Vol. 18, No. 2, pp. 173-180 (2018)
Journal of Agricultural Physics
ISSN 0973-032X
http://www.agrophysics.in

Research Article

Comparison of Artificial Neural Network and Multi-linear Regression for Prediction of Field Capacity Soil Moisture Content

PRIYA BHATTACHARYA[1], P. PRAMANIK[1]*, MRINMOY RAY[2] AND P. KRISHNAN[1]

[1]*Division of Agricultural Physics, ICAR- Indian Agricultural Research Institute, New Delhi-110012*
[2]*ICAR-Indian Agricultural Statistical Research Institute, New Delhi-110012*

ABSTRACT

Soil moisture content at field capacity (SWC_{FC}) is vital hydraulic soil parameters essential for modelling water and solute transport and irrigation scheduling in soil. SWC_{FC} can be measured directly in the laboratory using pressure plate apparatus but, their direct measurements are labor intensive, time taking and expensive. Pedotransfer functions (PTF) are usually used as indirect method of measuring soil hydraulic properties from some easily measurable soil properties. PTFs based on multiple-linear regression model (MLR) which considered linearity between the input and output variables, is the most commonly used indirect method of estimating for SWC_{FC}. In the recent years, Artificial Intelligence (AI) techniques based on artificial neural networks (ANN) are fetching more attention of the researchers for their capability of modeling complex "input-output" relationship. In this study, an attempt has been made to estimate SWC_{FC} from sand, silt, clay, bulk density (BD) and organic carbon (OC) using MLR and ANN and compare the performance of these two methods using statistical evaluation criteria. The developed MLR model showed that SWC_{FC} was positively correlated with clay, silt and OC (%) whereas, it was negatively correlated with sand (%) and BD of the soil. Results showed that ANN with four hidden layer was better in predicting SWC_{FC} because of its non-linear function. For SWC_{FC}, ANN resulted a decrease of RMSE by 25.6% as compared to MLR. Performance of ANN was better than MLR model in prediction of SWC_{FC} because ANN can capture non-linear functions.

Key words: Multilinear regression, Artifificial neural network, Soil water content at field capacity, Pedo-transfer function

Introduction

Soil moisture content at field capacity (SWC_{FC}) is an important hydraulic property of soil required in modelling of water and solute transport (Rab *et al*., 2011), biophysical model (Coºkun and Candemir, 2014) and determining the suitable time and rate of irrigation water to be applied (Jafar Nejadi *et al*., 2012). SWC_{FC} is the maximum water content of soil after 48 or 72 hours being watered and when the free drainage is insignificant. SWC_{FC} can be measured directly in the laboratory using pressure plate extractor technique (Richards, 1948) but, their direct measurements are labor intensive, time taking and expensive (Mohanty *et al*., 2015). Also, the results may not be precise because of spatial (Hopmans *et al*., 2002) and temporal variability of soil physical and hydraulic properties (Merdun *et al*., 2006). Precise measurement of hydraulic property like SWC_{FC} is very much required for hydrological modelling and optimizing agro-nomical practice as it controls the availability of soil water for plant growth (Lamorski *et al*.,

*Corresponding author,
Email: pragati.iari@gmail.com

2014). In several cases, hydraulic parameters of soil can be estimated instead of measured. Pedotransfer functions (PTF) are usually used in such conditions (Wösten *et al*., 2001) and permit the estimation of SWC_{FC} from some easily measurable soil properties. Lately, PTFs has gained substantial importance (Minasny and McBratney, 2002 and Minasny *et al*., 2004) as an alternative and indirect estimation of these properties from extensively obtainable or more easily measurable basic soil properties like sand, silt, clay, bulk density (BD), and organic matter (OM). Easily measurable soil properties are used as input variables in PTF and produces hydraulic parameters as output variable (Bouma, 1989). The most common method of estimation of soil hydraulic parameters using PTFs is multiple-linear regression model (MLR) (Wösten *et al*., 1995; Salchow *et al*., 1996; Mayr and Jarvis, 1999; Tomasella *et al*., 2000) where it is assumed that the relationships between input and output variables are linear.

Artificial Intelligence (AI) techniques based on artificial neural networks (ANN) are fetching more attention of the researchers for their capability of modeling complex "input-output" relationship (Jafarzadeh *et al*., 2015). The benefit of ANN is their capability to model the structure of complex systems by changing the weightage of each network component and it can also adjust the interconnections structure and they do not require a priori concept for the relations between the input variables and output data (Gociæ *et al*., 2015). After forming a network structure with coefficients values for describing the strength of influence of each network components, an ANN turn into a superior complex formula which can easily link input variables to output variables (Alexander and Morton, 1990). The multifaceted relationships between basic soil properties and hydraulic parameters have enabled ANN to become a good tool for estimating difficult soil parameters. Goh (1995) reported the superior efficiency and reliability of ANN in prediction as compared to statistical method. Erzin *et al*. (2009) also observed the superiority of ANN as compared to statistical method. So, the objective of this study is to estimate the SWC_{FC} using MLR and ANN and compare the performance of these two methods using statistical evaluation criteria.

Materials and Methods

Study area, collection of soil samples and soil analysis

The study was conducted in Nilokheri block of Karnal district of Haryana, India. The climate of the district is characterized by dry air with intense hot summer and cold winter. The cold season is from middle of November to middle of March. The hot season continues to about June. The summer monsoon arrives at the end of June and South west monsoon prevails from July to September. Altogether 120 soil samples were collected from the study area. The input parameters considered for the prediction of SWC_{FC} were soil texture i.e sand, silt and clay percentage, Walkley–Black organic carbon (OC,%) and soil bulk density (BD, Mg m^{-3}). Sand, silt and clay percentages were estimated by Bouyocous hydrometer method (Bouyocous, 1962). Organic carbon percentage was measured by following the wet oxidation method of Walkley and Black (1934). BD was determined by core auger method (Blake and Hartge, 1986).

Multi-linear regressions

Multiple linear regression model can be written as (equation 1)

$$y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} + \varepsilon_i \quad \text{... (1)}$$

Where, y is output variable and x is input variables and β's are the coefficient values and ε_i is a random error term with mean zero and constant variance.

Artificial neural network (ANN)

A neural network is made out of various interconnected simple processing elements called neurons or nodes (Hastie *et al*., 2001). Each node receives an input signal which is the aggregate ''information'' from other nodes or external stimuli, processes it locally through an activation or transfer function and produces a transformed output signal to other nodes or external outputs.

The ANN performs the following nonlinear function mapping between the input and output (equation 2)

$$y_i = f(x_{i1} + y_{i2}, ..., y_{ip}, w) + \varepsilon_i \quad ...(2)$$

Therefore, the neural network resembles a nonlinear regression model. In general, an ANN can be partitioned into three sections, named layers, which are known as: i)input layers, ii) hidden layers and iii) output layers. Schematic diagram of ANN has been presented in figure 1.

Evaluation criteria of ANN and MLR

Root mean square error (RMSE) =

$$\frac{1}{N}\sqrt{\sum_{i=1}^{N}(Pi - Oi)^2} \quad ...(3)$$

Correlation Coefficient (r) =

$$\frac{\sum(xi-\bar{x})(yi-\bar{y})}{\sqrt{\sum(xi-\bar{x})^2\sum(yi-\overline{y})^2}} \quad ...(4)$$

where

r = correlation coefficient of the linear relationship between the variables x and y

xi = values of the x-variable in a sample

x̄ = mean of the values of the x-variable

yi = values of the y-variable in a sample

ȳ = the mean of the values of the y-variable

Mean absolute percent error (MAPE) =

$$\frac{100}{N}\sum_{i=1}^{N}\frac{|Pi-Oi|}{Pi} \quad ...(5)$$

Mean absolute error (MAE) =

$$\frac{\sum_{i=1}^{N}|Pi-Oi|}{N} \quad ...(6)$$

All the analysis of data has been done in R software and MS-Excel.

Results and Discussion

The value of SWC_{FC} varied from 10.76 to 35.18% w/w for whole dataset (Table 1). In table 1, mean values of training data of SWC_{FC} was 20.07% w/w with a range varied from 10.76 to 30.51% w/w. The SD for SWC_{FC} was 3.97% w/w. The mean values of training data of sand, silt and clay were 53.7, 25.0 and 21.2%, respectively and corresponding SD were 9.71, 10.47 and 8.32%, respectively. The mean and SD values of training data of BD and OC were 1.47 Mg m^{-3}; 0.50%, and 0.17 Mg m^{-3}; 0.22% respectively. For testing data of SWC_{FC}, mean and SD values were 20.13 and 4.43% w/w, respectively. The mean values of sand, silt and clay content of testing dataset were 53.27, 26.07 and 20.77%, respectively and corresponding SD were 11.76, 11.34 and 9.44%, respectively. The mean values of BD and OC in testing data were

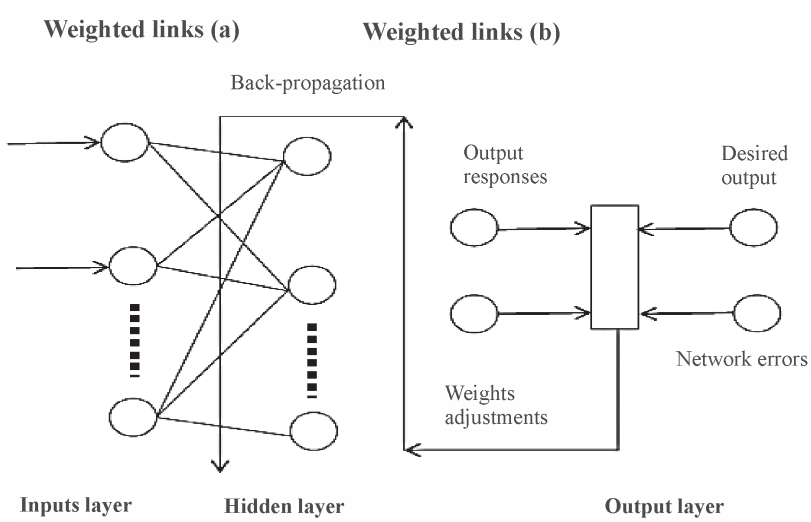


Fig. 1. Schematic diagram of AAN (Khlosi *et al.*, 2016)

Table 1. Descriptive statistics used to predict SWC_{FC}, sand, silt, clay BD and OC for both training and testing dataset

Parameters	Training dataset				Testing dataset			
	Minimum	Maximum	Mean	SD	Minimum	Maximum	Mean	SD
SWC_{FC} (%w/w)	10.76	30.51	20.07	3.97	12.76	35.18	20.13	4.43
SAND (%)	36	80.12	53.76	9.71	24	68	53.27	11.76
SILT (%)	4	48	25.00	10.47	4	48	26.07	11.34
CLAY (%)	1.88	43.88	21.22	8.32	7.88	48	20.77	9.44
BD (Mgm^{-3})	1.12	1.85	1.47	0.17	1.24	1.78	1.47	0.13
OC (%)	0.02	1.1	0.50	0.222	0.22	1.2	0.561	0.23

found to be 1.47 Mgm^{-3} and 0.56%, respectively, whereas, the corresponding SD values were 0.13 and 0.23 (Table 1).

Prediction of Soil water content at SWC_{FC} using multi-linear regression (MLR) model and Artificial neural network (ANN)

In case of predicting SWC_{FC} using MLR sand, silt, clay, BD and OC were used as independent variables. The developed model for prediction of SWC_{FC} was represented in equation 7.

$$SWC_{FC} = 18.72+0.06clay+0.08Silt-0.18\ sand-2.39BD+2.60OC \quad \ldots(7)$$

The developed model showed that SWC_{FC} is positively correlated with clay, silt and OC (%) whereas, it is negatively correlated with sand (%) and BD of the soil. The observed and predicted values of SWC_{FC} both for training and testing data using MLR have been presented in Fig. 2.

Soil SWC_{FC} (output variable) was predicted using ANN and sand, silt, clay, BD and OC as input variables. The optimal number of hidden layers was selected using a trial-and-error process by changing the number of hidden neurons from 1 to 10. In this study, it was observed that for prediction of SWC_{FC}, 4 no. of hidden layer performed better than MLR for both training and testing dataset with the value of r, being higher and RMSE, MAE and MAPE values being lower (Fig. 3). Among all the hidden layers in ANN, ANN with 4 number of hidden layers, were selected based on the RMSE, r, MAE and MAPE values that were found to be 2.81,0.70, 2.09 and 10.79, respectively (Fig. 3), for training dataset and corresponding values for testing dataset were 4.55,0.47, 3.87, 19.84, respectively (Fig 3). Results indicated that ANN with 4 number of hidden layer performed best both for training and testing dataset among all the other hidden layers. Fig. 5 showed the developed ANN model for prediction of SWC_{FC} with 4 numbers of hidden layers. The model run for 3073 steps to train the classifier error was of 0.97. Results showed that artificial neural network with four neurons in

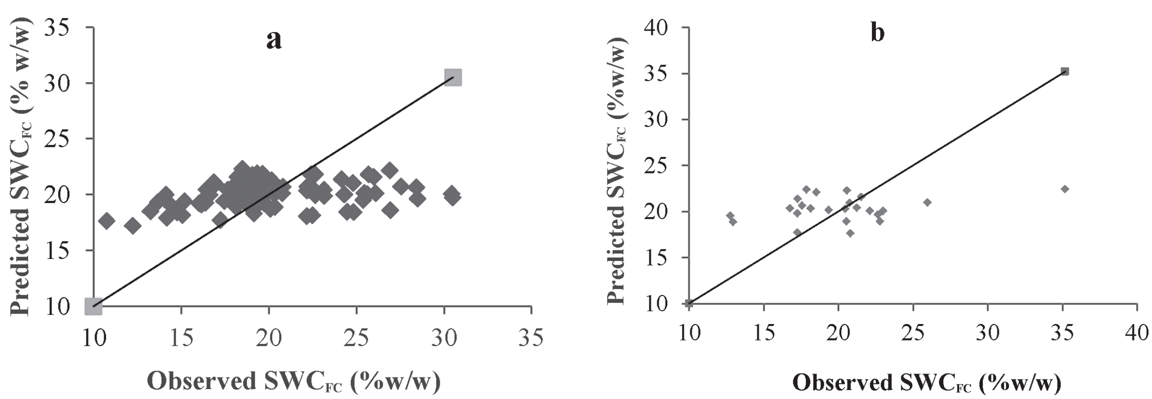


Fig. 2. Observed SWC_{FC} vs. predicted SWC_{FC} (a) training dataset (b) testing data set using MLR

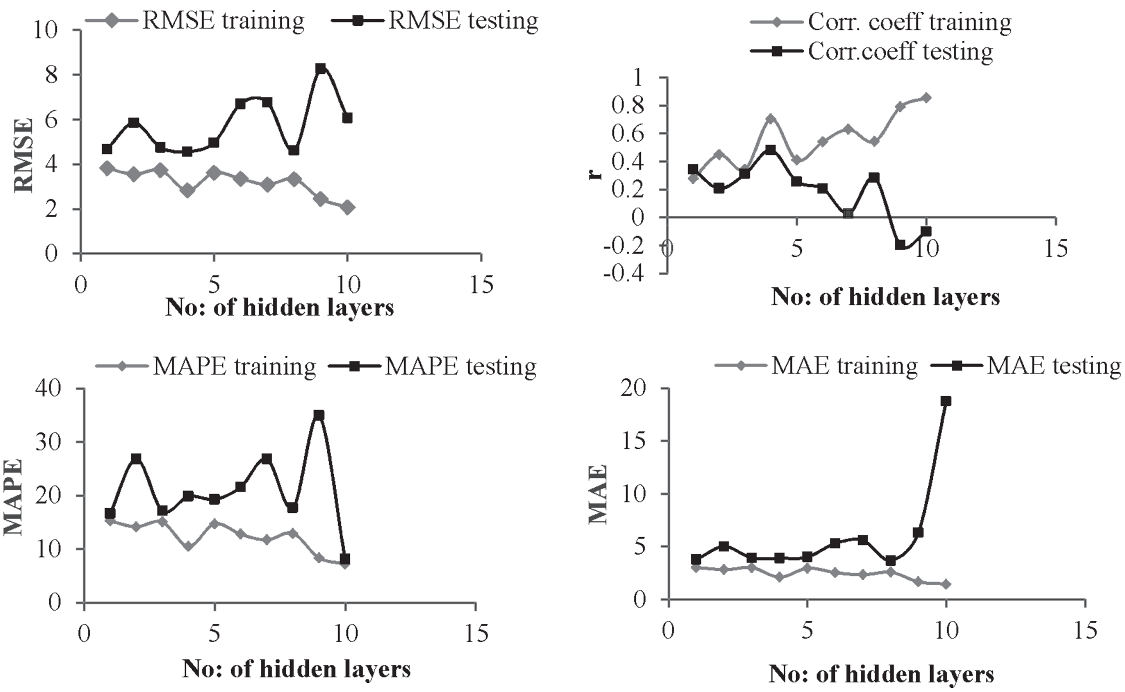
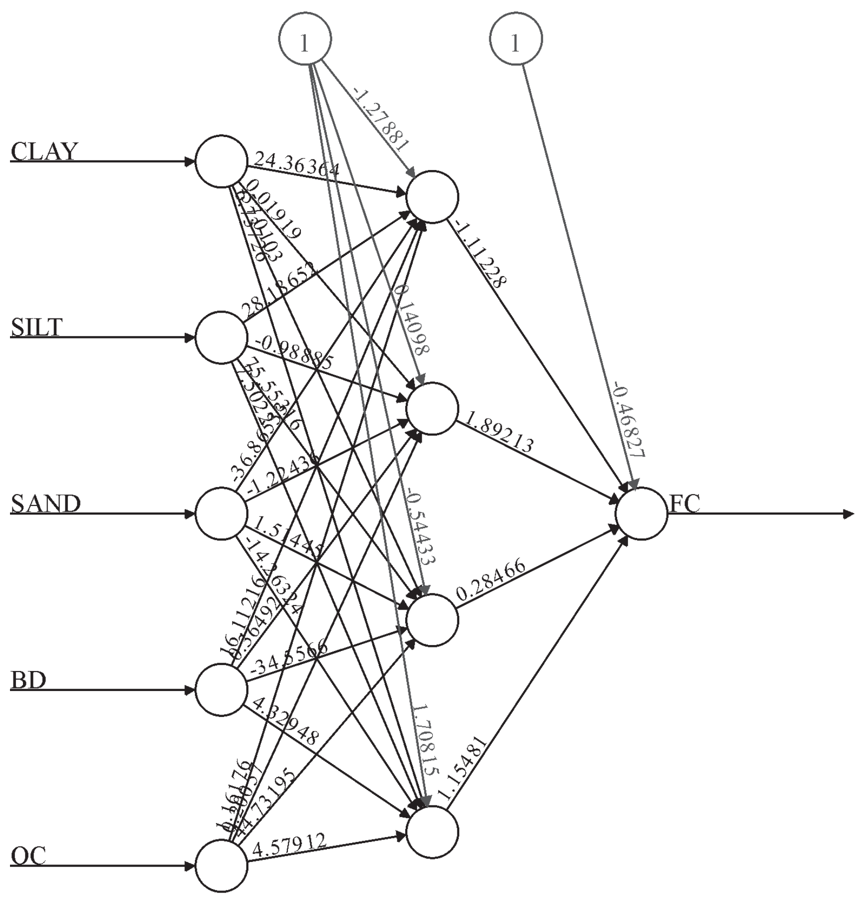


Fig. 3. Performance evaluation of ANN with 1 to 10 numbers of hidden layers



Error: 0.978137 Steps: 3073

Fig. 4. ANN model with 4 hidden layers

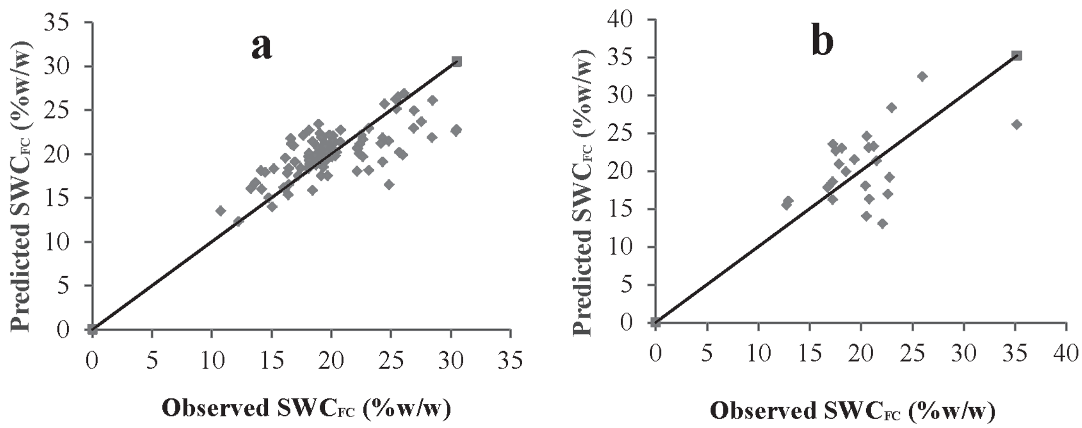


Fig. 5. Observed SWC_{FC} vs. predicted SWC_{FC} (a) training dataset (b) testing data set using ANN with 4 hidden layers

Table 2. Comparative performance evaluation of MLR and ANN in prediction of SWC_{FC}

Parameters		MLR				ANN			
		MAE	r	RMSE	MAPE	MAE	r	RMSE	MAPE
SWC_{FC}	Training dataset	3.02	0.28	3.78	15.2	2.09	0.70	2.81	10.4
	Testing dataset	3.11	0.32	4.11	15.2	3.87	0.47	4.55	19.8

hidden layer had better performance in predicting soil properties than training multivariate regression. Fig. 5 showed the observed and predicted SWC_{FC} for both training and testing dataset using ANN with 4 hidden layers.

Performance evaluation of MLR model was done for both training and testing data set (Figure 3 and Table 2). Correlation coefficient (r), RMSE, MAE and MAPE of training dataset were 0.28, 3.78, 3.02 and 15.22, respectively (Table 2), whereas for testing dataset corresponding values were 0.32, 4.11, 3.11, 15.23, respectively. For SWC_{FC}, ANN resulted in a decrease of RMSE by 25.6% as compared to MLR. Similar trend of decrease in RMSE by 25.66% as compared to MLR was observed in case of ANN for testing data. Among the two models, highest r value of 0.70 and 0.47 for training and testing dataset was obtained in ANN. ANN showed lowest MAPE and MAE values as compared to MLR. Thus, table 2 indicated that the performance of ANN was better than MLR in prediction of SWC_{FC}.

In the present study, the performance of MLR was poor as compared to ANN. Work done by Pramanik *et al.* (2013) showed the effectiveness of regression based PTFs in estimating soil penetration resistance from soil bulk density and soil water content. On contrary, regression-based PTFs (Rawls and Brakensiek, 1985; Minasny *et al.*, 1999) have often performed poorly because they require adequate prior knowledge of the input-output relationships. In the present study, it was found that ANN performed better than MLR which was in accordance to the work done by Amini *et al.*(2005), Erzin *et al*. (2009), Akbulut (2005) and Mohanty *et al*. (2015). This could be attributed to the ability of neural networks to extract maximum information out of the data (Amini *et al.*, 2005).

Conclusion

From the above study, it can be concluded that, ANN with four hidden layer was best in predicting SWC_{FC} because of its non-linear function. For SWC_{FC}, ANN resulted a decrease of RMSE by 25.6% as compared to MLR. Performance of ANN was better than MLR model in prediction of SWC_{FC} because ANN can capture non-linear functions. ANN should be tried for different combination of training and testing dataset as it also has lots of potential in prediction of other soil properties.

Acknowledgement

Financial assistance provided by NRDMS, DST and I.A.R.I in the form of I.A.R.I Fellowship for carrying out the study is gratefully acknowledged.

References

Akbulut, S. 2005. Artificial neural networks for predicting the hydraulic conductivity of coarse grained soils. *Eurasian Soil Science* **38**: 392-398.

Alexander, I. and Morton, H. *1990. An introduction to neural computing*. Chapman and Hall, London.

Amini, M., Abbaspour, K.C., Khademi, H., Fathianpour, N., Afyuni, M. and Schulin, R. 2005. Neural network models to predict cation exchange capacity in arid regions of Iran. *European Journal of Soil Science* **56**: 551-559.

Blake, G.R. and Hartge, K.H. 1986. Bulk density *Methods of soil analysis: part 1—physical and mineralogical methods* **28(6)**: 363-375.

Bouma, J. 1989. Land qualities in space and time. In J. Bouma and A.K. Bregt (ed.) Land qualities in space and time. Pudoc, Wageningen, The Netherlands. pp. 3-13.

Bouyoucos, G.J. 1962. Hydrometer method improved for making particle size analyses of soils 1. *Agronomy Journal* **54(5)**: 464-465.

Coºkun, G., Candemir, F., 2014. Using soil moisture constants and physical properties to predict saturated hydraulic conductivity. *European Journal of Soil Science* **3**: 77-81.

Erzin, Y., Gumaste, S. D., Gupta, A. K. and Singh, D. N. (2009). Artificial neural network (ANN) models for determining hydraulic conductivity of compacted fine-grained soils. *Canadian Geotechnical Journal* **46(8)**: 955-968.

Gociæ, M., Motamedi, S., Shamshirband, S., Petkoviæ, D., Ch, S., Hashim, R. and Arif, M. 2015. Soft computing approaches for forecasting reference evapotranspiration. *Comput. Electron. Agric.* **113**: 164-173. http://dx.doi.org/10.1016/j.compag.2015.02.010.

Goh, A.T.C. 1995. Modeling soil correlations using neural network. *J. Comput. Civ. Eng.* **9**(4): 275-278.

Hastie, T., Tibshirani, R.J. and Friedman, J. 2001. The elements of statistical learning: Data Mining, Inference and Prediction. Springer, New York.

Hopmans, J.W., D.R. Nielsen, and K.L. Bristow. 2002. How useful are small-scale soil hydraulic property measurements for large-scale vadose zone modeling. p. 247–258. In D. Smiles *et al* (ed.) Heat and mass transfer in the natural environment. The Philip Volume. Geophysical Monogr. Ser. 129. AGU, Washington, DC.

Jafarnejadi, A.R., Abbssayyad, G., RezaeArshad, R. and Davami, A. 2012. Pedotransfer function development for field capacity and permanenent wilting points using artificial neural networks and regression models. *International Journal of Agriculture Res. Rev.* **2**: 1079-1084.

Lamorski Krzysztof, Cezary S Bawi Nski, Felix Moreno, Gyöngyi Barna, Wojciech Skierucha, and José L. Arrue. 2014. Modelling Soil Water Retention Using Support Vector Machines with Genetic Algorithm Optimisation. *Scientific World Journal*. DOI. http://dx.doi.org/10.1155/2014/740521

Mayr, T. and Jarvis, N.J. 1999. Pedotransfer functions to estimate soil water retention parameters for a modified Brooks-Corey type model. *Geoderma* **91**: 1-9.

Merdun, Hasan, Ozer Cýnar, Meral Ramazan and Apan Mehmet. 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. *Soil and Tillage Research* **90**: 108-116.

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

Minasny, B., Hopmans, J.W., Harter, T., Eching, S.O., Tuli, A. and Denton, M.A. 2004. Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. *Soil Science Society of America Journal* **68**, 417–429. DOI: 10.2136/sssaj2004.4170

Minasny, B. and McBratney, A.B. 2002. The neuro-m methods for fitting neural network parametric pedotransfer functions. *Soil Science Society of America Journal* **66**: 352-361.

Mohanty, M., Nishant, K.S., Painuli, D.K., Bandyopadhyay, K.K., Hati, K.M., Reddy, K.S. and Chaudhary, R.S. 2015. Modelling soil water contents at field capacity and permanent wilting point using artificial neural network for Indian soils. *Natl. Acad. Sci. Lett*. **38** (5): 373-377.

Pramanik, P., Aggarwal, P. and Dey, P. 2013. Model for calculation of penetration resistance from easily measurable soil physical properties. *Indian Journal of Agricultural Sciences* **83**(3): 294-299.

Rab, M.A., Chandra, S., Fisher, P.D., Robinson, N.J., Kitching, M., Aumann, C.D. and Imhof, M. 2011. Modelling and prediction of soil water contents at field capacity and permanent wilting point of dryland cropping soils. *Soil Res.* **49**: 389-407.

Rawls, W.J. and D.L. Brakensiek. 1985. Prediction of soil water properties for hydrologic modelling. pp. 293–299. In E. Jones and T.J. Ward ed. Proc. Of Symp. ASCE, Denver, CO, 30 April–2 May 1985. ASCE, New York.

Richards, L.A. 1948. Porous plate apparatus for measuring moisture retention and transmission. *Soil Science* **66**(2): 105-110.

Salchow, E., Lal, R., Fausey, N.R. and Ward, A. 1996. Pedotransfer functions for variable alluvial soils in southern Ohio. *Geoderma* **73**: 165-181.

Tomasella, J., Hodnett, M.G. and Rossato, L. 2000. Pedotransfer functions for the estimation of soil water retention in Brazilian soils. *Soil Science Society of America Journal* **64**: 327-338.

W¨osten, J. H.M., Pachepsky, Y.A. and Rawls, W.J. 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics, *Journal of Hydrology* **251**(3-4): 123-150.

Walkley, A. and Black, I.A. 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science* **37(1)**: 29-38.

Wösten, J.H.M., Finke, P.A. and Jansen, M.J.W. 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. *Geoderma* **66**(3–4): 227–237. doi:10.1016/0016-7061(94)00079-P.

Received: October 10, 2018; Accepted: November 29, 2018