Application of modern methods: modeling of sedimentary soil ESP content

PROFESSOR DR. SORUSH NIKNAMIAN

BOARD MEMBER OF WESTON A PRICE FOUNDATION, WASHINGTON

DC, USA

EMAIL: SO.NIKNAMIAN@GMAIL.COM

ABSTRACT

Knowing the exchangeable sodium percentage (ESP) variations and its values in sodic or saline-sodic soils is essential in order to estimate the amount of soil amendments and better land management. ESP calculated from cation exchange capacity (CEC), and since CEC measurement is difficult and time-consuming, ESP computation is costly and subject to error. Thus, presenting a method to estimate ESP indirectly, by an easily available index is much more efficient and economical. In this study, 296 soil samples collected and analyzed from Sistan plain, southeastern Iran. Soil ESP were predicted by using artificial neural networks (ANNs), comprising radial basis functions (RBFN) and multilayer perceptron (MLP) and adaptive neuro-fuzzy inference systems (ANFIS), and results compared with stepwise linear regression (SLR) method. Results indicated that the SLR models performed poorly in order to estimate ESP ($R^2 \leq 0.58$ and root mean square error (RMSE) ≥ 4.31). Applying fewer inputs (electrical conductivity (EC and pH), ANFIS showed better results ($R^2=0.80$, RMSE=2.34), while increasing inputs (clay and organic carbon) decreased the accuracy ($R^2=0.82$, RMSE=4.20). Using more inputs, RBFN resulted in better performance in comparison with other methods ($R^2=0.83$, RMSE=2.85). Sensitivity analysis using the connection weight method demonstrated that EC, pH, clay percentage and bulk density are the most important variables in order to explain ESP variability in the region, respectively. Generally, considering the evaluation criteria and the number of used variables of models, ANFIS (with EC and pH as inputs) is the most appropriate method for estimating ESP in Sistan plain.

- 1 *Keywords:* Saline-sodic soils, Exchangeable sodium percentage, PTFs, Computational
- 2 Intelligence, Sistan plain

1. Introduction

Soil salinity and sodicity are main factors limiting crop growth in irrigated lands. Salinization is the accumulation of solution salts in soil profiles, which limit yield production. When salt accumulation exceeds crop threshold, its importance and effects would reveal. Too amount of salt accumulation in soils is the result of low irrigation and high evaporation rate (Wang et al., 2008). Currently, near 230 Mha of fields are irrigated worldwide, that 45 Mha of them (20%) affected by salinity (FAO, 2008). Sodicity is one of the most important attributes in saline soils, and influences their physical and chemical properties (Farahmand et al., 2011). Sodicity is considerable when exchangeable sodium percentage (ESP) exceeds 15% in soil. Soil structure degradation, decreasing hydraulic conductivity, soil aeration and permeability, and increasing pH up to 8.5, are main reasons of lowering yield in sodic soils (Richards, 1954). Drainage, irrigation and changing crop pattern are some measures to control soil salinity (Cetin and Kirda, 2003). Approximately 10% of soils suffer from saline-sodic problem. Therefore, identifying and managing regions with high salinity limitations are of the most important priorities in agriculture (Barzegar, 2001). Exchangeable sodium to other exchangeable cations ratio is one of the most imperative parameters for evaluating soil salinity and sodicity (Rohoades, 1968). In this regard, ESP usually is the best index.

$$ESP = \frac{Na}{CEC} \times 100 \quad (1)$$

Where:

ESP = Exchangeable sodium percentage, %

Na^+ = Measured exchangeable Na , C mol kg⁻¹

CEC = Cation exchange capacity, C mol kg⁻¹

1 Knowing the ESP variations and its values in sodic or saline-sodic soils, especially in agricultural
2 farms, is essential in order to estimate the amount of soil amendments and better land management.
3 ESP computation is time-consuming, costly and subject to error. Errors of ESP measurement relate
4 to CEC and exchangeable Na. Various error sources reported in measuring CEC by Bower method
5 (Bower et al., 1952), including remaining excess indicator cations during washing step, and the
6 existence of zeolite mineral in soil which resulted in overestimation of CEC and consequently ESP
7 would calculate less than its actual value. Moreover, not saturating exchanging parts with indicator
8 cation, in a full manner, soil wasting and hydrolysis of exchangeable indicator cation while
9 washing, not full replacing of sodium by ammonium and solution of gypsum, caused
10 underestimating CEC and overestimating ESP (Rhoades 1982). On the other hand, CEC
11 measurement is time-consuming and costly. When $EC_e \geq 10 \text{ dS m}^{-1}$ (EC_e : EC for saturation extract),
12 exchangeable sodium measurement in soils would prone to errors related to anion expulsion effect.
13 Due to this errors, exchangeable sodium underestimates (Jurinak et al., 1984). Thus, presenting a
14 method to estimate ESP indirectly, by an easily available index is much more efficient and
15 economical way to overcome mentioned problems.

16 Knowledge of relations and correlations among different soil properties and expressing them by
17 statistical models are one of important issues in soil study. These models called pedo-transfer
18 functions (PTFs) and comprising regression and artificial neural network models (Minasny et al.,
19 2004). PTFs calculate soil attributes which are costly and time-consuming in measurement as
20 function of other properties that easily obtained. Primarily, PTFs used linear regression but
21 gradually it was replaced by nonlinear regression. Statistical regression assumes observations and
22 variables are exact; however, in natural systems such as soils they are not. Thus, it is imperative
23 to use methods for fitting functions that are capable of explaining vague structure of systems and

1 producing actual patterns (Mohamadi and Taheri, 2005). Artificial neural networks (ANNs) are a
2 powerful tool for complicated computations and its easy applicability have been led to predictions
3 in various fields (Fortin et al., 2011; Jingwen et al., 2013 and Kurtulmus et al., 2013). In this regard,
4 nowadays ANNs that are inspired in a way that biological nervous systems works, applied widely.
5 An advantage of modeling methods which are based on calculation intelligence compared to
6 regression PTFs is that they do not need to previous information about relations between inputs
7 and outputs, and also their sensitivity to error in input data are less (Agyare and Park, 2007). In
8 other words, using minimum measured parameters, these models are able to predict target variables
9 variation, precisely. Sadrmomtazi et al. (2013) demonstrated that intelligent models predicted
10 more accurately than conventional regression models.

11 There have been developed many PTFs in order to estimate different soil properties till now. For
12 instance, Robbins and Meyer (1990) presented a model to predict ESP from pH and EC in sodic
13 soils of Australia. See Eq. (2).

$$14 \quad ESP = [(pH - A) \times (1 + C \times EC)/B]^2 \quad (2)$$

15 which A, B and C are soil specific coefficients. Furthermore, they proposed a second-order form
16 equation that its coefficients should modify for different soils. These researchers calculated A, B
17 and C ranges for different soil textures as follows: 4.62-6.95, 0.46-1.20 and 0.004-0.35,
18 respectively. Values of R^2 varies in range of 0.22 to 0.91.

19 Sistan is one of the interior and flat plains of Iran plateau located in southeastern Iran, with
20 elevation ranging from 475 to 500 meters above sea level and covered by alluvial delta of the
21 Hirmand River and its surrounding floods. Considering dry climate, high groundwater levels, poor
22 annual rainfall, high evapotranspiration rate, and unsuitable water quality which utilized to irrigate
23 agricultural farms, salinity is a serious problem in this region and is expanding. Unfortunately, due

to vastness of Sistan plain and difficult conditions for field investigations, soil studies and findings in this area are very little. Planning for preventing and solving the problem of salinity in order to improve soil quality and sustainable agricultural development is necessary and inevitable in this area. The purpose of current study is investigating and suggesting relationships and models to predict the amount of soil ESP from easily obtained properties of soil. Considering statistical regression methods that have been used in previous studies, computational-intelligence-based methods performance assessed. It was tried to present an accurate model with minimum input variables and acceptable precision that does not need to costly and time-consuming laboratory measurements for estimating ESP.

2. Material and Methods

2.1. Description of the study area

The study area is the Sistan plain located in the southeast of Iran, one of the driest regions of Iran and famous for its "120 day wind", a highly persistent dust storm in the summer which blows from north to south with velocities of nearly 20 knots. The Sistan delta has a very hot and dry climate. In summer, the temperature exceeds 50°C. Rainfall is about 55 mm year⁻¹ and occurs only in autumn and winter. Surface evapotranspiration of the area is 4500 to 5000 mm year⁻¹. Strong winds in the region are quite unique and are an important contributing factor for the high evaporation (Fig. 1).

2.2. Field and Laboratory Analyses

Soil samples were collected from 296 points throughout study area. Soil samples were taken in land with a high risk of salinization and/or sodification. Air-dried soil samples were passed through a 2-mm sieve for selected chemical and physical measurements.

The 1:5 and 1:1 soil to water extracts were prepared by adding 20 mL distilled water to 4g and 20g soil in a 100 mL bottles respectively. The bottles were sealed with a stopper, agitated for 15 min on a mechanical shaker (100 r min^{-1}), allowed to stand for 1 h, and then agitated again for 5 min before a sample was obtained by filtration (Chi and Wang, 2010). Organic carbon (OC), bulk density (Bd), calcium carbonate equal (CCE), pH, EC, concentrations of Na^+ , Ca^{2+} and Mg^{2+} , CEC and soil sample textures were determined (USDA-NRCS, 1996). The ESP was determined using Eq. (1).

Figure1[near here]

Physical and chemical properties of soil samples were used for modeling ESP and evaluating the accuracy of obtained model (Table 1).

Table1[near here]

2.3. Stepwise linear regression (SLR) model

An example of a SLR model is shown in the Eq. (3).

$$Y = K_1X_1 + \dots + K_nX_n + d \quad (3)$$

Where:

Y: dependent variable, for example, ESP of soil samples

X_1 to X_n : independent variables, for example the EC of soil samples,

K_1 to K_n : regression coefficient,

d: intercept.

In order to estimate and modeling the ESP by EC and sodium adsorption ratio (SAR) of soil, a SLR model was used. Before modeling, 85% of data assigned to simulation and 15% to test the model. All SLR equations were fitted to the data using the Datafit software and the best regression equations were selected.

2.4. Artificial Neural Networks (ANNs) modeling

ANNs is useful computational way for predicting and modeling abstruse relationships among parameters, especially when there is no explicit relation among parameters (Smith 1993; Gallant 1993). The ANN comprised of three layers: the input layer that all the data are imported to the network and calculation the weight for each input variables are done, the hidden layer or layers, that data are computed, and the output layer, that the ANN results are obtained. Every single layer includes one or more fundamental section(s) called a node or a neuron (Dreyfus et al., 2011). The problem is the key factor that determines the number of neurons in the layers. The small number of hidden neurons is a limiting factor to learn the process carefully, however, too high number can be very time-consuming and the network may overfit the data (Karunanithi et al., 1994).

In this study, three-layer multilayer perceptron (MLP) networks were constructed for computation of the PTFs. All the computations were performed using the Excel 2003 and MATLAB (Version 7.12, Mathworks, Inc., Natick, MA, USA).

2.5. MLP description

The MLP network includes an input layer, hidden layers and an output layer (Fig. 2). In this study, the inputs were pH and EC. The scaled values have been passed into the input layer and after that propagated from the input layer to the next layer which is called hidden layer, before reaching the output layer (Hussain et al., 2002). Each node in both hidden or output layer acts as a summing junction. Using the following equation, inputs combine and modify from the previous layer (Razavi et al., 2003; Jorjani et al., 2008).

$$Y_i = \sum_{j=1}^i X_i W_{ij} + b_j \quad (4)$$

Which:

1 Y_i is the net input to node j in hidden or output layer,

2 W_{ij} is the weight related to neuron i and neuron j ,

3 X_i is the input of neuron j ,

4 b_j is the bias connected to node j .

5 Figure2[near here]

6 To avoid reduction in network speed and accuracy and to make data values equal, it is necessary
7 to normalize input data (Torrecilla et al., 2007). Normalization was done so that the mean of the
8 data series became 0.5 (Kumar et al., 2002). The following equation is used for normalizing data:

9
$$x_n = 0.5 \left[\frac{x - \bar{x}}{x_{max} - x_{min}} \right] + 0.5 \quad (5)$$

10 where:

11 x_n : normalized value,

12 x : actual value,

13 $\bar{x}$: mean value,

14 x_{min} : minimum value,

15 x_{max} : maximum value of parameter.

16 MLP network need sample series of input and output data for designing and training. In this study,
17 PTFs have been considered as network inputs and ESP as output data. Seventy (70), 15 and 15%
18 of data were used to train, validate and test of MLP model, respectively.

19 MLP network applied for ANN modeling using MATLAB 7.12 software. Marquardt-Levenberg
20 learning rule and hyperbolic tangent function were used for training (Haykin 1994). Number of
21 neurons in hidden layer was computed by trial-and-error method and finally the best structure for
22 ESP was selected considering the greatest R^2 value and the least RMSE.

23 *2.6. Radial basis function network (RBFN) model*

1 RBFNs is based on supervised learning. RBFN were independently proposed by many researchers
2 and are a popular alternative to the MLP and RBFN are also good at modeling nonlinear data and
3 can be trained in one stage rather than using an iterative process as in MLP and also learn the given
4 application quickly (Venkatesan and Anitha, 2006).

5 The structure of RBFN is similar to that of MLP. It consists of layer of neurons. The main
6 distinction is that RBFN has a hidden layer which contains nodes called RBFN units. Each RBFN
7 has two key parameters that describe the location of the function's center and its deviation or width.
8 The hidden unit measures the distance between an input data vector and the center of its RBF. The
9 RBFN has its peak when the distance between its center and that of the input data vector is zero
10 and declines gradually as this distance increases. There is only a single hidden layer in a RBF
11 network with two sets of weights, one connecting the hidden layer to the input layer and the other
12 connecting the hidden layer to the output layer. Those weights connecting to the input layer contain
13 the parameters of the basis functions. The weights connecting the hidden layer to the output layer
14 are used to form linear combinations of the activations of the basis functions (hidden units) to
15 generate the network outputs. Since the hidden units are nonlinear, the outputs of the hidden layer
16 may be combined linearly and so processing is rapid (Foody, 2004).

17 *2.7. Adaptive neuro-fuzzy inference system (ANFIS)*

18 The advantage of the fuzzy inference system is that it can deal with linguistic expressions and the
19 advantage of a neural network is that it can be trained and also can self-learn and self-improve.
20 Jang (1993) took both advantages, combining the two techniques, and proposed the ANFIS. The
21 idea behind neural network and fuzzy inference combination is to design a system that uses a fuzzy
22 logic to represent knowledge in an interpretable manner and has the learning ability derived from
23 a neural network that can adjust the membership functions parameters and linguistic rules directly

from data in order to enhance the system performance (Wang et al., 2006). The ANFIS architecture contains a five-layer feed forward neural network (Fig. 3). ANFIS is a hybrid intelligent system which implements a Sugeno fuzzy inference system for a systematic approach to generate fuzzy rules from a given input-output dataset (Negnevitsky, 2005).

A hybrid ANFIS algorithm based on the Sugeno system improved by Jang (1993) was used for acquiring optimal output data in the study. The algorithm consists of the least-squares method and the back-propagation algorithm. The first method was used for optimizing the consequent parameters, while the second method in relation to fuzzy sets was employed to arrange the premise parameters (Ubeyli and Guler, 2005).

Figure3[near here]

In this study in order to predict ESP using MLP and ANFIS, EC_{1:5} and pH considered as inputs. MLP had three layers (an input layer, one hidden layer and an output layer) and six neurons in the hidden layer. Furthermore, ANFIS had five layers (input layer, input membership function layer, rules layer, consequent layer and output layer) with three gaussian membership functions (GaussMF) for input function.

2.8. Analysis

The R², RMSE and model efficiency factor (MEF) used to compare models predicted soil ESP and measured values and assess the performance of models.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_{di})^2}{\sum_{i=1}^n (y_{di} - y_m)^2} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{di} - y_i)^2} \quad (8)$$

$$MEF = 1 - \frac{\sum_1^n [y_{di} - y_i]^2}{\sum_1^n [y_{di} - y_m]^2} \quad (9)$$

1 where:

2 n: number of points,

3 y_i : output value got from the neural network model,

4 y_{di} : experimental value,

5 y_m : average of the experimental values.

6 2.9. Sensitivity analysis for quantifying variable importance

7 Prediction accuracy is a major benefit of ANN models, but the ANN models of any physical
8 processes are purely black box models, which do not explain the process being simulated and
9 whose utility is limited, without information regarding the relative importance of the parameters
10 in the system. The development of a method to couple input factors to meaningful outputs in ANN
11 models is of critical importance (Kemp et al., 2007). The data employed for developing ANN
12 models do contain important information regarding the physical process being simulated (Jain et
13 al., 2008).

14 A connection weight approach was used to evaluate the importance of inputs (soil moisture and
15 salinity) relative to output (crop yield) in ANNs. The connection weight method is to sum the
16 products of the input-hidden and the hidden-output connection weights between each input neuron
17 and output neuron for all input variables (Olden et al., 2004). The relative contributions of the
18 inputs to the output are dependent on the magnitude and direction of the connection weights. When
19 the signs of the input-hidden and hidden-output connection weights are the same (i.e., either both
20 are positive or negative), the input has a positive impact on the output. Contrarily, if the signs of
21 these connection weights are opposite, the specific input has a negative effect on the output. The
22 overall contribution of the input to the output depends on its sum of the positive and negative effect
23 across all different hidden nodes. The larger the sum of the connection weights, the greater the

importance of the variable. The relative importance of input variable i is determined through the following formula:

$$RI_i = \frac{\sum_{j=1}^m W_{ij} W_{jk}}{\sum_{i=1}^n \sum_{j=1}^m W_{ij} W_{jk}} \times 100 \quad (10)$$

where:

RI_i : relative importance of the variable i ($i=1,2,3,\dots,n$) in the input layer on the output variable (%).

j : index number of the hidden node ($j = 1, 2, 3, \dots, m$), W_{ij} : connection weight between input variable i and hidden node j , W_{jk} : connection weight between hidden node j and the output node k . The whole computation was repeated for each output neuron.

3. Results and Discussion

3.1. Descriptive statistics

Descriptive statistics related to ESP and other soil properties have calculated (Table 1). Soil ESP for study area are high (with an average of 16.3) which demonstrates the necessity for investigating ESP variation as sodicity index of soils in Sistan plain. Correlation analysis (Perason coefficient) applied amongst measured attributes and ESP using SAS software (Cary, NC., USA), and $EC_{1:1}$ and $EC_{1:5}$, clay percentage and soil pH were the most correlated variables (Table 2).

3.2. Modifying Robbins and Meyer Equation Coefficients

Table2[near here]

In this study, ESP modeled by the second-order form equation of Robbins and Meyer (1990) and A, B and C coefficients computed for dry alluvial soils of Sistan plain (Table 3).

Table3[near here]

According to statistic parameters of modified equation of Robbins and Meyer (1990), it could be concluded that nonlinear form of this equation is not able to explain ESP variation in study area, accurately. This conclusion supported in Fig. 4.

Figure4[near here]

3.3. Suggested Regression Models to Predict ESP in Dry Alluvial Soils

As Robbins and Meyer (1990) used EC and pH to predict ESP in sodic soils of Australia and reported that their model is economic, time-efficient and potentially able to calculate ESP from easily obtained data, in current study firstly, all parameters which measured in laboratory (comprising EC_{1:5}, pH, clay, OC, CCE, Bd, silt and sand) considered as inputs for model. Final model used only EC_{1:5} and pH as required data (Table 4).

Although the second regression model (Table 4) used more inputs than the first one, it is not able to explain more than 56% of ESP variations. In other words, 44% of ESP variability refers to factors that were not considered in regression model. Parts of this discrepancy can ascribe to nonlinear relations among ESP, EC and other soil properties which SLR models have not sufficient capability to recognize them.

Table4[near here]

Comparing measured and predicted values of ESP using regression models revealed that variation ranges of outputs in model 1 is narrower than model 2, while results of second model showed some overestimation (Figs. 5 and 6).

Figure5[near here]

Figure6[near here]

The purpose of this research was to present a model requiring as low as possible data which are easily obtainable. Since using soil parameters (other than EC and pH) did not improve model and also these parameters obtained easily, EC and pH are appropriate inputs for ESP estimate.

Regression models obtained, are different from Robbins and Meyer (1990). Therefore, relations between ESP and soil attributes are not consistent and influenced by several factors such as EC (Jurinak et al., 1984), ionic solution concentration (Shainberg et al., 1980), soil salinity (Frenkel and Alperovitch, 1983), and clay minerals and its components (Endo et al., 2002).

3.4. The Artificial Neural Networks and ANFIS modeling

In order to predict ESP by artificial intelligence methods (MLP, RBFN and ANFIS) using easily-obtained soil properties and comparing results with SLR, two models considered. First model comprised EC and pH as inputs, while in the second one, all measured parameters considered (EC, pH, clay, OC, CCE, Bd, silt and sand).

During training phase, the best numbers of neurons in hidden layer and the best function in neurons of hidden layer for improving precision of training phase selected by trial and error. Modeling results illustrate in Table 5.

Table5[near here]

Results showed that aiming at estimating ESP by model 1 (inputs: EC and pH), ANFIS is the most efficient model ($R^2=0.80$, RMSE=2.34 and MEF=0.81). MLP and RBFN are suitable, too, however, as shown in Fig. 7, ANFIS outputs are more comparable with input data and ESP predicted better. Generally, computational intelligence methods (MLP, RBFN and ANFIS) were more capable to predict ESP from EC and pH (model 1). Erzin and Gunes (2011) estimated swell percent and swell pressure of soil by using ANN and SLR and reported that ANN performed significantly better than SLR. They presented ANN as an inexpensive and rapid alternative for

laboratory methods to predict swell percent and swell pressure of soil. Estimating soil parameters from more readily available soil data in Ziyaran region, Keshavarzi et al. (2010) concluded that the ANN model with five neurons in hidden layer gives better estimates of field capacity and permanent wilting point than the multivariate regression model. Singh and Deo (2007) in their study to forecast daily river flows along river Narmada in India, using ANFIS, generalized regression neural network (GRNN) and RBFN, found out ANFIS and RBFN are more precise than GRNN and MLP. Amutha and Prochelvan (2011) after studying the seasonal ground water levels in Malattar sub-watershed, located in Vellore district, Tamilnadu, India, assessed performances of ANFIS and RBFN. Both models had 3 inputs. The results showed that the ANFIS is better when compared to RBFN.

Figure7[nere here]

Increasing inputs changed results in which RBFN has better outcomes in comparison with ANFIS and MLP ($R^2=0.83$, RMSE=2.85 and MEF=0.80). Outputs of RBFN are more consistent with input data and ESP predicted better (Fig. 8). Yilmaz and Kaynar (2011) in order to determine the swell potential of clay soils, applied MLP, RBFN, ANFIS and (SLR) models. They reported that MLP and ANFIS have quite the same results and RBFN is the best model. Increasing inputs in ANFIS resulted in increasing error, however, for MLP and RBFN this was vice versa, and although R^2 improved to some extent, RMSE improved significantly (Table 5).

Figure8[nere here]

General comparing among artificial intelligence methods aiming at ESP modeling, implied that ANFIS outperforms RBFN and MLP. Despite the fact that statistical parameters showed good performance of RBFN (model 2), ANFIS (model 1) is more desirable because of lesser inputs (compared with 8 inputs in RBFN 2 which required more time and cost to determine), easy

measurement and obtaining. Karami and Afiuni-Zadeh (2012) for modeling of sizing of rock fragmentation due to bench blasting by estimation of 80% passing size (K80) of Golgohar iron ore mine of Sirjan, Iran, found out that ANFIS is superior to RBFN. They expressed that using only two input parameters in ANFIS is the reason of its superiority over RBFN with seven inputs. Investigating histogram curves for the best models of ESP prediction shows better estimates of ANFIS (model 1) using EC and pH as inputs and RBFN (model 2) using all parameters (Fig. 9). Frequencies of error percent in RBFN (model 2) is closer to zero and has the least variation range and also SD (5.76), which implies its ability for modeling ESP employing more inputs, compared to other models. Furthermore, regression model has the widest curve and the most error variations (Fig. 9), however, SD for ANFIS (model 1) (2.62) is lesser than others that resulted to lesser error (RMSE=2.34) and more accuracy to estimate ESP (Fig. 9-d).

Figure9[near here]

Table6[near here]

In order to find the importance of used variables, the best model including all inputs (RBFN 2) considered. Then, connection weights between input variables and hidden nodes (W_{ij}) and connection weights between hidden nodes and the output nodes (W_{jk}) derived (Table 6). Finally, sensitivity of all inputs of RBFN 2 calculated using equation 9 (Fig. 10). This figure illustrates that $EC_{1:5}$, pH, clay percent and Bd, sort by relevance, are the most important parameters regarding ESP estimation in alluvial soils of Sistan plain.

Figure10[near here]

4. Conclusions

In this study, the feasibility of predicting ESP by soil easily-obtained properties assessed using different methods. ESP estimated by regression models and results implied that they have not

1 acceptable performance ($R^2 \leq 0.58$, $RMSE \geq 4.31$ and $MEF \geq 39$). However, using more inputs
2 improved estimation in regression model 2 and partial R^2 values showed the effect of EC, clay, pH
3 and OC, sort by relevance, in approximating ESP. Then, artificial intelligence models utilized with
4 the same inputs, which demonstrated better results than regression.

5 When used less inputs (model 1), ANFIS are the most efficient model ($R^2=0.80$, $RMSE=2.34$,
6 $MEF=0.81$ and $SD=2.62$), while increasing inputs (model 2) lowered the accuracy ($R^2=0.82$,
7 $RMSE=4.20$ and $MEF=0.71$).

8 Increase in number of input data beside control number of neurons in middle layer, made the RBFN
9 (model 2) the most powerful model ($R^2=0.83$, $RMSE=2.85$ and $MEF=0.80$). Results showed that
10 RBFN, ANFIS and MLP are able to predict ESP fro easily-obtained properties of soils, accurately.

11 Considering results of suggested models (1 and 2) for estimating ESP and according to number of
12 input data beside evaluation criteria, model 1 (inputs: EC and pH) proposes. ANFIS reported the
13 best estimates by this model. Moreover, the other advantage is less inputs that require less time
14 and cost to obtain compared to required data in model 2. Sensitivity analysis results for applied
15 variables regarding ESP estimation revealed that EC, pH, clay percentages and bulk density are
16 the most important data.

17 In total, due to superiority of artificial intelligence models compare to SLR, it is possible to use
18 soil easily-obtained properties such as EC and pH to estimate ESP. It is imperative to conduct
19 similar researches in different soils.

References

- Agyare, W.A., Park, S. J., 2007. Artificial neural network estimation of saturated hydraulic conductivity. *Vadose Zone Journal*, 6, 423-431.
- Amutha, R., Porchelvan, P., 2011. Seasonal Prediction of Groudwater levels using ANFIS and Radial Basis Neural Network. *International Journal of Geology, Earth and Environmental Sciences*, 1, 98-108.
- Barzegar, A., 2001. Saline and Sodic Soils: Productivity and Efficiency. Urmia University Press. (in Persian)
- Bower, C.A., Reitemeier, R.F., Fireman, M., 1952. Exchangeable cation analysis of saline and alkali soils. *Soil Science*, 73, 251-261.
- Cetin, M., Kirda, C., 2003. Spatial and temporal changes of soil salinity in a cotton field irrigated with low-quality water. *Journal of Hydrology*, 272, 238-249.
- Chi, C.M., Wang, Z.C., 2010. Characterizing salt-affected soils of Songnen Plain using saturated paste and 1:5 soil-to-water extraction methods. *Arid Land Research and Management*, 24, 1-11.
- Dreyfus, G., Martinez, J.M., Samuelides, M., Gordon, M.B., Badran, F., Thiria, S., 2011. Apprentissage statistique: Réseaux de neurones-Cartes topologiques-Machines à vecteurs supports: Eyrolles.
- Endo, T., Yamamoto, S., Honna, T. Eneji, A.E., 2002. Sodium-calcium exchange selectivity as influenced by clay minerals and composition. *Soil Science*, 167, 117-125.
- Erzin, Y., Güneş, N., 2011. The prediction of swell percent and swell pressure by using neural networks. *Mathematical and Computational Applications*, 16, 425-436.
- Farahmand, A., Oustan, S.H., Jafarzadeh, A.J., Asgarzad, A.N., 2011. The parameters of sodium and salinity in some salt affected soils of the Tabriz Plain. *Journal of Soil and Water*, 22, 1-15. (In persian)
- Food and Agriculture Organization of the United Nations 2008. Land Resources, Management, Planning and Use. <http://www.fao.org/ag/agl/agll/spush>.
- Foody, G.M., 2004. Supervised image classification by RBFN and MLP and neural networks with and without an exhaustively defined set of classes. *International Journal of Remote Sensing*, 25, 3091-3104.
- Fortin, J.G., Anctil, F., Parent, L.E., Bolinder, M.A., 2011. Site-specific early season potato yield forecast by neural network in Eastern Canada. *Precision Agriculture*. 12, 905-923.
- Frenkel, H., Alperovitch, N., 1983. Factors affecting the estimation of exchangeable sodium percentage in soils from Israel. *Hassadeh*, 63, 1291-1296.
- Gallant, S.I., 1993. Neural network learning and expert systems: MIT press. Cambridge, MA.
- Haykin, S. 1994. Neural Networks: A Comprehensive Foundation. Macmillan, New York.
- Hussain, M., Shafiur Rahman, M., Ng, C., 2002. Prediction of pores formation (porosity) in foods during drying: generic models by the use of hybrid neural network. *Journal of Food Engineering*, 51, 239-48.
- Jain, S.K., Nayak, P.C., Sudheer, K.P., 2008. Models for estimating evapotranspiration using artificial neural network, and their physical interpretation. *Hydrological Processes*, 22, 2225-2234.

- 1 Jang, J.S.R., 1993. ANFIS-Adaptive-network-based fuzzy inference system. *IEEE Transactions*
2 *System Man Cybernetics*, 23, 665-658.
- 3 Jingwen, Z., Zhenghong, X., Shouwen, Ch., 2013. Simulation and prediction of the thuringiensin
4 abiotic degradation processes in aqueous solution by a radius basis function neural network model.
5 *Chemosphere*, 91, 442–447.
- 6 Jorjani, E., Chehreh Chelgani, S., Mesroghli, S., 2008. Application of artificial neural networks to
7 predict chemical desulfurization of Tabas coal. *Fuel*, 87, 2727-2734.
- 8 Jurinak, J.J., Amrhein, C., Wagenet, R.J., 1984. Sodic hazard: The effect of SAR and salinity in
9 soils and overburden materials. *Soil Science*, 137, 152 -158.
- 10 Karami, A., Afiuni-zadeh, S., 2012. Sizing of rock fragmentation modeling due to bench blasting
11 using adaptive neuro-fuzzy inference system and radial basis function. *International Journal of*
12 *Mining Science and Technology*, 22, 459–463.
- 13 Karunanithi, N., Grenney, W.J., Whitley, D., Bovee, K., 1994. Neural networks for river flow
14 prediction. *Journal of Computing in Civil Engineering*, 8, 201-20.
- 15 Kemp, S., Zaradic, P., Hansen, F., 2007. An approach for determining relative input parameter
16 importance and significance in artificial neural networks. *Ecological Modelling*, 204, 326-334.
- 17 Keshavarzi, A., Sarmadian, F., Sadeghnejad, M., Pezeshki, P., 2010. Developing Pedotransfer
18 Functions for Estimating some Soil Properties using Artificial Neural Network and Multivariate
19 Regression Approaches. *Pro Environment*, 3, 322-330.
- 20 Kumar, M., Raghuwanshi, N.S., Singh, R., Wallender, W.W., Pruitt, W.O., 2002. Estimating
21 evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering*,
22 128, 224-233.
- 23 Kurtulmus, F., Lee, W.S., Vardad, A., 2013. Immature peach detection in colour images acquired
24 in natural illumination conditions using statistical classifiers and neural network. *Precision*
25 *Agriculture*, 15, 57-79.
- 26 Minasny, B., Hopman, J., Harter, W.T., Eching, S.O., Toli, A., Denton, M.A., 2004. Neural
27 networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. *Soil*
28 *science Society American Journal*, 68. 417– 429.
- 29 Mohamadi, J., Tahri, S.M., 2005. Fitting the pedotransfer functions by using the fuzzy regression.
30 *Science and Technology of Agriculture and Natural Resources*, 2, 51-60. (In Persian)
- 31 Negnevitsky, M., 2005. *Artificial Intelligence: a guide to intelligent systems*. Pearson Education
32 Limited, Essex, England.
- 33 Olden, J.D., Joy, M.K., Death, R.G., 2004. An accurate comparison of methods for quan- tifying
34 variable importance in artificial neural networks using simulated data. *Ecol. Model.* 178, 389-397.
- 35 Razavi, M.A., Mortazavi, A., Mousavi, M., 2003. Dynamic modelling of milk ultrafiltration by
36 artificial neural network. *Journal of Membrane Science*. 220, 47-58.
- 37 Rhoades, J.D., 1968. Mineral weathering correction for estimating the sodium hazard of irrigation
38 waters. *Soil Science Society of America Proceedings*, 32, 648-652.
- 39 Rhoades, J.D., 1982. Cation exchange capacity. In A. L. Page et al., (Ed.), *Methods of Soil*
40 *Analysis*. Soil Science Society of American, Madison, WI.
- 41 Richards, L.A., 1954. Diagnosis and improvement of saline and alkali soils. In *USDA Handbook*
42 60. U. S. Department of Agriculture, Washington, DC.

- 1 Robbins, C.W., Meyer, W.S., 1990. Calculating pH from EC and SAR values in salinity models
2 and SAR from soil and bore water pH and EC data. *Australian Journal of Soil Research*, 28, 1001-
3 1011.
- 4 Sadrmmomtazi, A., Sobhani, J., Mirgozar, M.A., 2013. Modeling compressive strength of EPS
5 lightweight concrete using regression, neural network and ANFIS. *Journal of Construction and*
6 *Building Materials*, 42, 205–216.
- 7 Shainberg, I., Oster, J. D., Wood, J.D., 1980. Sodium calcium exchange in montmorillonite and
8 illite suspensions. *Soil Science Society of America Journal*, 44, 960-964.
- 9 Singh, P., Deo, M.C., 2007. Suitability of Different Neural Networks in Daily Flow Forecasting.
10 *Applied Soft Computing*, 7, 968–978.
- 11 Smith, M., 1993. *Neural networks for statistical modeling*: Thomson Learning.
- 12 Soil Survey Staff 1996. *Soil survey laboratory methods manual*. Soil Survey Investigations Rep.
13 42. Version 3.0. U.S. Gov. Print. Washington, DC.
- 14 Torrecilla, J., Otero, L., Sanz, P., 2007. Optimization of an artificial neural network for
15 thermal/pressure food processing: Evaluation of training algorithms. *Computers and Electronics*
16 *in Agriculture*, 56: 101-10.
- 17 Ubeyli, E. D., Guler, I., 2005. Adaptive Neuro-Fuzzy Inference Systems for Analysis of Internal
18 Carotid Arterial Doppler, *Signals Computers in Biology and Medicine*, 35, 687-702.
- 19 USDA-NRCS, 1996. *Soil Survey Laboratory Methods Manual*. Soil Survey Investigations Report,
20 No. 42. Version 3.0. Nebraska.
- 21 Venkatesan, P., Anitha, S., 2006. Application of a radical basis function neural network for
22 diagnosis of diabetes mellitus. *Current Science*, 91, 1195-1199.
- 23 Wang, Y.G., Xiao, D.N., Li, Y., Li, X.Y., 2008. Soil salinity evolution and its relationship with
24 dynamics of groundwater in the oasis of inland river basins: case study from the Fubei region of
25 Xinjiang province. *China Environmental Monitoring and Assessment*, 140, 291-302.
- 26 Wang, Z., Palade, V., Xu, Y., 2006. Neuro-fuzzy ensemble approach for microarray cancer gene
27 expression data analysis. In *Evolving Fuzzy Systems*. 2006 International Symposium on, 241-246.
- 28 Yilmaz, I., Kaynar, O., 2011. Multiple regression, ANN (RBF, MLP) and ANFIS models for
29 prediction of swell potential of clayey soils. *Expert Systems with Applications*, 38, 5958–5966.

Variable	Unit	Mean	Minimum	Maximum	SD	Variance	CV
Clay	%	23.2	7.1	49.5	9.20	84.9	39.8
Silt	%	42.5	0.0	82.0	14.6	212	34.2
Sand	%	34.6	2.6	85.0	18.2	330	52.9
Bd	Gr cm ⁻³	1.39	0.8	1.89	0.36	0.13	25.4
OC	%	0.52	0.04	1.26	0.25	0.06	47.7
EC	dS m ⁻¹	4.61	0.2	115.8	9.75	95.1	211.5
pH	log[H ⁺]	8.84	7.7	10.3	0.53	0.28	5.98
Na ⁺	Cmol kg ⁻¹	3.08	1.03	9.32	1.33	1.77	43.3
ESP	Cmol kg ⁻¹	23.9	5.86	59.7	9.20	84.5	38.4
CEC	Cmol kg ⁻¹	13.3	5.92	34.0	4.47	19.9	33.55

SD: standard deviation, CV: Coefficient of variation, Bd: bulk density, OC: organic carbon,

CEC: cation exchange capaci

	Clay	CCE	Silt	Sand	Bd	OC	EC _{1:1}	EC _{1:5}	pH
ESP	0.32**	0.14*	0.27**	-0.37**	0.05	-0.05	0.68**	0.70**	0.35**

** : significant at the 0.01 level, * : significant at the 0.05 level

Equation form	A	B	C	RMSE	R ²	MEF
ESP = [(pH - A) × (1 + C × EC)/B] ²	7.22	4.28	0.0199	4.53	0.50	0.33

Models	Step	Variable Entered	Partial R-Square	Model R-Square	F Value	Pr > F
Reg Model 1	1	EC _{1:5}	0.49	0.49	286.74	<.0001
	2	pH	0.02	0.52	14.01	0.0002
Reg Model 2	1	EC _{1:5}	0.49	0.49	286.73	<.0001
	2	Clay	0.04	0.53	20.19	<.0001
	3	pH	0.02	0.56	14.91	0.0001
	4	OC	0.01	0.58	3.65	0.0572
Final Results						
Models	Pedotransfer function			RMSE	R ²	MEF
Reg 1	ESP = 0.748EC _{1:5} + 1.98pH – 3.858			4.91	0.52	0.39
Reg 2	ESP = 0.7EC _{1:5} + 0.155Clay + 1.764pH – 2.161OC – 4.1994			4.34	0.58	0.50

All variables Entered left in the model are significant at the 0.1000 level

Model	Input	MLP				RMSE	R ²	MEF
		Layer No.	Hidden layer neurons	HLF*	OLF*			
1	EC _{1:5} +pH	3	4	tribas	Purelin	3.95	0.74	0.69
2	8 parameters**	3	6	tansig	Purelin	3.65	0.76	0.75
Model	Input	RBFN				RMSE	R ²	MEF
		Spread	Hidden layer neurons	HLF*	OLF*			
1	EC _{1:5} +pH	0.7	5	gaussian	Liner	3.55	0.77	0.74
2	8 parameters**	1	10	gaussian	Liner	2.85	0.83	0.80
Model	Input	ANFIS				RMSE	R ²	MEF
		Layer No.	Rules	IMF*	OMF*			
1	EC _{1:5} +pH	5	9	gaussMF	liner	2.34	0.80	0.81
2	8 parameters**	5	4	gaussMF	liner	4.20	0.82	0.71

*HLF: hidden layer function, OLF: output layer function, IMF: input membership function, OMF:

output membership function, LIF: input layer function

** 8 parameters: EC_{1:5}, pH, clay, silt, sand, organic carbon, carbonate calcium equivalent, density

Input weight										
	Neuron1	Neuron2	Neuron3	Neuron4	Neuron5	Neuron6	Neuron7	Neuron8	Neuron9	Neuron10
CCE	0.4814	0.591	0.6458	0.4745	0.454	0.4951	0.4198	0.2622	0.4951	0.3
Bd	0.7087	0.7087	0.4655	0.4334	0.6674	0.6903	0.3555	0.6353	0.4472	0.6
OC	0.4324	0.4404	0.4598	0.4657	0.5221	0.4316	0.509	0.6081	0.4689	0.6
Clay	0.7059	0.6163	0.3757	0.4182	0.4418	0.3686	0.4182	0.5125	0.4076	0.5
Silt	0.5875	0.5631	0.4351	0.4656	0.5814	0.679	0.4107	0.3497	0.5326	0.4
Sand	0.3065	0.3769	0.6281	0.576	0.4485	0.3891	0.6306	0.6427	0.5147	0.5
EC _{1:5}	0.6792	0.5089	0.4616	0.5025	0.4651	0.815	0.4588	0.6624	0.4661	0.4
pH	0.7391	0.6741	0.4832	0.5127	0.5009	0.4635	0.4969	0.5521	0.493	0.5
Layer weight										
	Neuron1	Neuron2	Neuron3	Neuron4	Neuron5	Neuron6	Neuron7	Neuron8	Neuron9	Neuron10
	2.0593	-0.6136	3.8075	-18.0898	-0.6824	0.5658	5.564	4.4402	8.4618	-4.3

Table1

Statistical parameters of soil physical and chemical properties

Table2

Correlations between organic carbon (OC) and some soil properties

Table3

ESP modeling results by modified equation of Robbins and Meyer (1990) in alluvial soils of Sistan plain

Table4

The Root Mean Squer Errorr (RMSE) of the independent variable, Coefficient of Determination (R^2) and Model Efficiency Factor (MEF) of the soil ESP pedotransfer function

Table 5

Properties of resulted MLP, RBFN and ANFIS models

Table 6

Neuron weights used for sensitivity analysis

Fig. 1. Map showing the geographical setting of the study area, Sistan plain, Iran

Fig. 1. Schematic representation of the best ANN architecture

Fig. 2. ANFIS architecture of two input and nine rules

Fig. 4. Measured vs. predicted values of ESP using equation of Robbins and Meyer (1990)

Fig. 5. Measured vs. estimated values of ESP using easily obtained soil properties in regression model 1

Fig. 6. Measured vs. estimated values of ESP using easily obtained soil properties in regression model 2

Fig. 7. Measured vs. estimated values of ESP using easily obtained soil properties in model 1 by RBFN, MLP and ANFIS

Fig. 8. Measured vs. estimated values of ESP using easily obtained soil properties in model 2 by RBFN, MLP and ANFIS

Fig. 9. Histogram curves of the best models: (a) Regression 2, (b) MLP 2, (c) RBFN 2 and (d) ANFIS 1 (St.D: standard deviation)

Fig. 10. Sensitivity coefficients histogram for some soil properties