IJST, Transactions of Civil Engineering, Vol. 39, No. C1, pp 143-165
Printed in The Islamic Republic of Iran, 2015

PREDICTION OF SOIL-WATER CHARACTERISTIC CURVE USING GENE EXPRESSION PROGRAMMING*

A. JOHARI** AND A. HOOSHMAND NEJAD

Dept. of Civil and Environmental Engineering, Shiraz University of Technology, Shiraz, I. R. of Iran
Email: johari@sutech.ac.ir

Abstract– Soil–Water Characteristic Curve (SWCC) is one of the most important parts of any model that describes unsaturated soil behavior as it explains the variation of soil suction with changes in water content. In this research, Gene Expression Programming (GEP) is employed as an artificial intelligence method for modelling of this curve. The principal advantage of the GEP approach is its ability to generate powerful predictive equations without any prior assumption on the possible form of the functional relationship. GEP can operate on large quantities of data in order to capture nonlinear and complex relationships between variables of the system. The selected inputs for modelling are the initial void ratio, initial gravimetric water content, logarithm of suction normalized with respect to atmospheric air pressure, clay content, and silt content. The model output is the gravimetric water content corresponding to the assigned input suction. Sensitivity and parametric analyses are conducted to verify the results. It is also shown that clay content is the most influential parameter in the soil–water characteristic curve. The results illustrate that the advantages of the proposed approach are highlighted.

Keywords– Unsaturated soil, soil–water characteristic curve, artificial intelligence, gene expression programming

1. INTRODUCTION

Soil-water characteristic curve has considerable importance in unsaturated soil behavior such as shear strength, volume change, diffusivity and absorption, as well as most soil properties such as specific heat, permeability and thermal conductivity which can be also related to the soil-water characteristic curve [1]. This curve can be depicted as a continuous function describing the water storage capacity of a soil as it is subjected to various suctions. The SWCC contains significant information with respect to the amount of water contained in the pores at various suctions of soil and the pore size distribution corresponding to the stress state in the soil [2].

SWCCs are affected by various factors such as the pore shape and pore size distribution, the particle size distribution, the specific surface area, the chemo-physical properties of the soil phases, the soil density and the temperature. The effects of pores and particles are studied in many researches [3-5].The effect of temperature on the SWCC is studied by relatively limited researchers [6-8]. Grant and Salehzadeh [8] incorporated the thermal effects into van Genuchten's equation [9] to obtain a temperature-dependent SWCC equation. Recently, Zhou et al. [10] presented an approach to modelling the effect of temperature on the SWCC of deformable soils.

SWCC can be determined directly or indirectly in the laboratory through different tests. Direct methods include pressure plate, Buchner funnel, tensiometers, and pressure membranes. These methods measure the pore-water pressure in the soil or impose a known air pressure to soil and allow water content to come to equilibrium with the imposed air pressure. Among these methods, conventional pressure plate

*Received by the editors January 25, 2014; Accepted May 7, 2014.
**Corresponding author

based on American Society for Testing and Materials (ASTM) D698-70 [11]is the most common method. Indirect methods include filter paper and heat dissipation sensors. These methods use measurements or indicators of water content or a physical property that is sensitive to changes in water content. However, these mentioned experiments are usually costly and time consuming. Therefore, several empirical methods have been proposed in the literature to cope with it.

The methods for predicting the SWCC of a particular soil can be classified into five groups as follows:

1. Fitting type equations for the SWCC. In this group of equations a simple mathematical equation is fitted to the experimental data, and the unknown parameters are determined [9, 12–14].
2. Water contents at different suctions are correlated to specific soil properties such as D_{10} (sieve size for 10% passing) and porosity. This process generally requires a regression analysis followed by a curve fitting procedure [15, 16].
3. Correlating parameters of an analytical equation with basic soil properties such as grain size distribution and dry density, using a regression analysis [17, 18].
4. Physico-empirical modelling of the SWCC. This approach converts the grain size distribution into a pore size distribution, which is in turn related to a distribution of water content and associated pore pressure [19–22].
5. Artificial Intelligence (AI) methods such as neural network, gene expression programming and other machine learning methods have been used in various disciplines of civil engineering [23–25]. Prediction of the SWCC, using artificial intelligence falls into the fifth group [26–28].

Among these categories second and third groups have great similarity but they differ in the parameters that are correlated with soil properties. Furthermore, in this categorization artificial intelligence methods are known to have a better accuracy and a relatively straight forward approach [26, 28].

There has been considerable work in the literature to come up with a suitable closed form relationship for the SWCC [e.g. 2, 9]. However, the approaches employed so far make certain assumptions in order to arrive at the desired equation.

The main objective of this paper is to employ a powerful approach called GEP, a branch of artificial intelligence method, to propose a suitable relationship for the SWCC. The main advantage of the GEP approaches over the regression and other soft computing techniques is their ability to generate predictive equations without assuming prior form of the existing relationship. In this study, soil water retention parameters such as initial void ratio, initial gravimetric water content, logarithm of suction normalized with respect to atmospheric air pressure, clay content, and silt content are considered as independent variables.

2. ARTIFICIAL INTELLIGENCE METHODS FOR DETERMINING SWCC

There are several artificial intelligence methods such as Artificial Neural Network (ANN), Genetic-Based Neural Network (GBNN), Genetic Programming (GP) and Evolutionary Polynomial Regression (EPR) which are employed to model the SWCC. These models are presented briefly, below:

a) ANN model

Neural Network (NN) is a computer-based modelling technique for computation and knowledge representation inspired by the neural architecture and operation of the human brain. NNs have experienced a considerable resurgence of interest in recent years, though they were initially developed during the early 1940s.

ANN is constructed directly from experimental data. This is a fundamentally different approach to modelling of the material behavior, and because of their ability to learn and generalize interactions among many variables ANNs have the potential to model various aspects of material behavior. The basic architecture of ANN has been covered by Rumelhart and McClelland [29]. Each processing unit (neuron), acting as an idealized neuron in human brain, receives input from the units to which it is connected, computes an activation level, and transmits that activation to other processing units. A multi-layer perceptron NN has an input layer, an output layer, and a number of hidden layers connected to each other. Recently, researchers have employed this method in various majors of civil engineering such as predicting ground water level, streamflow or beam deflection [30-32]. However, in the case of the SWCC, a computer program coded using MATLAB was developed for training the network by Johari and Javadi [28]. The optimum neural network structure is shown in Fig. 1, and corresponding optimal connection weights of the model are presented in Table 1.

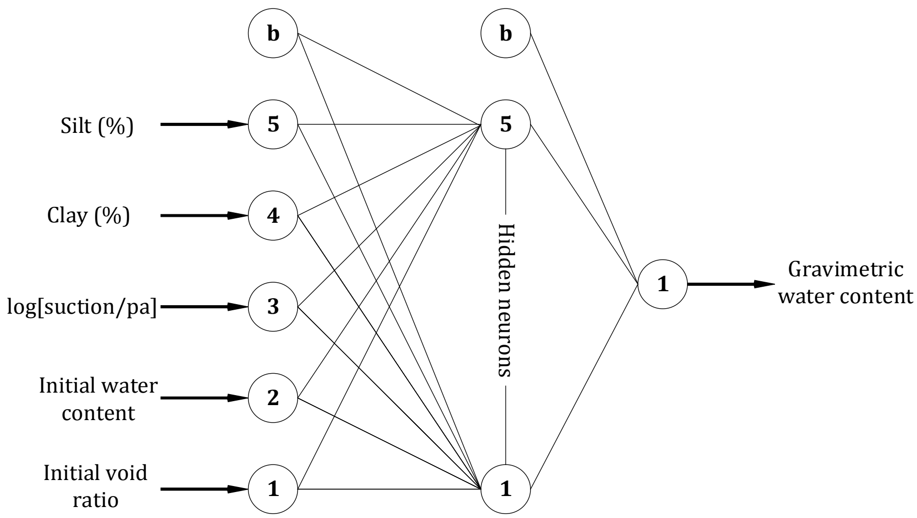

Fig. 1. The neural network model for the prediction of SWCC

Table 1. Optimal connection weights of ANN [28]

Hidden Neuron	Initial Void Ratio	Initial Water Content	$[\log(U_a\text{-}U_w)/p_a]$	Clay(%)	Silt(%)	Input bias	Output Neuron
1	1.529	-0.692	0.968	-0.343	0.115	-0.382	-0.620
2	0.999	0.123	-0.528	-0.300	-0.078	-1.311	0.959
3	1.075	0.444	0.083	0.404	0.120	0.798	0.430
4	0.375	0.416	0.833	0.629	1.617	0.589	-0.045
5	-0.974	-1.155	-0.797	0.639	-0.793	-1.622	-0.083
bias	-	-	-	-	-	-	0.011

b) GBNN model

In genetic-based neural network model, based on the optimization features of the Genetic Algorithm (GA), it is used for determining the optimal weights of a NN for predicting SWCC. The optimum number of hidden neurons is determined by trial and error. The hidden layer of this NN consisted of five neurons with one output neuron yielding the gravimetric water content corresponding to the assigned input suction. The configuration of this model is the same as the ANN model [28]. Figure1 shows the proposed network configuration, and Table 2 indicates the optimal connection weights of the model proposed by Johari et al. [26].

In order to match the starting point of the model with the laboratory results, all the outputs of the model were corrected using equation 2.

Table 2. Optimal connection weights of GBNN [26]

Hidden neuron	Initial void ratio	Initial water content	[log (u_a -u_w)/pa]	Clay(%)	Silt(%)	Input bias	Output neuron
1	-12.92	17.83	19.87	7.58	17.05	-34.6	-10.75
2	-10.20	25.41	-28.80	26.54	0.15	6.80	1.07
3	8.67	10.06	-20.52	-3.74	-10.24	-8.29	7.78
4	-18.14	-0.13	-12.80	4.09	2.56	1.19	-2.72
5	4.71	-10.83	8.30	-2.54	-0.89	0.05	-4.79
Bias	-	-	-	-	-	-	-0.04

c) GP model

GP, a branch of GA [32], is a method for learning the most "fit" computer programs by means of artificial evolution. In other words, its behavior forms a metaphor of the processes of evolution in nature. GP, similar to GA, initializes a population that combines the random members known as chromosomes. Afterward, fitness of each chromosome is evaluated with respect to a target value. The principle of Darwinian natural selection is used to select and reproduce "fitter" programs.Johari et al. [27] proposed a model to estimate SWCC for soils using GP method as shown below:

$$\omega = 0.794\ \left(w + 0.215\right)\left\{\left[\left(0.116^{Su} \times Cl^{Si}\right)^{(e+0.234)} + \left(Cl^{0.368\ (Si/Cl)}\right) \times \left(Su^{e} - Su\right)\right] Cl\right\}^{Su^2} \tag{1}$$

This formula was then scaled based on initial water content to yield

$$\bar{\omega} = \omega \times \left(w / \omega_0\right) \tag{2}$$

Where:

e= Initial void ratio
w= Initial water content
Su= log (suction(kPa)/pa)where pa= Atmospheric pressure (taken as 100 kPa)
Cl= Clay content (%)
Si= Silt content (%)
ω= Predicted gravimetric water content
$\bar{\omega}$= Adjusted gravimetric water content
ω_0= Predicted initial water content (at 0.2 kPa)

d) EPR model

EPR is a data driven method based on evolutionary computing, aimed at searching polynomial structures representing a system. A general EPR expression may be presented as [22]:

$$y = \sum_{j=1}^{n} F\left(X, f(X), a_j\right) + a_0 \tag{3}$$

Where **y** is the estimated vector of output of the process; $\mathbf{a_j}$ are model parameters; **F** is a function constructed by the EPR process; **X** is the matrix of input variables; **f** is a function defined by the user; and **n** is the number of terms of the target expression. The general functional structure represented is constructed from elementary functions by EPR using a GA strategy. Giustolisi and Savic [34] and Javadi and Rezania [35] introduced the detailed explanation of this method.

Ahangar-Asr et al. [36] developed a model for predicting the SWCC using EPR as shown below:

$$\omega = \frac{1.48\times10^{-6}Su^3}{e^3 . Cl . Si} + \frac{1.8Su . Cl^3 \text{ -}1.79e . Su^2. Cl . Si}{w} - \frac{4.07\times10^{-3}Su. Si + 0.25e. Cl^2}{Cl} - 1.7\times10^3 w. Su$$
$$+2.25\times10^{-3}w^2 - 0.17e . w + 3.11e^2 - \frac{2.15e^2.w^3}{Cl. Si^2} + 0.10214 \qquad (4)$$

3. GENE EXPRESSION PROGRAMMING

GEP is a branch of artificial intelligence and recent extension to GP that develops computer programs of different sizes and shapes encoded in linear chromosomes of fixed length [37]. The main advantage of the GP-based approaches over the regression and other soft computing techniques is their ability to generate prediction equations without assuming prior form of the existing relationship. Ferriera [37] makes a comparison between the GEP technique and GP as shown below:

- In GEP Population individuals (chromosomes) are linear and fixed length, converted to non-linear with varying sizes and lengths (expression trees or computer programs) at a later stage but, Population individuals are non-linear, varying in length as well as shape (also known as 'parse trees')
- GEP always produces valid expressions but sometimes GP obtain invalid expressions.
- GEP has totally separated genomes and phenome but GP uses a single entity working as genome and phenome at the same time.
- GEP is well established beyond the replicator threshold in contrast to GP, which is not yet established.

There have been some scientific efforts directed at applying GEP to the civil engineering tasks [e.g. 38-44].GEP is an evolutionary algorithm for learning the most fit computer programs. It incorporates both the simple, linear chromosomes of fixed length similar to GA and the ramified structures of different sizes and shapes similar to the parse trees of GP [37, 45, 46].

These computer programs can take many forms: they can be conventional mathematical models, neural networks, decision trees, sophisticated nonlinear regression models, logistic regression models, nonlinear classifiers, complex polynomial structures, logic circuits and expressions, and so on. But irrespective of their complexity, all GEP programs are encoded in very simple linear structures – the chromosomes. These simple linear chromosomes are a breakthrough because, no matter what, they always encode valid computer programs. So we can mutate them and then select the best ones to reproduce and then create better programs and so on, endlessly. This is one of the prerequisites for having a system evolving efficiently, searching for better and better solutions for all kinds of problems.

4. APPLICATION OF GEP FOR MODELLING OF SWCC

The GEP software, GeneXproTools 4.0 [47] was used in this study to perform symbolic regression using GEP to find a formulation for soil-water characteristic curve. GeneXproTools is a predictive analytics suite developed by Gep soft. GeneXproTools modeling frameworks include logistic regression, classification, regression, time series prediction, and logic synthesis. It is the unique software that employs the GEP method in function finding.

A large number of generations were needed to find a formula with minimum error. The formulation selection was based on simplicity and its relevance to the nature of the problem, thus ensuring a simple and efficient final GEP model. To set the model parameters a performance analysis was done. In the GEP, values of setting parameters have significant influence on the fitness of the output model. These include the number of chromosomes, number of genes, gene's head size, and the rate of genetic operators. This approach involved using different settings and conducting runs in steps. During each step, runs were carried out and the values of one of the above mentioned parameters were varied, whereas the values of the other parameters were kept constant. At the end of each run, the Mean Sum Squared Errors (MSSE) for both training and testing sets was recorded in order to identify the values that give the least MSSE.

The linking function must be chosen as ‘‘addition’’or ‘‘multiplication’’ for algebraic sub trees [47].However, in this research all linking functions were analysed. Figure2 shows the effect of the linking function on the performance of the GEP model. It can be seen that the GEP model performs best when linking function is addition.

Table 3. Input parameters used for the GEP models

Parameter	Achieved functions, values and rates
Fitness function	MSSE
Linking function	Addition (+)
Function set	+, -, ÷,×, Inv, X^2
Number of chromosomes	15
Number of genes	2
Gene head size	9
Recombination rate	0.2
Mutation rate	0.044

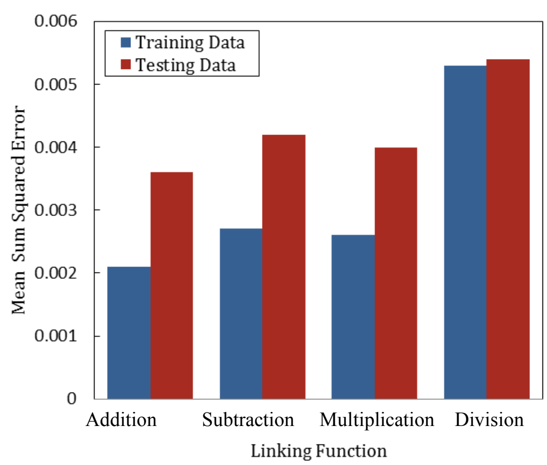

Fig. 2. Effect of linking function on the performance of the GEP model

5. DATABASE

Results from pressure plate tests performed on clay, silty clay, sandy loam, and loam soil reported by various researchers and compiled by SoilVision [49] were adopted for the analysis. Table 4 indicates the range of the properties of the soil used in this study. This database consists of the results from 186 pressure plate tests, together with their grain size distributions. Final suction values typically ranged from 800 to 1,700 kPa with few tests having suction values as large as 10^5kPa. Pressure plate test results were then digitized to obtain the necessary database. For digitization, an increasing incremental value of suction was adopted. Hence, the suction value was doubled in each increment. Initial suction value was fixed at 0.2 kPa. The database thus developed had a total of 2,694 patterns. From previous works on the topic, it is well understood that the SWCC is dependent on the grain size distribution, soil density, suction and water content of the soil. Hence, any of the SWCC parameters and/or any combination of these parameters may be considered as appropriate candidates for inputs of the model. Therefore, five parameters namely, void ratio, initial water content, logarithm of suction normalized with respect to atmospheric air pressure, clay content, and silt content, were selected as input. The output parameter was the gravimetric water content corresponding to the assigned input suction. For normalization each component of the data set was normalized to lie in an interval of [0, 1] using a max-min approach

Table 4. Range of Basic Soil Properties for Samples from SoilVision (2002) adopted for Developing GEP Model

Property	Range
Initialvoid ratio	0.458-2.846
Suction(kPa)	0.2-104857.6
Specific gravity	2.28-2.92
Water content (%)	0.18-98.27
Dry density (kg/m3)	702-1811
Initial water content (%)	17.34-105.41
Clay (%)	4.4-76.7
Silt (%)	10.3-87.5
Sand (%)	0.1-55.3

6. MODEL DEVELOPMENT

The optimum GEP program which is the optimum formulation for SWCC was obtained by developing the programs towards the formulation with minimum error compared with the actual experimental results. In this process, the input parameters and suitable required parameters were assigned and the sum of absolute differences between the predicted value of water content and actual value which had been obtained from experiment was monitored. Iterations continued until this error measure did not decrease considerably for training and testing data. Figure 3 indicates the variation of average relative absolute error during model development. The model training error descends from 1.77 in the incipient generations to about 0.33 after 100,000 generations and, in testing, the error declines from 2.82 to about 0.37 in the same generations. The average relative error is defined as:

$$\text{Average Relative Error (ARE)} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{A_i - P_i}{A_i}\right| \times 100 \tag{5}$$

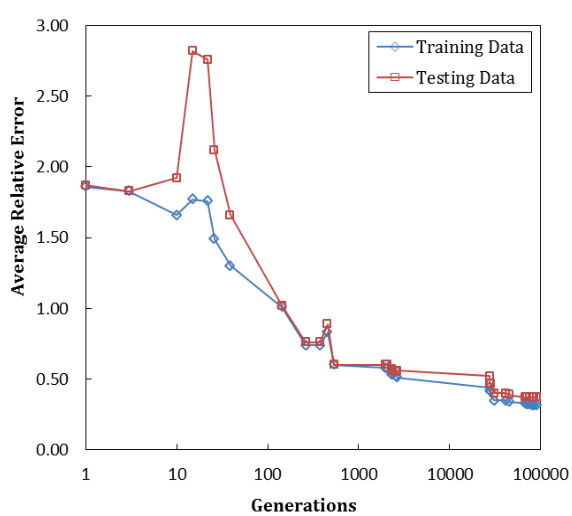

Fig. 3. Variation of error measured during training and testing generations

As mentioned earlier, one of the advantages of the GEP technique is that the relationship between the inputs of the model and the corresponding outputs is automatically constructed in the Expression Trees (ET). In this research the appropriate ET (ET1 and ET2) that are linked to each other to produce the final model are presented in Fig. 4.

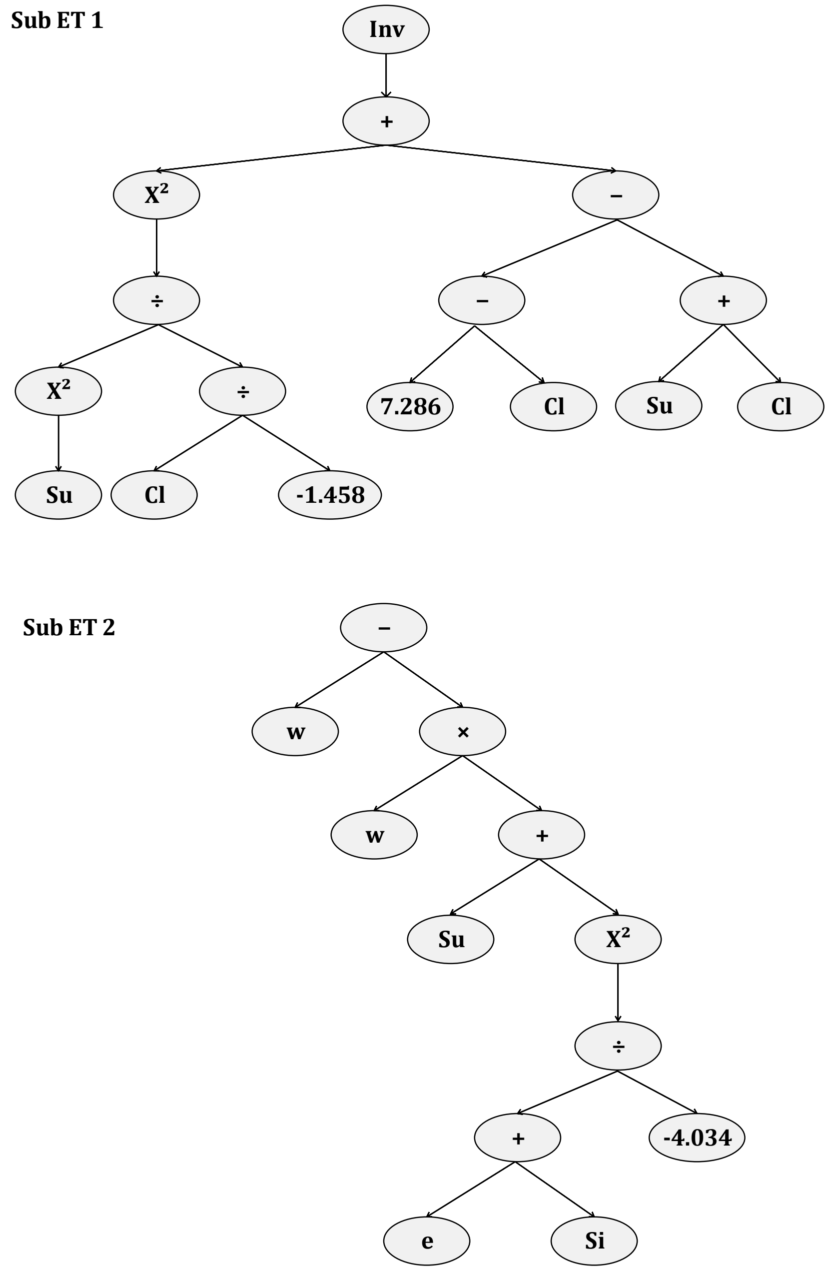

Fig. 4. Expression tree of the developed GEP model

The depicted trees in Fig. 4 can easily formulate into mathematical equation as in the following model:

$$\omega = \frac{-1}{Su + 2Cl - 2.202\dfrac{Su^4}{Cl^2} - 7.285} - W\left(Su + 0.062\left(Si + e\right)^2 - 1\right) \tag{6}$$

Where:

e = Initial void ratio;
w = Initial water content;
Su = log[suction(kPa)/p_a]where pa= Atmospheric pressure (taken as 100 kPa);
Cl = Clay content (%);
Si = Silt content(%);
ω =Predicted gravimetric water content

Afterward, this formula was scaled based on initial water content using equation 2.Eqs. (6) and (2) were used to simulate the SWCCs of all 131 pressure plate tests in the modelling set and all 55 tests in the validation set. The proposed model may be used to predict SWCC solely from basic soil parameters without resorting to a sophisticated experimental test. The procedure includes the following steps:

a) Choose some value of suctions, beginning from 0.2(kPa).
b) Normalize the input parameters(e, w, log [su/p_a], Cl and Si) using max-min approach.
c) Calculate gravimetric water contentat selected suctions using Eq. (6).
d) De-normalize the predicted gravimetric water contents.
e) Adjust the predicted water contents regarding Eq. (2).
f) Draw the SWCC using adjusted water content versus corresponding selected suctions.

Estimation of SWCC using the proposed model and following the mentioned procedure requires input parameters that may be determined using simple laboratory tests that may take only 1 or 2 days in contrast to the lengthy laboratory procedure (20–30 days) needed for SWCC determination.On the other hand, this approach in assessment of SWCC is utterly straightforward and there is no complex calculation, therefore, SWCC is promptly obtained through some simple mathematical calculations.

Figures 5 to 7 show the SWCC for 3 specimens used in developing the model. In these figures, the GEP prediction and actual experimental data are shown. From these figures it may be concluded that the formulation has a good potential for predicting the SWCC with reasonable accuracy. Similarly, Figs. 8 to 10present the prediction of the GEP for 3 typical tests not used in developing the model (testing). From these figures, it may also be concluded that the proposed method is also capable of simulating new test results, though it was not used in developing the corresponding model.

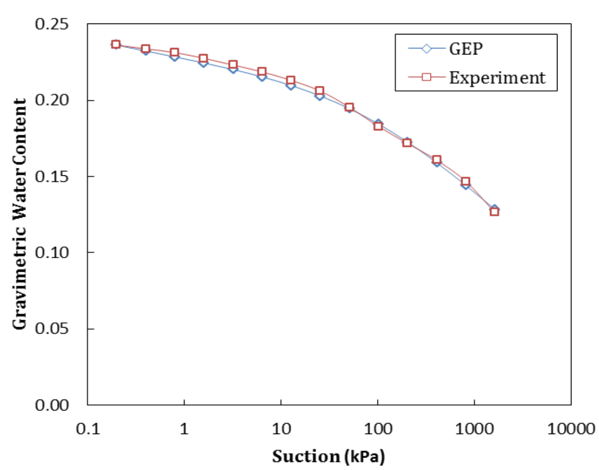

Fig. 5. Best simulation results among tests used for trainingGEP model.(MSSE= 5.2×10^{-6}) Void ratio: 0.69; Initial water content: 26.33%; Clay content 32.01%; Silt content: 66.01%

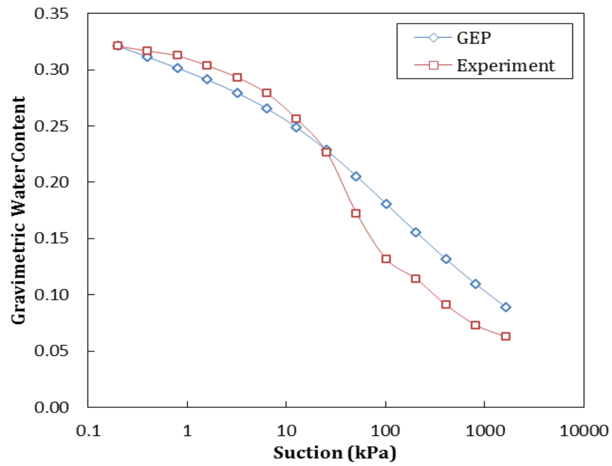

Fig. 6. Average simulation results among tests used for trainingGEP model. (MSSE= 6.9×10^{-4})
Void ratio: 0.98; Initial water content: 36.84%; Clay content 17.16%; Silt content: 79.95%

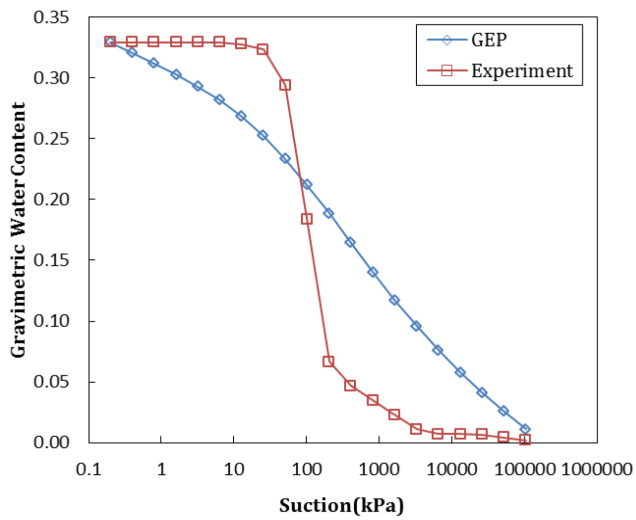

Fig. 7. Worst simulation results among tests used for trainingGEP model. (MSSE= 4.1×10^{-3})
Void ratio: 0.96; Initial water content: 32.90%; Clay content 21.86%; Silt content: 77.41%

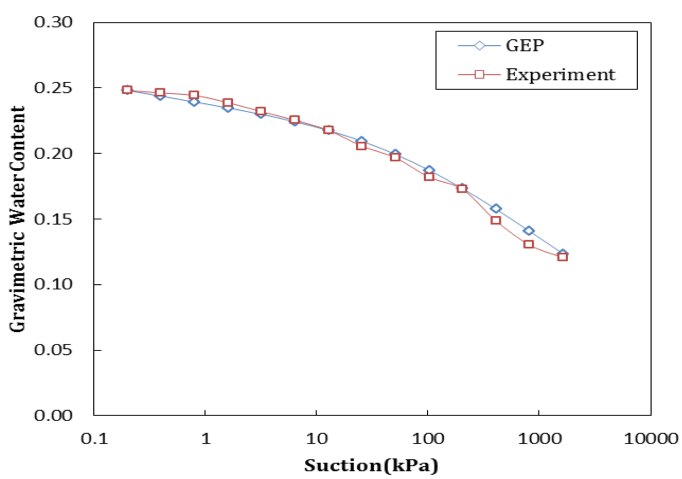

Fig. 8. Best simulation results among tests used for testing GEP model. (MSSE= 2.1×10^{-5})
Void ratio: 0.71; Initial water content: 26.89%; Clay content 28.71%; Silt content: 63.57%

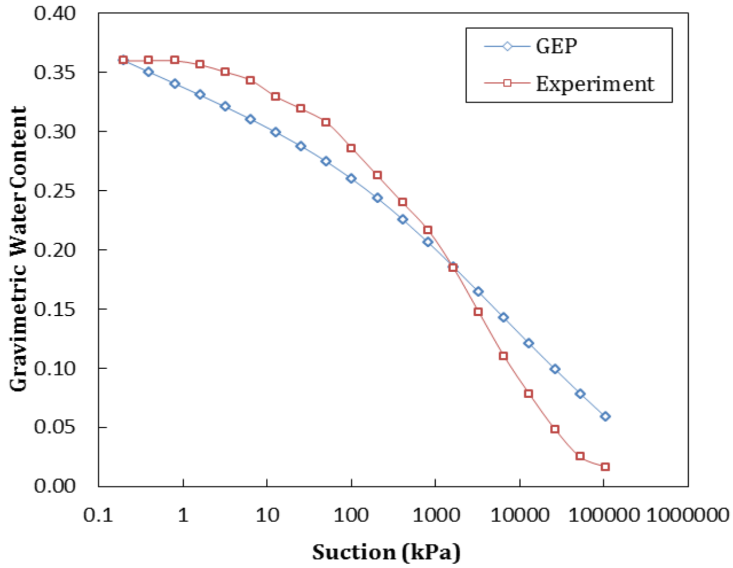

Fig. 9. Average simulation results among tests used for testing GEP model. (MSSE= 8.9×10^{-4})
Void ratio: 0.95; Initial water content: 35.99%; Clay content 37.68%; Silt content: 39.41%

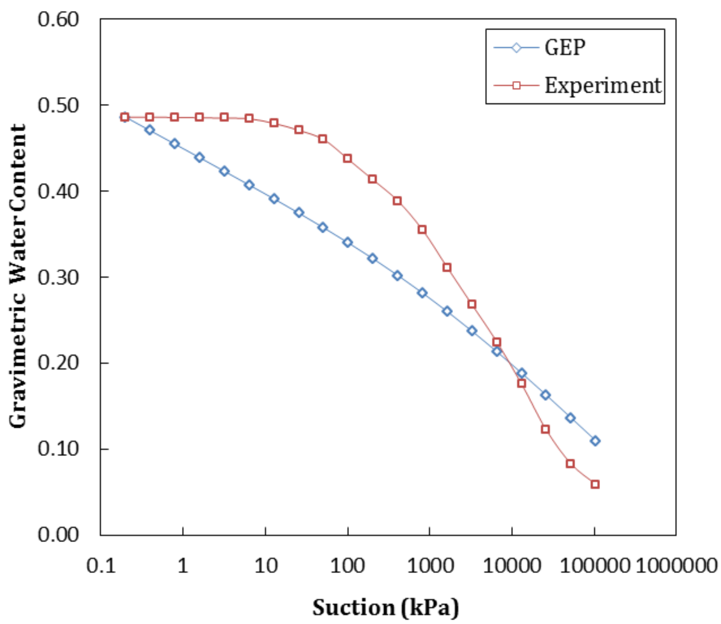

Fig. 10. Worst simulation results among tests used for testing GEP model.(MSSE= 4.1×10^{-3})
Void ratio: 1.37; Initial water content: 48.61%; Clay content 63.30%; Silt content: 29.54%

Several GEP models were developed using different arrangements of input variables. The performance of seven GEP models and effect of input parameter on accuracy of training and testing data-set are shown in Table 5.It can be seen that model 7 has a significantly superior performance.Therefore, the SWCC strongly depends on the whole selected input parameters clay content, silt content, water content, suction and void ratio.

Table 5. Mathematical expressions of some sample GEP models

No.	Mathematical Expression of the model	Number of inputs	R^2 Training	R^2 Testing
1	$\frac{w}{2.262\times e}-w\times(Su+0.832)$	3	0.82	0.79
2	$\frac{1}{(e+9.261)}-e\times Su\times(1+Si^2)+e$	3	0.90	0.85
3	$w-(-0.216\times Si\times(Cl-Su))+\frac{1}{(w\times Si-4.479)^3}-Su\times w$	4	0.92	0.89
5	$\frac{1}{(w^3+1.905)^3-e\times Su}+w+0.296\times Su\times(Cl-Su)$	4	0.93	0.91
6	$\frac{1}{\left((Su+Cl)\times\frac{Su}{Cl}\right)^2+7.146-Su-2Cl}-w\times\left(Su+\left(\frac{Si+e}{3.995}\right)^2-1\right)$	5	0.94	0.92
7	$\frac{-1}{Su+2Cl-2.202\frac{Su^4}{Cl^2}-7.285}-w\times\left(Su+0.062(Si+e)^2-1\right)$	5	0.94	0.94

7. VALIDATION ASSESSMENT OF THE PROPOSED MODEL

In this section, the experimental results of Johari et al. [50] tests are used to demonstrate the suitability of the proposed GEP model for more validation. In these tests Soil samples from 14 different locations in Shiraz city in Fars province of Iran were tested and their SWCCs were established, using a pressure plate apparatus.Details of the soil propertiesand experimental program are presented in [50].

Properties of the tested soils that are used in the GEP model as input parameters to predict SWCC are available in Table 6. Figures 11 to 13 show SWCC for 3 specimens used in developing the model. In these figures, GEP prediction and actual experimental data are shown. From these figures it may be also concluded that the formulation has good potential for predicting SWCC with reasonable accuracy.

Table 6. Properties of tested soils [50]

Series	Group	Initial void ratio	Initial water content	Clay (%)	Silt (%)
1	ML[a]	0.582	20.80	12.57	52.05
2	CL[a]	0.788	27.40	26.94	69.31
3	CL	0.818	26.73	34.65	60.90
4	ML	0.577	21.03	19.30	54.02
5	CL-ML[a]	0.791	28.94	33.90	64.38
6	ML	0.768	24.69	23.28	65.16
7	ML	0.675	25.19	16.81	77.78
8	ML	0.564	19.21	8.64	57.48
9	ML	0.669	22.74	21.92	54.77
10	CL	0.767	25.66	31.47	62.75
11	CL-ML	0.591	21.51	26.28	66.36
12	CL	0.623	22.91	27.40	60.01
13	CL	0.614	22.49	38.21	53.46
14	CL	0.715	21.29	13.86	71.15

[a] ML: Low plasticity silt, CL: Low plasticity clay, CL–ML: Low plasticity silty clay.

To illustrate the correlation quality of model, the experimental results and their corresponding predictions were plotted against each other in Fig. 14. It is to be noted that the models were calibrated using SoilVision [49] database, and their accuracy was evaluated using the experimental data from Shiraz data [50]. Total error of the GEP prediction is 2.92×10^{-4} and R^2 is equal to 0.92, these factors show good accuracy of proposed model in predicting SWCCs.

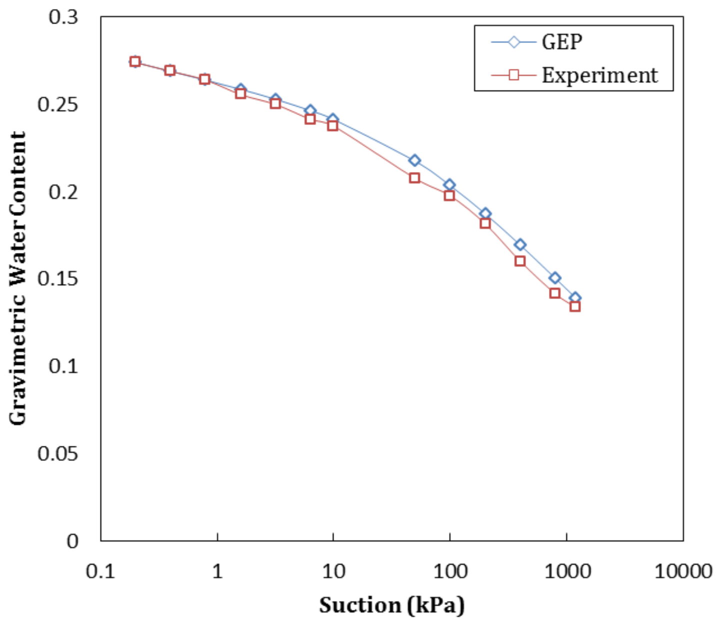

Fig. 11. Best simulation results among tests [50] used for validation GEP model. (MSSE= 2.8×10^{-5})
Void ratio: 0.79; Initial water content: 27.40%; Clay content 26.94%; Silt content: 69.31%

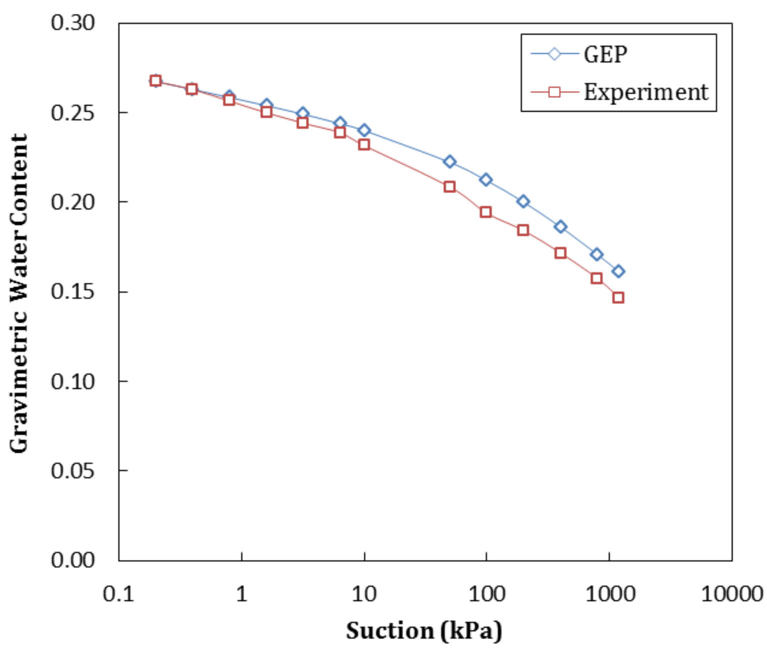

Fig. 12. Average simulation results among tests [50] used for validation GEP model. (MSSE= 1.1×10^{-4})
Void ratio: 0.82; Initial water content: 26.73%; Clay content 34.65%; Silt content: 60.90%

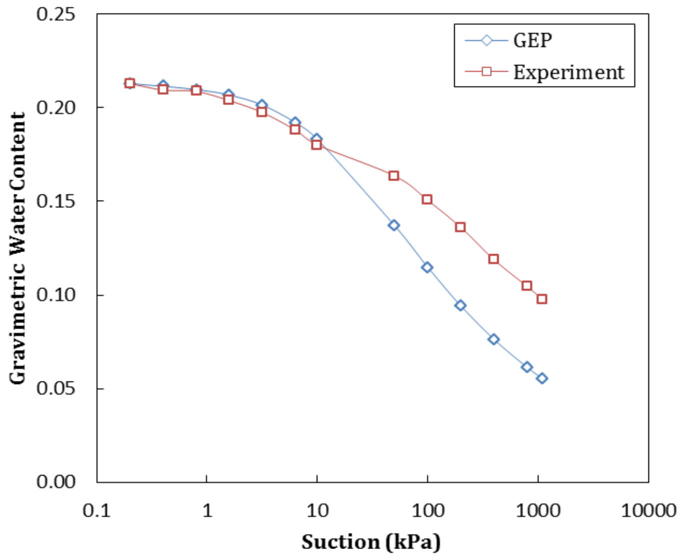

Fig. 13. Worst simulation results among tests [50] used for validation GEP model. (MSSE= 2.9×10^{-4})
Void ratio: 0.72; Initial water content: 21.29%; Clay content 13.86%; Silt content: 71.15%

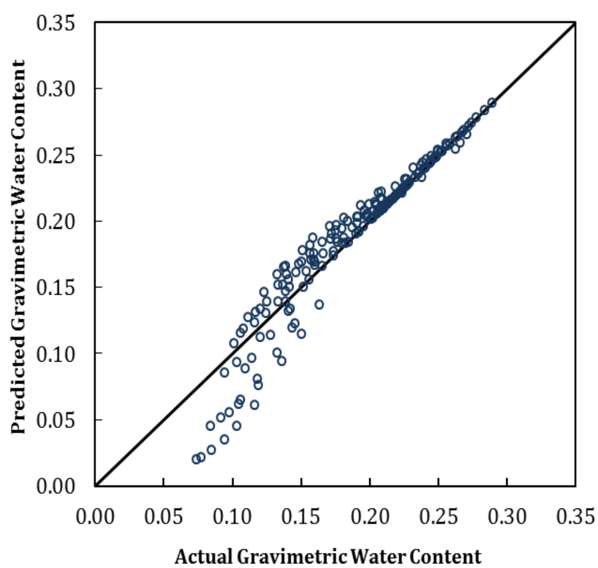

Fig. 14. Actual versus predicted gravimetric water content using GEP model (R^2= 0.92)

8. COMPARISION WITH PREVIOUS MODEL

As mentioned, a number of models have been presented by various investigators for estimating SWCC. Among these models in other groups, the approach presented by Fredlund et al. [51] was considered to give a more reasonable estimate of the SWCC approach [2, 21]. Figures 15 and 16 compare predicted gravimetric water content by GEP and the Fredlund et al. model[50]respectively, versus actual data for training data. Similarly, Figs. 17 and 18 compare predicted gravimetric water content by GEP and the Fredlund et al. model [51] respectively, versus actual data for testing data. These figures show a good correlation between the predictions made using GEP formulation and the actual data both for training and testing data.

Furthermore, Table7 presents the error in GEP prediction compared with the aforementioned approach. In this table, ARE defined by Eq. (5) and the Mean Sum Squared of the Error (MSSE) is defined by:

$$\mathrm{MSSE} = \frac{1}{N}\sum_{i=1}^{N}\left(A_i - P_i\right)^2 \tag{7}$$

R^2= correlation coefficient (square of the Pearson product moment correlation coefficient) where Ai= actual value for i^{th} data; Pi= predicted output data for i^{th} data; and N= total number of data available in the database.

Table 7. Performance of GEP and Fredlund et al. model[50]

	Training Data			Testing Data		
Model	ARE (%)	MSSE	R^2	ARE (%)	MSSE	R^2
GEP	25.59	0.0014	0.94	28.73	0.0014	0.94
Fredlund et al.	34.73	0.0071	0.85	35.39	0.0047	0.89

Additionally, an analysis was performed to compare performance of described AI methods in section 2. In Table 8 the value of R^2 and MSSE of the methods for training, testing and validation data are shown. It is worth mentioning that, besides the reasonable accuracy, the simplicity of the model is another important key factor for developing the model. The proposed model has both of these properties together.

Table 8. Comparing performance of artificial intelligence methods

Model	Training		Testing		Validation	
	R^2	MSSE	R^2	MSSE	R^2	MSSE
GEP	0.94	0.0014	0.94	0.0014	0.92	0.0003
GP	0.94	0.0016	0.93	0.0015	0.95	0.0002
ANN	0.94	0.0013	0.91	0.0018	0.93	0.0002
GBNN	0.94	0.0014	0.91	0.0019	0.96	0.0001
EPR	0.98	0.0010	0.96	0.0012	0.94	0.0003

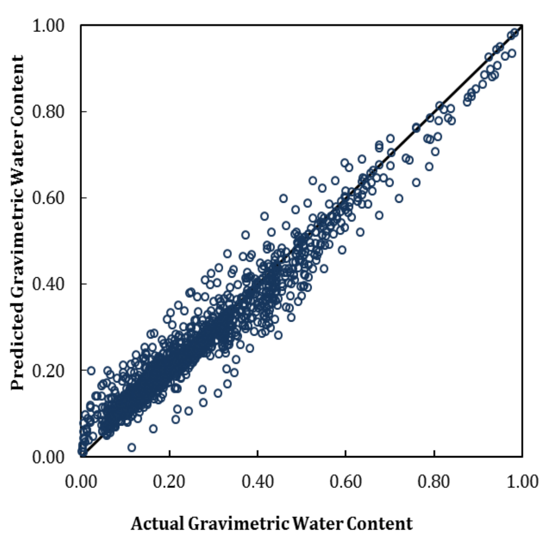

Fig. 15. Actual versus predicted gravimetric water content for training data in GEP approach (R^2= 0.94)

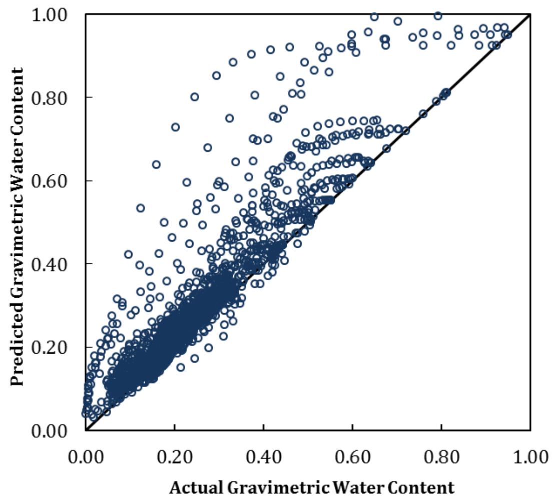

Fig. 16. Actual versus predicted gravimetric water content for training data in Fredlund et al. approach (R^2= 0.85)

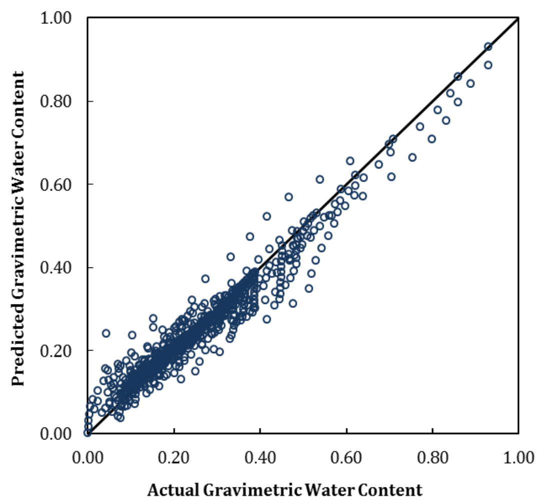

Fig. 17. Actual versus predicted gravimetric water content for testing data in GEP approach. (R^2= 0.94)

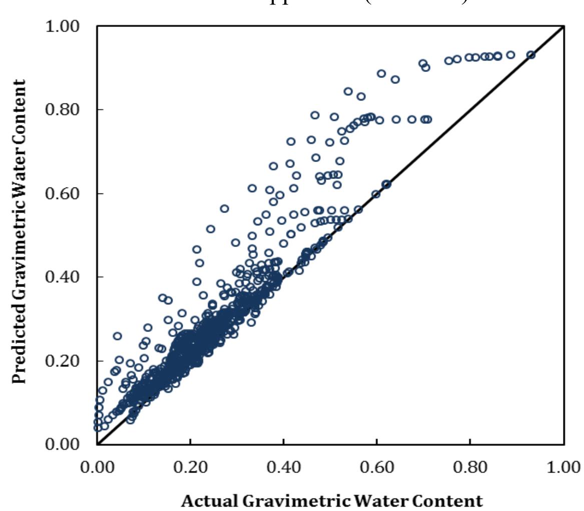

Fig. 18. Actual versus predicted gravimetric water content for testing data in Fredlund et al. approach. (R^2= 0.89)

9. SENSITIVITY ANALYSIS

To consider the model response to changes in input parameters, a sensitivity analysis was carried out. For this purpose, all input parameters, initial void ratio, initial water content, log [suction (kPa)/pa], clay and silt content were considered. In this procedure, the influence of each parameter on predicted gravimetric water content was assessed. In this analysis, the value of the input parameters of a random sample was increased approximately 30% while the ranges of the other input parameters were kept constant. The results are given in Table 9. In this table negative change means reduction and positive means increasing effect in gravimetric water content. It is shown that, with an increase in initial water content and clay content the Predicted gravimetric water content increases. On the other hand, Table 9 shows that with an increase in initial void ratio, log [suction (kPa)/pa] and silt content the gravimetric water content decreases. This table shows that the clay content is the most effective parameter in predicted gravimetric water content.

Table 9. The change in gravimetric water content corresponding to 30% increase in the value of the input parameters

Parameter	e	w	Su	Cl	Si
Change (%)	-0.60	0.50	-6.06	14.60	-2.05

10. PARAMETRIC ANALYSIS

For further verification of the proposed GEP model, a parametric analysis was performed in this research. The main goal was to find the effect of each parameter on the predicted gravimetric water content. Figures 19 to 23 present the predicted values of the gravimetric water content as a function of each parameter where the other parameters were constant. For this purpose, several arbitrary data sets from training and testing data-sets were considered for the parametric analysis. Response from typical datasets, given in Table 10, was selected to investigate influence of various parameters. Among these figures when suction, initial void ratio and silt content increase, the gravimetric water content decreased. However, during increasing clay content and initial water content, predicted water content increased.

Figure 19 shows that when initial void ratio is increased, predicted water content decreased. It means that the soil with higher density has higher water content. Figure 20 illustrates suction and predicted water content have reverse relation, therefore water content decreased with increasing suction. Figure 21 shows that in each soil sample, if soil has higher initial water content then the predicted water content in SWCC will be a higher value. Figure 22 illustrates that, as expected, increase in the clay content causes an increase in predicted water content significantly. Finally, assessment of model prediction indicated that influence of silt content is close to coarse grade particle, although it has a slight effect on predicted water content. This behavior is shown in Fig. 23.

Table 10. The arbitrary selected data from training and testing sets for parametric analysis

Parameter	e	w	Su(kPa)	Cl (%)	Si (%)
Training data	0.69	26.33	750.00	32.01	66.01
Testing data	0.71	25.46	750.00	41.08	13.46

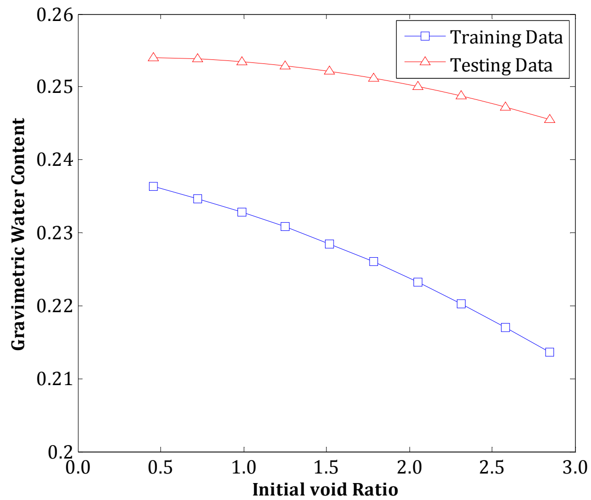

Fig. 19. Parametric analysis of the model with respect initial void ratio

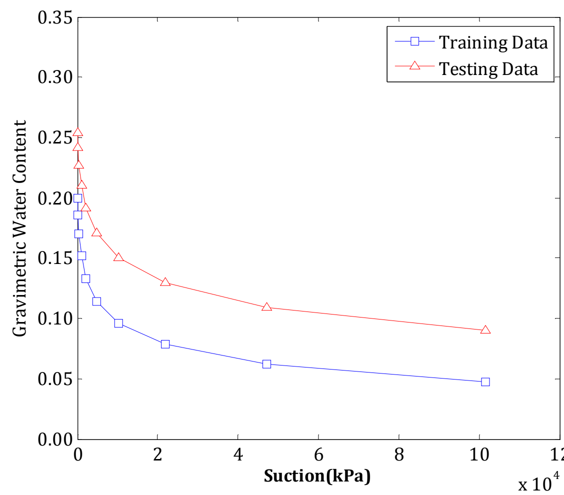

Fig. 20. Parametric analysis of the model with respect to suction

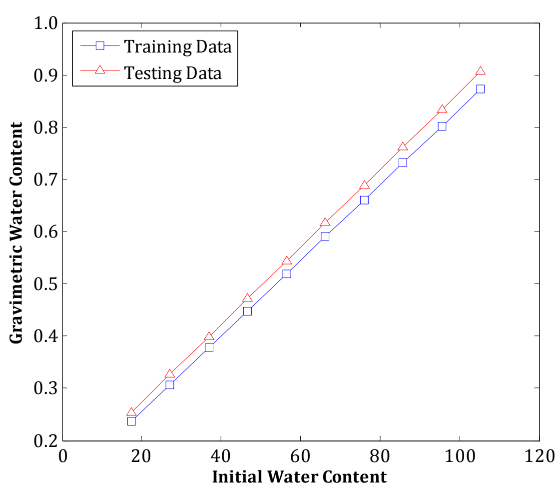

Fig. 21. Parametric analysis of the model with respect to initial water content

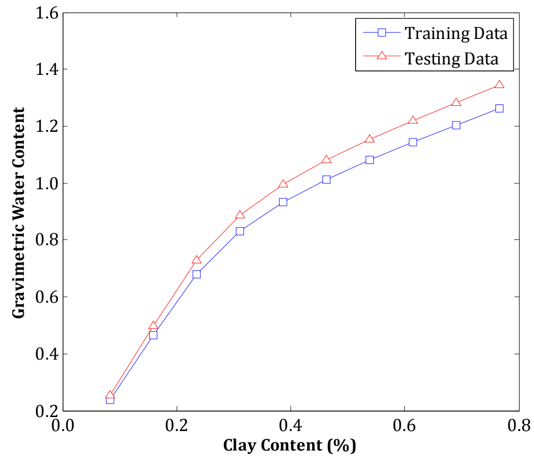

Fig. 22. Parametric analysis of the model with respect to clay content

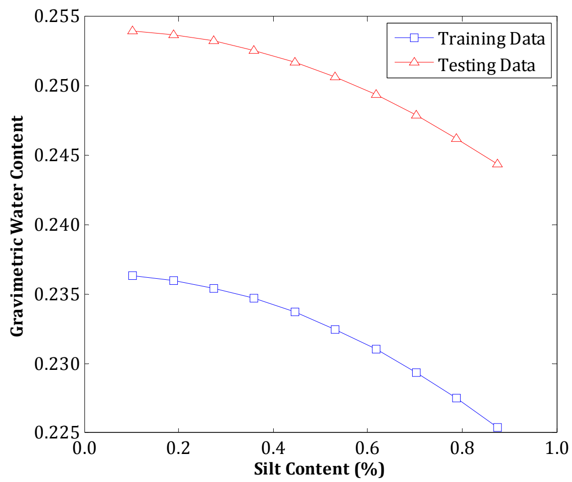

Fig. 23. Parametric analysis of the model with respect to silt content

11. CONCLUSION

A model based on GEP was proposed to estimate the SWCC for soils. A database containing the results of pressure plate tests carried out on a wide variety of fine grained soils was employed to develop the model. Test results were then digitized and normalized to obtain the necessary database. During the first phase, the model was developed using the results from 131 pressure plate tests. In the second phase, it was validated using 55 additional test results that the model had not been exposed to during the first phase. The model prediction indicated a reasonable accuracy, both for the results used in the first phase, as well as results in the validation phase. The model prediction had some discrepancies compared to the actual test data; however, comparison of the results from the proposed model with conventional methods indicated its superior performance for prediction of SWCCs. Furthermore, sensitivity analysis showed that the clay content is the most effective parameter in predicted gravimetric water content of model.

These models have certain limitations in that they do not take into account the hysteresis phenomena and soil fabric effects. The authors suggest the following future works for further improvements and extension on the topic:

-Studying other types of AI systems such as neuro-fuzzy networks and radial basis function.

-Extending the AI systems to include hysteresis phenomena, soil fabric and stress state effects.

-Extending the AI systems to include hysteresis phenomena and soil fabric.
-Reliability assessment of the developed model

REFERENCES

1. Fredlund, D. G. & Rahardjo, H. (1993). *Soil mechanics for unsaturated soils*. Wiley-IntersciencePublication.
2. Fredlund, M. D., Wilson, G. W. & Fredlund, D. G. (2002). Use of grain-size distribution for estimation of the soil water characteristic curve. *Canadian Geotechnical Journal*, Vol. 39, No. 5, pp. 1103–1117.
3. Arya, L. M. & Paris, J. F. (1981). A physicoempirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data. *Soil Science Society of America Journal*, Vol. 45, No. 6, pp. 1023–30.
4. Burger, C. A. & Shackelford, C. D. (2001). Soil–water charateristic curves and dual porosity of sand diatomaceous earth mixtures. *Journal of Geotechnical and Geoenvironmental Engineering*, Vol. 127, No. 9, pp. 790–800.
5. Chateau, X. & Viet, T. B. (2010). *Influence of the temperature on the water content curves: a micromechanical approach*. In: Buzzi, O., Fityus, S. and Sheng, D. editors. Unsaturated soils – theoretical and numerical advances in unsaturated soil mechanics. Newcastle (Australia): CRC Press, pp. 849–54.
6. Aubertin, M., Mbonimpa, M., Bussiere, B. & Chapuis, R. P. (2003). A model to predict the water retention curve from basic geotechnical properties. *Canadian Geotechnical Journal*, Vol. 40, No. 6, pp. 1104–22.
7. Romero, E., Gens, A. & Lloret, A. (2001). Temperature effects on the hydraulic behaviour of an unsaturated clay. *Geotechnical &Geological Engineering*, Vol. 19, No. 3, pp. 311–32.
8. Grant, S. A. & Salehzadeh, A. (1996). Calculation of temperature effects on wetting coefficients of porous solids and their capillary pressure functions. *Water Resources Research*, Vol. 32, No. 2, pp. 261–70.
9. Van Genuchten, M. Th. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society* of America *Journal*, Vol. 44, pp. 892–898.
10. Zhou, A., Sheng, D. & Li, J. (2014). Modelling water retention and volume change behaviours of unsaturated soils in non-isothermal conditions. *Journal of Computers and Geotechnics*, Vol. 55, pp. 1–13.
11. American Society for Testing and Materials (ASTM), (2000). Test method for laboratory compaction characteristics of soil using standard effort, ASTM D698.
12. Brooks, R. H. & Corey, A. T. (1964). Hydraulic properties of porous media. Colorado State University, Fort Collins, CO., Hydrology Paper No. 3.
13. Pedroso, D. M. & Williams, D. J. (2010).A novel approach for modeling soil water characteristic curves with hysteresis. *Computers and Geotechnics*, Vol. 37, No. 3, pp. 374–380.
14. Johari, A. & Javadi, A. A. (2011). Prediction of soil water characteristics curves using SPSS.*Proceedings of the Fifth Asia-Pacific Conferences on Unsaturated Soils,* Vol. 1, pp. 345–350.
15. Hutson, J. L. & Cass, A. (1987). A retentively function for use in soil water simulation models.*Soil Science Journal*, Vol. 38, No. 1, pp. 105–113.
16. Aubertin, M., Ricard, J. F. & Chapuis, R. P. (1998). A predictive model for the water retention curve: application to tailings from hardrock mines. *Canadian Geotechnical Journal*, Vol. 35, No. 1, pp. 55–69.
17. Cresswell, H. P. & Paydar, Z. (1996). Water retention in Australian soil, description and prediction using parametric functions. *Australian Journal of Soil Research*, Vol. 34, No. 2, pp. 195–212.
18. Tomasella, J. & Hodnett, M. G. (1998). Estimating soil water retention characteristics from limited data in Brazilian Amazonia. *Soil Science Journal*, Vol. 163, No. 3, pp. 190–202.
19. Pereira, J. H. F. & Fredlund, D. G. (2000). Volume change behavior of collapsible compacted gneiss soil. *Journal* of *Geotechnical and Geoenvironmental Engineering*, Vol. 126, pp. 907–916.

20. Fredlund, M. D., Wilson, G. W. & Fredlund, D. G. (2002). Use of grain-size distribution for estimation of the soil water characteristic curve. *Canadian Geotechnical Journal*, Vol. 39, No. 5, pp. 1103–1117.
21. Zapata, C. E., Houston, W. N. & Walsh, K. D. (2003). Soil-water characteristic curve variability, In Advances in Unsaturated Geotechnics. *Geotechnic*. Special Pub., Vol. 99, pp. 84–124.
22. Fredlund, D. G. & Pham, H. Q. (2006).A volume–mass constitutive model for unsaturated soils in terms of two independent stress state variables. *4thInternationalconference on unsaturated soils*, ASCE, Vol. 1, Carefree, Arizona, pp. 105–134.
23. Cheng, C. T., Ou, C. P. & Chau, K. W. (2002). Combining a fuzzy optimal model with a genetic algorithm to solve multiobjective rainfall-runoff model calibration. *Journal of Hydrology*, Vol. 268, No. 1–4, pp. 72–86.
24. Xie, J. X., Cheng, C. T., Chau, K. W. & Pei, Y. Z. (2006). A hybrid adaptive time delay neural network model for multistepaheadprediction of sunspot activity. *International Journal of Environmental Pollution*, Vol. 28, No. 3–4, pp. 364–381.
25. Muttil, N. & Chau, K. W. (2006). Neural network and genetic programming for modelling coastal algal blooms. *International Journal of Environmental Pollution*, Vol. 28, No. 3–4, pp. 223–238.
26. Johari, A., Habibagahi, G. & Ghahramani, A. (2006). Prediction of soil-water characteristic curve using a genetic based neural network. *ScientiaIranica*, Vol. 13, No. 3, pp. 284–294.
27. Johari, A., Habibagahi, G. & Ghahramani, A. (2006). Prediction of soil-water characteristic curve using genetic programming. ASCE, *Journal of Geotechnical* and *Geoenvironmental Engineering*, Vol. 132, No. 5, pp. 661–665.
28. Johari, A. & Javadi, A. A. (2011). Prediction of a soil-water characteristic curve using neural network. *5thInternationalConference on Unsaturated Soils*, Barrcelona, Spain, Vol. 1, pp. 461–466.
29. Rumelhart, D. E. & McClelland, J. L. (1988). Explorations in parallel distributed processing. The MIT Press, Boston, Mass.
30. Wu, C. L., Chau, K. W. & Li, Y. S. (2009). Predicting monthly streamflow using data-driven models coupled with data-preprocessingtechniques. *Water Resources Research*, Vol. 45, W08432.
31. Taormina, R., Chau, K. W. & Sethi, R. (2012). Artificial Neural Network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Engineering Applications of Artificial Intelligence*. Vol. 25, No. 8, pp. 1670-1676.
32. Chaudhary, S., Pendharkar, U. & Nagpal, A. K. (2014). Neural Networks for deflections in continuous composite beams considering concrete cracking. *Iranian Journal of Science & Technology, Transactions of Civil Engineering*, Vol. 38, No. C1^{+}, pp. 205-221.
33. Holland, J. H. (1975). *Adaptation in natural and artificial system*. University of Michigan Press, Ann Arbor, Mich.
34. Giustolisi, O. & Savic, D. A. (2006).A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics,* Vol. 8, No. 3, pp. 207–222.
35. Javadi, A. A. & Rezania, M. (2009).Intelligent finite element method; an evolutionary approach to constitutive modeling. *Journal of Advanced Engineering*, Vol. 23, No. 4, pp. 442–451.
36. Ahangar-Asr, A., Johari, A. & Javadi, A. A. (2012). An evolutionary approach to modelling the soil–water characteristic curve in unsaturated soils. *Journal of Computers & Geosciences*, Vol. 43 , pp. 25–33.
37. Ferreira, C. (2001). Gene expression programming: A new adaptive algorithm for solving problems. *Complex Syst*., Vol. 13, pp. 87–129.
38. Mollahasani, A., Alavi, A. H. & Gandomi, A. H. (2011). Empirical modeling of plate load test moduli of soil via gene expression programming. *Computers and Geotechnics*, Vol. 38, pp. 281-286.
39. Allavi, A. H., Aminian, P., Gandomi, A. h. & Arabesmaeili, M. (2011). Genetic-based modeling of uplift capacity of suction caissons. *Expert Systems with Applications,* Vol. 38, pp. 12608-12618.

40. Calabar, A. F. & Cevik, A. (2011). Triaxialbehavior of sand–mica mixtures using genetic programming. *Expert Systems with Applications*, Vol. 38, pp. 10358–10367.
41. Alkroosh, I., & Nikraz, H. (2012). Predicting axial capacity of driven piles in cohesive soils using intelligent computing. *Engineering Applications of Artificial Intelligence*, Vol. 25, pp. 618-627.
42. Shiri, J., Sadraddini, A. A., Nazemi, A. H., Kisi, O., Landeras, G., FakheriFard, A. & Marti, P. (2014). Generalizability of gene expression programming-based approaches for estimating daily reference evapotranspiration in coastal stations of Iran. *Journal of Hydrology*, Vol. 508, pp.1–11.
43. Ozbek, A., Unsal, M. & Dikec, A. (2013). Estimating uniaxial compressive strength of rocks using genetic expression programming. Journal of Rock Mechanics and Geotechnical Engineering, Vol. 5, pp. 325–329.
44. Alkroosh, I. & Nikraz, H. (2014). Predicting pile dynamic capacity via application of an evolutionary algorithm. Journal of Soils and Foundations, http://dx.doi.org/10.1016/j.sandf.2014.02.013
45. Ferreira, C. (2006). Gene *expression programming: mathematical modeling by an artificial intelligence*. Second ed. Springere_Verlag, Germany.
46. Ferreira, C. (2010). What is GEP? FromGeneXprotools Tutorials-A Gepsoft Web Resource. http://www.gepsoft.com/tutorial002.htm.Retrieved 23rdAugust.
47. GEPSOFT. GeneXproTools. Version 4.0, (2006), Available online: http://www.gepsoft.com
48. Ferreira, C., (2001). Gene expression programming in problem solving. *6th Online World Conference on Soft Computing in Industrial Applications* (invited tutorial).
49. Soil Vision (2002). *Soil Vision System Ltd*. Sask, Saskachewan, Canada.
50. Johari, A., Habibagahi, G. & Ghahramani, A. (2011). Prediction of SWCC using artificial intelligent systems: A comparativestudy. *Scientia Iranica,* Vol. 18, No. 5, pp. 1002–1008.
51. Fredlund, M. D., Fredlund, D. G. & Wilson, G. W. (1997). Predictionof the soil-water characteristic curve from grain size distribution andvolume-mass properties. *Procedure, 3rdBrazilian Symp. On Unsaturated Soils*, Rio de Janeiro, Brazil, pp. 13–23.

APPENDIX

In this Appendix, an example was presented to illustrate the procedure for prediction of SWCCs, using GEP model. For this purpose, the laboratory results of a sample were employed with the following inputs value. Void ratio= 0.98; Initial water content= 36.84%; Clay content= 17.16%; Silt content= 79.95%, and the value of suctions was started from 0.2 (kPa) and multiplied by 2 until 1638.4 (kPa).

1. The input parameters of the model were normalized to lie in an interval of [0, 1], using a max–min approach. The results are shown in Table A. 1.

$$e = \frac{e_1 - e_{min}}{e_{max} - e_{min}} = \frac{0.976 - 0.458}{2.846 - 0.458} = 0.217$$

$$w = \frac{w_1 - w_{min}}{w_{max} - w_{min}} = \frac{36.84 - 17.34}{105.41 - 17.34} = 0.221$$

$$Cl = \frac{Cl_1 - Cl_{min}}{Cl_{max} - Cl_{min}} = \frac{17.16 - 4.4}{76.7 - 4.4} = 0.176$$

$$Si = \frac{Si_1 - Si_{min}}{Si_{max} - Si_{min}} = \frac{79.95 - 10.3}{87.5 - 10.3} = 0.901$$

The magnitude of e, w, Cl and Si were constant in sample while the magnitude of Su was changed. For each arbitrary suction point, for instance 25.6(kPa), the normalization was done as follows:

$$Su = \frac{\log\left(Su_1 / 100\right) - \log\left(Su_{min} / 100\right)}{\log\left(Su_{max} / 100\right) - \log\left(Su_{min} / 100\right)} = \frac{\log\left(25.6 / 100\right) - \log\left(0.2 / 100\right)}{\log\left(104857.6 / 100\right) - \log\left(0.2 / 100\right)} = 0.368$$

2. By placing the above normalized value into the Eq. (6), the result will be:

$$\omega_{(Normilized)} = 0.250$$

$$\omega_{(Denormalized)} = 0.250 \times \left(0.9827 - 0.0018\right) + 0.0018 = 0.247$$

3. Based on the results of steps 2 and 3, the de-normalized water content for suction of 25.6(kPa) is calculated using Eq.(2) as:

$$\bar{\omega} = 0.247 \times \frac{0.321}{0.344} = 0.230$$

Table A.1. GEP model predictions for selected sample

e	w	Cl	Si	Su	Experiment water content	Predicted water content (Normalized)	Predicted water content (De-normilized)	$\bar{\omega}$
0.217	0.221	0.176	0.901	0	0.321	0.348	0.344	0.321
0.217	0.221	0.176	0.901	0.052	0.316	0.338	0.333	0.311
0.217	0.221	0.176	0.901	0.105	0.312	0.327	0.323	0.301
0.217	0.221	0.176	0.901	0.157	0.303	0.316	0.312	0.291
0.217	0.221	0.176	0.901	0.210	0.293	0.303	0.299	0.280
0.217	0.221	0.176	0.901	0.263	0.278	0.289	0.285	0.266
0.217	0.221	0.176	0.901	0.315	0.256	0.271	0.268	0.250
0.217	0.221	0.176	0.901	0.368	0.226	0.250	0.247	0.230
0.217	0.221	0.176	0.901	0.421	0.172	0.225	0.223	0.208
0.217	0.221	0.176	0.901	0.473	0.131	0.199	0.197	0.184
0.217	0.221	0.176	0.901	0.526	0.114	0.172	0.171	0.159
0.217	0.221	0.176	0.901	0.578	0.091	0.146	0.145	0.135
0.217	0.221	0.176	0.901	0.631	0.073	0.121	0.121	0.113
0.217	0.221	0.176	0.901	0.684	0.062	0.099	0.099	0.092