

Comparative Study of Statistical, Numerical and Machine Learning-Based Pedotransfer Functions of Water Retention Curve with Particle Size Distribution Data

S. Amanabadi^{a, b, *}, M. Vazirinia^a, H. Vereecken^c, K. Asefpour Vakilian^d, and M. H. Mohammadi^b

^aDepartment of Soil Science, Faculty of Agriculture and Natural Resources, Science and Research Branch, Islamic Azad University, Tehran, Iran

^bDepartment of Soil Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran

^cAgrosphere Institute (IBG-3), Forschungszentrum Jülich GmbH, Jülich, Germany

^dDepartment of Agrotechnology, College of Abouraihan, University of Tehran, Tehran, Iran

*e-mail: amanabadi@ut.ac.ir

Received August 22, 2019; revised September 10, 2019; accepted September 27, 2019

Abstract—The water retention curve (WRC) describes the nonlinear relation of soil water content (SWC) and matric potential. Since direct measurement of SWC and matric potential is difficult and time consuming, indirect approaches including statistical, numerical, and pattern recognition-based pedo-transfer functions (PTFs) that relate basic soil properties to the WRC have been developed during the last few decades. Although several studies reporting the performance of these models can be found in literature, it seems that an extensive investigation which compares the available models and introduces a reliable method to soil hydrologists can be useful. Therefore, in this study, the performance of multiple linear regressions (MLR) models, scaled numerical models and machine learning methods including artificial neural networks (ANN) and adaptive neuro-fuzzy inference systems (ANFIS) are compared using 98 UNSODA codes with various soil textures to estimate WRC. Results showed that regardless of the soil texture, ANN (RMSE = 0.029) predicts the WRC more accurately than ANFIS (RMSE = 0.035), scaled model (RMSE = 0.060) and MLR (RMSE = 0.071), respectively. Considering the soil texture, ANFIS performance is the best in the moderate and fine textured soils, while scaled numerical model predicts with acceptable performance in sandy soils. WRC prediction using easily available soil characteristics particularly when there is a lack of data, shows that newly developed machine learning methods are capable of predicting WRC considerably accurate for sustainable water flow and solute transport management.

Keywords: Adaptive Neuro-Fuzzy Inference System, artificial neural networks, multiple linear regression, soil water content

DOI: 10.1134/S106422931930001X

INTRODUCTION

Modeling water flow and solute transport in vadose region is commonly carried out using Richards equation and convection-dispersion equation (CDE), respectively, which require soil information including unsaturated soil hydraulic functions and water retention curve (WRC) [63]. WRC is a fundamental soil hydraulic function which controls water flow and solute transport in the unsaturated zone. The relationship between soil water content (SWC) and matric potential is nonlinear. Indeed, WRC modeling is difficult using simple or even complicated mathematical formulas since pedo-transfer functions (PTFs) require many input variables [29]. Due to temporal and spatial variability, direct measurement of hydraulic properties is labor-intensive, costly and time-consuming [38, 68]. Therefore, efforts have

been made to predict the WRC indirectly employing different PTFs [13].

Indirect estimation of soil hydraulic conductivity functions and WRC have gathered the interest of researchers in recent decades [33]. On the other hand, several studies have shown that WRC can be estimated efficiently using soil easily available characteristics such as soil texture, bulk density, organic matter, porosity and particle size distribution (PSD) [27, 31]. PSD is capable of significantly improving WRC prediction particularly in sandy soils. The macro-pores and void ratio of sandy soils make prediction of WRC more accurately using PSD. The micro-pores and physicochemical characteristics of clay soils limit WRC estimation using PSD curve [17, 23].

Machine-learning methods such as artificial neural networks (ANN) and adaptive neuro-fuzzy inference

system (ANFIS) which are newly developed for PTFs are another group of models which have been successfully used to predict hydraulic processes including WRC [61], hydraulic conductivity, and other soil hydrological functions [3, 43, 20]. These methods can predict soil hydraulic characteristics accurately having high flexibility in soil process evaluation [1].

The performance of numerical, multiple linear regression (MLR) and machine learning methods to predict the WRC have been widely discussed in previous studies [8, 25, 52]. This study is a fundamental development on the work carried out by Meskini-Vishkaee et al. [23] who predicted WRC using PSD. To do this, several models including MLR, scaled numerical PTFs, ANN and ANFIS are compared in various soil textures to clarify how soil texture can produce a significant effect on models' performance. The objectives of this study are (a) to evaluate the proposed scaled numerical model using PSD data to predict the WRC, and (b) to compare MLR, scaled numerical PTFs and machine learning-based methods in the WRC prediction in various soil textures selected from UNSODA database.

PRELIMINARIES: STATISTICAL, NUMERICAL AND MACHINE LEARNING MODELS

Progress in Modeling WRC

Statistical, numerical PTFs and neural network models have been used to determine the correlation of basic soil properties (for instance, sand, silt and clay percentages and organic matter content) and WRC parameters [23, 30, 56, 67]. These models can be classified into three groups: the first group describes the relationship of SWC and matric potential by close-form equations [50]. The second group are physico-empirical models to predict WRC PTFs which are derived from physical concept. Multi linear regression and machine learning methods are among these models [30]. Third group is dedicated to conceptual models which are based on some conceptual assumptions without any empirical coefficients. Since several WRC models have empirical coefficient and they are data and scale-dependent, some researchers have attempted to predict WRC using conceptual models. As an example, Mohammadi and Vancloster [44] developed a robust and conceptual transfer function to estimate WRC from PSD. The efficiency of such models does not decrease in case of changing the model scale [60].

Multiple Linear Regression (MLR)

Some statistical methods such as multiple regressions have been used to find a mathematical equation relating the model inputs and outputs [5]. Although these models are based on some hypothesis such as data normality and model structure, they cannot estimate WRC accurately [69] since these models have no

essential flexibility for the modeling of complicated hydraulic process of soil [5].

Regression methods are suggested to be used for the description of quantitative relations between inputs and outputs. According to Pearson correlation coefficient, inputs enter the MLR models to estimate WRC [22]. To construct an MLR model, soil physical characteristics including sand, silt, clay, saturated water content (θ_s), matric potential (h), bulk and particle densities (BD, PD) (in some models, porosity (n)) are considered as input variables, while soil water content (SWC) (θ) of the corresponding matric potential is considered as the output variable. Therefore, the equation of MLR can be defined as Eq. (1)

$$Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \dots + \alpha_n x_n. \quad (1)$$

where Y and $x_1 - x_n$ are the output and input variables, respectively. α_0 is the intercept of MLR PTFs and $\alpha_1 - \alpha_n$ are the multiplication coefficients of MLR PTFs. Available and reliable databases provide a variety of inputs for statistical models and therefore, these models can be widely used to predict the WRC [22, 58].

Khodaverdiloo et al. [26] derived MLR PTFs in some calcareous soils showing that the WRC predictions using MLR PTFs were more accurate than the Rosetta PTFs. ROSETTA software uses modified neural network analysis to estimate soil hydraulic parameters with hierarchical pedo-transfer functions. They also showed that MLR PTFs were not capable of distinguishing between calcareous and non-calcareous soils in the WRC prediction. In another study, comparison of MLR PTF and other implemented PTFs revealed that MLR PTF of Hydraulic Properties of European Soil (HYPRES) database was less accurate than the developed PTFs for Belgian soils, while the MLR PTF of HYPRES database resulted in better performance than Rosetta PTF for European and US soils [51]. Vereecken et al. [30] reviewed PTFs to estimate the van Genuchten–Mualem soil hydraulic properties. Results indicated that MLR PTFs were not capable of predicting the WRC accurately in dry range and model error increased by increasing the absolute matric potential.

Scaled Numerical Models

Although modeling the nonlinear relationships of SWC and matric potential is a matter of importance, recently developed numerical PTFs are capable thoroughly of predicting the WRC. Karup et al. [16] predicted the WRC of undisturbed structured soil using bulk density, organic matter and PSD fraction using a combination of hydrometer and wet-sieving methods. They indicated that silt and clay fractions and organic matter significantly improved the WRC prediction in dry range. They also reported that the proposed PTF predictions were sensitive to bulk density variation.

Mohammadi and Vanclooster [44] developed a robust and conceptual transfer function to estimate the WRC from PSD (MV model). They showed that conceptual models can predict soil hydraulic properties based on some conceptual assumptions without any requirements to empirical coefficients. Mohammadi and Meskini-Vishkaee [45] used the MV model to predict a continuous form of the WRC (MM model). They showed that MV and MM models under-predicted the SWC in dry ranges of WRC. To overcome this limitation, Meskini-Vishkaee et al. [23] used a scaling approach to predict the continuous form of WRC based on PSD data, the model introduced by Van Genuchten [50] (VG model) and particle packing state parameter (MV-VG model).

Meskini-Vishkaee et al. [23] showed that the scaling approach improves the WRC prediction by 30% for all the selected soils compared to MV model, MM model and Rosetta PTF. Empirical parameters of WRC and database dependent PTFs are the error sources of numerical models which describing soil hydraulic functions. Correction of such systematic errors using scaling approaches can significantly develop the performance of WRC model. Meskini-Vishkaee et al. [23] proposed a WRC scaled model based on VG model assuming that the residual water content equals to zero ($\theta_r = 0$) (Eq. (2))

$$\theta = \frac{\theta_s}{\left(1 + (\alpha h)^{n^*}\right)^m}, \quad (2)$$

where θ (L^3L^{-3}) is the SWC corresponding to matric potential h , θ_s (L^3L^{-3}) is the saturated SWC, h (L) is the matric potential, m and α (L^{-1}) are fitting coefficients and n^* is the scaled pore size distribution index (Eq. (3))

$$n^* = \lambda c_n. \quad (3)$$

where c_n is a fitting coefficient obtained from PSD data and λ is defined by Eq. (4).

$$\lambda = \frac{\xi}{\xi_{\max}}. \quad (4)$$

The parameter ξ_{\max} (–) equals to 1.41432 and ξ (–) is a coefficient depending on the arrangement state of soil particles and is defined by Eq. (5)

$$\xi = \frac{1.9099}{e + 1}, \quad (5)$$

where e (–) is the void ratio given by Eq. (6)

$$e = \frac{PD - BD}{BD}. \quad (6)$$

Mohammadi and Vanclooster [44] introduced a conceptual robust model (MV model) to predict the soil matric potential from the particle size based on the pore space geometry (Eq. (7))

$$h_i = \frac{0.543 \times 10^{-4} \xi}{r_i}, \quad (7)$$

where h_i (L) is the matric potential of the i -th fraction size and r_i (L) is the radius of the i -th particle fraction size. Simplified assumptions of MV model which ignore the considerable effects of clay surface forces lead to under-predictions in a dry range of the WRC, despite the fact that MV model can predict WRC accurately since WRC does not depend on database and empirical parameters. Following Arya and Paris model (AP model), θ_i can be calculated by Eq. (8)

$$\theta_i = \theta_s \sum_1^i w_i, \quad (8)$$

where w_i (–) is the particle mass fraction of the i -th fraction. When $\theta_r = 0$, combining Eq. (8) and VG model gives

$$\sum_1^i w_i = \frac{1}{\left(1 + \left(\alpha \frac{0.543 \xi}{r_i}\right)^n\right)^m}. \quad (9)$$

Eq. (9) finally gives VG model parameters (Eq. 10).

$$\theta_i = \frac{\theta_s}{\left(1 + (\alpha h_i)^{n^*}\right)^m}. \quad (10)$$

Eq. (9) is fitted to the PSD data to estimate VG model parameters (m , n^* and α) which are used as the input parameters in Eq. (10) as the WRC predictor model. Since Eq. (9) includes three fitting parameters, it should be used to fit on the full range of PSD data containing at least four measured points. In summary, the fitting of Eq. (9) allows the estimation of WRC parameters (m , n^* , and α). Considering that BD is known and the scaling factor and subsequently n^* can be calculated, the continuous form of WRC will be predicted using Eq. (2).

Artificial Neural Networks (ANNs)

ANN is one of the machine learning methods which can act similar to human neural system to predict the output of the network based on a trained dataset. In ANNs, the learner neurons learn a dataset in the form of sigmoid functions with certain weights and biases. Although there should be a dataset to train ANN models [40], it does not require prior knowledge about the relationships between the model inputs and output [37, 47]. The disadvantage of ANN models is that a large database is generally required for training an ANN [40].

Machine learning methods such as ANNs can be used for WRC prediction in unsaturated soils [56]. ANN determines the correlation of basic soil properties (e.g., the percentage of sand, silt, and clay and organic matter content) and WRC [30, 57]. ANNs can learn highly nonlinear behavior of hydraulic process of soil to predict WRC by finding complicated relations between inputs and output [41, 65, 66]. ANNs have been used to

evaluate WRC for HYPRES and international soil reference and information center (ISRIC) [55]. Haghverdi et al. [6] used ANN model for accurate prediction of SWC in a desirable matrix potential having easily available soil characteristics including sand, silt, clay, bulk density, and organic matter.

An ANN model usually has three layers, namely input, hidden, and output layers. Neurons in input and output layers are responsible for model input and output variables, respectively. The neurons in hidden layer discover the relationships between input and output variables. The optimal number of neurons in hidden layer is usually determined by trial and error technique to obtain the highest performance of ANN [18]. Although assessing hydraulic functions of unsaturated zone are complex, ANN is capable of modeling soil and water interactions. However, as a limitation, ANN is among black-box models which it does not provide us with the existing structural behavior among the input and output variables [7, 47]. Uncertainties of soil physical characteristics due spatial and temporal variability can be explained using ANN models [6].

Adaptive Neuro-Fuzzy Inference Systems (ANFIS)

ANFIS is another machine learning method in which hidden layer neurons learn the relationships between input and output variables by a set of fuzzy rules instead of sigmoid functions [28]. ANFIS model contains a five-layer network layers of which are called input layer, membership function layer, fuzzy rules layer, defuzzification layer, and output layer, respectively [28]. The first layer of ANFIS takes inputs and provides inputs to the next layer. The second layer fuzzifies inputs using different membership functions to indicate membership degree of each input. The third layer of ANFIS performs inference fuzzy rules. The fourth layer defuzzifies the fuzzy values to crisp numbers. The fifth layer simply generates model output using a linear equation [53, 59]. Two learning algorithms can be used to train ANFIS models, i.e., back-propagation and hybrids algorithms. ANFIS is capable of describing complex nonlinear relations of input and output variables with acceptable accuracy. Essentially, ANFIS is a rule-based fuzzy logic model whose rules are developed during the training phase [10, 39]. The trained machine can reliably predict the output even when there are uncertainties and variations in input variables.

Previous studies have indicated that ANFIS can predict WRC [22, 28]. Although ANN is not always suitable for accurate prediction of the WRC and other soil physical characteristics [56], ANFIS is introduced as a reliable method which uses the learning capacity of fuzzy inference system in ANN model [22, 28, 32]. Intensive nonlinear relationship of SWC and matrix potential data pairs and variability of soil physical characteristics raises a broad range of uncertainties of soil hydraulic phenomena. As a solution, ANFIS is

capable of tolerating the uncertainties in soil properties and minimizing the errors of the WRC prediction [15, 32].

ANFIS is widely used to predict physical and chemical soil characteristics [43, 69]. For example, soil cation exchangeable capacity (CEC) has been successfully predicted in paddy lands using ANFIS model [24, 34].

As another work, Si et al. [35] modeled the SWC in extreme arid area using ANFIS. The proposed ANFIS models were compared by different ANN models. The comparison results indicated that ANFIS models were more efficient in predicting SWC than ANN models. Ghorbani et al. [24] compared ANN and ANFIS models with MLR to predict CEC. They showed that performances of ANFIS and ANN were similar while their accuracy was better than MLR. Hosseini et al. [46] predicted the soil phosphorus using different statistical and machine learning methods including ANN, ANFIS and genetic algorithm. They showed that genetic algorithm and ANN performed better than ANFIS for the estimation of soil phosphorus.

Given that ANFIS is a combination of ANN and fuzzy logic, ANN shortcomings such as over-fitting and local minima do not concern ANFIS models [57]. Moreover, the ANFIS PTF can be implemented to predict the WRC without determination of specific soil physical relationships before modeling even when the input data are limited for numerical modeling [9, 14, 35].

MATERIALS AND METHODS

To create a dataset, 98 soil codes including 2187 soil samples with at least four PSD data fraction and continuous water retention curve were selected from UNSODA database to compare the models' performance in different soil textures. UNSODA database contains unsaturated hydraulic characteristics of 780 soil codes from all over the world, especially Europe and America [2]. It provides researchers with reliable and accurate soil hydraulic functions to develop the estimations of water flow and solute transport.

The distribution of the soil textural classes is shown in Fig. 1 and the descriptive statistics of all soil variables for the soil codes are given in Table 1. In MLR modeling, all input and output variables followed normal distribution, input variables had linear correlation with output, and no multi-collinearity was observed among model's input variables. In MLR, ANN and ANFIS models, porosity was considered as one of the input variables for some of UNSODA codes. The MLR and machine learning models were implemented in SPSS 21 (IBM, Armonk, NY, USA) and MATLAB R2015a (The Mathworks Inc., Natick, MA, USA) programming environments, respectively.

The e , ξ and λ parameters in scaled numerical model were calculated using available bulk and particle densities. In most UNSODA soil codes, θ_s is avail-

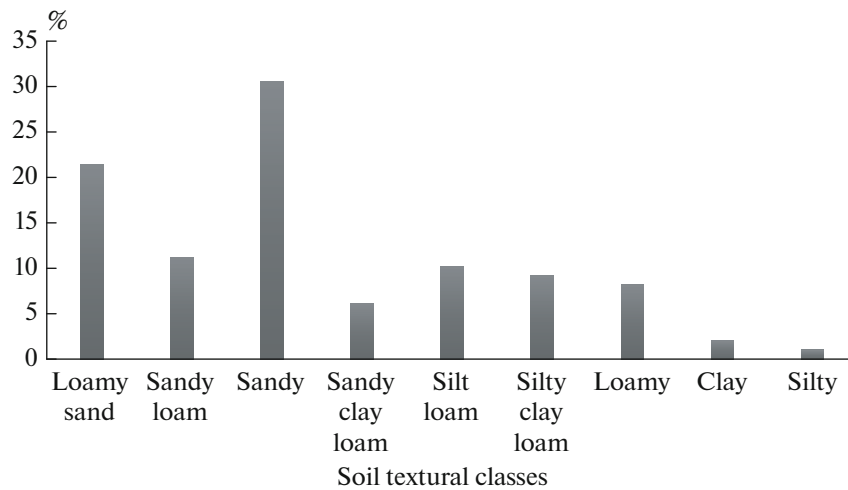


Fig. 1. Distribution of soil textural classes in the selected dataset.

able; for those samples with no θ_s data, the suggestion of Chan and Govindaraju [62] who assumed saturation water content to be equal to the corresponding water content of the lowest measured matric potential was used.

In ANN PTFs, the number of neurons in network hidden layer changed from 5 to 15 and the performance of each network evaluated to obtain the optimum ANN model in prediction of the WRC. In ANFIS modeling, the Gaussian function was used to construct the learning network. This function is the most commonly used membership function in the literature [21, 53] since it is fully defined with only two parameters and its smoothness.

Before the modeling process, the values of input and output variables were normalized using Eq. (11) to make it possible to compare the models based on their

performances. Eq. (11) maps all values of each variable between 0 and 1

$$X_n = \frac{X_m - X_{\min}}{X_{\max} - X_{\min}}, \quad (11)$$

where X_n is the normalized value, X_m is the measured value, and X_{\min} and X_{\max} are the minimum and maximum measured values of each variable, respectively.

All soil codes were used to construct the scaled numerical models. In MLR models, due to the complex nonlinear relationships among SWC, soil physical characteristics and matric potential, dataset was divided into two subgroups to assess the multi regression method (80% of data for training and remaining 20% for test). Similarly, since machine learning methods require test data to evaluate the performance of the model, the database were randomly divided into two groups for ANFIS model: 80% (1750 soil samples) for

Table 1. Descriptive statistics of soil physical properties input* and output** variables

Variable	Minimum	Maximum	Mean	Standard deviation
BD*, kg m ⁻³	1.14	1.97	1.53	0.141
PD*, kg m ⁻³	2.54	2.82	2.64	0.034
n^*	0.257	0.55	0.42	0.051
Sand*, %	0.9	98	63	32.2
Silt*, %	0.5	87	24	22.9
Clay*, %	0.2	54	12.8	12.4
θ_s^*	0.2	0.797	0.428	0.131
K_{sat} , m s ⁻¹	0.091	716.0	184.05	196.12
$ h $, m	0.0	15850	91.78	351.02
$\log h ^*$, m	0.0	4.2	1.743	0.430
θ^{**}	0.0	0.82	0.261	0.104

BD and PD, n , θ_s , K_{sat} , h , and θ indicate bulk and particle densities, porosity, saturated water content, saturated hydraulic conductivity, matric potential, and SWC, respectively.

training and 20% (437 soil samples) for test. One of the limitations of ANNs is being trapped in local minima. To untie this knot, a part of training set is usually dedicated for validation in ANN models and validation performance determines either the training process is completed or not [7]. Therefore, 65% (1421 soil samples) for training the network, 20% (438 soil samples) for validation and 15% (328 soil samples) for test, respectively.

Mean square error (MSE) [61], root mean square error (RMSE) [16, 22, 62], mean absolute error (MAE) [35, 59], mean bias error (MBE) [22, 62], and coefficient of determination (R^2) [9, 17] are widely used to evaluate the efficiency of WRC models. Performance evaluation parameters were computed using Eqs. (12) to (16) between the measured and predicted water content of given matric potential

$$\text{MSE} = \frac{\sum_i^N (\theta_{i(p)} - \theta_{i(m)})^2}{N}, \quad (12)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\theta_{i(p)} - \theta_{i(m)})^2}, \quad (13)$$

$$R^2 = 1 - \frac{\sum_i^N (\theta_{i(m)} - \theta_{i(p)})^2}{\sum_i^N (\theta_{i(m)} - \bar{\theta}_{(m)})^2}, \quad (14)$$

$$\text{MAE} = \sum_{i=1}^N \frac{|\theta_{i(m)} - \theta_{i(p)}|}{N}, \quad (15)$$

$$\text{MBE} = \sum_i^N \frac{\theta_{i(m)} - \theta_{i(p)}}{N}, \quad (16)$$

where N is the number of samples, $\theta_{i(p)}$ and $\theta_{i(m)}$ are predicted and measured water content in the i -th matric potential, respectively, and $\bar{\theta}_{(m)}$ is the mean measured water content. The lower the error parameters and higher the R^2 , the better performance of the model. MAE is a statistical criterion to show the average of error magnitude and MBE is used to show the average bias of each method. The positive value of MBE indicates over prediction.

Because of the random data dividing, the machine learning methods were run for five times to be assured that each soil sample is tested by the models for at least one time and then, the mean values of evaluation criteria for the test samples are reported.

RESULTS AND DISCUSSION

MLR PTFs

Performance evaluation of MLR PTFs according to the input variables is presented in Table 2. RMSE of MLR PTFs ranges from 0.071 to 0.089. MLR5 and

MLR8 have the maximum and minimum RMSEs (0.089, 0.071) among the MLR models, respectively. According to the table, MLR5 and MLR8 have only one different input variable. In other words, replacing particle density with porosity in MLR8 has significantly increased the accuracy of MLR model. MSE, RMSE and R^2 of MLR8 as the most accurate MLR PTF which predicts WRC through easily available soil physical characteristics are 0.005, 0.071, and 0.69, respectively. The negative MBE values of the MLRs reveal that all implemented MLRs under-predict the WRC.

Similar multiplication coefficients of $\log |h|$ in MLR PTFs indicate the important role of matric potential in the prediction of WRC regardless of the kind and number of input variables. MLR4 with six input variables and MLR8 with seven input variables perform similarly. In other words, using silt fraction along with particle density as input variables in MLR8 instead of porosity does not significantly improve the WRC prediction.

The unbiased prediction of WRC is attributed to nature of MLR models, simplified hypothesis of soil hydraulic characteristics modeling, complex relation of SWC and matric potential and limitations of this type of model. Results indicate that increasing the number of input variables do not necessarily lead to improve WRC prediction. This issue indicates that MLR models are not sensitive to the number of input variables. MLRs are based on several assumptions including model structure, normality of residuals and errors, and no or little collinearity of input variables, which limit the MLRs applications. Moreover, predicting WRC using MLR models creates systematic errors because of extensive nonlinear relationships of SWC and matric potential. Another shortcoming is that the MLRs efficiencies considerably depend on dataset. Several studies have shown that MLR models have no essential flexibility for the modeling of complicated soil hydraulic processes [5, 69].

Figures 3a–3h compares the measured vs. predicted soil water content using MLR models. The mean slope and R^2 values of linear regression of various MLR models are 0.6126 and 0.6499, respectively. The $y = x$ line (solid black line) shows the most desirable fitting line and the error of the model will be zero if all the data pairs lie on this line. The area between 1 : 1 plot ($y = x$) and fitted regression lines shows the systematic error [23]. The general trend of MLR models show that WRC is overestimated when the soil water content is less than mean saturated water content of all soil samples, while WRC using MLR models is underestimated when soil water content is raised to more than mean saturated water content of all soil samples. Measured and predicted WRC in scatter plots of Figure 3 shows the undesirable predictions of the WRC by MLR PTFs. This is because of the complex nonlinear relationships of SWC and matric

Table 2. Performance evaluation parameters of MLR PTFs according different input variables

Model	Input variables *	MSE	RMSE	R^2	MAE	MBE	Equation
MLR1	$\log h $, θ_s , Sand	0.0057	0.075	0.65	0.0581	-0.0013	$0.26\theta_s - 0.08\text{Sand} - 0.08\log h + 0.15$
MLR2	$\log h $, θ_s , Sand, n	0.0052	0.072	0.68	0.0558	-0.0013	$0.25\theta_s - 0.14\text{Sand} - 0.08\log h - 0.05n + 0.15$
MLR3	$\log h $, θ_s , Sand, n , BD	0.0052	0.072	0.68	0.0555	-0.0026	$0.25\theta_s - 0.13\text{Sand} - 0.08\log h - 0.04n - 0.05\text{BD} + 0.15$
MLR4	$\log h $, θ_s , Sand, n , BD, Clay	0.0051	0.071	0.69	0.0546	-0.0006	$0.26\theta_s - 0.08\text{Sand} - 0.08\log h - 0.04n - 0.06\text{BD} + 0.08\text{Clay} + 0.11$
MLR5	$\log h $, θ_s , Sand, n , BD, Clay, Silt	0.0081	0.089	0.50	0.0734	-0.0547	$0.26\theta_s + 0.71\text{Sand} - 0.08\log h - 0.04n - 0.06\text{BD} + 0.52\text{Clay} + 0.71\text{Silt} - 0.7$
MLR6	$\log h $, θ_s	0.0075	0.086	0.54	0.0694	-0.0020	$0.38\theta_s - 0.06\log h + 0.028$
MLR7	$\log h $, θ_s , BD	0.0069	0.083	0.57	0.0654	-0.0007	$0.38\theta_s - 0.07\log h - 0.18\text{BD} + 0.11$
MLR8	$\log h $, θ_s , Sand, BD, Clay, Silt, PD	0.0050	0.071	0.69	0.0540	-0.0019	$0.26\theta_s + 0.71\text{Sand} - 0.08\log h - 0.13\text{PD} - 0.06\text{BD} + 0.52\text{Clay} + 0.71\text{Silt} - 0.66$

* h , θ_s , n , BD and PD indicate matric potential, saturated water content, porosity, bulk and particle densities, respectively.

Table 3. Performance evaluation parameters of scaled numerical PTFs for several soil textures

Soil texture	Number of soil	MSE	RMSE	R^2	MAE	MBE
Sandy	30	0.0071	0.084	0.96	0.026	-0.009
Loamy sand	21	0.0040	0.063	0.96	0.039	-0.008
Sandy loam	11	0.0010	0.032	0.98	0.028	-0.010
Silt loam	10	0.0024	0.049	0.97	0.035	0.090
Silty clay loam	9	0.0014	0.038	0.96	0.032	0.010
Loam	8	0.0035	0.059	0.91	0.054	0.045
Sandy clay loam	6	0.0021	0.046	0.95	0.024	-0.013
Clay	2	0.0014	0.038	0.97	0.019	0.055
Silt	1	0.0005	0.023	0.96	0.023	0.005
Weighted average		0.0040	0.060	0.95	0.032	0.017

potential and MLR models are not flexible enough to simulate such nonlinear behavior.

Scaled Numerical PTFs

The results of scaled numerical PTFs with respect to the soil texture are presented in Table 3. Scaled numerical PTF predicts the WRC (weighted average RMSE = 0.060) more accurately than the MLR8 PTFs (RMSE = 0.071). Although the input variables of both MLR8 and scaled numerical PTF are PSD, bulk and particle densities, and h , their predictions are considerably different. The RMSE of scaled numerical PTFs ranges from 0.023 (silty soils) to 0.084 (sandy soils). The comparison of RMSEs for sandy (0.084) and clay (0.038) soils indicates that scaled numerical PTF predicts the WRC of fine textured soil significantly better than that of the coarse textured soils.

Furthermore, RMSE for loamy soils (0.059) is similar to that of the weighted average (0.060) showing that the performance of the scaled numerical PTF is still acceptable in medium textured soils. It should be noticed that RMSE for sandy soils (0.084) is approximately 1.5 times higher than that of the loamy soils (0.059). The average MBE value for all soil textures shows that the results of scaled numerical PTFs are over-predicted (Table 3). Maximum over-prediction is seen in clay soils. However, in sandy, loamy sand, sandy loam and sandy clay loam soils, slight under-prediction is observed (Table 3). MV model assumes that all soil particles are spherical, and the soil structure can only influence the soil bulk density. The effects of soil organic matter content, particle surface energy, and lens and film water volume are not supported by this model [44, 45]. Therefore, the prediction error obtained by the scaled numerical model can

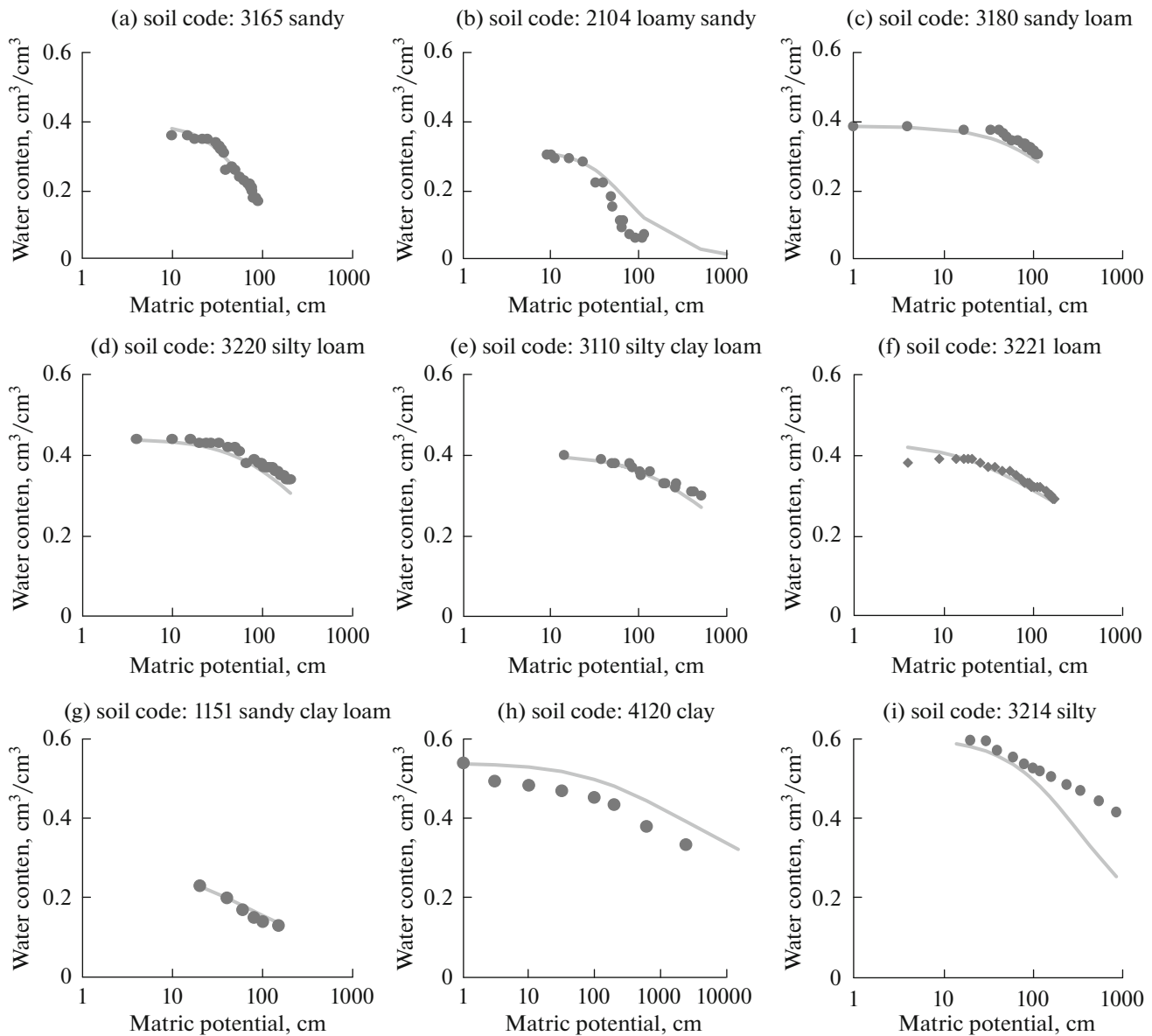


Fig. 2. (a–i). The experimental and estimated WRC using scaled numerical model for various soil textures. Black points are measured data while the curves indicate the predicted values.

be partially attributed to the suggestions made by the MV model. Weighted average RMSEs (0.06) of scaled numerical PTF is similar to that obtained by Meskini-Vishkaee et al. [23].

Figure 2 represents the measured and predicted WRC for clay (code: 4120), loamy (code: 3221), sandy (code: 3165), sandy clay loam (code: 1151), sandy loam (code: 3180), silty loam (code: 3220), silty (code: 3214), silty clay loam (code: 3110) and loamy sand (code: 2104) soils using scaled numerical PTF. Scaled numerical PTF slightly under-predicts the dry range of WRC in sandy loam soil while the model slightly over-predicts the dry range in loamy sand soils. The scaled numerical model in clay over-predicts the full range of WRC

while the full range of WRC in sandy soils is predicted accurately. In sandy loam soil, slight under-prediction of the WRC is seen in dry range while the WRC is predicted accurately in wet range using scaled numerical model. Lack of soil structure in most coarse textured soil reduces the performance of WRC prediction.

Soil structure creates bias towards measured and predicted WRC near saturation which reduces the prediction performance of WRC in the wet range. The inaccurate prediction of WRC at the end regions of the curves can be addressed to linear nature of WRC in the dry range [15]. The WRC of silty clay loam and sandy clay loam soils are properly predicted by scaled numerical PTF. Similar results are reported by Li et al. [17]

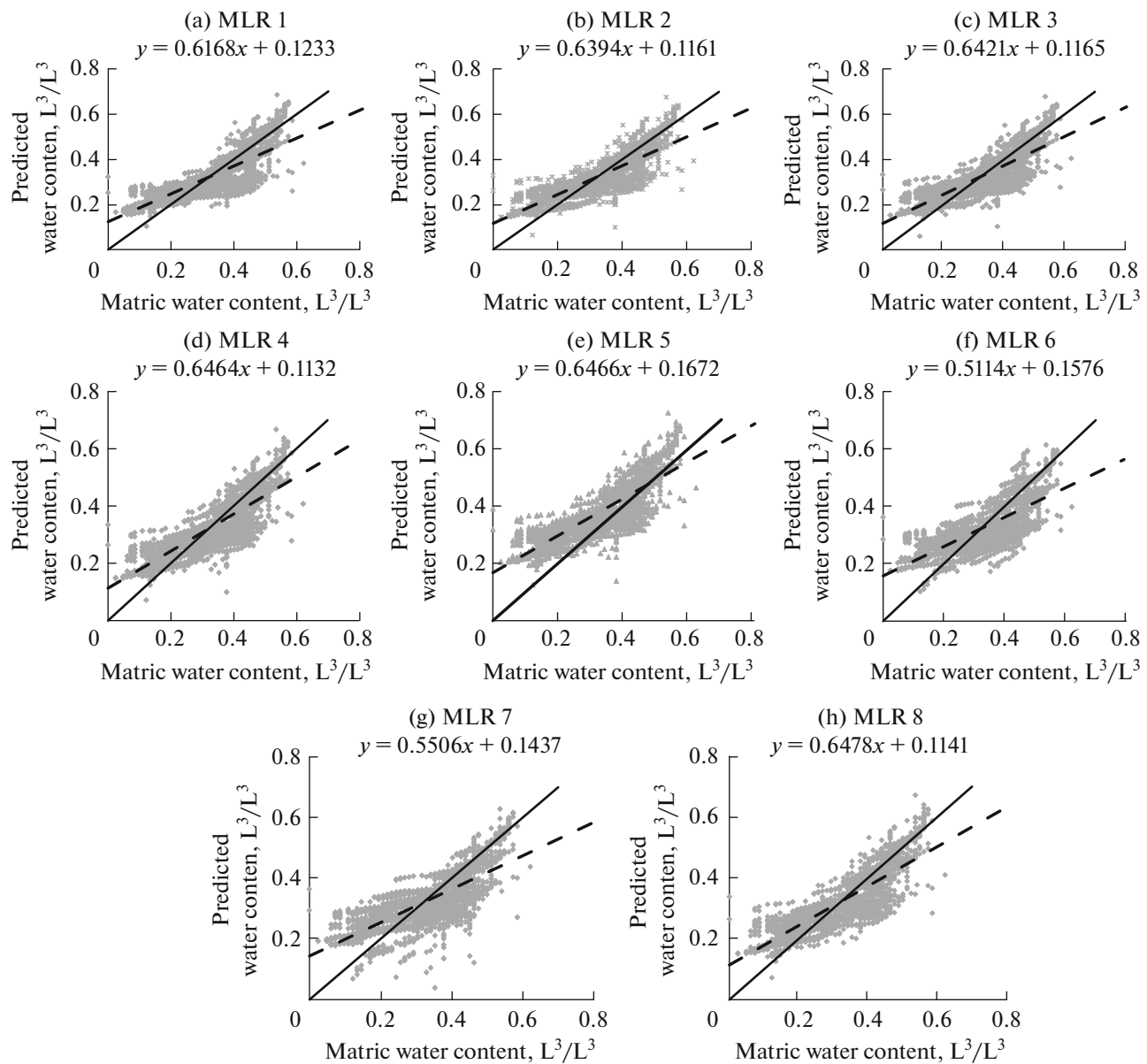


Fig. 3. (a–h). Comparison of measured vs. predicted soil water content of given matric potential using MLR models. Dashed line shows the fitted equation to the predicted values while the solid line indicates the $y = x$.

who compared the performance of AP, MV and Rosetta PTFs to predict available WRC of the corresponding matric potential. They indicated that MV model can provide the best fitted PTFs to measured data in comparison with AP and Rosetta PTFs. The WRC of silty soil is under-predicted using scaled numerical model particularly in dry range. The under-prediction of scaled numerical model can be partially attributed to the assumptions of the MV model.

The predicted full ranges of WRC using scaled numerical PTF in loamy and loamy sand soils have similar trends. The scaled numerical PTF slightly over-predicts the WRC for clay soil. The knowledge of clay mineralogy is capable of justifying over-predic-

tion of WRC for clay soils which is not considered in this study. The organic carbon (OC), clay and silt fraction contents determine the amount of adsorbed water to solid surfaces which cause systematic errors for the WRC prediction.

ANN PTFs

Table 4 shows the results of ANN performance evaluation parameters including MSE, RMSE, R^2 , MAE and MBE. The number of neurons in network hidden layer is changed from 5 to 15 and the highest performance for each model is brought in Table 4. Results reveal that the RMSE of ANN models ranges

Table 4. Performance evaluation parameters of ANN PTFs according different input variables

Model	Input variables*	Neurons N	MSE	RMSE	R^2	MAE	MBE
ANN1	$\log h $, θ_s , Sand	10	0.0026	0.051	0.83	0.055	-0.0014
ANN2	$\log h $, θ_s , Sand, n	6	0.0021	0.046	0.86	0.031	0.0003
ANN3	$\log h $, θ_s , Sand, n , BD	9	0.0012	0.035	0.92	0.027	0.0009
ANN4	$\log h $, θ_s , Sand, n , BD, Clay	11	0.0012	0.035	0.93	0.026	-0.0005
ANN5	$\log h $, θ_s , Sand, n , BD, Clay, Silt	11	0.0008	0.029	0.96	0.026	-0.0006
ANN6	$\log h $, θ_s	10	0.0014	0.037	0.91	0.028	-0.0003
ANN7	$\log h $, θ_s , BD	6	0.0034	0.058	0.78	0.047	0.0004
ANN8	$\log h $, θ_s , Sand, BD, Clay, Silt, PD	10	0.0010	0.032	0.93	0.054	0.0019

* h , θ_s , n , BD and PD indicate matric potential, saturated water content, porosity, bulk and particle densities, respectively.

from 0.029 (ANN5) to 0.058 (ANN7). The comparison of ANN8 (RMSE = 0.032) and ANN5 (RMSE = 0.029) shows that in contrast with MLR PTFs, using porosity instead of PD as an input variable can increase the performance of the ANN model.

The general trend of ANN RMSE indicates that by increasing the number of input variables up to seven including sand, silt, clay, saturated water content, $\log|h|$, and bulk and particle densities, the RMSE moderately decreases. For instance, the RMSE of ANN1 (0.051) with three input variables is considerably more than RMSE of ANN5 (0.029) with seven input variables. Similar results are reported by Nemes et al. [12] and Vereecken et al. [30]. The negative MBE values of ANN1, ANN4, ANN5 and ANN6 indicate that some of the ANN PTFs under-predict the WRC while ANN2, ANN3, ANN7 and ANN8 slightly over-predict the WRC. The ANN5 model is significantly capable of describing the relationships of SWC and matric potential. In along with findings of this study, Moreira and Pedrollo [63] reported that ANN models resulted in under-prediction in some soil textures.

Although input variables of ANN3 and ANN4 are different in clay fraction, according to RMSE values, ANN3 has similar performance to ANN4 in prediction of the WRC. This shows that clay fraction does not have remarkable effects on the WRC prediction using ANN models, while Wuddivira et al. [48] stated that SWC is directly related to clay content and clay particles control soil hydraulic characteristics. It is likely because of the complexity of hydraulic functions and ANN nature. Although the amount of adsorbed water is highly correlated with clay fraction content [36, 48] and clay mineralogy [15], slight under-prediction is observed in ANN4 with clay fraction as one of input variables.

Comparison of input variables of ANN3, ANN4 and ANN5 indicates that silt fraction has fundamental role in the prediction of the WRC. ANN predicts the WRC more accurately using full PSD fraction data than partial PSD fraction data. Moreover, previous

studied stated that organic matter and silt fraction as well as clay fraction play considerable role in WRC predictions [15, 19].

Formation and stability of soil aggregation as soil structure index is attributed to organic matter. Fine silt (2–20 μm) and coarse silt (20–50 μm) fraction remarkably influence the WRC predictions [15]. Positive correlation of organic matter and clay fraction in WRC prediction has already reported in previous studies [1, 64].

Figures 4a–4h visually illustrates the measured vs. predicted WRC using ANN models. Fractional over-estimation in prediction of WRC using ANN models can be observed in ANN2, ANN3, ANN4, ANN5 and ANN6 when soil water content is less than mean saturated water content of all soil samples. Slight underestimation of WRC in ANN2, ANN3, ANN4, ANN5 and ANN6 occurred when the soil water content reached more than mean saturated soil water content. Comparison of various ANN and MLR models for the given dataset reveals more accurate prediction of full range of WRC using ANN models than MLR models.

ANFIS PTFs

Table 5 demonstrates the performance evaluation parameters of Gaussian ANFIS PTFs with regard to different input variables. The RMSE of ANFIS PTFs ranges from 0.035 (ANFIS8) to 0.082 (ANFIS6). Results show that among the implemented ANFIS PTFs, ANFIS8 with seven input variables including sand, silt, clay, saturated water content, $\log|h|$, bulk and particle densities results in the most accurate prediction according to the least RMSE (0.035) and the most R^2 (0.92). The RMSE of ANFIS5 (0.036) is almost equal to the RMSE of ANFIS8 (0.035) which shows that in contrast with MLR PTF, ANFIS PTF is not sensitive to replacing porosity with particle density as an input variable.

The silt fraction is the only input variable which makes difference between ANFIS4 and ANFIS5.

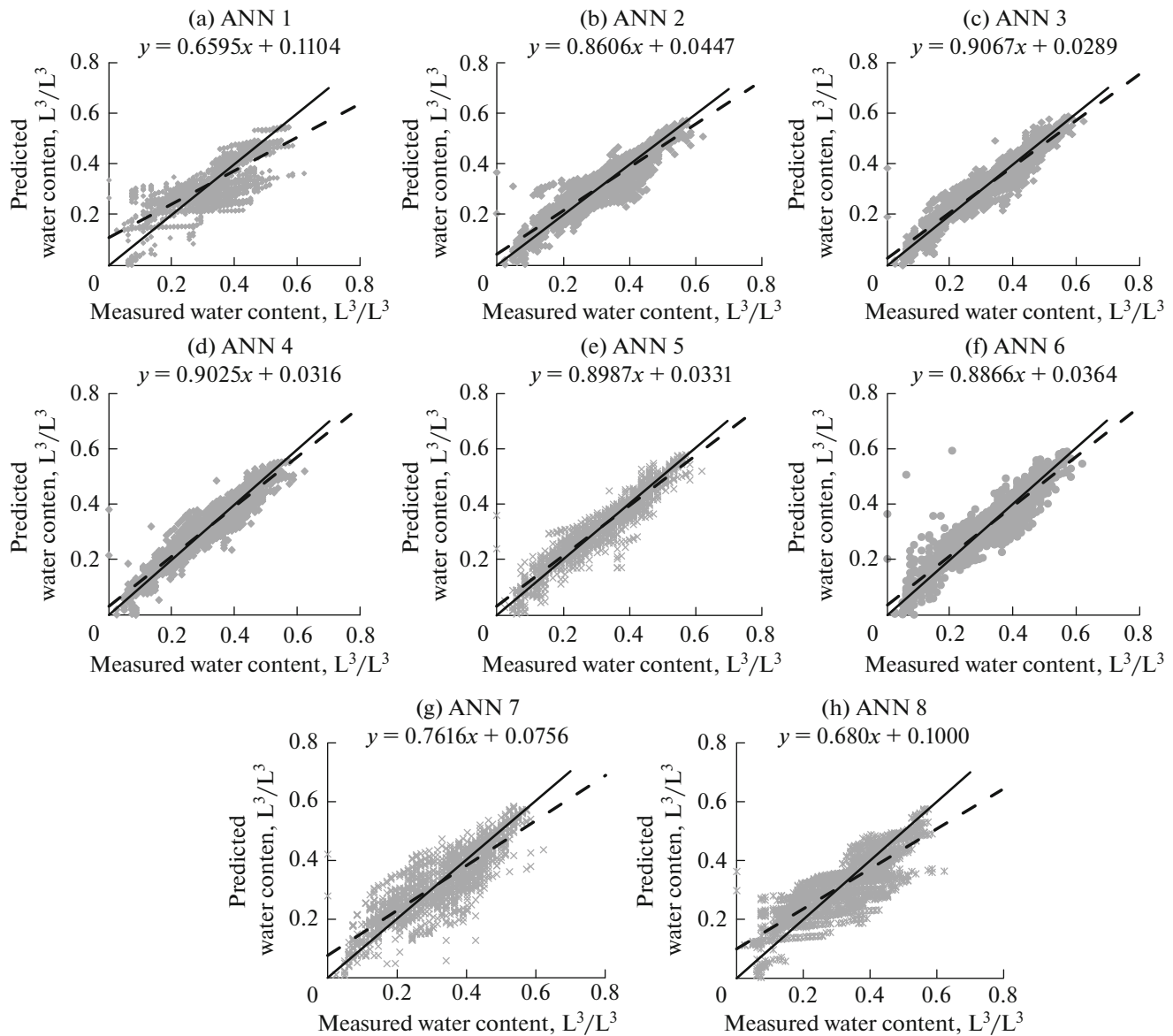


Fig. 4. (a–h). Comparison of measured vs. predicted soil water content of given matric potential using ANN models. Dashed line shows the fitted equation to the predicted values while the solid line indicates the $y = x$.

Although ANFIS is known to be good at generalization and being tolerant to missing data [11, 15], this model was not capable of indicating silt fraction role in improving accuracy prediction of WRC. This can be attributed to systematic ANFIS error and fuzziness of input variables [32]. Differences in predicted and measured soil water content using ANFIS model is attributed to ANFIS systematic error, clay mineralogy, organic matter, simplified assumptions of soil aggregate and pore system. Furthermore, the comparison of ANFIS4, ANFIS5 and ANFIS8 indicates that ANFIS likely cannot distinguish heterogeneity of PSD fraction data.

The comparison of the performance of ANFIS8 with seven input variables and ANFIS6 with two vari-

ables indicates that the input variables significantly affect the performance of ANFIS PTFs. The RMSE trend shows gradual reduction by increasing the number of input variables. However, Fashi [21] showed that the number of input variables in ANFIS model does not influence the WRC prediction.

Most of the ANFIS PTFs, with the exception of ANFIS3 and ANFIS7, over-predict the WRC based on their positive MBE values. Results of ANFIS PTFs show that ANFIS predicts the WRC considerably accurate similar to earlier findings [4, 42, 49, 54].

Figures 5a–5h compares measured vs. predicted soil water contents using various ANFIS models. Most of ANFIS models show slight inaccurate predictions of WRC when the soil water content reached less or

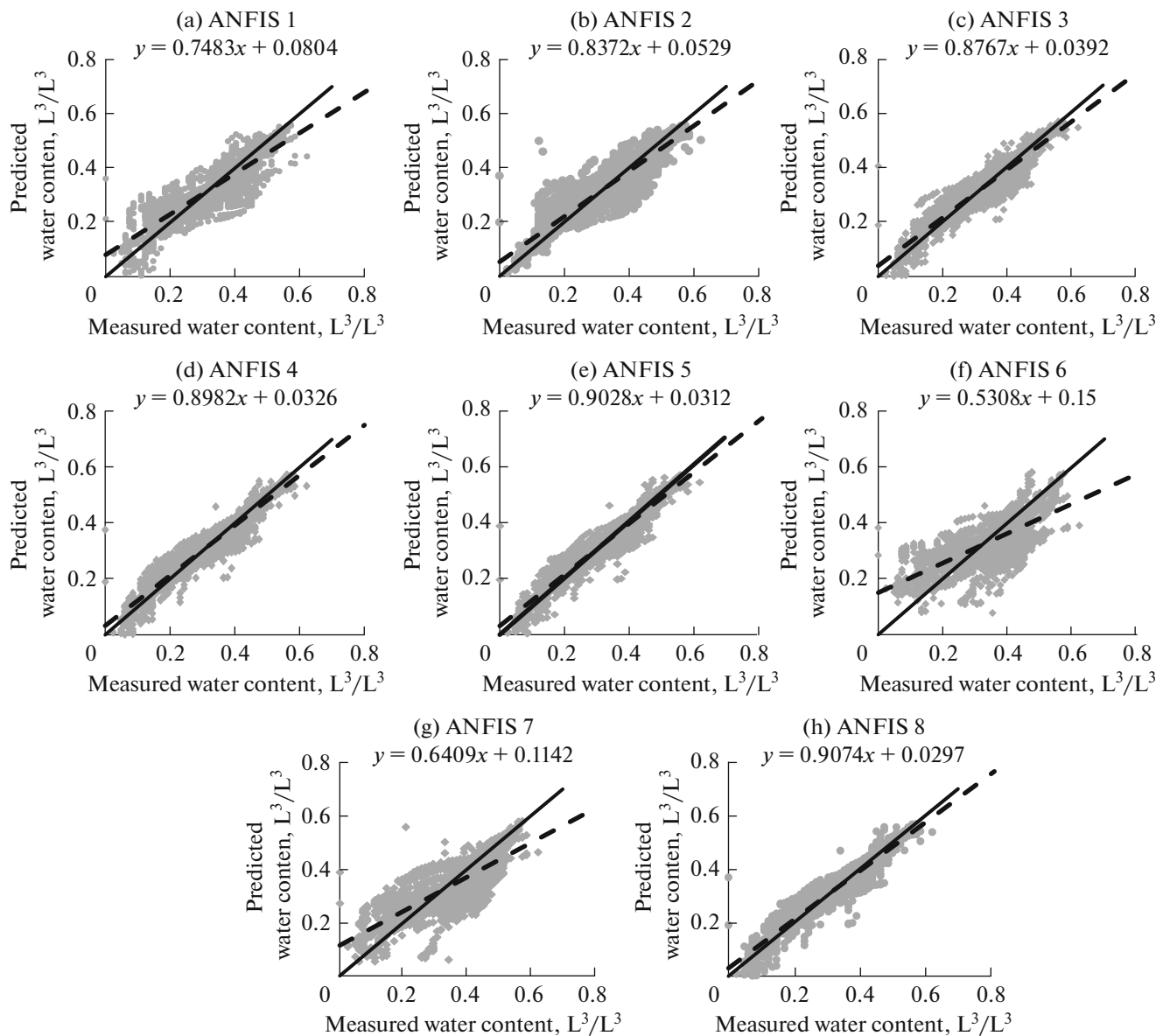


Fig. 5. (a–h). Comparison of measured vs. predicted soil water content of given matric potential using ANFIS models. Dashed line shows the fitted equation to the predicted values while the solid line indicates the $y = x$.

Table 5. Performance evaluation parameters of ANFIS PTFs according different input variables

Model	Input variables*	N MF**	MSE	RMSE	R^2	MAE	MBE
ANFIS1	$\log h $, θ_s , Sand	3, 2, 2	0.0041	0.064	0.74	0.046	-0.0003
ANFIS2	$\log h $, θ_s , Sand, n	3, 2, 2, 2	0.0027	0.052	0.83	0.033	-0.0010
ANFIS3	$\log h $, θ_s , Sand, n , BD	3, 2, 2, 2, 2	0.0021	0.046	0.87	0.029	0.0001
ANFIS4	$\log h $, θ_s , Sand, n , BD, Clay	3, 2, 2, 2, 2, 2	0.0014	0.038	0.91	0.025	-0.0001
ANFIS5	$\log h $, θ_s , Sand, n , BD, Clay, Silt	3, 2, 2, 2, 2, 2, 2	0.0013	0.036	0.92	0.023	-0.0002
ANFIS6	$\log h $, θ_s	3, 3	0.0068	0.083	0.55	0.068	-0.0006
ANFIS7	$\log h $, θ_s , BD	3, 3, 3	0.0058	0.076	0.63	0.057	0.0001
ANFIS8	$\log h $, θ_s , Sand, BD, Clay, Silt, PD	3, 2, 2, 2, 2, 2, 2	0.0012	0.035	0.92	0.024	-0.0002

* h , θ_s , n , BD and PD indicate matric potential, saturated water content, porosity, bulk and particle densities, respectively.

** N MF indicates the number of membership functions of input variables.

Table 6. Comparison of the best PTFs performances in various soil textures

Soil texture	Best method	MSE	RMSE	R^2	MAE	MBE
Sandy	MLR8	0.017	0.130	0.46	0.069	-0.0001
	Scaled PTF	0.007	0.084	0.96	0.026	-0.009
	ANFIS5	0.012	0.110	0.80	0.036	-0.0002
	ANN4	0.011	0.105	0.83	0.035	-0.003
Loamy sand	MLR8	0.007	0.084	0.89	0.070	-0.02
	Scaled PTF	0.004	0.063	0.97	0.039	-0.008
	ANFIS8	0.001	0.032	0.98	0.028	-0.0028
	ANN5	0.001	0.032	0.98	0.030	-0.0025
Sandy loam	MLR8	0.001	0.032	0.98	0.030	-0.006
	Scaled PTF	0.0010	0.032	0.98	0.028	-0.01
	ANFIS3	0.0002	0.014	0.99	0.023	-0.004
	ANN6	0.0005	0.022	0.99	0.017	-0.004
Silt loam	MLR3	0.002	0.045	0.99	0.040	-0.03
	Scaled PTF	0.0024	0.049	0.97	0.035	0.09
	ANFIS8	0.00005	0.007	1	0.005	0.0003
	ANN5	0.0002	0.014	0.99	0.011	-0.0001
Silty clay loam	MLR1	0.001	0.032	0.74	0.033	0.016
	Scaled PTF	0.0014	0.038	0.96	0.032	0.01
	ANFIS5	0.0001	0.010	0.97	0.009	-0.001
	ANN2	0.0004	0.020	0.92	0.015	-0.001
Loamy	MLR8	0.004	0.063	0.98	0.040	0.02
	Scaled PTF	0.0035	0.059	0.91	0.054	0.045
	ANFIS5	0.001	0.032	0.99	0.020	0
	ANN5	0.002	0.045	0.99	0.020	0
Sandy clay loam	MLR1	0.002	0.045	0.98	0.043	0.031
	Scaled PTF	0.0021	0.046	0.95	0.024	-0.013
	ANFIS8	0.00001	0.003	0.99	0.009	-0.0013
	ANN4	0.00003	0.005	0.99	0.015	0.0039
Clay	MLR4	0.0004	0.020	0.99	0.020	-0.011
	Scaled PTF	0.0014	0.038	0.97	0.020	0.055
	ANFIS4	0.00008	0.009	0.99	0.008	0
	ANN5	0.00006	0.008	1	0.006	0.001
Silty	MLR6	0.0001	0.010	1	0.010	0.002
	Scaled PTF	0.0005	0.023	0.96	0.023	0.005
	ANFIS8	0.00001	0.003	1	0.003	0.0001
	ANN7	0.00002	0.004	1	0.004	-0.0032

more than mean saturated water content of all soil samples except ANFIS6 and ANFIS7. Inaccurate predictions of WRC using ANFIS6 and ANFIS7 are also proved by higher RMSEs of these two ANFIS models than the other implemented ANFIS models (Table 5).

Comparison of Implemented PTFs in Various Soil Textures

Table 6 compares the models' performance in several soil textures according to the performance evaluation criteria including MSE, RMSE, R^2 , MAE and MBE. A glance at the results shows that ANFIS PTFs with different number of input variables have the best

PTF efficiency in sandy loam, silty loam, silty clay loam, loamy, sandy clay loam and silty soils. In loamy sand soils, ANFIS and ANN have similar performance with RMSE being equal to 0.032.

ANN PTFs with different input variables are the best fitted PTFs in clay soils showing that ANN method is capable of predicting the WRC of fine textured soils. However, in sandy soils, it seems that traditional numerical scaled PTFs have performed more accurately in the prediction of WRC. ANN4 with six input variables (RMSE = 0.105) and ANFIS5 with seven input variables (RMSE = 0.110) are not suitable compared to scaled numerical PTF (RMSE = 0.084) in this soil texture.

All the fitted PTFs under-predict the WRC in sandy soils regarding MBE values. Slight under-predictions of WRC are also seen in loamy sand, sandy loam, and sandy soils. Slight prediction error occurs in sandy soils due to its macro pores and void ratio [17, 23]. The RMSE of ANFIS5 (0.032) and ANN5 (0.045) with the same number of input variables in loamy soils show that the ANFIS PTF predicts the WRC more accurately in moderate textured soils.

Scaled numerical PTF has better PTF performance than MLR8 with RMSEs equal to 0.059 and 0.063, respectively, in loamy soils. ANFIS8, MLR6 and scaled numerical PTF over-predict the WRC (positive MBE) while ANN7 under-predicts the WRC in silty soils. According to the RMSE criterion, ANFIS8 in both sandy clay loam and silty soils is the best fitted PTF. ANFIS8 with seven input variables under-predicts the WRC in sandy clay loam while it over-predicts the WRC in silty soils.

Comparison of ANN in several soil textures indicates under-prediction of WRC in most soil textures with high sand and low clay percentages including sandy, loamy sand, sandy loam, silty loam, silty clay loam and silty soils. Lack of soil structure can be a source of prediction errors in coarse textured (sandy) soils [15] which deviates the predicted from the measured SWC.

In fine textured soils, since the fine silt and clay fraction content adsorb soil water to soil solid surface by surface adsorptive forces [15], inaccurate WRC prediction is expected. Micro-pores and physicochemical characteristics of clay fraction play an inevitable role in water flow and soil hydraulic conductivity [17, 23]. It is worth mentioning that clay mineralogy and the considerable effect of silt on water adsorption in WRC predictions should not be neglected [16]. The results of the current study are consistent with the findings reported by Moreira and Pedrollo [63].

CONCLUSIONS

Comparative study is a well-established approach to identify weaknesses and strengths of several data-modeling methods, particularly when they are different in nature and assumptions. To achieve the objectives of this work, performances of different models including MLR, ANN and ANFIS to predict WRC are compared regardless to soil texture. Recently developed scaled numerical PTF using PSD data is also proposed to predict WRC for several soil textures. Comparative results, regardless of the soil texture, indicate that the ANN having $\log|h|$, θ_s , sand, clay, and silt fractions, porosity, and bulk density as its input variables results in the highest performance ($R^2 = 0.96$). This offers machine learning methods for field applications due to their robustness on nonlinearities, out of range input data and/or noise in inputs.

In comparison with the MLR method, scaled numerical approach is capable of predicting WRC more accurately especially in the dry range. Considering the soil texture, machine learning models do not perform well in sandy soils while numerical scaled PTF has resulted an acceptable performance. In other soil textures, the high accuracy of machine learning methods including ANN and ANFIS is obvious. The findings of this study can provide researchers with the best method of the WRC prediction and help them to improve other methods based on the advantages and disadvantages of the studied models.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. A. C. Resurreccion, P. Moldrup, M. Tuller, T. Ferré, K. Kawamoto, T. Komatsu, and L. W. de Jonge, "Relationship between specific surface area and the dry end of the water retention curve for soils with varying clay and organic carbon contents," *Water Resour. Res.* **47**, W06522 (2011). <https://doi.org/10.1029/2010WR010229>
2. A. D. Nemes, M. Schaap, F. Leij, and J. Wösten, "Description of the unsaturated soil hydraulic database UNSODA version 2.0," *J. Hydrol.* **251**, 151–162 (2001). [https://doi.org/10.1016/S0022-1694\(01\)00465-6](https://doi.org/10.1016/S0022-1694(01)00465-6)
3. A. Elshorbagy and K. Parasuraman, "On the relevance of using artificial neural networks for estimating soil moisture content," *J. Hydrol.* **362**, 1–18 (2008). <https://doi.org/10.1016/j.jhydrol.2008.08.01>
4. A. Erofeev, D. Orlov, A. Ryzhov, and D. Koroteev, "Prediction of porosity and permeability alteration based on machine learning algorithms," *Transp. Porous Med.* **128**, 677–700 (2019).
5. A. Garg, A. Garg, W.-H. Zhou, K. Tai, and M. Deo, "A new simulation approach of genetic programming in modelling of soil water retention property of unsaturated soil," *Eng. Comput.* **32**, 914–930 (2015). <https://doi.org/10.1108/EC-05-2014-0110>
6. A. Haghverdi, W. Cornelis, and B. Ghahraman, "A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data," *J. Hydrol.* **442**, 46–54 (2012). <https://doi.org/10.1016/j.jhydrol.2012.03.036>
7. A. Hashemi, K. Asefpour Vakilian, J. Khazaei, and J. Massah, "An artificial neural network modeling for force control system of a robotic pruning machine," *J. Inf. Org. Sci.* **38**, 35–41 (2014).
8. A. H. Alavi, A.H. Gandomi, A. Mollahassani, A. A. Heshmati, and A. Rashed, "Modeling of maximum dry density and optimum moisture content of stabilized soil using artificial neural networks," *J. Plant Nutr. Soil Sci.* **173**, 368–379 (2010). <https://doi.org/10.1002/jpln.200800233>
9. A. Jafarzadeh, M. Pal, M. Servati, M. FazeliFard, and M. Ghorbani, "Comparative analysis of support vector machine and artificial neural network models for soil

- cation exchange capacity prediction,” *Int. J. Environ. Sci. Technol.* **13**, 87–96 (2016).
<https://doi.org/10.1007/s13762-015-0856-4>
10. A. Kamari, A. H. Mohammadi, M. Lee, and A. Bahadori, “Decline curve based models for predicting natural gas well performance,” *Petroleum* **3**, 242–248 (2017).
<https://doi.org/10.1016/j.petlm.2016.06.006>
 11. A. Keshavarzi, F. Sarmadian, M. Sadeghnejad, and P. Pezeshki, “Developing pedotransfer functions for estimating some soil properties using artificial neural network and multivariate regression approaches,” *Proenvironment* **3**, 322–330 (2010).
 12. A. Nemes, M. Schaap, and J. Wösten, “Functional evaluation of pedotransfer functions derived from different scales of data collection,” *Soil Sci. Soc. Am. J.* **67**, 1093–1102 (2003).
<https://doi.org/10.2136/sssaj2003.1093>
 13. C. Antinoro, V. Bagarello, V. Ferro, G. Giordano, and M. Iovino, “A simplified approach to estimate water retention for Sicilian soils by the Arya–Paris model,” *Geoderma* **213**, 226–234 (2014).
<https://doi.org/10.1016/j.geoderma.2013.08.004>
 14. C. T. Silveira, C. Oka-Fiori, L. J. C. Santos, A. E. Sirtoli, C. R. Silva, and M. F. Botelho, “Soil prediction using artificial neural networks and topographic attributes,” *Geoderma* **195**, 165–172 (2013).
<https://doi.org/10.1016/j.geoderma.2012.11.016>
 15. D. K. Jensen, M. Tuller, L. W. de Jonge, E. Arthur, and P. Moldrup, “A new two-stage approach to predicting the soil water characteristic from saturation to oven-dryness,” *J. Hydrol.* **521**, 498–507 (2015).
<https://doi.org/10.1016/j.jhydrol.2014.12.018>
 16. D. Karup, P. Moldrup, M. Tuller, E. Arthur, and L. Jonge, “Prediction of the soil water retention curve for structured soil from saturation to oven dryness,” *Eur. J. Soil Sci.* **68**, 57–65 (2017).
<https://doi.org/10.1111/ejss.12401>
 17. D. Li, G. Gao, M. Shao, and B. Fu, “Predicting available water of soil from particle-size distribution and bulk density in an oasis–desert transect in northwestern China,” *J. Hydrol.* **538**, 539–550 (2016).
<https://doi.org/10.1016/j.jhydrol.2016.04.046>
 18. D. Zhao, W. Ni, and Q. Zhu, “A framework of neural networks based consensus control for multiple robotic manipulators,” *Neurocomputing* **140**, 8–18 (2014).
<https://doi.org/10.1016/j.neucom.2014.03.041>
 19. E. Arthur, M. Tuller, P. Moldrup, D. K. Jensen, and L. W. de Jonge, “Prediction of clay content from water vapor sorption isotherms considering hysteresis and soil organic matter content,” *Eur. J. Soil Sci.* **66**, 206–217 (2015).
<https://doi.org/10.1111/ejss.12191>
 20. E. Olyaie, H. Banejad, K.-W. Chau, and A. M. Mellesse, “A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States,” *Environ. Monit. Assess.* **187**, 189 (2015).
<https://doi.org/10.1007/s10661-015-4381-1>
 21. F. H. Fashi, “Evaluation of adaptive neural-based fuzzy inference system approach for estimating saturated soil water content,” *Model. Earth Syst. Environ.* **2**, 197 (2016).
<https://doi.org/10.1007/s40808-016-0255-y>
 22. F. Karandish and J. Šimůnek, “A comparison of numerical and machine-learning modeling of soil water content with limited input data,” *J. Hydrol.* **543**, 892–909 (2016).
<https://doi.org/10.1016/j.jhydrol.2016.11.007>
 23. F. Meskini-Vishkaee, M. H. Mohammadi, and M. Vanclooster, “Predicting the soil moisture retention curve, from soil particle size distribution and bulk density data using a packing density scaling factor,” *Hydrol. Earth Syst. Sci.* **18**, 4053–4063 (2014).
<https://doi.org/10.5194/hess-18-4053-2014>
 24. H. Ghorbani, H. Kashi, N. Hafezi Moghadas, and S. Emamgholizadeh, “Estimation of soil cation exchange capacity using multiple regression, artificial neural networks, and adaptive neuro-fuzzy inference system models in Golestan Province, Iran,” *Commun. Soil Sci. Plan.* **46**, 763–780 (2015).
<https://doi.org/10.1080/00103624.2015.1006367>
 25. H. Jiang and W. R. Cotton, “Soil moisture estimation using an artificial neural network: a feasibility study,” *Can. J. Remote Sens.* **30**, 827–839 (2004).
<https://doi.org/10.5589/m04-041>
 26. H. Khodaverdiloo, M. Homaei, M. T. van Genuchten, and S. G. Dashtaki, “Deriving and validating pedotransfer functions for some calcareous soils,” *J. Hydrol.* **399**, 93–99 (2011).
<https://doi.org/10.1016/j.jhydrol.2010.12.040>
 27. H. Puhmann and K. von Wilpert, “Pedotransfer functions for water retention and unsaturated hydraulic conductivity of forest soils,” *J. Plant Nutr. Soil Sci.* **175**, 221–235 (2012).
<https://doi.org/10.1002/jpln.201100139>
 28. H. Tabari, O. Kisi, A. Ezani, and P. H. Talaei, “SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment,” *J. Hydrol.* **444**, 78–89 (2012).
<https://doi.org/10.1016/j.jhydrol.2012.04.007>
 29. H. Vereecken, A. Schnepf, J. W. Hopmans, M. Javaux, D. Or, T. Roose, J. Vanderborght, M. H. Young, W. Amelung, M. Aitkenhead, and S. D. Allison, “Modeling soil processes: review, key challenges, and new perspectives,” *Vadose Zone J.* **15**, 1–57 (2016).
<https://doi.org/10.2136/vzj2015.09.0131>
 30. H. Vereecken, M. Weynants, M. Javaux, Y. Pachepsky, M. Schaap, and M. T. Genuchten, “Using pedotransfer functions to estimate the van Genuchten–Mualem soil hydraulic properties: a review,” *Vadose Zone J.* **9**, 795–820 (2010).
<https://doi.org/10.2136/vzj2010.0045>
 31. H. Vereecken, J. Maes, J. Feyen, and P. Darius, “Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content,” *Soil Sci.* **148**, 389–403 (1989).
 32. I. Yilmaz and O. Kaynar, “Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils,” *Exp. Syst. Appl.* **38**, 5958–5966 (2011).
<https://doi.org/10.1016/j.eswa.2010.11.027>
 33. J. Deng, X. Chen, Z. Du, and Y. Zhang, “Soil water simulation and predication using stochastic models based on LS-SVM for red soil region of China,” *Water*

- Resour. Manage. **25**, 2823–2836 (2011).
<https://doi.org/10.1007/s11269-011-9840-z>
34. J. Seyedmohammadi, L. Esmaelnejad, and H. Ramezanzpour, “Determination of a suitable model for prediction of soil cation exchange capacity,” *Model. Earth Syst. Environ.* **2**, 156 (2016).
<https://doi.org/10.1007/s40808-016-0217-4>
 35. J. Si, Q. Feng, X. Wen, H. Xi, T. Yu, W. Li, and C. Zhao, “Modeling soil water content in extreme arid area using an adaptive neuro-fuzzy inference system,” *J. Hydrol.* **527**, 679–687 (2015).
<https://doi.org/10.1016/j.jhydrol.2015.05.034>
 36. J. Wäldchen, I. Schöning, M. Mund, M. Schruppf, S. Bock, N. Herold, K. U. Totsche, and E. D. Schulze, “Estimation of clay content from easily measurable water content of air dried soil,” *J. Plant Nutr. Soil Sci.* **175**, 367–376 (2012).
<https://doi.org/10.1002/jpln.201100066>
 37. K. Asefpour Vakilian and J. Massah, “An artificial neural network approach to identify fungal diseases of cucumber (*Cucumis sativus* L.) plants using digital image processing,” *Arch. Phytopathol. Pflanzenschutz* **46**, 1580–1588 (2013).
<https://doi.org/10.1080/03235408.2013.772321>
 38. K. Christiaens and J. Feyen, “Analysis of uncertainties associated with different methods to determine soil hydraulic properties and their propagation in the distributed hydrological MIKE SHE model,” *J. Hydrol.* **246**, 63–81 (2001).
[https://doi.org/10.1016/S0022-1694\(01\)00345-6](https://doi.org/10.1016/S0022-1694(01)00345-6)
 39. K. Yetilmezsoy, M. Fingas, and B. Fieldhouse, “An adaptive neuro-fuzzy approach for modeling of water-in-oil emulsion formation,” *Colloid. Surf. A* **389**, 50–62 (2011).
<https://doi.org/10.1016/j.colsurfa.2011.08.051>
 40. L. Baker and D. Ellison, “Optimization of pedotransfer functions using an artificial neural network ensemble method,” *Geoderma* **144**, 212–224 (2008).
<https://doi.org/10.1016/j.geoderma.2007.11.016>
 41. M. A. Joe and P. R. Prasanna, “Simulation of soil water retention curve using artificial neural networks with pseudocontinuous pedotransfer functions,” *Int. J. Appl. Eng. Res.* **11**, 4700–4706 (2016).
 42. M. Aqil, I. Kita, A. Yano, and S. Nishiyama, “A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behavior of runoff,” *J. Hydrol.* **337**, 22–34 (2007).
<https://doi.org/10.1016/j.jhydrol.2007.01.013>
 43. M. Behnia, H. A. Valani, M. Bameri, B. Jabalbarez, and H. E. Damaneh, “Potential assessment of ANNs and adaptive neuro fuzzy inference systems (ANFIS) for simulating soil temperature at different soil profile depths,” *Int. J. Adv. Biol. Biomed. Res.* **5**, 52–59 (2017).
<https://doi.org/10.18869/IJABBR.2017.419>
 44. M. H. Mohammadi and M. Vanclooster, “Predicting the soil moisture characteristic curve from particle size distribution with a simple conceptual model,” *Vadose Zone J.* **10**, 594–602 (2011).
<https://doi.org/10.2136/vzj2010.0080>
 45. M. Mohammadi and F. Meskini-Vishkaee, “Predicting soil moisture characteristic curves from continuous particle-size distribution data,” *Pedosphere* **23**, 70–80 (2013).
[https://doi.org/10.1016/S1002-0160\(12\)60081-2](https://doi.org/10.1016/S1002-0160(12)60081-2)
 46. M. Hosseini, S. R. Agereh, Y. Khaledian, H. J. Zoghhalchali, E. C. Brevik, and S. A. R. Movahedi Naeini, “Comparison of multiple statistical techniques to predict soil phosphorus,” *Appl. Soil Ecol.* **114**, 123–131 (2017).
<https://doi.org/10.1016/j.apsoil.2017.02.011>
 47. M. Mukhlisin, A. El-Shafie, and M. R. Taha, “Regularized versus non-regularized neural network model for prediction of saturated soil-water content on weathered granite soil formation,” *Neural Comput. Appl.* **21**, 543–553 (2012).
<https://doi.org/10.1007/s00521-011-0545>
 48. M. N. Wuddivira, D. A. Robinson, I. Lebron, L. Bréchet, M. Atwell, S. de Caires, M. Oatham, S. B. Jones, H. Abdu, and A. K. Verma, “Estimation of soil clay content from hygroscopic water content measurements,” *Soil Sci. Soc. Am. J.* **76**, 1529–1535 (2012).
<https://doi.org/10.2136/sssaj2012.0034>
 49. M. T. Dastorani, A. Moghadamnia, J. Piri, and M. Rico-Ramirez, “Application of ANN and ANFIS models for reconstructing missing flow data,” *Environ. Monit. Assess.* **166**, 421–434 (2010).
<https://doi.org/10.1007/10661-009-1012-8>
 50. M. T. van Genuchten, “A closed-form equation for predicting the hydraulic conductivity of unsaturated soils,” *Soil Sci. Soc. Am. J.* **44**, 892–898 (1980).
 51. M. Weynants, H. Vereecken, and M. Javaux, “Revisiting Vereecken pedotransfer functions: Introducing a closed-form hydraulic model,” *Vadose Zone J.* **8**, 86–95 (2009).
<https://doi.org/10.2136/vzj2008.0062>
 52. N. K. Twarakavi, J. Šimůnek, and M. Schaap, “Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines,” *Soil Sci. Soc. Am. J.* **73**, 1443–1452 (2009).
<https://doi.org/10.2136/sssaj2008.0021>
 53. N. Talpur, M. N. M. Salleh, and K. Hussain, “An investigation of membership functions on performance of ANFIS for solving classification problems,” *IOP Conf. Ser.: Mater. Sci.* **226**, 012103 (2017).
<https://doi.org/10.1088/1757-899X/226/1/012103>
 54. P. C. Nayak, K. Sudheer, D. Rangan, and K. Ramasastri, “A neuro-fuzzy computing technique for modeling hydrological time series,” *J. Hydrol.* **291**, 52–66 (2004).
<https://doi.org/10.1016/j.jhydrol.2003.12.010>
 55. P. Tempel, N. Batjes, and V. van Engelen, *IGBP-DIS Soil Data Set for Pedotransfer Function Development*, ISRIC working paper 96/05 (International Soil Reference and Information Centre, Wageningen, 1996).
 56. P. Zou, J. Yang, J. Fu, G. Liu, and D. Li, “Artificial neural network and time series models for predicting soil salt and water content,” *Agric. Water Manage.* **97**, 2009–2019 (2010).
<https://doi.org/10.1016/j.agwat.2010.02.01>
 57. S. G. Dashtaki, M. Homaei, and H. Khodaverdiloo, “Derivation and validation of pedotransfer functions for estimating soil water retention curve using a variety of soil data,” *Soil Use Manage.* **26**, 68–74 (2010).
<https://doi.org/10.1111/j.1475-2743.2009.00254.x>

58. S. I. Hwang and S. I. Choi, "Use of a lognormal distribution model for estimating soil water retention curves from particle-size distribution data," *J. Hydrol.* **323**, 325–334 (2006).
<https://doi.org/10.1016/j.jhydrol.2005.09.005>
59. S. Uzuner and D. Cekmecelioglu, "Comparison of artificial neural networks (ANN) and adaptive neuro-fuzzy inference system (ANFIS) models in simulating polygalacturonase production," *BioResources* **11**, 8676–8685 (2016).
60. S. Zacharias and G. Wessolek, "Excluding organic matter content from pedotransfer predictors of soil water retention," *Soil Sci. Soc. Am. J.* **71**, 43–50 (2007).
<https://doi.org/10.2136/sssaj2006.0098>
61. T. Asefa, M. Kemblowski, M. McKee, and A. Khalil, "Multi-time scale stream flow predictions: the support vector machines approach," *J. Hydrol.* **318**, 7–16 (2006).
<https://doi.org/10.1016/j.jhydrol.2005.06.001>
62. T. Chan and R. Govindaraju, "Estimating soil water retention curve from particle-size distribution data based on polydisperse sphere systems," *Vadose Zone J.* **3**, 1443–1454 (2004).
<https://doi.org/10.2113/3.4.1443>
63. T. Moreira de Melo and O. C. Pedrollo, "Artificial neural networks for estimating soil water retention curve using fitted and measured data," *Appl. Environ. Soil Sci.* **2015**, 535216 (2015).
<https://doi.org/10.1155/2015/535216>
64. V. Balland, J. A. Pollacco, and P. A. Arp, "Modeling soil hydraulic properties for a wide range of soil conditions," *Ecol. Model.* **219**, 300–316 (2008).
<https://doi.org/10.1016/j.ecolmodel.2008.07.009>
65. V. Vijayaraghavan, A. Garg, C. H. Wong, K. Tai, and Y. Bhalerao, "Predicting the mechanical characteristics of hydrogen functionalized graphene sheets using artificial neural network approach," *J. Nanostruct. Chem.* **3**, 83 (2013).
<https://doi.org/10.1186/2193-8865-3-83>
66. V. Vijayaraghavan, A. Garg, C. Wong, K. Tai, and S. Mahapatra, "Measurement of properties of graphene sheets subjected to drilling operation using computer simulation," *Measurement* **50**, 50–62 (2014).
<https://doi.org/10.1016/j.measurement.2013.12.028>
67. X. Dai, Z. Huo, and H. Wang, "Simulation for response of crop yield to soil moisture and salinity with artificial neural network," *Field Crop. Res.* **121**, 441–449 (2011).
<https://doi.org/10.1016/j.fcr.2011.01.016>
68. Y. Abbasi, B. Ghanbarian-Alavijeh, A. Liaghat, and M. Shorafa, "Evaluation of pedotransfer functions for estimating soil water retention curve of saline and saline-alkali soils of Iran," *Pedosphere* **21**, 230–237 (2011).
[https://doi.org/10.1016/S1002-0160\(11\)60122-7](https://doi.org/10.1016/S1002-0160(11)60122-7)
69. Z. Zolfaghari, M. Mosaddeghi, and S. Ayoubi, "ANN-based pedotransfer and soil spatial prediction functions for predicting Atterberg consistency limits and indices from easily available properties at the watershed scale in western Iran," *Soil Use Manage.* **31**, 142–154 (2015).
<https://doi.org/10.1111/sum.12167>