the training and testing dataset that were generated by ANN model

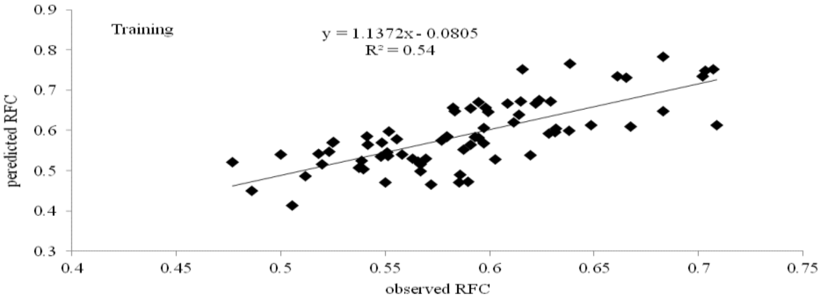
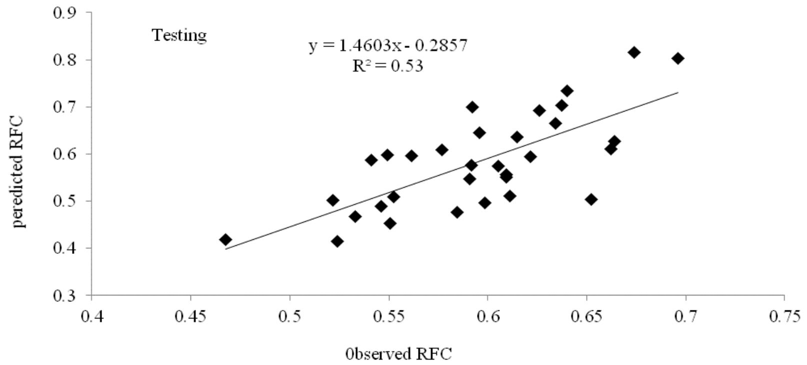


Fig. 8. The R^2 between measured and predicted RFC parameter in the training and testing dataset that were generated by SVR model



Continued Fig. 8. The R^2 between measured and predicted RFC parameter in the training and testing dataset that were generated by SVR model

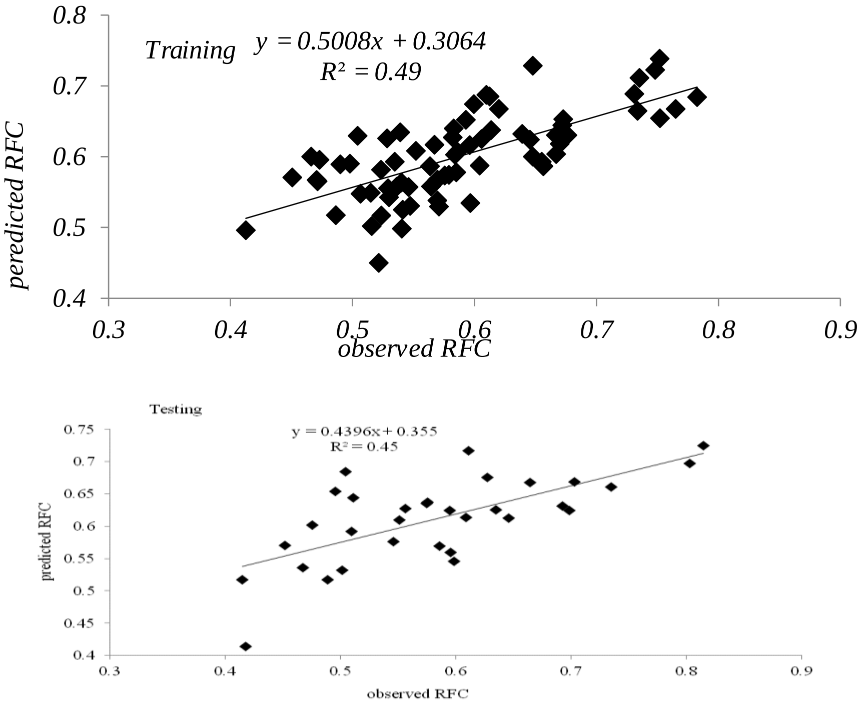


Fig. 9. The R^2 between measured and predicted RFC parameter in the training and testing dataset that were generated by MLR model

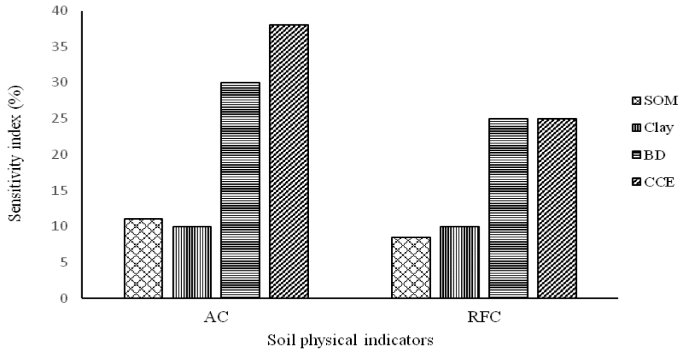


Fig. 10. The sensitivities of RMSE derived by ANN model to removals of soil physicochemical properties

4. Conclusion

This study introduced a new method, an advanced ACO-ANN combination, for selecting the best subset from soil properties that are easily measurable and influence SPQIs. Our method could also be applied to other sites under arid and semi-arid conditions. According to the results from running ABACO-ANN, the highest effect on SPQIs belonged to a subset with features including SOM, BD, clay, and PD values. Following the selection of the best features, the results obtained from ANN, SVR, and MLR techniques implied that the ANN technique is a more accurate and robust tool for predicting SPQIs compared to the others.

References

Aghdam, M.H., N. Ghasem-Aghaee, M.E. Basiri, 2009. Text feature selection using ant colony optimization. Expert systems with applications 36, 6843-6853.

Ahmed, A.-A., 2005. Feature subset selection using ant colony optimization. International Journal of Computational Intelligence.

Allmaras, R., V. Fritz, F. Pfleger, S. Copeland, 2003. Impaired internal drainage and Aphanomyces euteiches root rot of pea caused by soil compaction in a fine-textured soil. Soil and Tillage Research 70, 41-52.

Andrews, S.S., D.L. Karlen, C.A. Cambardella, 2004. The soil management assessment framework. Soil Science Society of America Journal 68, 1945-1962.

Boser, B.E., I.M. Guyon, V.N. Vapnik, 1992. A training algorithm for optimal margin classifiers, Proceedings of the fifth annual workshop on Computational learning theory. ACM, pp. 144-152.

Brejda, J.J., T.B. Moorman, D.L. Karlen, T.H. Dao, 2000. Identification of regional soil quality factors and indicators I. Central and Southern High Plains. Soil Science Society of America Journal 64, 2115-2124.

Da Silva, A.P., B. Kay, 2004. Linking process capability analysis and least limiting water range for assessing soil physical quality. Soil and Tillage Research 79, 167-174.

Da Silva, A.P., B. Kay, E. Perfect, 1997. Management versus inherent soil properties effects on bulk density and relative compaction. Soil and Tillage Research 44, 81-93.

Dexter, A., 2004. Soil physical quality: Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. Geoderma 120, 201-214.

Doran, J., L. Mielke, J. Power, 1990. Microbial activity as regulated by soil water-filled pore space, Transactions 14th International Congress of Soil Science, Kyoto, Japan, August 1990, Volume III., pp. 94-99.

Dorigo, M., G.D. Caro, 1999. Ant colony optimization: A new meta-heuristic, IEEE Congress on Evolutionary Computing.

Dorigo, M., V. Maniezzo, A. Colorni, 1996. Ant system: optimization by a colony of cooperating agents. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 26, 29-41.

Drury, C., T. Zhang, B. Kay, 2003. The non-limiting and least limiting water ranges for soil nitrogen mineralization. Soil Science Society of America Journal 67, 1388-1404.

Gee, G.W., J.W. Bauder, 1986. Particle size analysis, In: Klute, A. (Ed.), Methods of Soil Analysis: Part 1, American Society of Agronomy and Soil Science Society of America, Madison, WI, pp. 383–411.

Goldberg, D., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Professional. Reading, Massachusetts, US.

Hati, K., A. Swarup, D. Singh, A. Misra, P. Ghosh, 2006. Long-term continuous cropping, fertilisation, and manuring effects on physical properties and organic carbon content of a sandy loam soil. Soil Research 44, 487-495.

Karlen, D.L., D.E. Stott, 1994. A framework for evaluating physical and chemical indicators of soil quality. Defining soil quality for a sustainable environment, 53-72.

Kashef, S., H. Neazamabadi-pour, 2015. An Advanced ACO Algorithm for Feature subset Selection. Neurocomputing 147, 271-279.

Klute, A., 1986. Methods of soil analysis. Part 1. American Society of Agronomy, Inc. Soil Science Society of America, Madison, Wisconsin, USA.

Meena, M.J., K. Chandran, A. Karthik, A.V. Samuel, 2012. An enhanced ACO algorithm to select features for text categorization and its parallelization. Expert Systems with Applications 39, 5861-5871.

Michalewicz, Z., 1994. GAs: What are they?, Genetic algorithms+ data structures= evolution programs. Springer, pp. 13-30.

Moncada, M.P., D. Gabriels, W.M. Cornelis, 2014. Data-driven analysis of soil quality indicators using limited data. Geoderma 235, 271-278.

Mueller, L., B.D. Kay, B. Deen, C. Hu, Y. Zhang, M. Wolff, F. Eulenstein, U. Schindler, 2009. Visual assessment of soil structure: Part II. Implications of tillage, rotation and traffic on sites in Canada, China and Germany. Soil and Tillage Research 103, 188-196.

Nelson, D., L.E. Sommers, 1982. Total carbon, organic carbon, and organic matter. Methods of soil analysis. Part 2. Chemical and microbiological properties, 539-579.

Nelson, R.E., 1982. Carbonate and gypsum, In: Page, A.L. (Ed.), Methods of Soil Analysis: Part 1. Agronomy Handbook 9. American Society of Agronomy and Soil Science Society of America, Madison, WI, pp. 181–197.

Pal, M., G.M. Foody, 2010. Feature selection for classification of hyperspectral data by SVM. Geoscience and Remote Sensing, IEEE Transactions on 48, 2297-2307.

Reynolds, W., B. Bowman, C. Drury, C. Tan, X. Lu, 2002. Indicators of good soil physical quality: density and storage parameters. Geoderma 110, 131-146.

Reynolds, W., C. Drury, C. Tan, C. Fox, X. Yang, 2009. Use of indicators and pore volume-function characteristics to quantify soil physical quality. Geoderma 152, 252-263.

Reynolds, W., C. Drury, X. Yang, C. Fox, C. Tan, T. Zhang, 2007. Land management effects on the near-surface physical quality of a clay loam soil. Soil and Tillage Research 96, 316-330.

Reynolds, W., C. Drury, X. Yang, C. Tan, 2008. Optimal soil physical quality inferred through structural regression and parameter interactions. Geoderma 146, 466-474.

Rhoades, J.D., 1996. Salinity: electrical conductivity and total dissolved solids, In: Page, A.L. (Ed.), Methods of Soil Analysis: Part 2. Agronomy Handbook 9. American Society of Agronomy and Soil Science Society of America,, Madison,WI, pp. 417–435.

Schaap, M.G., F.J. Leij, M.T. Van Genuchten, 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. Soil Science Society of America Journal 62, 847-855.

Shirani, H., M. Habibi, A. Besalatpour, I. Esfandiarpour, 2015. Determining the features influencing physical quality of calcareous soils in a semiarid region of Iran using a hybrid PSO-DT algorithm. Geoderma 259, 1-11.

Shekofteh, H., F. Ramazani, H. Shirani, 2017. Optimal feature selection for predicting soil CEC: Comparing the hybrid of ant colony organization algorithm and adaptive network-based fuzzy system with multiple linear regression. Geoderma 298, 27-34.

Staff, S.S., 2014. Keys to Soil Taxonomy. Twelfth ed. NRCS, USDA, USA.

Thomas, G.W., 1996. Soil pH and soil acidity, In: Page, A.L. (Ed.), Methods of Soil Analysis: Part 2. Agronomy Handbook 9. American Society of Agronomy and Soil Science Society of America, Madison, WI, pp. 475–490.

Topp, G., W. Reynolds, F. Cook, J. Kirby, M. Carter, 1997. Physical attributes of soil quality. Developments in Soil Science 25, 21-58.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer, New York, USA.

Vieira, S.M., J.M. Sousa, T.A. Runkler, 2010. Two cooperative ant colonies for feature selection using fuzzy models. Expert Systems with Applications 37, 2714-2723.

White, R., 2006. Principles and practice of soil science 4th ed. Blackwell Publishing.

Wohlberg, B., D.M. Tartakovsky, A. Guadagnini, 2006. Subsurface characterization with support vector machines. Geoscience and Remote Se