

Prediction of saturated hydraulic Conductivity of Iraqi soil using multiple regression, ANN (RBF, MLP) models

Sahar Esmail Mahmood

Abstract---Saturated hydraulic conductivity (Ks), is one of the soil hydraulic properties which is important and necessary in environmental studies such as subsurface ground water, water and mass transport models and irrigation and drainage studies. Since, its direct measurement is time consuming and therefore costly, indirect measurement such as pedotransfer functions (PTFs). Provide an alternative way to estimate the Ks from readily available soil properties such as sand, silt, clay content, organic matter and bulk density.

This study was done to estimate the Ks in northern and middle of Iraq. The objective of this study was to develop and compare between three intelligence models radial basis function neural networks (RBF-ANN), multilayer perceptron neural networks (MLPN), and multiple-linear regression (MLR) to predict the Ks. Input variable included sand%, clay% and bulk density. The total of 75 soil samples were divided into three groups as 40 for the training, 10 for the testing and 26 for the validation of PTFs. The results indicated that ANNs and RBF ANN are effective methods compared with the MLR models. The correlation between predicted and measured Ks value using artificial Neural Network (ANN) was better than Radial base function (RBF) and (MLR). Root mean square error values for ANN, RBF and MLR were 0.0116, 0.021, and 0.037 respectively, which shows that multilayer perceptron (MLPN) model is a powerful tool and has better performance than (RBF) and (MLR) in prediction of Ks.

Keywords: Saturated hydraulic conductivity, pedotransfer function, soil physical properties, neural network.

I. INTRODUCTION

Soil hydraulic properties such as saturated hydraulic conductivity (Ks) govern many soil hydrological processes; there for, they are very important and necessary in water infiltration mass transport models, irrigation and drainage practices [1]. Direct measurement of soil hydraulic properties including Ks is difficult costly and time-consuming and becomes impractical due to spatial and temporal variabilities when hydrologic predictions are needed for large areas. Also it requires skilled operators and sophisticated measurement devices [2]. In the past few decades, as an alternative, indirect approximation of hydraulic properties from some basic and easily measured soil properties (such as clay, sand, and silt content and bulk density) using pedotransfer functions (PTFs) has received considerable acceptance [3]-[7]. "Pedotransfer function was first introduced for empirical regression equations relating water and solute transport parameters to the basic soil properties that are available in soil survey [8].

The Ks is an important soil hydraulic property often estimated using PTFs. Different methods such as regression models [3],[9]-[11] and artificial neural

networks (ANN) are available for derivation of PTFs. In recent years, PTFs created by using artificial neural network (ANN) (especially feed forward ANN) have established widespread with many scientists. ANN-PTFs have been developed by researchers such as Minasny et al., Minasny and McBratney, and Pachepsky et al. [5],[6],[11]. The total decision made by these (and other) investigators was that when the number of input parameters is greater than three, ANN usually completes better than regression procedures, mostly when qualms in the quality of the data were small [12]. Radial basis function (RBF) and Multilayer perceptron (MLP) are two of the supreme usually used neural network structure. Overall difference between MLP and RBF is that RBF is a localist type of learning which is reactive only to a limited segment of input space. RBF employs a local learning tactic versus MLP worldwide learning and this leads to higher rate of precision and faster training of RBF [13].

State of the art shows that, in most previous studies, Ks was predicted using regression and ANN models [3],[14],[15]. Many comparisons of PTFs have been made with respect to different data sets, different mathematical procedures (regression versus ANN models), and different input parameters. Wieland and Mirschel compared a feed-forward Neural Network (NN), a radial basis function network (RBF), and trained fuzzy algorithm for regional yield estimation of agricultural crops (winter rye, winter barley) [16]. There for the aims of this study were (a) to deal the procedures of RBF, ANN, and MLR methods to expect saturated hydraulic conductivity using easily existing soil properties (b) to appraise the performance of these techniques and determine the best methodology for estimate of saturated hydraulic conductivity.

An ANN is highly interrelated network of many simple dealing out units called neurons, which are equivalent to the biotic neurons in the human brain. Neurons having similar features in an ANN are settled in groups called layers. The neurons in one layer are connected to those in the next to layers but not to those in the same layer. The linking between the two neurons in adjacent layers is signified by what is known as joining strength or weight. An ANN normally contains three layers, an input layer, a hidden layer, and an output layer. In a feed-forward Network, the weighted connections feed activations only in the forward direction from an input layer to the output layer. On the other hand, in a persistent network other weighted connections are used to feed previous

stimulations back into the network. The structure of a feed-forward ANN is shown in Figure 1 [17].

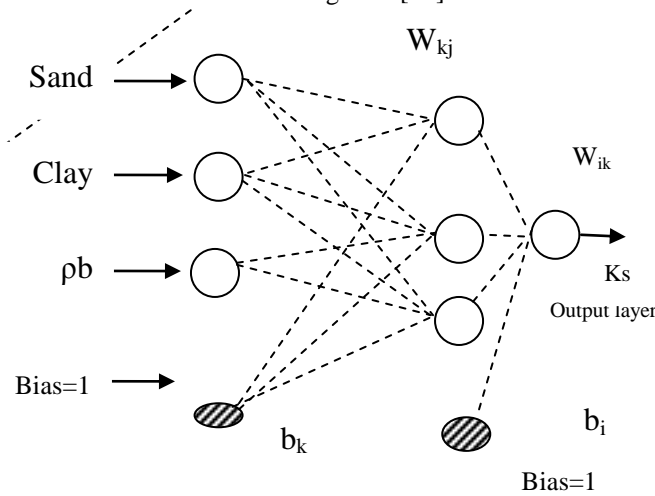


Fig.1. MLP-ANN Neural network structure used in this study

In Figure (1), the circles represent neurons; the lines joining the neurons represent weights, sand, clay and P_b are input variables; K_s represents the output variables; W_{kj} and W_{ik} represent the weights between input and hidden and hidden and output layers, respectively. Parameters b_k and b_i represent the bias of the corresponding hidden and output layer neurons. The role of bias in a neuron is to displace the original function domain by magnitude equal to that of the bias and there by translate the area influence to its activation state [17].

An important step in developing an ANN mode is training of its weight matrix. The weights are initialized randomly in suitable range and then updated using certain training mechanism. There are primarily two types of training mechanisms supervised and unsupervised. A supervised training algorithm requires an external teacher to guide the training process this typical involves a large number of examples (or patterns) of inputs and output for training. The inputs in an ANN are the cause variable and outputs are the effect variables of the physical system being modeled. The primary goal of training is to minimize the objective (error) function by searching for a set of connection strengths that cause the ANN to produce outputs that are equal to or closer to target data [17].

Radial basis Function (RBF) and multilayer perceptron (MLP) are two of the greatest generally used neural network construction in literature for classification or regression problems. They are strong classifiers with the ability to specify for vague input data.

II. MATERIALS AND METHOD

Data collections: The data used in this study were obtained from Ms.c thesis I, Khalil M. T. 2007 [18] and Msc Thesis H. Bagweta, (1983) [19] the location of samples taken from three location in mosul city (Al Fathelia, KaraTapa and Al Hamdania) and at other location in the middle of Iraq at many cities (Heat, Hathramout, Saqlawia, Dora, Hashmia and Tabra. The

data set which was used to develop the PTFs included the data from 75 soil profiles from different part of mosul city and Baghdad, Erbil, Kirkuke soil sample were taken (0-30 cm) (30-50 cm) (50-80 cm) (80-120cm) layers.

The data consisted of four soil properties (three as independent variable and one as dependent variable) Independent variable were the percentage of clay $\text{\textcircled{C}}$, sand (s), and bulk density(pb). Saturated hydraulic conductivity (K_s) was considered as dependent output variable, soil sample were air-dried and sieved by a 2 mm sieve for physical analysis particle size distribution was determined using pipet method. Pb was determined on undisturbed sample using cylinder method being made of 10 cm and 100 cm^3 cylinders after drying 24 h in 105 $^\circ$ C(ovens). Saturated hydraulic conductivity was measured using inversed auger hole method the principle of the auger hole test above the water table consists of boring a hole to given depth, filling it with water and measuring the rate of fall of the water level [20]

MLP,RBF, and MLR of pedotransfer function were used to predict th K_s . The data set was divided in to three separate data set; the training one (53%) testing (13%) and validation (33%). The training data set was used to train the ANN, RBF and (MLR) models, where the testing data set was used to verify the accuracy and the effectiveness of the train models. The weight of the neural networks was applied on validation sample to predict the K_s value. A computer program was run in MATLAB (ver:10b) to train , test and validate the data using ANN and RBF structure of K_s prediction the performances of new techniques (ANN and RBF) were compared with MLR method.

Regression Analysis

In the multiple linear regression (MLR) analysis, first, the most essential input variables were selected using backwards stepwise method, and then linear, quadratic, and possible interaction terms of these basic soil properties were investigated using the SPSS software. The same data sets were used in the derivation (N=76) and validation (N=26) and testing (N=10) of PTFs developed using ANN, RBF and Regression Method for reliable comparison. The resulting function was tested using data sets not included in the derivation procedure (validation data).

Developing PTFs Using Artificial Neural Network Model (ANN)

Pedotransfer functions, have recently become a popular topic in soil science research, Different types of function have been developed to predicted either physical or chemical properties of soil .Most pedotransfer functions have been developed to predict soil hydraulic properties, especially saturation hydraulic conductivities, water retention curves . This is mainly as the response to the urgent need for soil hydraulic properties as input to soil-water models.

The usual step in deriving PTFs by forming empirical relationships between basic soil properties and other soil properties. This can be achieved by various mathematical

methods. Such as multiple linear regression (Wosten et al., 1995). A recent approach for fitting PTFs is use artificial neural networks (ANN). Most researchers have found that ANN perform better than multiple regression (Koekkoek and Boolting, 1999). An advantage of using the neural network approach is that no relationships need to be assumed beforehand. Instead the network is trained to find the relationship.

Then the best ANN architecture is determined by finding the optimal number of hidden neurons through training of the various architectures using a trial and error method once the best ANN architecture is trained, it is validate using the validation data set.

In this study two different kinds of ANN were established the first ANN model is multilayer perceptron (MLP) which is the most usually used neural network construction in ecological modeling and soil science [21] were the second ANN model is radial basis function (RBF).

Performance Evaluation Criteria: Four standard statistics were used as evaluation criteria to evaluate the performance of ANN models (MLP and RBF) and MLR method.

These were, values account for (VAF) Equation (1) and root mean square error (RMSE) Equation (2), and indices were calculated to control the performance of the prediction capacity of predictive model developed in the study as employ by Al Varez and Babuskg (1999) Finol, Gvo, and Jiny (2001)

$$VAF = 1 - \frac{\text{var}(y - \hat{y})}{\text{var}(y)} * 100 \quad \dots\dots\dots(1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y - \hat{y})^2} \quad \dots\dots\dots(2)$$

Where y and \hat{y} are the measured and predicted values respectively. If the VAF is 100 and RMSE is 0 the model will be excellent.

Mean absolute percentage error (MAPE) which is a measure of accuracy in a fitted series value in statistics was also used for comparison of the prediction performances of the models. MAPE usually expresses accuracy as a percentage Equation (3)

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{Ai - Pi}{Ai} * 100 \quad \dots\dots\dots(3)$$

Where Ai is the actual value and Pi is the predicted value. Also the coefficient of determination (R^2) used to evaluate the performance of ANN & MLR models Equation (4).

$$R^2 = 1 - \frac{\sum_{i=1}^N (yi - yi')^2}{\sum_{i=1}^N (yi - \bar{y})^2} \quad \dots\dots\dots(4)$$

Where N is the number of data set, yi is the measured value of output variable yi' the predicted value of output

variable and \bar{y} is the average of predicted value of output variable.

III. RESULT AND DISCUSSION

Before developing PTFs using MLR and ANN models, convincing statistics of data were derived using SPSS (version 18) Descriptive statistics for the soil physical and hydraulic strictures which were used in the development (train), (validation) & (test) of PTFs using models are summarized in table (1).

The studied soils have wide ranges of physical properties for instance, the ranges of clay, sand and bulk density are 24.89-59.23, 8.96-51.52%, 1.19-1.37gm/cm³ for training or derivation data set respectively, validation & testing data set had similar range. The studied soil are mostly originated from rivers alluvial process which probably cause such wide range of soil physical properties the soils are on average with medium to fine texture, where the mean clay percent and ρ_b are 45.58 and 1.24gm/cm² respectively.

The Ks values of the studied soils mostly low ranging form 0.11-0.68 cm/h with stander deviation of 0.097, 0.064 and 0.049 for training, validation and testing data sets respectively an existence of micro pores and high percent of clay might be also responsible for low value of Ks.

Table 1. Descriptive statistics of the data sets used for training, validation & testing ANN and MLR

<i>Training</i>					
Variable	Units	Max	Min	Mean	S.D.
Clay	%	59.23	24.89	48.24	9.88
Sand	%	51.52	8.96	18.98	7.409
Bulk	Gm/cm ³	1.37	1.19	1.24	0.0444
Ks	Cm/h	0.68	0.11	0.244	0.09705

Validation

Variable	Units	Max	Min	Mean	S.D.
Clay	%	58.5	32.0	41.43	6.52
Sand	%	22.0	2.0	12.66	5.16
Bulk	Gm/cm ³	1.3	1.19	1.25	0.0247
Ks	Cm/h	0.425	0.175	0.313	0.0648

Testing

Variable	Units	Max	Min	Mean	S.D.
Clay	%	56.37	37.5	45.69	5.44
Sand	%	18.63	2.0	11.0	5.188
Bulk	Gm/cm ³	1.26	1.19	1.23	0.0211
Ks	Cm/h	0.375	0.225	0.27	0.497

Table 2. Performance indices (RMSE, VAF, MAPE & R^2) for models

Model	RMSE	VAF%	MAPE	(R^2)
Multiple Liner Regression	0.037	65.5	12.07	0.708
Radial Base function	0.021	91.9	6.49	0.90
ANN - MLP	0.012	97.16	3.41	0.97

Multiple regression models

Multiple linear regression procedure going back to Pearson's 1908 use of it, is established to estimate the variance in an interval dependent, based on linear combinations of interval, dichotomous, or dummy self-determining variables. The aim of multiple regression is to learn more about the connotation between several independent or predictor variables and a dependent or criterion variable. The multiple regression equation takes the form

$$Y = b_1x_1 + b_2x_2 + \dots + b_nx_n + c.$$

b_1, b_2, \dots, b_n are the regression coefficients representing the amount the regression variable Y changes when the equivalent independent change 1 unit.

c is the constant, where the regression line intercepts the y axis, representing the amount the dependent y will be when all the independent variables are 0.

The standardized versions of the b coefficients are the beta weights and the ratio of the beta coefficients is the ratio of the relative predictive power of the independent variables.

In order to develop PTFs for predicting K_s through MLR model, first the essential input variables were selected using stepwise method, and then linear interaction terms of these basic soil properties were investigated by means of SPSS 18 software.

After training the regression model with training data set the equation applied on validation data set. The regression equation was as follows.

$$\text{Predicted } K_s = -2.80617 - 0.000277 * \text{clay} - 0.009137 * \text{sand} + 2.603 * \text{pb}$$

In derived equation clay, sand and bulk density were chosen as the independent variables. After determining regression equation, the accuracy of MLR model was evaluated through comparing its predicted K_s with experimental data. The obtained value RMSE, VAF%, MADE and R^2 using MLR model for predicting K_s are tabulated in table (2).

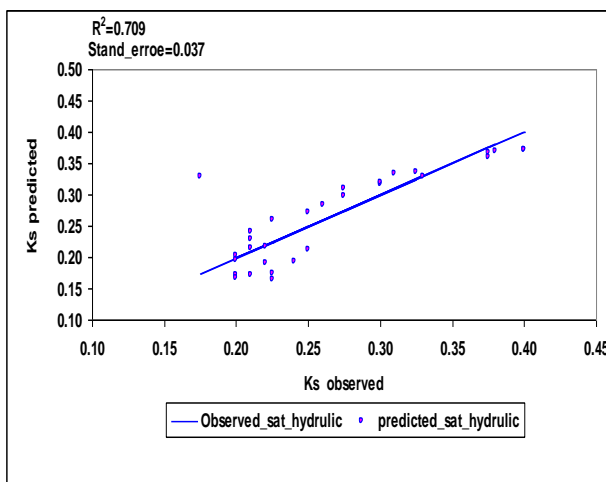


Fig 2. The scatter plot of measured versus predicted of saturated hydraulic conductivity of soil using Multiple Linear Regression Model with validation samples

As observed from the table for test data set the R^2 and RMSE value that have been obtained are 0.708 and 0.037, respectively. Merdum et al. obtained higher R^2 and RMSE value varied from 0.637 to 0.979 and from 0.013 to 0.938 for regression method respectively [3]. The scatter plot of the measured against predicted K_s values obtained from MLR model for the validation data set with medium coefficient of determination is illustrated in Figure (2)

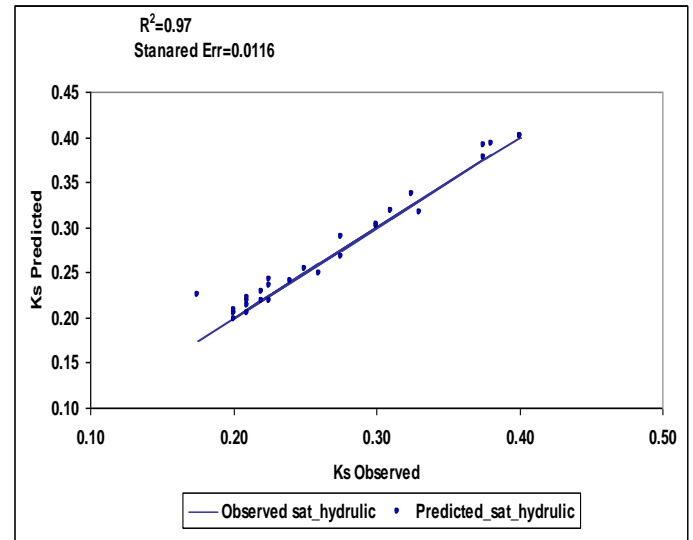


Fig 3. The scatter plot of measured versus predicted of saturated hydraulic conductivity of soil using Neural Network Model with validation samples

PTFs Development Using ANN Models

In current study two different algorithms of ANNs including radial basis function (RBF) and multilayer perceptron (MLP) models were investigated to predict K_s through employing the same data set which was used by MLR. A three layer feed-forward ANN architecture with an input layer one hidden layer and an output layer was developed for predicting K_s by means of both ANN model.

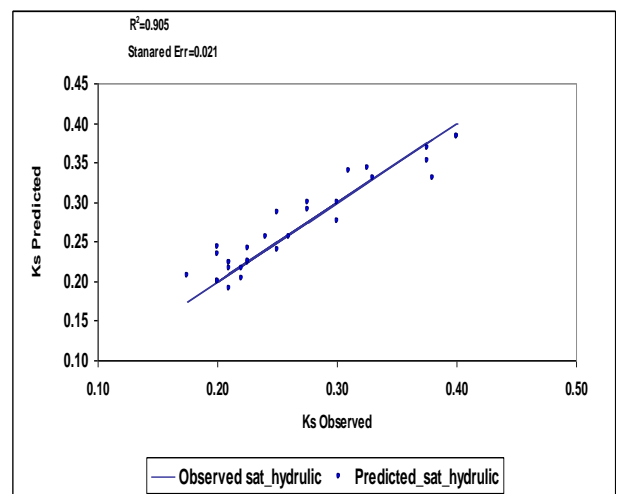


Fig 4. The scatter plot of measured versus predicted of saturated hydraulic conductivity of soil using Radial Base Function Model with validation samples

MLP Network table (2), shows the results of statistical analyses between the observed and MLP ANN-simulated values of Ks for validation stage. For MLP network the best architecture consist of three neurons in the input layer, three neurons in the hidden layer and one neuron in the output layer with tan sigmoid and pure line three should functions for hidden and output layer respectively, gave the best results. The R^2 and RMSE value among the observed and predicted Ks are 0.97 and 0.012 respectively.

The values are in accordance with the previous studies using ANN methods. Agyare et al. Which estimating Ks obtained.

R^2 and RMSE about 0.6 and 0.42 respectively [26] merdun et al. Obtained R^2 ranges and RMSE. Varied from 0.44 to 0.952 and from 0.02 to 3.511, respectively [3]. Figure (3) shows the relationship between the measured and predicted Ks value for validation stage indicated that MLP network can predict Ks with acceptable accuracy.

In order to employ RBF. Gaussian function that is the most widely used in applications was chosen as three should function for hidden layer. In the next step a regression analysis of the network response between ANN outputs and the corresponding targets was performed.

Table 2 tabulated. That the provided predictions between ANNs outputs and the corresponding targets using RBF model ($R^2=0.9$, $RMSE = 0.021$).

The levels of R^2 and RMSE derived by both ANN models had higher accuracy than those derived by multiple linear regression for predicted Ks which was support for those previous studies conducted by merdun et al.

Tamari et al. Yilmaz et al. and other researchers [3],[22],[23] this is due to that, un like the traditional regression PTFs ANNs do not require a priori regression model which relates input and output data that in general is difficult because these model are not known [24]. Also Minasny and McBratney Pachepsky et al. and Tameri et al. stated that when the number of input parameters is greater than three ANNs usually perform better than regression techniques. Particularly when uncertainties in the quality of the data were small [11],[22],[25] Additionally many investigations have indicated that the a neural network with one hidden layer is capable of approximating any finite nonlinear function with very high accuracy [26].

The scatter plots between measured and predicted Ks using MLP and RBF model for validation stage with acceptable accuracy are indicated in Figure (3) and Figure (4) respectively.

IV. CONCLUSION

Expecting the saturated hydraulic conductivity (Ks) of soil is one of the important subjects in showing water flow and solute transport process in Mosul Zone. This paper showed the development and validation of PTFs for

guesses of Ks from basic soil properties by using MLR and ANN model in Mosul Zone northern of Iraq. The predictive abilities of these methods were also compared. using some assessment criteria of validation data. The results revealed that the MLP-ANN had advantage to the RBF model for Ks prediction Both applied ANNs algorithms revealed acceptable accuracy. The evaluation of the RBF and MLP networks indicate, that they had approximately similar performance for prediction Ks, as obtained results using MLP were slightly better than RBF ones but their differences were not found significant Among the employed ANN models MLP prediction was more accurate than RBF network prediction.

REFERENCES

- [1] V. Bagarello, S. Sferlazza, and A. Sgroi, "Comparing two methods of analysis of single-ring infiltrometer data for a sandy-loam soil" *Geoderma*, vol. 149, no. 3-4, pp. 415-420, 2009.
- [2] N. Islam, W. W. Wallender, J. P. Mitchell, S. Wicks, and R. E. Howitt, "Performance evaluation of methods for the estimation of soil hydraulic parameters and their suitability in a hydrologic model," *Gepderma*, vol. 134, no. pp. 135-151, 2006.
- [3] H. Merdun, Ö. Cinar, R. Meral, and M. Apan, "Comparison of artificial Neural Network and regression pedo transfer functions for prediction of soil water retention and saturated hydraulic conductivity," *Soil and Tillage Research*, vol. 90, no. 1-2, pp. 108-116, 2006.
- [4] J.Tomasella, Y.Pachepsky,S. Crestan,and W.J.Rawls, "Comparison of two techniques to develop pedotransfer function for water retention," *Soil Science Society of America Journal*,Vol. 67,no.4,pp. 1085-1092 , 2003.
- [5] B. Minasny, J. W. Hopmans, T. Harter, S. O. Eching, A. Tuli, and M. A. Denton, "Neural Networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data," *Soil Science Society of Americal Journal*, vol. 68, no. 2, pp. 417-429, 2004.
- [6] Y. A. Pachepsky, W. J. Rawls, and H. S. Lin, "Hydropedology and pedotransfer functions," *Geoderma*, vol. 131, no. 3-5, pp. 308-316, 2006.
- [7] J. H. M. Wösten, P. A. Finke, and M. J. W. Jansen, "Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristic," *Geoderma*, vol. 66, no. 3-4, pp. 227-237, 1995.
- [8] J. Bouma and J. A. van Lanen, "Transfer functions and threshold values: from soil characteristics to land qualities," in proceedings of the International Workshop on Quantified Land Evaluation procedures, K. J. Beek, P. A. Burrough, and D. E. McCormack, Eds., pp. 106-111, International Inst. For Aerospace Survey and Earth Sciences, Washington, DC, USA, May 1987.
- [9] H. S. Lin, K. J. McInnes, L. P. Wilding, and C. T. Hallmark, "Effects of soil morphology on hydraulic properties, II. Hydraulic pedotransfer functions," *Soil Science Society of America Journal*, vol. 63, no. 4, pp. 955-961, 1999.
- [10] T. Mayr and N. J. Jarvis, "Pedo transfer function to estimate soil water retention parameter for modified

- Brooks -Corey type model, Geoderma, vol. 91, no. 1-2, pp. 1-9 1999.
- [11] B. Minasny and A. B. McBratney, "The neuroin method for fitting neuralociety of Americal Journal, vol. 66, no. 2, pp. 352-361, 2002.
- [12] L. Baker and D. Ellison, "Optimisation of pedotransfer functions using an Artificial neural network ensemble method," Geoderma, vol. 144, no. 1-2, pp. 212-224, 2008.
- [13] F. Z. Panagou, V. Kodogiannis, and G. J. F., Nychas, "Modelling fungal growth using radial basis function neural networks networks: the cases of the ascomycetous funhus *Monascus rubber van Tieghem*," International Journal of food Microbiology, vol. 117, no. 3, pp. 276-286, 2007.
- [14] J. Tomasella, M. G. Hodnett, and L. Rossato, "Pedotrasfer functions for the estimation of soil water retention in Brazilian soils," Soil Science Society of America Journal, vol. 64, no. 1, pp. 327-338, 2000.
- [15] J. H. M. Wosten, Y. A. Pachepsky, and W. J. rawls "Pedotransfer function: bridging the gap between available basic soil data and missing soil hydraulic characteristics," Journal of Hydrology, vol. 251, no.3-4, pp.123-150, 2001.
- [16] R. Wieland and W. Mirschel, "Adaptive fuzzy modelling versus artificial neural networks, "Environmental modelling and software, vol. 23, no. 2, pp. 215-224, 2008.
- [17] K. Parasuraman, A. Elshorbagy, and B. C. Si, "Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles," Soil Science Society of America Journal, vol. 70, no. 6, pp. 1851-1859, 2006.
- [18] I. Khalil, M. T. Distribution of Manganese oxide and hydroxide in some calcareous soils ,Nineveh province , Mosul University, college of Agriculture and Forestry Msc, in soil and water science, 2007.
- [19] H. Bagweta, A. A. "Distribution of trace element in some alluvial Iraqi soil "Msc Thesis Baghdad University, College of Agriculture ,1983.
- [20] H. P. Ritzema, Drainage Principles and Applications, ILRI, The Netherlands, 1994.
- [21] W. A. Agyare, S. J. Park, and P. L. G. Vlek, "Artificial neural network estimation of saturated hydraulic conductivity," Vadose Zone Journal, vol. 6, no. 2, pp. 423-431, 2007.
- [22] S. Tamari, J. H. M. Wosten, and J. C. Ruiz-Suarez, "Testing an artificial neural network for predicting soil hydraulic conductivity," Soil Science Society of America Journal, vol. 60, no. 6, pp. 1732-1741, 1996.
- [23] I. Yilmaz, M. Marschalko, M. Bednarik, O. Kaynar, and L. Fojtova, "Neural computing models for prediction of permeability coefficient of coarse-grained soil," Neural computing and Applications, vol. 21, no. 5, pp. 957-968, 2012.
- [24] M. G. Schaap and F. J. Leij, "Using neural networks to predict soil water retention and soil hydraulic conductivity," Soil and Tillage Research, vol. 47, no. 1-2, pp. 37-42, 1998.
- [25] Y. A. Pachepsky, D. Timlin, and G. Varallyay, "Artificial neural networks to estimate soil water retention from easily measurable data," Soil Science Society of America Journal, vol. 60, no. 3, pp. 727-733, 1996.
- [26] M. Kim and J. E. Gilley, "Artificial Neural Network estimation of soil erosion and nutrient concentrations in runoff from land application areas," Computers and Electronics in Agriculture, vol. 64, no. 2, pp. 268-275, 2008.