Improving Infiltration Prediction by Point-based PTFs for Some Soils in Southern of Iran

Maryam Molayem (mina_molayem2020@yahoo.com)
Islamic Azad University Marvdasht

S A. Abtahi
Islamic Azad University Marvdasht

M. Jafarinia
Islamic Azad University Marvdasht

J. Yasrebi
Islamic Azad University Marvdasht

Original Manuscript

Keywords: Artificial neural network, Cumulative infiltration, Multiple linear regression, hydraulic properties of the soil, Support vector machines

Posted Date: February 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-186783/v1

Improving infiltration prediction by point-based PTFs for some soils in southern of Iran

M. Molayem[a], S A. Abtahi[a*], M. Jafarinia[a], J. Yasrebi[a]

[a] *Department of agriculture management, Marvedasht branch, Islamic Azad University, Marvedasht, Iran.*

[*]Corresponding Author: Seyed Ali Abtahi

Email: seyedaliabtahi@yahoo.com

Phone: 071-32257824

Improving infiltration prediction by point-based PTFs for some soils in southern of Iran

Abstract

Modeling soil water infiltration at the field scale with ruler of calcareous, saline and sodic conditions is important for a better understanding of infiltration processes in these soils and future of infiltration modeling. The aim of the present study was to derive and evaluate soil water infiltration models for some calcareous, saline and sodic soils in Marvdasht plain, southern of Iran. The infiltration data was measured in 72 locations at the regional scale with 3 replications. In each location, the basic soil properties were also measured. The multiple linear regression (MLR) and feed-forward multilayer perceptron artificial neural networks (ANN) model were used to estimate cumulative soil water infiltration at different time. The results performance of water infiltration models such as Kostiakov, Kostiakov–Lewis, USDA-NRCS, Philip, Horton and Green-Ampt models according to the mean R^2, ME, RMSE and SDRMSE indices for all soils showed the Kostiakov–Lewis model provided the most accurate predictions. Moreover, the results showed that the derived ANN models at different times with a R^2 of 0.438-0.661 and a RMSE of 0.977-17.111 performed better than MLR model. There would be great interest to improve the cumulative soil water infiltration in site-specific soil utilization, management and protection of the environment by MLR and ANN methods.

Key words: Artificial neural network, Cumulative infiltration, Multiple linear regression, hydraulic properties of the soil, Support vector machines

1. Introduction

Soil water infiltration is the entry of water into the soil, and plays an important role in the hydrological cycle (Hillel 2003). Additionally, water infiltration has an important influence on water run-off generation, soil erosion, and nutrient uptake by plant and leaching to the deeper layers. Moreover, many components related to hydrology and soil erosion models can be derived from infiltration data using Pedo-transfer functions (PTFs). Hence, the accurate quantification of water infiltration to the soil is the first step for the sustainable soil utilization, management and protection of the environment in semi-arid region soils (Parchami-Araghi et al. 2013; Mirzaee et al. 2014).

Traditionally, there are two types of water infiltration models including physically-based and empirically-based models, which were categorized by Rawls et al. (1993). Most physically-based water infiltration models are approximately obtained by solutions of the Richards' flow equation (Ghorbani Dashtaki et al. 2009; Cui and Zhu 2018). The empirical-based models built from water infiltration measured data are listed in Table 1. The performance of physical and empirical-based water infiltration models has been examined extensively in soil science (Machiwal et al. 2006; Turner 2006; Neshat and Parehkar 2007; Ghorbani Dashtaki et al. 2009; Angelaki et al. 2013; Mirzaee et al. 2014; Sihag and Singh 2018). However, these models for the estimation of the cumulative infiltration data at a specific region and time need to be calibrated for their parameters. The direct infiltration measurement at the field large-scale is time-consuming, costly and difficult; the results of parameters estimation have not been successful in different conditions (Ghorbani Dashtaki et al. 2016), therefore, it is necessary to focus on the prediction of cumulative soil water infiltration at specific times through more easily available hydraulic properties of the soil.

Insert Table 1

The PTFs may be an alternative procedure for estimating cumulative water infiltration by correlation with more easily available hydraulic properties of the soil. The PTFs based on the nature of prediction are comprised of two important types, including point- and parametric-based PTFs, which predict soil hydraulic properties and model parameters describing difficult to measure hydraulic properties of the soil, respectively (Vereecken et al. 2010). Multiple linear regression (MLR) (Pahlavan-Rad et al. 2020) and artificial neural networks (ANNs) are the two best methods widely used to estimate difficult-to-measure hydraulic properties of the soil. Several recent studies have shown that the ANNs models are powerful for predicting soil properties such as water retention parameters and hydraulic conductivity (Schaap et al. 1998; Baker and Ellison 2008), water infiltration process (Parchami-Araghi et al. 2013), soil-water retention (Haghverdi et al. 2012), Atterberg consistency limits (Zolfaghari et al. 2015), predicting soil organic matter (Somaratne et al. 2005; Mirzaee et al. 2016), simulating soil erosion (Licznar and Nearing 2003; Mirzaee et al. 2017) and Fe and Mn (Sihag et al. 2019b). An advantage of ANNs models in comparison with regression models is that ANNs models require no a priori model concept and functional form to relate PTFs input to output (Schaap et al. 1998). The ANNs models are well-known for salving nonlinear relationships between hydraulic properties of the soil. In addition, this method could be an exact method to map the soil hydraulic properties in a nonlinear method. Moreover, support vector machines (SVMs) are considered a progressive data mining method due to their strong theoretical background. The SVMs as a type of machine-learning algorithm (Vapnik 1995) have been used to estimate the soil–water retention curve (Nguyen et al. 2015; Khlosi et al. 2016), and soil properties including clay, sand, pH, total organic carbon, and permanganate oxidizable carbon contents (Deiss et al. 2020). Additionally, it was indicated that the radial basis kernel function-based SVR approach works well in predicting the cumulative

infiltration of sandy soils in India (Sihag et al. 2020) and in semi-arid region of Iran (Sihag et al. 2019a). Generally, the advantages of the soft computing techniques such as ANNs and SVMs require only a few user-defined parameters (Sihag et al. 2020) and, are much faster, robust to noise and easily adapt to unknown situations. These models are prone to limitations due to local minimum and maximum data, excess complexity of the network structure, and the need for more memory space to store the derived model (Iounousse et al. 2015; Zhang et al. 2016).

A semi-arid climate describes climatic regions that receive precipitation below potential evapotranspiration. The moisture deficits in semi-arid regions, such as in Iran, cause the development of calcium carbonate ($CaCO_3$) (Baumhardt and Lascano 1993). The presence of carbonates in soils display alkaline pH and an exchange complex dominated by Ca^{2+} (Ostovari et al. 2017; Ostovari et al. 2020). In addition, carbonates can promote soil structure (Blume et al. 2016). The calcareous, saline and sodic soils are known for high calcium carbonate equivalent and pH (pH>7.5), high electrical conductivity (EC>4 $dS.m^{-1}$) and high sodium (ESP>15), respectively (George et al. 2012).

Overall, the hydraulic properties of the soils best suited to describing the cumulative soil water infiltration remain unclear under saline, sodic and calcareous conditions. The problem is further complicated by the cumulative soil water infiltration difference at specific times. Therefore, the present study focuses on predicting cumulative soil water infiltration at specific times steps throughout the infiltration times by MLR-, ANN- and SVM-based PTFs.

The aims of this study were to: (i) derive accurate PTFs models to predictcumulative soil water infiltration at specific times (i.e. 5, 10, 20, 45, 150, 210 and 270 minutes (I_5, I_{10}, I_{20}, I_{45}, I_{150}, I_{210} and I_{270})) using MLR, ANN and SVM models, and (ii) evaluate the efficacy of more easily

measured hydraulic properties of the soil for the estimation of cumulative soil water infiltration in calcareous, saline and sodic soils in southwest Iran.

2. Material and methods

2.1. Study site

The study region is the intensive agriculture regions of Marvdasht plain, which is located in a semi-arid part of Fars province, Iran (Fig. 1). The study area covers approximately 197,923 ha (29° 44′-30° 26′ N lat., 52° 17′-53° 30′ E long.). The region is surrounded mainly by hills, however the central region is predominantly a flat plain, the elevation of the study area ranges from 1574 to 3121 m. The native vegetation has vanished because of intensive human activities. The current agriculture productions in this region are generally maize and wheat crop rotation. Soil parent materials of the study area are calcareous materials derived from surrounding limestone escarpments. In addition, alluvial soils have formed by the accumulation of transported materials through concentrated flows. The most common soil orders are Inceptisols, Entisols and Mollisols according to the American Soil Taxonomy (USDA 2010).

Insert Fig 1

2.2. Climate factors

The average monthly precipitation and temperature distribution during the period of 2001-2018 are shown in Fig. 2 (a) and (b). The climate of the semi-arid region is characterized by irregular heavy rainfall events; the average annual precipitation and air temperature are reported as 291.7 mm and 17.5 °C, respectively. More than 50% of the rainfalls occurs in the form of

intense, short duration rainstorms during the winter season (December, January, February months) (Fig. 2 (a)).

Insert Fig 2

2.3. Soil survey

Seventy-two soil surface samples were sampled based on the stratified random method from top horizons, up to a 30 cm deep (Fig. 1). The soil samples were air-dried, passed through a 2 mm sieve and analyzed. The content of initial soil moisture was analyzed by the gravimetric procedure. The soil samples were analyzed for soil texture inclusive sand (0.05 –2 mm), silt (0.05 - 0.002 mm) and clay (< 0.002 mm) contents by hydrometer procedure (Gee and Bauder 1986), soil organic carbon (SOC) based on the modified Walkley–Black procedure (Nelson 1982), and soil calcium carbonate equivalent (CCE) by back-titration procedure (Nelson and Sommers 1986). Sand, silt and clay data were then applied to calculate the geometric mean (Dg) (mm) and geometric standard deviation (Sg) of soil particle diameters as follows (Shirazi and Boersma 1984):

$$\mathrm{Dg} = \exp(\mathrm{a})\,. \qquad \mathrm{a} = 0.01\sum_{\mathrm{i}=1}^{\mathrm{n}} f_i \ln(M_i) \tag{1}$$

$$\mathrm{S_g} = \exp(\mathrm{b})\,. \qquad \mathrm{b} = 0.01\sum_{\mathrm{i}=1}^{\mathrm{n}} f_i \ln^2 M_i - \mathrm{a}^2 \tag{2}$$

Where f_i is the mass percentage of particles with arithmetic mean diameter M_i and n is the number of particle fractions (sand, silt, clay, …). The descriptive statistics for all hydraulic properties of the soil are presented in Table 2. The soil textural classes at present study were sandy loam (5%), sandy clay loam (6%), clay loam (43%) and loam (46%).

Insert Table 2

154

155 *2.4. Infiltration experiments*

156 The double ring procedure was used to measure water infiltration to soil for 72 selected
157 points (Fig. 1). The infiltration experiment was replicated three times at each point with a triangular
158 pattern distancing 5 m. The diameters inner and outer ring infiltrometer are 30 and 70 cm,
159 respectively, which were used to obtain infiltration data. Water infiltration to the soil continued
160 until a steady state condition was observed (USDA-NRCS 2005). The basic infiltration rate was
161 determined when the variations in the infiltration rate is about 10% (USDA-NRCS 2005).
162 However, each replication of infiltration experiments took time around 270 min for each soil. The
163 time of basic infiltration rate was determined by Eq. (3) for t (Walker et al. 2006).

$$\frac{dI}{dt} = -600\left(\frac{d^2I}{dt^2}\right) \tag{3}$$

164 The model parameters listed in Table 1 were obtained by fitting process in MatLab v.8.5 (The Math
165 Works, Inc. 2007). In this way, the least-squares method was used to obtain the infiltration models
166 parameters. The sum of the squares minimizes the differences between the predicted and observed
167 infiltration data at each studied point (Parhi et al. 2007). The following equation (4) was used to
168 obtain the infiltration models parameters:

$$SSE_i = \sum_{i=1}^{n} (I_{obs}(i) - \widehat{I_{pre}}(i))^2 \tag{4}$$

169 Where $\widehat{I_{pre}}$ is the predicted and I_{obs} is the observed infiltration data.

170

171 *2.5.Predictor variables for developing PTFs models*

172 For this study, input data were easily measurable basic hydraulic properties of the soil
173 including sand, silt, clay, Dg, Sg, CCE and SOC variables. The hydraulic properties of the soil

were used to predict cumulative soil water infiltration at specific times (i.e., 5, 10, 20, 45, 150, 210 and 270 min (I_5, I_{10}, I_{20}, I_{45}, I_{150}, I_{210} and I_{270})).

2.6. Development of Multiple linear regression (MLR)- and artificial neural networks (ANNs)-based PTFs models

2.6.1. MLR model

MLR was used to estimate difficult-to-measure hydraulic properties (i.e. cumulative soil water infiltration) at specific times i.e. 5, 10, 20, 45, 150, 210 and 270 min (I_5, I_{10}, I_{20}, I_{45}, I_{150}, I_{210} and I_{270}). The MLR models are described as (Ho 2006):

$$I = \alpha + X^T\beta + \varepsilon \quad (5)$$

Where I is the cumulative soil water infiltration at specific times, α is the intercept, $X = (X_1, \ldots, X_n)$ are the easily measurable basic hydraulic properties of the soil, $\beta = \{\beta_1, \ldots, \beta_n\}$ are the regression coefficients and ε shows random measured errors that the linear regression model is not able to explain. The step-wise technique was used to relate the cumulative water infiltration at different times to the easily measurable basic hydraulic properties of the soil. The least squares algorithm was used to determine $\beta = \{\beta_1, \ldots, \beta_n\}$ (regression coefficients) (Cressie 1993).

2.6.2. ANN model

Different types of ANNs have been widely used in soil science. The architecture of network and determination of weights are the important differences among the various types of ANNs (Mirzaee et al. 2020). However, several studies have indicated that the feed-forward multilayer perceptron (MLP) network with an input layer, a hidden layer and an output layer is able to predict and recognizes any kind of relationships between inputs and outputs soil variables (Minasny and McBratney 2002; Khalilmoghadam et al. 2009; Wen et al. 2010; Mirzaee et al. 2017). Hence, this

type of neural network was used to predict cumulative soil water infiltration at specific times in the present study (Fig. 3). The following equation (6) explains the MLP network that was applied to estimate soil water infiltration at the present study:

$$I = f_2\left\{b_k + \sum_{k=1}^{n}\left[W_k f_1\left(b_j + \sum_{i=1}^{m}[W_{ij}X_i]\right)\right]\right\} \quad (6)$$

where; $f_1(h_j)$ and $f_2(h_k)$ are the transfer functions of the neurons in the hidden layer and the transfer functions of the neuron in the output layer, respectively. Both transfer functions (i.e. $f_1(h_j)$ and $f_2(h_k)$) were adopted to purelin (linear transfer function), tansig and logsig (two sigmoid transfer functions). b_k and b_j are the bias at the output layer and the neuron k of the hidden layer (k =1,…,n), respectively; X_i is the input variable r; W_{ij} is the weight of connection between input variable I (I_1,…, I_r) and neuron j of the hidden layer; W_k is the weight of connection between neuron k of the hidden layer and the single output layer neuron (Hagan et al. 1996).

Insert Fig 3

In the present study, the ANN models were developed and tested with 1-40 neurons in the hidden layer through a trial and error process. A Bayesian-regularization algorithm was used to have a more stable response and overcome the problem of over-fitting in training of ANN model. The Bayesian-regularization algorithm minimizes errors at both of the network errors and the network weights and bias (MacKay 1992). The following equation (7) was applied to rescale all of the input hydraulic properties of the soil in a range of 0.1–0.9 (Mirzaee et al. 2017):

$$x_n = 0.1 + 0.8\left[\frac{x - x_{min}}{x_{max} - x_{min}}\right] \qquad x_{min} < x < x_{max} \quad (7)$$

where x_n is rescaled data, x_{min} is the minimum and x_{max} is the maximum of measured data. ANN models were constructed using MatLab v.8.5 (The MathWorks, Inc. 2007).

217

218 *2.6.2. SVM model*

219 The SVM method suggested by Vapnik (1995), is a statistical learning machine to solve the
220 classification problem. The main idea of this method is to project the input variables by means of
221 kernel functions with a higher dimensional feature. The optimum parameters value (i.e.,
222 regularization parameter (C) and regression precision (ε)) were predicted through a trial and error
223 process, with the best accuracy used for developing the SVM models. A range of 0 to 28 and 0.005
224 to 0.8 was set for C and ε parameters, respectively. A 10-fold cross-validation method was used to
225 obtain an optimum value for these meta-parameters. In the present study, the Gaussian Radial
226 Basis Function obtained the based performance for kernel.

227

228 *2.7. Statistical analysis of soil water infiltration models*

229 The 72-sampled soil data was randomly divided into two parts including a training dataset
230 (80 %) and test dataset (20 %). The relationship between predicted and measured water infiltration
231 was evaluated by the following statistical analysis:

$$\mathrm{ME} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\widehat{\mathrm{I}_\mathrm{i}} - \mathrm{I}_\mathrm{i}) \tag{8}$$

$$\mathrm{RMSE} = \left[\frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}\left(\widehat{\mathrm{I}_\mathrm{i}} - \mathrm{I}_\mathrm{i}\right)^2}{\mathrm{N}}\right]^{0.5} \tag{9}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{I}_\mathrm{i} - \hat{\mathrm{I}})^2}{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{I}_\mathrm{i} - \overline{\mathrm{I}_\mathrm{i}})^2} \tag{10}$$

232 Where ME is mean error, RMSE is root mean square error, R^2 is coefficient of determination, $\overline{I_i}$ is
233 the mean of the observed water infiltration and N is the number of data (i.e. 72). R^2 ranges between

234 0 and 1 ($0 < R^2 < 1$) and RMSE is higher 0 (RMSE > 0). It is obvious that the best model has the
235 ME and RMSE values close to 0 and the higher the R^2 values.
236
237 **3. Results and Discussion**
238 *3.1. Evaluation of current infiltration models*
239 The cumulative infiltration (I) data vs. time (t) for a soil is presented in Fig. 4. The infiltration
240 models were fitted to the infiltration data for each soil for obtaining the parameters of water
241 infiltration models. An example of the fitting of water infiltration models to the measured
242 cumulative infiltration data for a soil is given in Fig 4. Table 3 showed the information of obtained
243 parameter values for water infiltration models including minimum, maximum and mean values. In
244 addition, the mean values of R^2, ME and RMSE indices are given in Table 4. As can be seen in
245 Table 4, the minimum and maximum mean R^2 and RMSE values varied from 0.968 to 0.999 and
246 0.232 to 1.168, respectively. The SDRMSE values showed that the highest consistency was
247 observed for KL and NRCS models in different Marvdasht plain soils (Table 4). On the other hand,
248 KL and NRCS compared with other soil water infiltration models are less dependent to the
249 hydraulic properties of the soil.
250 The ME values for all models are presented in Table 4. Based on the ME criteria, KL, Ho, NRCS
251 and Ph over-estimated and Ko and GA under-estimated the cumulative water infiltration (Tables
252 4). However, the over-estimation and under-estimation were very small for all models. Generally,
253 the result performance of water infiltration models for all soils, according to the mean R^2, ME,
254 RMSE and SDRMSE indices listed in Table 4, showed that the KL model provided the most
255 accurate predictions.
256 **Insert Fig 4**

Insert Table 3

Insert Table 4

3.2. Prediction of cumulative soil water infiltration using MLR- and ANN-based PTFs models

3.2.1. By MLR models

The results in Table 5 present the MLR models for estimating cumulative infiltration at different times (5, 10, 20, 45, 90, 150, 210 and 270 min) and their performance indices in the soils affected by saline, sodic and calcareous materials. Based on the results in Table 5, the derived MLR models used organic carbon and soil particles data such as sand, clay and Dg data. The cumulative water infiltration at 5, 10, 20, 45, 90, 150 minutes was influenced by exchangeable sodium percent (ESP) (Table 5). The cumulative water infiltration decreased with increasing ESP factor at the studied area. In addition, the electrical conductivity (EC) and calcium carbonate equivalent (CCE) affected cumulative water infiltration at times 5, 10 min and 45, 90 min, respectively (Table 5). The R^2 criteria values ranged from 0.299 to 0.582, RMSE values ranged from 1.110 to 19.584 and ME values ranged from -0.0016 to 1.3231 at the testing phase (Table 5). The scatter plots of the observed vs. the estimated values of soil water infiltration for all cumulative water infiltration data at different times by MLR models are shown in Fig. 5(a).

Insert Table 5

3.2.2. By ANN models

The results of applying ANN models to estimate cumulative water infiltration at different times are shown in Table 6. As can be seen in Table 6, the ME values for the test data set showed that the ANN models consistently overestimated the cumulative water infiltration at 5, 10, 20, 45, 150, 210 and 270 minute times (Table 6). The R^2 and RMSE values ranged from 0.438 to 0.661

and 0.977 to 17.111, respectively (Table 6). The scatter plot of the measured and predicted values of cumulative water infiltration at all times are given in Fig. 5(b).

Insert Table 6

3.2.3. By SVM models

The results indicating the incorporating hydraulic properties of the soil for developing SVM-based PTF are presented in Table 7. The R^2 values of the derived PTFs for estimating cumulative infiltration at different times subset ranged from 0.496 to 0.720 at testing phase (Table 7). As well as, the ME and RMSE values of SVM-based PTFs at different times subset are given in Table 7. The scatter graph of the measured versus predicted values of cumulative water infiltration with respect to the 1:1 reference line at all times are presented in Fig. 5(c).

Insert Table 7

Insert Fig 5

4. Discussion

For predicting water infiltration, the KL model had the best performance for the majority of studied soils in the Marvdasht plain. The result of this study is consistent with Parlange and Haverkamp (1989), Mirzaee et al. (2014) and Babaei et al. (2018), which showed that the KL model was the best model describing cumulative soil water infiltration under ponding conditions.

Moreover, for predicting cumulative soil water infiltration at specific times, point-based PTFs models were used in this study. The ANN models outperformed the MLR models at different time intervals (Table 5, 6). This differences in the results of ANN- and MLR-based PTFs models performance could be related to the relationships between soil water infiltration at different times and easily measured hydraulic properties of the soil that might be a non-linear relationship.

Parchami-Araghi et al. (2013) declared that the ANN models had the best performance in estimating the cumulative infiltration for some calcareous soils. Kashi et al. (2014) reported the performance of MLP better than the MLR, RBF and ANFIS models in estimating the soil infiltration rate. Indeed, several studies for predicting hydraulic properties of the soil such as soil water retention (Pachepsky et al. 1996), surface shear strength (Khalilmoghadam et al. 2009) and erodibility parameters (Mirzaee et al. 2017; Mirzaee et al. 2020) have shown that the ANN models outperform MLR models.

Overall, the comparison of performance indices at testing phase in Table 5, 6 and 7 showed the SVM models at all times performed better than MLR and ANN models for the estimation cumulative water infiltration. The RMSE values for SVM (0.755-15.108) were smaller than those for ANN (0.977-17.111) and the MLR (1.110-19.584) for the test data-set of derived PTFs. In contrast, the R^2 values for SVM (0.496-0.720) were larger than those for ANN (0.438-0.661) and the MLR (0.299-0.582) (Table 5, 6, 7). The SVM method has some advantages over ANN: (i) SVM is not affected by the problem of local minima, (ii) needs low learning parameters and (iii) produces robust results (Khlosi et al. 2016).

The accurate prediction of cumulative soil water infiltration at different times under unique influence of saline, sodic and calcareous materials is of crucial importance for arid and semi-arid regions. Soil texture data and organic carbon are easily measured hydraulic properties of the soil, which were used in developed MLR- and ANN-based PTFs at present study. Lilly et al. (2008) also used soil particles and organic carbon to predict hydraulic properties of the soil. The presence of saline, sodic and calcareous materials in southern parts of Iran is a serious problem for agricultural production. The calcium carbonate equivalent content in the study area ranged from 11.5 to 66.5% (Table 2). The high calcium carbonate contents in these soils result in nutrient

deficiency of phosphorous, zinc and iron elements. In calcareous soils, the hydrolysis of carbonate ions controls soil pH (George et al. 2012). In calcareous soils in southern Iran, Ostovari et al. (2016) and Tashayo et al. (2020) recently observed that calcium carbonate has great influence on the stability of aggregates and soil permeability. Khalilmoghadam et al. (2009), Havaee et al. (2015) and Mirzaee et al. (2017) also showed that the calcium carbonate content is an important factor affecting hydraulic properties of the soil in arid regions. The salinity in the study region varied from 0.04 to 12.69 dS m^{-1}. It is clear that the salinity influences the water infiltration, uptake of nutrients and crop yield and growth (George et al. 2012). In addition, in sodic soil crop growth and yield negatively affects due to the high sodium levels. In these soils, soil particles swell after an irrigation or rainfall and resulting in convex surfaces. Soil water infiltration in sodic soils are very limited (George et al. 2012). Hence, it is important to calibrate and develop soil water infiltration models for different field conditions such as those found in the Marvdasht plain soils, which are affected by saline, sodic and calcareous materials.

5. Conclusion

This study attempts to show the potential of MLR-, ANN- and SVM-based PTFs models to estimate cumulative water infiltration at specific times for some calcareous, saline and sodic soils. The results of the present study indicate that the derived SVM-based PTFs models, linking more easily measured hydraulic properties of the soil including sand, silt, clay, Dg, Sg, CCE and SOC variables, performed better than MLR- and ANN-based PTFs models at all times. The variation of cumulative soil water infiltration at 5, 10, 20, 45, 150, 210 and 270 min explained up to 68.5, 72.0, 64.9, 66.5, 49.6, 63.4, 69.0 and 63.5%, respectively, by the best developed model (i.e. SVM-based

PTFs models). To save time and cost in large-scale, derived SVM-based PTFs models in this study could be applicable models for the cumulative infiltration prediction of optimal design and management of irrigation systems for calcareous, saline and sodic soils in semi-arid environments. However, optimal design irrigation systems according to the water infiltration information are more important to avoid over-irrigation or under-irrigation of crops. It would be a great interest to improve the cumulative soil water infiltration by MLR, ANN and SVM methods in the presence of soil structure information in future research.

6. References

Angelaki A, Sakellariou-Makrantonaki M, Tzimopoulos C (2013) Theoretical and experimental research of cumulative infiltration. Transport in Porous Media 100(2): 247-257.

Babaei F, Zolfaghari AA, Yazdani MR, Sadeghipour A (2018) Spatial analysis of infiltration in agricultural lands in arid areas of Iran. Catena 170: 25-35.

Baker L, Ellison D (2008) Optimization of pedo-transfer functions using an artificial neural network ensemble method. Geoderma 144: 212-224.

Baumhardt RL, Lascano RJ (1993) Physical and hydraulic properties of a calcic horizon. Soil Sci. 155 (6): 368-374.

Blume HP, Brümmer GW, Fleige H, Horn R, Kandeler E, Kögel-Knabner I (2016) Soil Science (Scheffer/Schachtschabel). Springer-Verlag, Berlin. pp. 175-284.

Cressie N (1993) Statistics for spatial data. John Wiley & Sons, New York.

Cui G, Zhu J (2018) Infiltration model based on traveling characteristics of wetting front. Soil Sci. Soc. Am. J. 82: 45-55.

374 Deiss L, Margenot AJ, Culman SW, Demyan MS (2020) Tuning support vector machines
375 regression models improves prediction accuracy of soil properties in MIR spectroscopy. Geoderma
376 365: 114227.
377 Gee GW, Bauder JW (1986) Particle size analysis. In: Klute, A. (Ed.), Methods of Soil
378 Analysis: Part 1 Agronomy Handbook No 9. American Society of Agronomy and Soil Science
379 Society of America, Madison, WI, pp. 383-411.
380 George E, Horst WJ, Neumann E (2012) Adaptation of plants to adverse chemical soil
381 conditions, in: Marschner, P. (Ed.), Marschner's mineral nutrition of higher plants. London,
382 Academic Press, pp. 409-472.
383 Ghorbani Dashtaki Sh, Homaee M, Mahdian MH, Kouchakzadeh M (2009) Sitedependence
384 performance of infiltration models. Water Resour. Manage. 23: 1573-1650.
385 Ghorbani-Dashtaki S, Homaee M, Loiskand W (2016) Towards using pedotransfer functions
386 for estimating infiltration parameters. Hydrological Sciences Journal 61(8): 1477-1488.
387 Green WH, Ampt CA (1911) Studies on soil physics, I. Flow of air and water through soils.
388 J Agric Sci. 4, 1-24.
389 Hagan MT, Demuth HB, Beale MH (1996) Neural Network Design. PWS, Boston, MA.
390 Haghverdi A, Cornelis WM, Ghahraman B (2012) A pseudo-continuous neural network
391 approach for developing water retention pedotransfer functions with limited data. J. Hydrol 442-
392 443: 46-54.
393 Havaee S, Mosaddeghi MR, Ayoubi S (2015) In situ surface shear strength as affected by
394 soil characteristics and land use in calcareous soils of central Iran. Geoderma 237-238: 137-148.
395 Hillel D (2003) Environmental soil physics. New York: Academic Press.

Ho R (2006) Handbook of univariate and multivariate data analysis and interpretation with SPSS. Chapman and Hall/CRC, Boca Raton.

Horton RE (1940) Approach toward a physical interpretation of infiltration capacity. Soil Sci Soc Am J. 5: 339-417.

Iounousse J, Er-Raki E, Motassadeq AE, Chehouani H (2015) Using an unsupervised approach of probabilistic neural network (PNN) for land use classification from multitemporal satellite images. Appl Soft Comput. 30: 1-1.

Kashi H, Emamgholizadeh S, Ghorbani H (2014) Estimation of soil infiltration and cation exchange capacity based on multiple regression, ANN (RBF, MLP), and ANFIS models. Commun. Soil Sci. Plant Anal. 45(9): 1195-1213.

Khalilmoghadam B, Afyuni M, Abbaspour KC, Jalalian A, Dehghani AA, Schulin R (2009) Estimation of surface shear strength in Zagros region of Iran - a comparison of artificial neural networks and multiple-linear regression models. Geoderma 153: 29-36.

Khlosi M, Alhamdoosh M, Douaik A, Gabriels D, Cornelis WM (2016) Enhanced pedotransfer functions with support vector machines to predict water retention of calcareous soil. Eur. J. Soil Sci. 67: 276-284.

Kostiakov AV (1932) On the dynamics of the coefficient of water percolation in soils and on the necessity for studying it from a dynamics point of view for purposes of amelioration. In: Fauser O, editor. Transactions of the Sixth Commission of International Society of Soil Science, Part A. Leiden: Martinius Nijhoff Publishers; p. 15-21.

Licznar P, Nearing MA (2003) Artificial neural networks of soil erosion and runoff prediction at the plot scale. Catena 51: 89-114.

Lilly A, Nemes A, Rawls W, Pachepsky YA (2008) Probabilistic approach to the identification of input variables to estimate hydraulic conductivity. Soil Sci. Soc. Am. J. 72(1): 16-24.

Machiwal D, Madan Kumar JHA, Mal BC (2006) Modelling infiltration and quantifying spatial soil variability in a watershed of Kharagpur, India. Biosyst Eng. 95: 569-582.

Mackay DJC (1992) Bayesian interpolation. Neural Computation 4(3): 415-447.

Mezencev VJ (1948) Theory of formation of the surface runoff. Meteorol. I gidrologia 3: 33-46 (in Russian).

Minasny B, McBratney AB (2002) The Neuro-m method for fitting neural network parametric pedo-transfer functions. Soil Sci. Soc. A. J 66: 352-361.

Mirzaee S, Zolfaghari AA, Gorji M, Dyck M, Ghorbani Dashtaki S (2014) Evaluation of infiltration models with different numbers of fitting parameters in different soil texture classes. Archives of Agronomy and Soil Science 60(5): 681-693.

Mirzaee S, Ghorbani-Dashtaki S, Mohammadi J, Asadzadeh F, Kerry R (2017) Modeling WEPP erodibility parameters in calcareous soils in northwest Iran. Ecological Indicators 74: 302-310.

Mirzaee S, Ghorbani-Dashtaki S, Kerry R (2020) Comparison of a spatial, spatial and hybrid methods for predicting inter-rill and rill soil sensitivity to erosion at the field scale. Catena 188: 104439.

Nelson DW, Sommers LP (1986) Total carbon, organic carbon and organic matter. In: Page, A.L. (Ed.), Methods of Soil Analysis: Part 2: Agronomy Handbook No 9. American Society of Agronomy and Soil Science Society of America, Madison, WI, pp. 539-579.

Nelson RE (1982) Carbonate and gypsum. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis. Part 2, second ed. Agron. Monogr. 9. ASA, Madison, WI, pp. 181-197.

Neshat A, Parehkar M (2007) The compaeison of methods for determining the vertical infiltration rate. Journal of Agriculture Science and Natural Resources 14(3): 186-195 (In Persian).

Nguyen PM, De Pue J, Le KV, Cornelis WM (2015) Impact of regression methods on improved effects of soil structure on soil water retention estimates. Journal of Hydrology 525: 598-606.

Ostovari Y, Ghorbani-Dashtaki S, Bahrami HA, Naderi M, Dematte JA., Kerry R (2016) Modification of the USLE K factor for soil erodibility assessment on calcareous soils in Iran. Geomorphology 273: 385-395.

Ostovari Y, Ghorbani-Dashtaki S, Bahrami HA, Naderi M, Dematte JAM (2017) Soil loss prediction by an integrated system using RUSLE, GIS and remote sensing in semi-arid region. Geoderma Reg. 11: 28-36.

Ostovari Y, Moosavi AK, Pourghasemi HR (2020) Soil loss tolerance in calcareous soils of a semiarid region: evaluation, prediction, and influential parameters. Land Degrad. Dev. 1-12.

Pachepsky YA, Timlin D, Varallyay G (1996) Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci. Soc. Am. J 60: 727-733.

Pahlavan-Rad MR, Dahmardeh K, Hadizadeh M, Keykha G, Mohammadnia N, Gangali M, Keikha M, Davatgar N, Brungardf C (2020) Prediction of soil water infiltration using multiple linear regression and random forest in a dry flood plain, eastern Iran. Catena 194: 10471.

Parchami-Araghi F, Mirlatifi M, Ghorbani Dashtaki Sh, Mahdian MH (2013) Point estimation of soil water infiltration process using Artificial Neural Networks for some calcareous soils. Journal of Hydrology 481: 35-47.

463 Parhi PK, Mishra SK, Singh R (2007) A modification to Kostiakov and modified Kostiakov
464 infiltration models. Water Resources Management 21: 1973-1989.
465 Parlange JY, Haverkamp R (1989) Infiltration and ponding time. In: Morel-Seytoux HJ,
466 editor. Unsaturated flow in hydrologic modeling, theory and practice. Boston (MA): Kluwer
467 Academic; p. 95-126.
468 Philip JR (1957) The theory of infiltration: 1. The infiltration equation and its solution. Soil
469 Science. 83: 345-357.
470 Rawls WJ, Ahuja LR, Brakensiek DL, Shirmohammadi A (1993) Infiltration and soil water
471 movement. In: Maidment, D.R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York.
472 Schaap MG, Leij FJ, Van Genuchten MTh (1998) Neural network analysis for hierarchical
473 prediction of soil hydraulic properties. Soil Sci. Soc. A. J 62: 847-855.
474 Shirazi MA, Boersma L (1984) A unifying quantitative analysis of soil texture. Soil Sci. Soc.
475 A. J 48: 142-147.
476 Sihag P, Singh B (2018) Field evaluation of infiltration models. Scientific and Technical
477 Journal 4: 1-12.
478 Sihag P, Keshavarzi A, Kumar V (2019a) Comparison of different approaches for modeling
479 of heavy metal estimations. SN Appl. Sci. 1: 780.
480 Sihag P, Singh VP, Angelaki A, Kumar V, Sepahvand A, Golia E (2019b) Modelling of
481 infiltration using artificial intelligence techniques in semi-arid Iran. Hydrological Sciences Journal
482 64(13): 1647-1658.
483 Sihag P, Tiwari NK, Ranjan S (2020) Support vector regression-based modeling of
484 cumulative infiltration of sandy soil. ISH Journal of Hydraulic Engineering 26(1): 44-50.

485 Somaratne S, Seneviratne G, Coomaraswamy U (2005) Prediction of soil organic
486 carbon across different land-use patterns: a neural network approach. Soil Sci. Soc. Am. J. 69:
487 1580-1589.
488 Tashayo B, Honarbakhsh A, Akbari M, Ostovari Y (2020) Digital mapping of Philip model
489 parameters for prediction of water infiltration at the watershed scale in a semi-arid region of Iran.
490 Geoderma Reg. 22: e00301.
491 The MathWorks Inc (2007) MATLAB: the language of technical computing. Version 7.5.
492 Ismaning: The MathWorks Inc.
493 Turner ER (2006) Comparison of infiltration equations and their field validation with rainfall
494 simulation [MSc. thesis]. College Park (MD): University of Maryland; 202 pp.
495 USDA (2010) Keys to Soil Taxonomy, 11th ed., USDA National Resources Conservation
496 Service, Washington, DC.
497 US Department of Agriculture Natural Resources and Conservation Service (NRCS) (2005)
498 National Engineering Handbook, Part 623, Surface Irrigation. National Technical Information
499 Service, Washington, DC (Chapter 4).
500 US Department of Agriculture, Natural Resources and Conservation Service (NRCS) (1974)
501 National Engineering Handbook. Section 15. Border Irrigation. National Technical Information
502 Service, Washington, DC (Chapter 4).
503 Vapnik VN (1995) The Nature of Statistical Learning Theory. J. Wiley & Sons, New York.
504 Vereecken H, Weynants M, Javaux M, Pachepsky Y, Schaap MG, Van Genuchten MTh
505 (2010) Using pedotransfer functons to estimate the van Genuchten–Mualem soil hydraulic
506 properties: A review. Vadose Zone J. 9: 795-820.

Walker WR, Prestwich C, Spofford T (2006) Development of the revised USDA-NRCS intake families for surface irrigation. Agric. Water Manage 8 (5): 157-164.

Wen X, Zhao Z, Deng X, Xiang W, Tian D, Yan W, Zhou X, Peng C (2014) Applying an artificial neural network to simulate and predict Chinese fir (Cunninghamia lanceolata) plantation carbon flux in subtropical China. Ecological Modelling. 294: 19-26.

Zhang X, Song Q, Gao Z, Zheng Y, Weng P, Jiao LC (2016) Spectral-spatial feature learning using cluster-based group sparse coding for hyperspectral image classification. IEEE J Sel Top Appl Earth Obs Remote Sens. 9(9): 4142-4159.

Zolfaghari Z, Mosaddeghi MR, Ayoubi S (2015) ANN-based pedo-transfer 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 (1): 142-154.

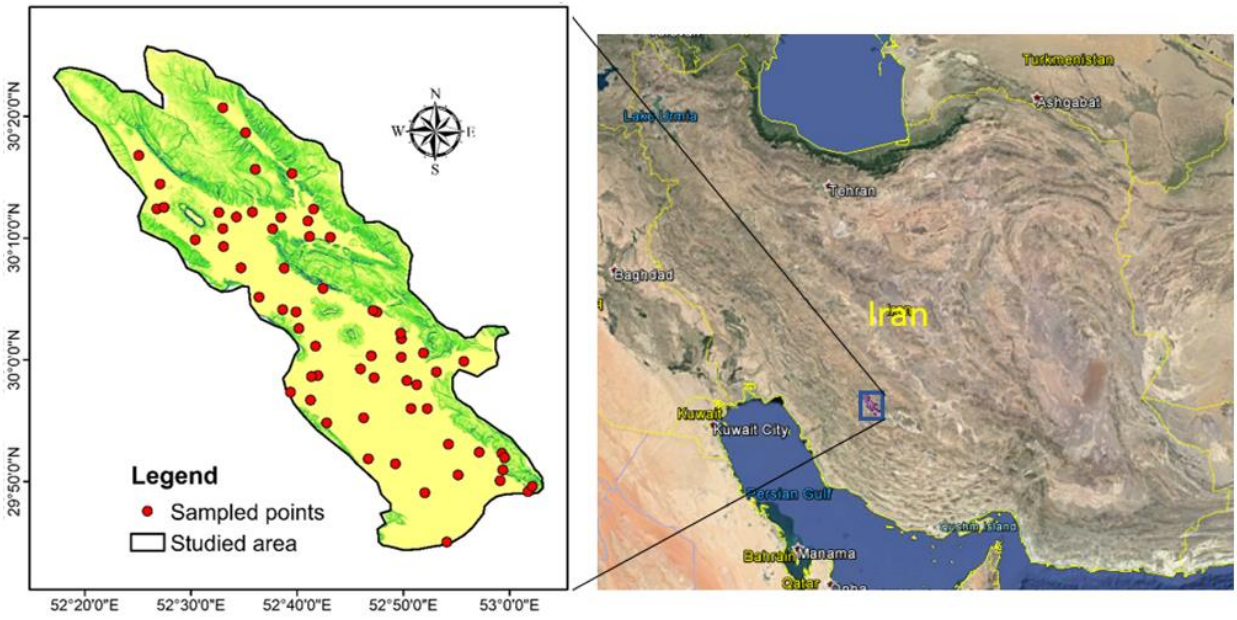

Fig 1. Location of study area in the northeast Fars province, Iran and distribution of sample points.

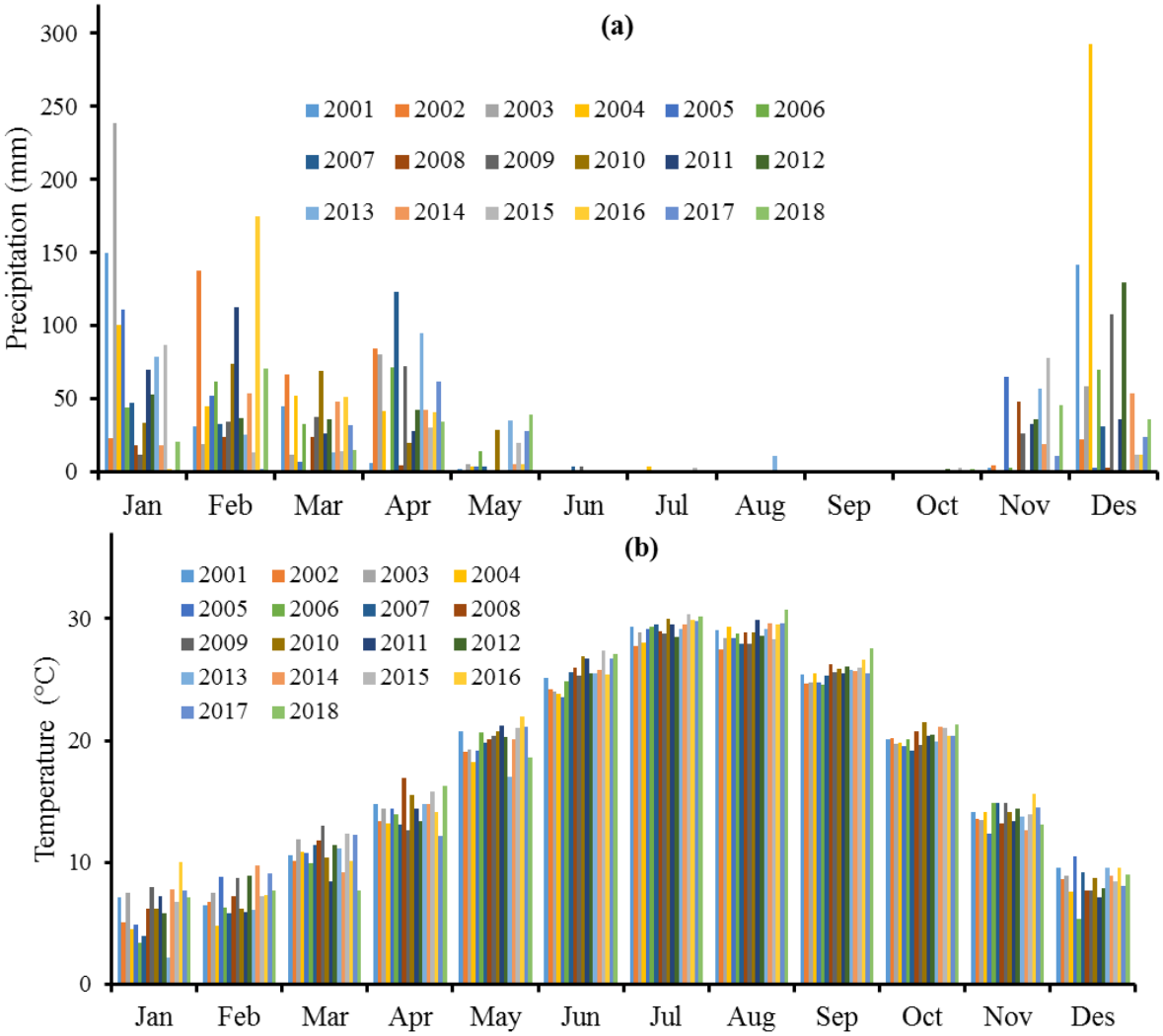

Fig 2. Average monthly precipitation and temperature (from 2001 to 2018)

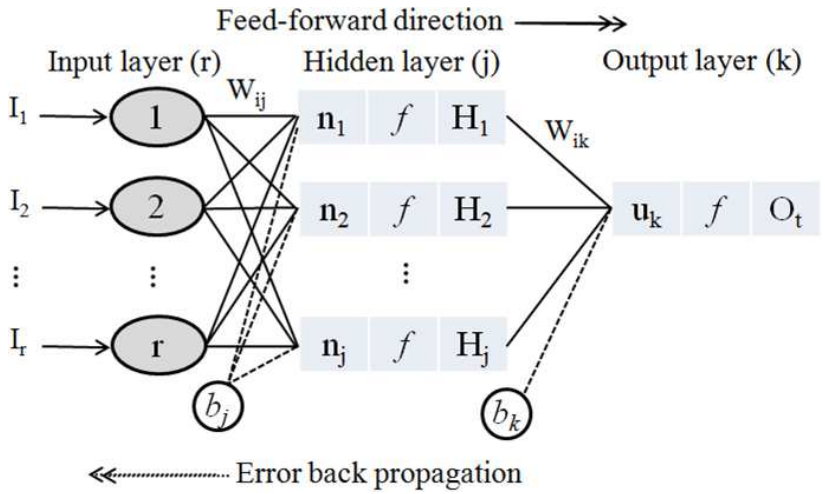

Fig. 3. Schematic of multilayer perceptron used to predict soil water infiltration.

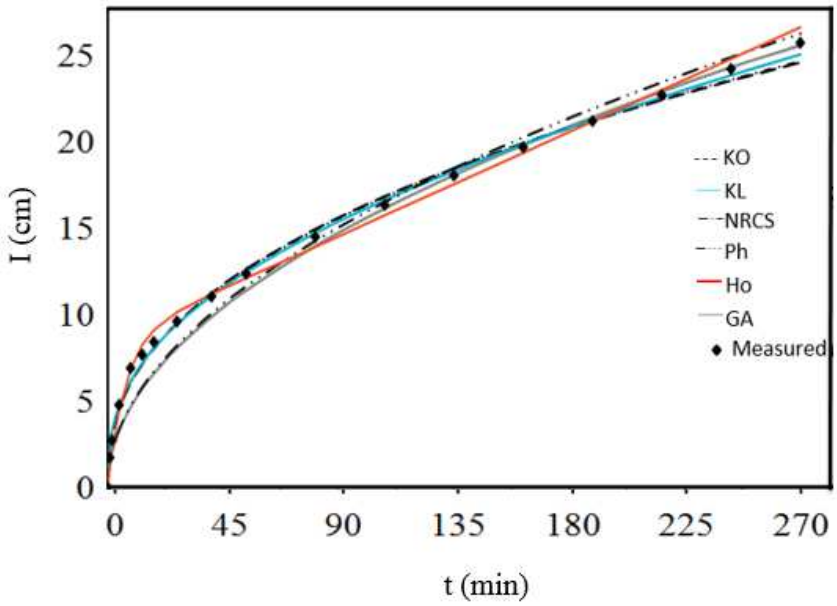

Fig 4. Fitted infiltration curve for a given soil, I is the cumulative infiltration (cm) and t is the time (min) (For abbreviations of models, refer to Table 1).

583

584

585

586

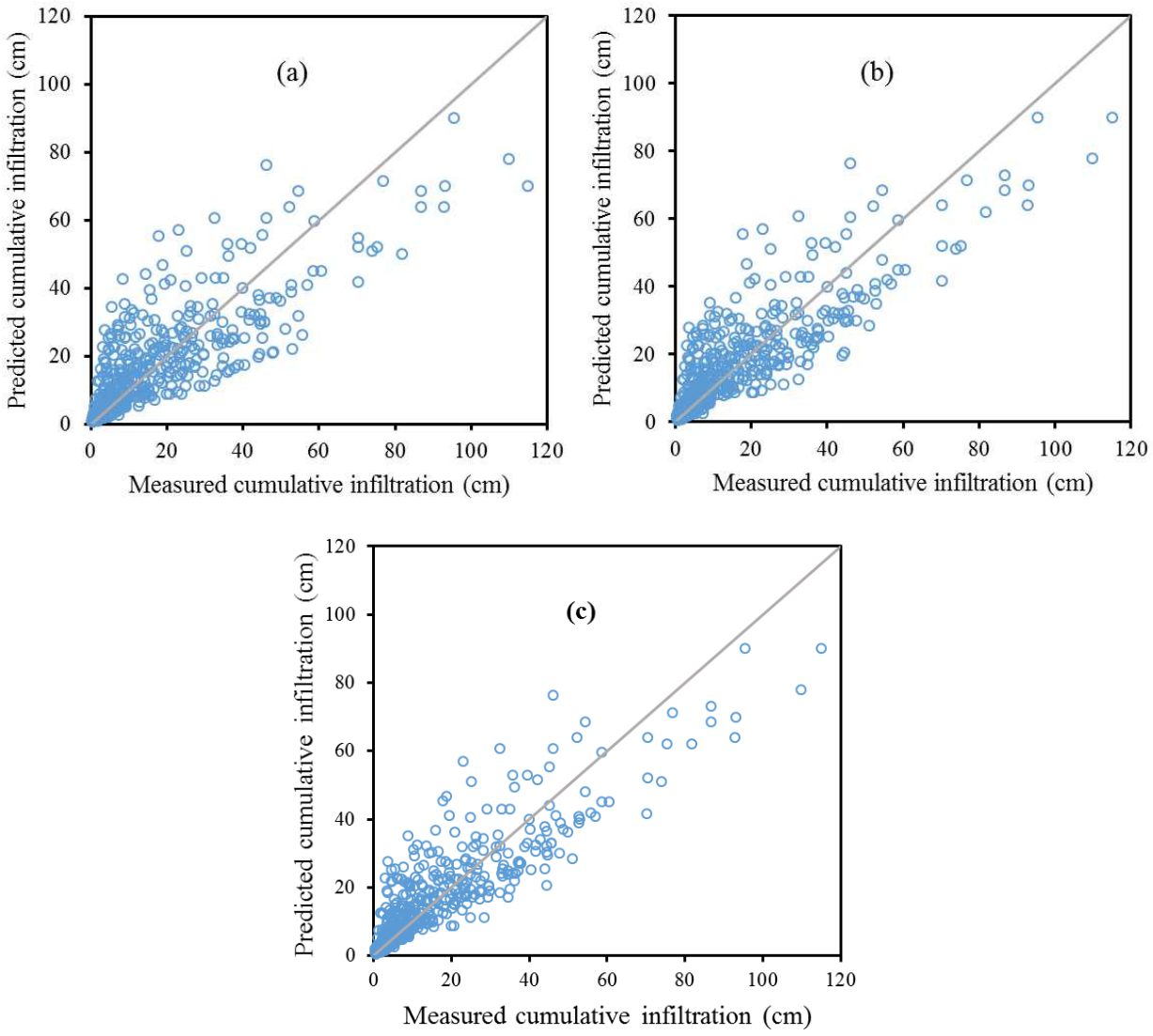

Fig 5. Measured versus predicted cumulative water infiltration data at regression models (a), ANN models (b) and SVM (c).

587

588

589

590

Table 1
The infiltration equation of physical and empirical-based models.

Model name	Abbreviation	Equation	Model parameters
Emperical-based model			
Kostiakov (1932)	Ko	$I = at^b$	a>0 and 0<b<1 are dimensionless constants
Kostiakov–Lewis (Mezencev, 1948)	KL	$I = at^b + kt$	0<b<1, k>0 and a>0 are dimensionless constants
USDA-NRCS (1974)	NRCS	$I = at^b + 0.698$	a and b are dimensionless constants
Physical-based model			
Philip (1957)	Ph	$I = St^{0.5} + At$	S is sorptivity ($LT^{-1/2}$) and A is transmissivity factor (LT^{-1})
Horton (1940)	Ho	$I = ct + m(1 - e^{-at})$	c (LT^{-1}) is the basic infiltration rate, m=(i_0-i_c)/a, where i_0 (LT^{-1}) is the initial infiltration rate at t = 0, a > 0 is a constant that determines the rate at which i_0, approaches i_c.
Green-Ampt (1911)	GA	$I = kt + Gln(1 + I/G)$	K(LT^{-1}) is field-saturated hydraulic conductivity of the transmission zone and G = (h_s-h_t) Δθ, where Δθ is soil moisture deficit, and h_s (L) and h_t (L) are soil water pressure head at the surface and at the wetting front, respectively.

608

609

610

611

612

613

Table 2
Descriptive statistics of study soil.

Properties	Min	Max	Mean	Std. dev
Clay (%)	16.3	38.9	25.4	4.98
Silt (%)	20.0	40.0	31.3	3.99
Sand (%)	30.1	62.8	43.4	6.95
OC (%)	0.99	2.16	1.73	0.28
CCE (%)	11.5	66.5	41.9	11.25
Dg (mm)	0.026	0.097	0.057	0.019
Sg	8.78	21.1	13.7	2.65
pH	7.03	7.89	7.59	0.197
EC (dS m^{-1})	0.04	12.69	2.14	2.21
ESP (%)	1.05	67.88	13.44	14.27

CCE: Calcium carbonate equivalent; Dg: Geometric mean diameter; Sg: Geometric standard deviation; EC: Electrical conductivity; ESP: Exchangeable sodium percentage.

614

615

616

617

618

619

620

621

622

623

624

Table 3
The parameter values of the water infiltration models for all soils

Model name	Parameters	Minimum	Maximum	Mean
GA				
	k	0.000	0.385	0.075
	G	0.255	3509	132.1
Ph				
	A	0.000	1.00	0.079
	S	0.000	3.16	0.901
Ho				
	a	0.000	1.00	0.123
	m	0.003	30.1	2.343
	C	0.002	44.9	4.869
NRCS				
	a	0.002	2.99	0.488
	b	0.300	1.32	0.804
Ko				
	a	0.278	1.00	0.645
	b	0.038	2.89	0.765
KL				
	a	0.000	1.00	0.071
	b	0.000	0.79	0.446
	k	0.000	2.99	0.929

640

641

642

643

644

645

Table 4
The evaluation indices of the water infiltration models for all soils

Statistics	GA	Ph	Ho	NRCS	Ko	KL
R^2	0.968 (6)[a]	0.989 (4)	0.982 (5)	0.991 (3)	0.996 (2)	0.999 (1)
ME	-0.005 (1)	0.034 (3)	0.201 (6)	0.036 (4)	-0.051 (5)	0.014 (2)
RMSE	1.167 (6)	0.373 (3)	0.824 (5)	0.384 (4)	0.370 (2)	0.232 (1)
SDRMSE	1.093 (5)	0.459 (4)	1.221 (6)	0.289 (2)	0.333 (3)	0.280 (1)
Final rank	(5)	(4)	(6)	(3)	(2)	(1)

[a]: Rank of models according to the evaluation indices is presented in the parenthesis.

646

647

648

649

650

651

652

653

654

655

656

657

658

659

Table 5
MLR–based models for estimation cumulative infiltration in 5, 10, 20, 45, 90, 150, 210 and 270 min.

MLR models	Train data			Test data		
	ME	RMSE	R^2	ME	RMSE	R^2
$I_5 = 1.321 + 0.011OC + 28.019Dg - 0.0519ESP + 0.1132EC$	-0.0018	1.115	0.396	-0.0016	1.110	0.426
$I_{10} = 1.873 + 0.081OC + 42.253Dg - 0.076ESP + 0.139EC$	0.2014	1.487	0.531	0.2008	1.454	0.561
$I_{20} = 2.539 + 0.321OC + 67.9919Dg - 0.089ESP$	0.2009	2.592	0.431	0.2000	2.599	0.423
$I_{45} = -9.102 + 2.262OC + 0.391Sand - 0.133ESP + 0.0036CCE$	0.1211	4.487	0.482	0.1206	4.481	0.519
$I_{90} = 25.182 + 4.031OC - 0.609Clay - 0.227ESP + 0.056CCE$	0.0036	9.083	0.292	0.0025	9.001	0.299
$I_{150} = -1.798 + 5.587OC + 348.861Dg - 0.324ESP$	-0.0603	12.923	0.402	-0.0700	12.999	0.400
$I_{210} = -28.017 + 17.963OC + 479.62Dg$	1.3828	15.329	0.545	1.3231	14.987	0.582
$I_{270} = -38.74 + 24.581OC + 588.17Dg$	0.6092	19.928	0.465	0.5675	19.584	0.495

I_5, I_{10}, I_{20}, I_{45}, I_{90}, I_{150}, I_{210} and I_{270} are cumulative water infiltration at 5, 10, 20, 45, 90, 150, 210 and 270 min, respectively.

Table 6
ANN–based models for estimation cumulative infiltration in 5, 10, 20, 45, 90, 150, 210 and 270 min.

ANNs models			Training data			Test data		
Time subset	Selected architecture	Transfer functions	ME	RMSE	R^2	ME	RMSE	R^2
I_5	2-5-1	logsig-purelin	0.1005	0.967	0.558	0.1015	0.977	0.533
I_{10}	3-7-1	logsig-purelin	0.2381	1.331	0.639	0.2250	1.310	0.645
I_{20}	2-4-1	logsig-purelin	0.3136	2.292	0.572	0.3211	2.255	0.579
I_{45}	3-5-1	logsig-purelin	0.6227	3.991	0.612	0.6255	3.998	0.608
I_{90}	5-6-1	logsig-purelin	-0.1525	8.403	0.418	-0.1400	8.356	0.438
I_{150}	5-6-1	logsig-purelin	0.2884	11.286	0.553	0.2810	11.199	0.583
I_{210}	5-8-1	logsig-purelin	1.8299	13.649	0.651	1.8123	13.576	0.661
I_{270}	5-5-1	logsig-purelin	1.8367	17.502	0.562	1.8412	17.111	0.613

I_5, I_{10}, I_{20}, I_{45}, I_{90}, I_{150}, I_{210} and I_{270} are cumulative water infiltration at 5, 10, 20, 45, 90, 150, 210 1nd 270 min, respectively.

696

697

698

699

700

701

702

Table 7
SVM–based models for estimation cumulative infiltration in 5, 10, 20, 45, 90, 150, 210 and 270 min.

Time subset	Training data			Test data		
	ME	RMSE	R^2	ME	RMSE	R^2
I_5	0.0089	0.766	0.690	0.0065	0.800	0.685
I_{10}	0.0564	1.102	0.725	0.1698	1.110	0.720
I_{20}	0.4532	1.998	0.667	0.3099	2.004	0.649
I_{45}	0.5564	3.645	0.673	0.4116	3.689	0.665
I_{90}	-0.1500	8.000	0.523	-0.1521	8.111	0.496
I_{150}	0.1265	10.187	0.624	0.1078	10.004	0.634
I_{210}	1.0086	12.657	0.694	1.0760	12.588	0.690
I_{270}	0.9703	15.009	0.640	1.0002	15.108	0.635

I_5, I_{10}, I_{20}, I_{45}, I_{90}, I_{150}, I_{210} and I_{270} are cumulative water infiltration at 5, 10, 20, 45, 90, 150, 210 1nd 270 min, respectively.

703

704

Figures

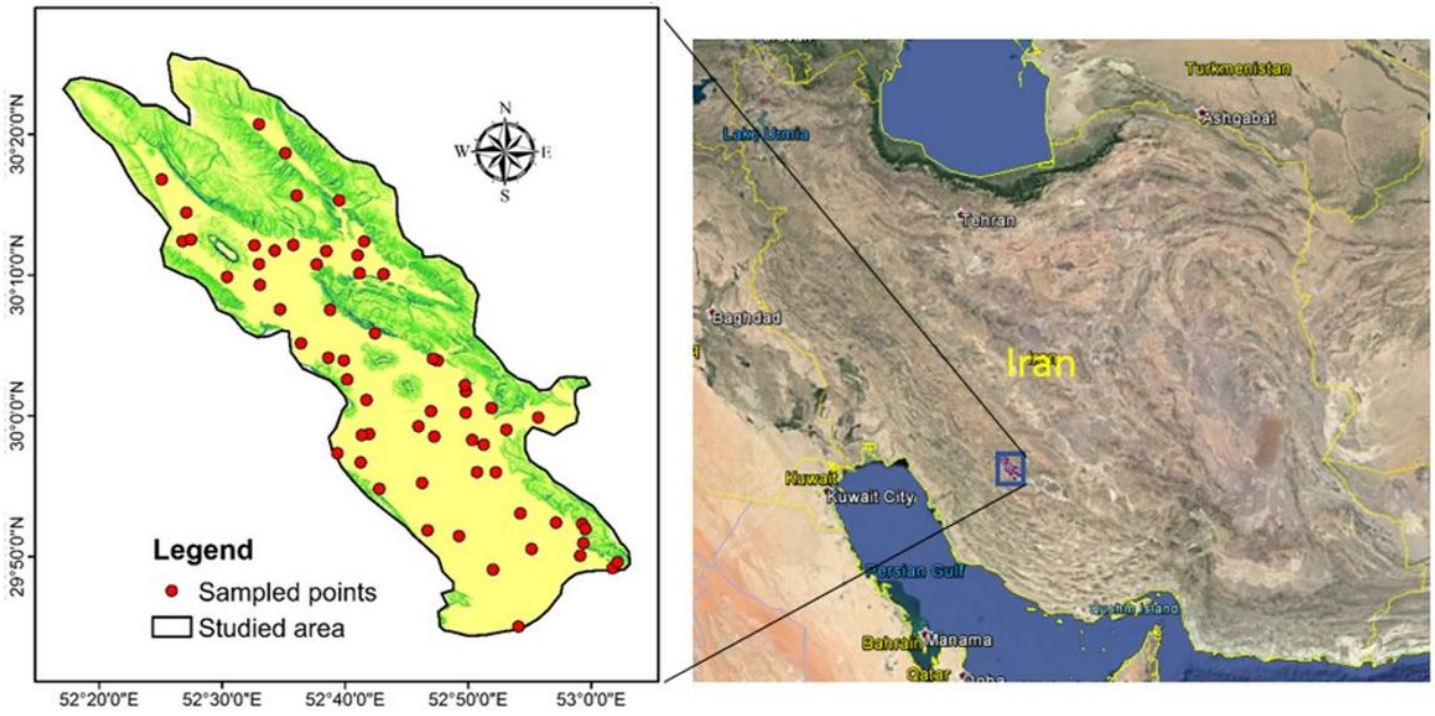

Figure 1

Location of study area in the northeast Fars province, Iran and distribution of sample points. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

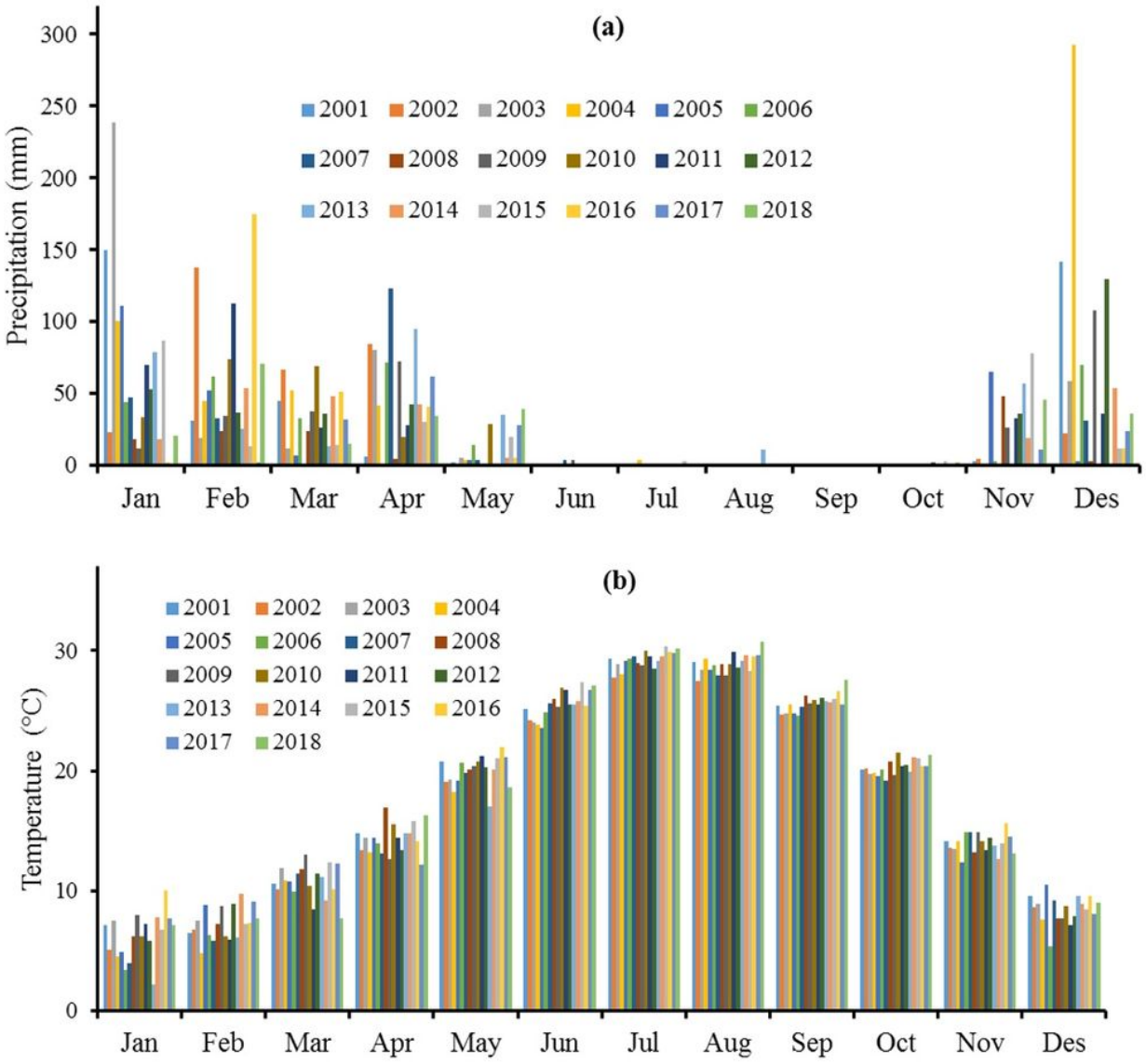

Figure 2

Average monthly precipitation and temperature (from 2001 to 2018)

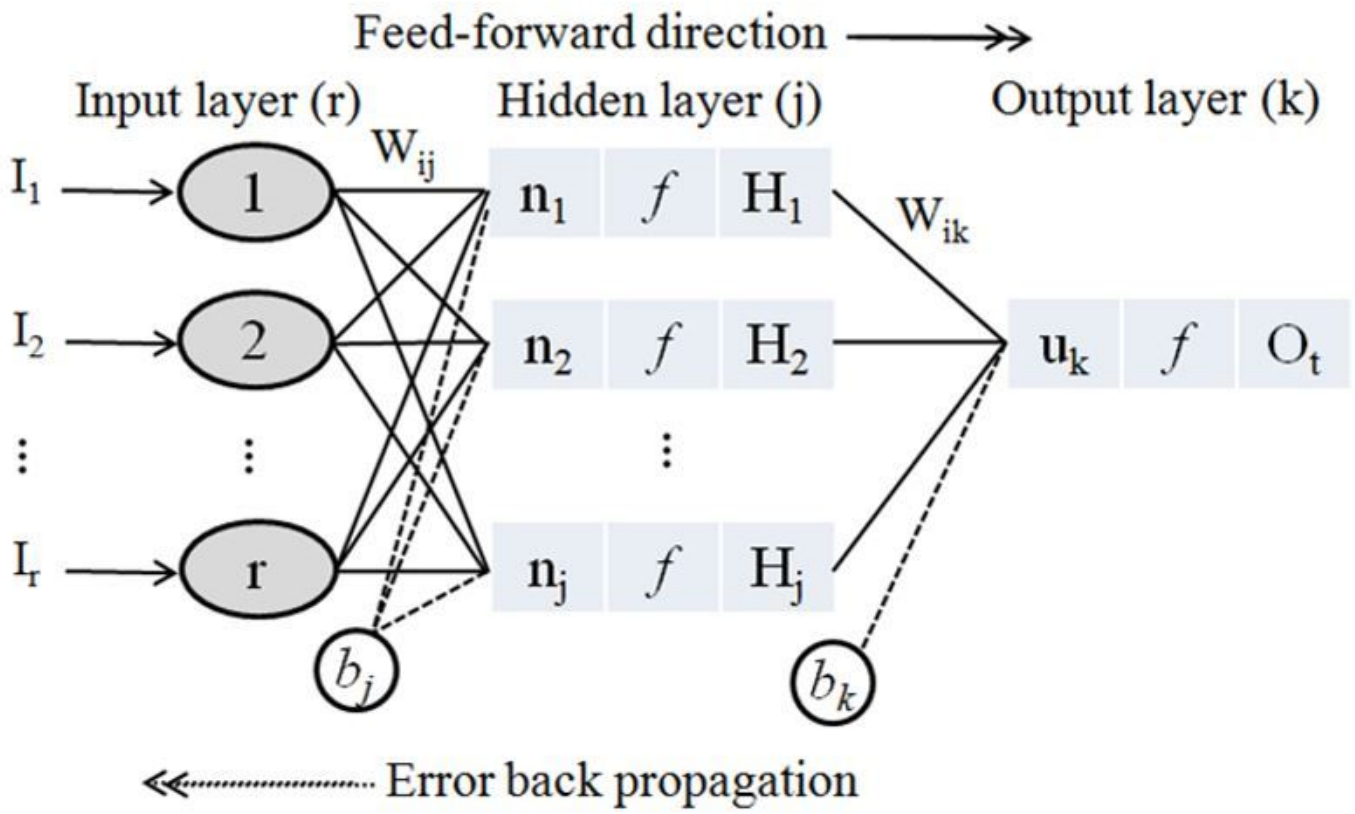

Figure 3

Schematic of multilayer perceptron used to predict soil water infiltration.

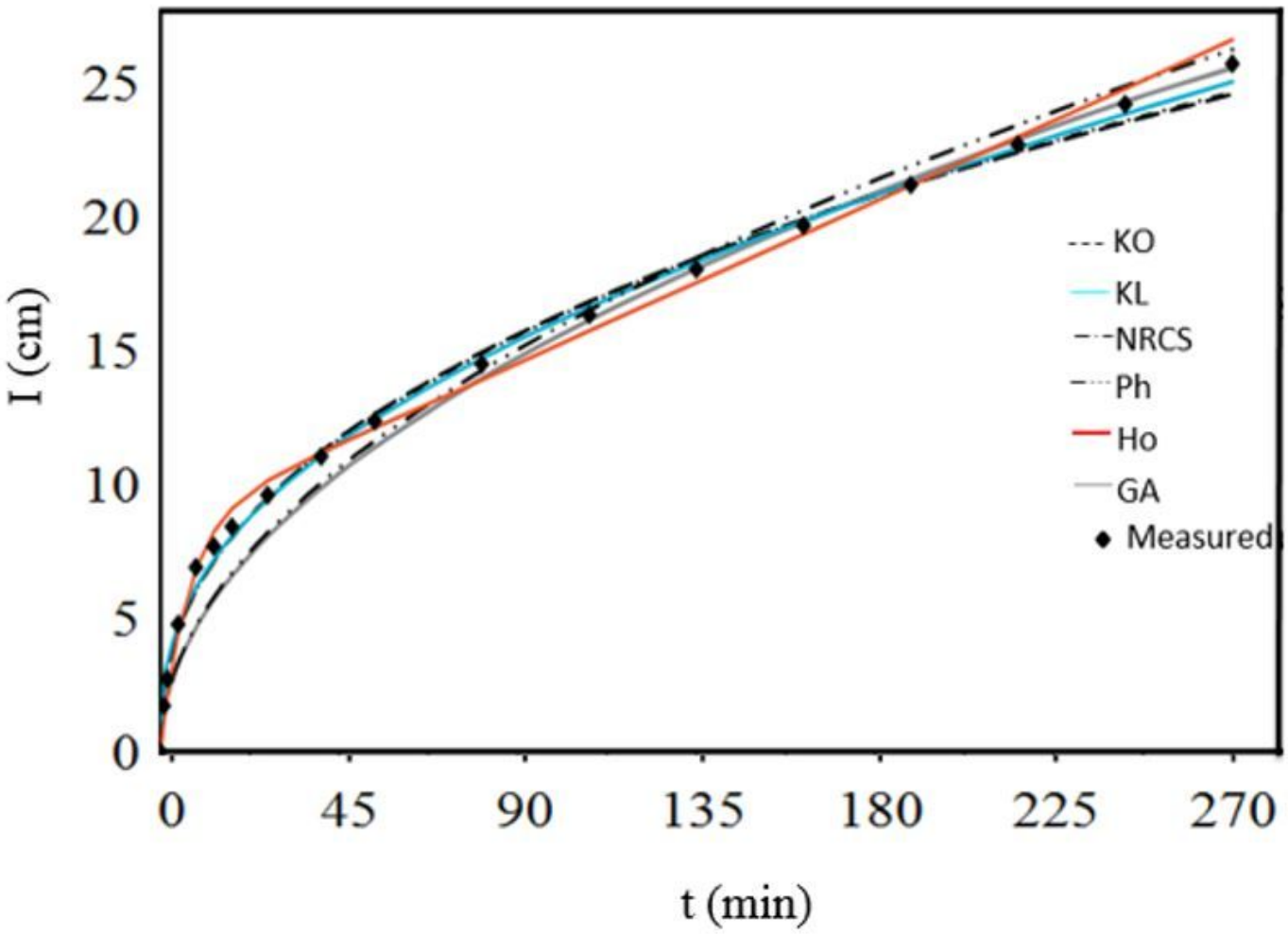

Figure 4

Fitted infiltration curve for a given soil, I is the cumulative infiltration (cm) and t is the time (min) (For abbreviations of models, refer to Table 1).

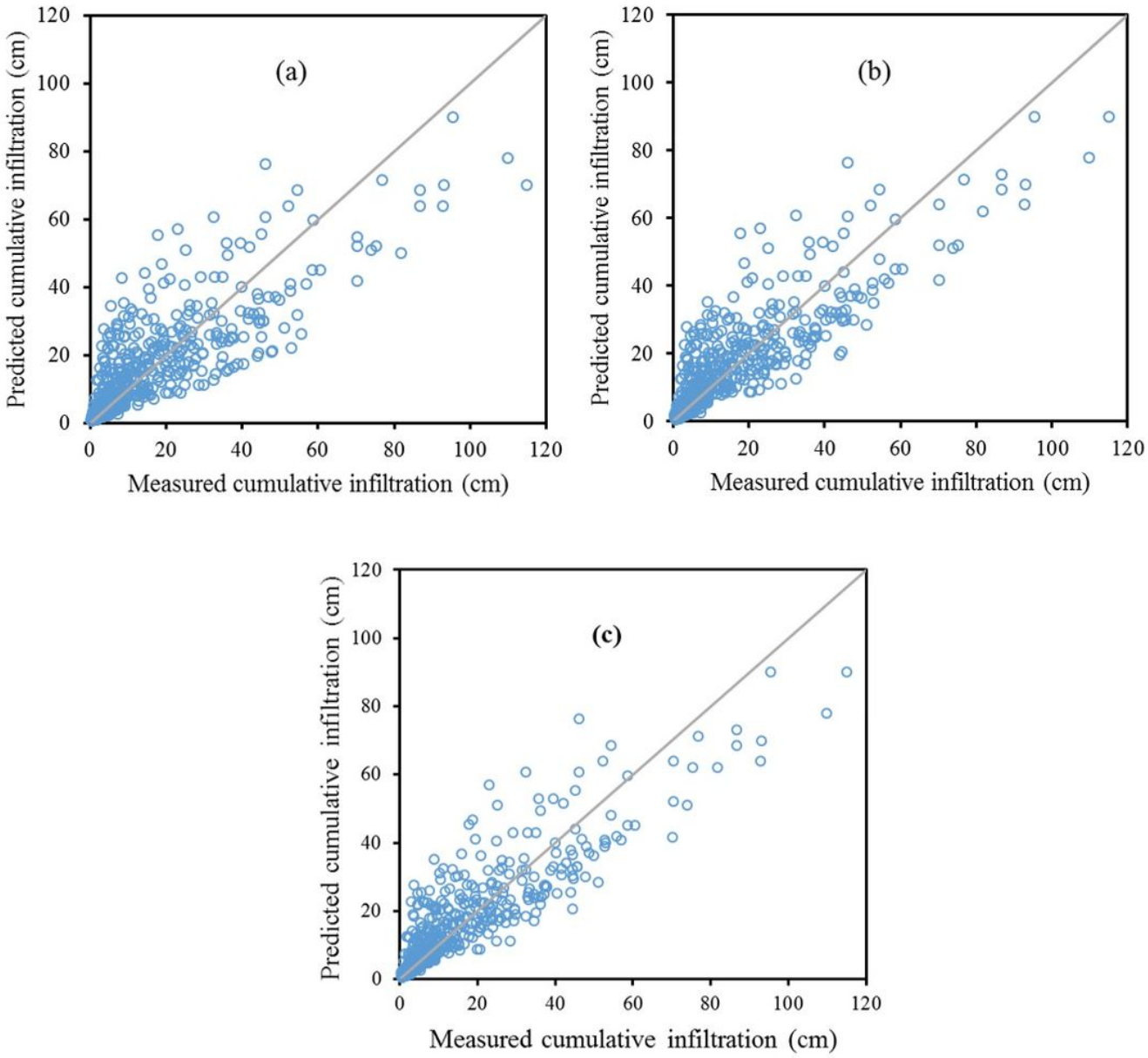

Figure 5

Measured versus predicted cumulative water infiltration data at regression models (a), ANN models (b) and SVM (c).