ORIGINAL ARTICLE

Regularized versus non-regularized neural network model for prediction of saturated soil-water content on weathered granite soil formation

**Muhammad Mukhlisin · Ahmed El-Shafie ·
Mohd Raihan Taha**

Abstract Modeling unsaturated water flow in soil requires knowledge of the hydraulic properties of soil. However, correlation between soil hydraulic properties such as the relationship between saturated soil-water content θ_s and saturated soil hydraulic conductivity k_s as function of soil depth is in stochastic pattern. On the other hand, soil-water profile process is also believed to be highly non-linear, time varying, spatially distributed, and not easily described by simple models. In this study, the potential of implementing artificial neural network (ANN) model was proposed and investigated to map the soil-water profile in terms of k_s and θ_s with respect to the soil depth d. A regularized neural network (NN) model is developed to overcome the drawbacks of conventional prediction techniques. The use of regularized NN advantaged avoid over-fitting of training data, which was observed as a limitation of classical ANN models. Site experimental data sets on the hydraulic properties of weathered granite soils were collected. These data sets include the observed values of saturated and unsaturated hydraulic conductivities, saturated water contents, and retention curves. The proposed ANN model was examined utilizing 49 records of data collected from field experiments. The results showed that the regularized ANN model has the ability to detect and extract the stochastic behavior of saturated soil-water content with relatively high accuracy.

1 Introduction

The study of rainwater infiltration on forested hillslopes is a very important area of research in forest hydrology. Rainwater infiltration produces positive pore water pressures in soil layers and is closely related to the occurrence of slope failures (e.g., [24, 25]). Soil data serve as an important initialization parameter for hydro-ecological and climatological modeling of water and chemical movement, heat transfer, or land-use change. Most soil hydraulic properties are difficult to measure and therefore have to be estimated in most cases. In addition, knowledge of the moisture content and water movement in a soil layer is fundamental to the analysis of biological reactions, plant activity, material transports in forest ecosystems, and stream water chemistry [30, 37]. To analyze rainwater infiltration into soil, it is important to have an understanding of the hydraulic properties of the soil, in particular the relationship between volumetric water content θ_s and soil capillary pressure ψ and the relationship between unsaturated hydraulic conductivity K and ψ. These relationships are known as the water retention curve and the hydraulic conductivity function, respectively. Many models for water retention (e.g., [7, 14, 19, 33, 39]) have been developed, incorporating earlier models by [8] and [23], with the aim of deriving analytical expressions that can be used to predict the relative hydraulic conductivity of soil. The resultant water retention curve is considered one of the most fundamentally important hydraulic characteristics of soil [2].

M. Mukhlisin · A. El-Shafie (✉) · M. R. Taha
Department of Civil and Structural Engineering,
Universiti Kebangsaan, Selangor, Malaysia
e-mail: elshafie@vlsi.eng.ukm.my

M. Mukhlisin
Department of Civil Engineering,
Polytechnic Negeri Semarang, Semarang, Indonesia

The relationship between saturated soil-water content θ_s and saturated soil hydraulic conductivity K_s as function of soil depth is in stochastic pattern. However, soil-water profile process is believed to be highly non-linear, time varying, spatially distributed, and not easily described by simple models. Most present model techniques are based upon conventional linear or non-linear statistical models, such as regression analysis. Although these models have been utilized for many years to provide relatively good prediction accuracy, they have several limitations to detect and mimic such highly stochastic pattern of soil-water profile.

Soil-water content is believed to be highly non-linear, time varying, spatially distributed, and not easily described by simple models. Two major approaches for modeling the soil-water content prediction process have been explored in the literature. These are the conceptual (physical) models and the system-theoretic models. Conceptual models are designed to approximate within their structures (in some physically realistic manner) the general internal subprocesses and physical mechanisms, which govern the water profile in soil. These models usually incorporate simplified forms of physical laws and are generally non-linear, time invariant, and deterministic with parameters that are representative of complex behavior of the soil-water content characteristics. Until recently, for practical reasons (data availability, calibration problems, etc.), most conceptual soil-water profile prediction model assumed lumped representations of the parameters. While such models ignore the stochastic properties of the soil-water profile process, they attempt to incorporate realistic representations of the major non-linearities inherent in the soil-water profile parameter relationships. Conceptual soil-water profile models are generally reported to be reliable in predicting. However, the implementation and calibration of such model can typically encounter various difficulties including sophisticated mathematical tools, significant amounts of calibration, and some degree of experience with the model.

While conceptual models are of importance in the understanding of soil-water processes, there are many practical situations such as soil-water profile where the main concern is with making accurate predictions. In such a situation, it is preferable to develop and implement a simpler system-theoretic model instead of developing a conceptual model. In the system-theoretic approach, models based on differential equations (or difference equations in case of discrete-time systems) are used to identify a direct mapping between the inputs and outputs without detailed consideration of the internal structure of the physical processes. The linear time series models such as ARMAX (autoregressive moving average with exogenous inputs) models developed by Box and Jenkins [5] have been usually used in such situations because they are relatively easy to develop and implement. They have been determined to provide satisfactory predictions in many applications [6, 34, 40]. However, such models do not attempt to represent the non-linear dynamics inherent in the soil-water process and therefore may not always perform adequately.

Recently, significant progress in the fields of non-linear pattern recognition and system theory using artificial neural networks (ANN) has been performed (e.g., [1, 3, 29]). An ANN is a non-linear mathematical structure that is capable of representing arbitrarily complex non-linear processes that relate the inputs and outputs of any system.

Agyare et al. [1] analyzed the measurement of soil properties together with terrain attributes in artificial neural networks (ANNs) to estimate saturated hydraulic conductivity (K_s), for two pilot sites in the Volta basin of Ghana. It was observed that good data distribution, range, and amounts are prerequisites for good ANN estimation and therefore data preprocessing is important for ANN. With adequate and sensitive data, ANN can be used to estimate K_s, using soil properties such as sand, silt, and clay content, bulk density, and organic carbon. Although the inclusion of terrain parameters can improve the estimation of K_s using ANN, they cannot be relied on as the sole input parameters as they yield poor results for the scale considered in this study. The source of training data was found to significantly influence the topsoil K_s, but the subsoil was not sensitive to training data source.

Parasuraman et al. [29] investigated the applicability of neural networks in estimating K_s at field scales and compared the performance of the field-scale pedotransfer functions (PTFs) with the published neural networks program Rosetta, also compared the performance of two different ensemble methods, namely bagging and boosting in estimating K_s. Data sets from two distinct sites are considered in the study. The performances of the models were evaluated when only sand, silt, and clay content (SSC) was used as inputs and when SSC and bulk density ρ_b (SSC + ρ_b) were used as inputs. The result showed that for both data sets, the field-scale models performed better than Rosetta. The comparison of field-scale ANN models employing bagging and boosting algorithms indicates that the neural network model employing the boosting algorithm results in better generalization by reducing both the bias and variance of the neural network models. Although artificial neural networks (ANNs)-based PTFs have been successfully adopted in modeling soil hydraulic properties at larger scales (national, continental, and intercontinental), the utility of ANNs in modeling saturated soil-water content θ_s and saturated soil hydraulic conductivity K_s as function of soil depth has rarely been reported. In addition, in Rosetta model, there are many parameters that supposed to be considered for the model

input, which is not necessarily to be available in most of the applications.

On the other hand, the major objective of training an ANN for prediction is to generalize, i.e., to have the outputs of the network approximate target values given inputs that were not in the training set. However, prediction based on ANN learning algorithms is fundamentally difficult and faces problems. One of the major shortcomings is that the ANN model experienced over-fitting problem while training session and occurs when a neural network loses its generalization [12, 13, 22, 28, 31].

1.1 Problem statement

In this study, the potential of implementing artificial neural network (ANN) model was proposed and investigated to map the soil-water profile in terms of K_s and θ_s with respect to the soil depth d. On the other hand, obviously, AI provides a viable and effective approach for developing input–output prediction models in situations that do not require modeling of the whole and/or part of the internal parameter of the soil-water profile. Although those models have proved to be efficient, its convergence tends to be very slow and yields suboptimal solutions. This may not be suitable for adaptive accurate prediction purpose. In fact, the major objective of training an ANN for prediction is to generalize, i.e., to have the outputs of the network approximate target values given inputs that were not in the training set. However, prediction based on ANN learning algorithms is fundamentally difficult and faces problems. One of the major shortcomings is that the ANN model experienced over-fitting problem while training session and occurs when a neural network loses its generalization.

A number of data sets on the hydraulic properties of weathered granite soils were collected from some catchment area in Japan. These data sets include the observed values of saturated and unsaturated hydraulic conductivities, saturated water contents, and retention curves. The proposed ANN model was examined utilizing 49 records of data collected from field experiments.

1.2 Objective

The objective of this paper is to analyze the soil-water profile and to develop a robust prediction model of the soil moisture content at different depths utilizing ANN. While such a model is not intended to substitute physically based conceptual models, it can provide an accurate prediction for soil moisture content using only the hydraulic conductivity and the depth as available input and output time series data. The anticipated impact of this model is that it can predict the soil moisture content without the need to explicitly consider the internal geologic or hydraulic parameters.

2 Experimental work

Granite soils are known to be very sensitive to weathering and vulnerable to landslides. In Japan, many disasters have been occurred in granite soil areas following heavy rains, resulting in a total of more than 1,000 casualties over the last 62 years. In all of these cases, the major disasters resulting from these rainstorms were owing to landslides that occurred on weathered granite slopes [9]. On the other hand, rainwater infiltration into soil also has great effect in slope instability. Therefore, this study analyzes an accurate prediction for soil moisture content with respect to the hydraulic conductivity and the depth using ANN model.

Forty-nine data sets on the hydraulic properties of weathered granite soils were collected from published studies (i.e., [18, 20, 21, 27, 36]). These data were observed from 5 catchment areas in Japan (i.e., Aichi, Enzan, Fudo, Kiryu, and Rokkou catchment areas). Observations were divided into surface and subsurface soil layers and comprised two sample sets of 33 and 16 samples of soil that were taken from 5- to 25-cm depths and from 70- to 170-cm depths, respectively.

3 Methodology

3.1 Artificial neural network

Artificial neural network is densely interconnected processing units that utilize parallel computation algorithms. The basic advantage of ANN is that they can learn from representative examples without providing special programming modules to simulate special patterns in the data set [16]. This allows ANN to learn and adapt to a continuously changing environment. Therefore, ANN can be trained to perform a particular function by tuning the values of the weights (connections) between these elements. The training procedure of ANN is performed so that a particular input leads to a certain target output as shown in Fig. 1.

Multilayer ANN has been reported as a powerful modeling tool [16]. The input and output layers of any network have numbers of neurons equal to the numbers of the inputs and outputs of the system, respectively. The architecture of a multilayer feed-forward neural network can have many layers between the input and the output layers where a layer represents a set of parallel processing units (or nodes), namely the hidden layer.
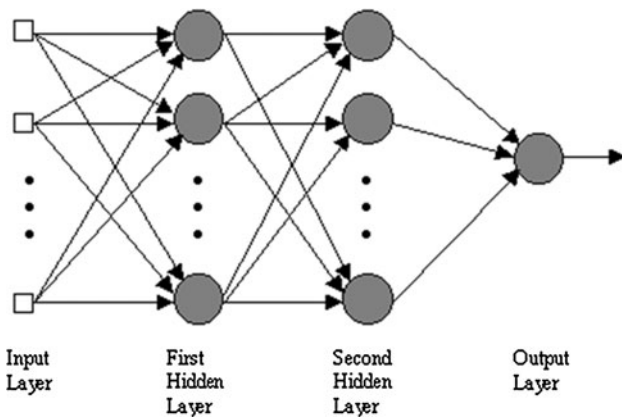
Fig. 1 Artificial neural network model diagram

The main function of the hidden layer is to allow the network to detect and capture the relevant patterns in the data and to perform complex non-linear mapping between the input and the output variables. The sole role of the input layer of nodes is to relate the external inputs to the neurons of the hidden layer. Hence, the number of input nodes corresponds to the number of input variables. The outputs of the hidden layer are passed to the last (or output) layer, which provides the final output of the network. Finding a parsimonious model for accurate prediction is particularly critical since there is no formal method for determining the appropriate number of hidden nodes prior to training. Therefore, here we resort to a trial-and-error method commonly used for network design.

In the prediction context, multilayer feed-forward neural network training consists of providing input–output examples to the network and minimizing the objective function (i.e., error function) using either a first-order or a second-order optimization method. Training can be formulated as one of minimizing a function of the weight, the sum of the non-linear least squares between the observed and the predicted outputs defined by the following equation:

$$E = \frac{1}{2} \sum_{P=1}^{n} (Y_O - Y_P)^2 \tag{1}$$

where n is the number of patterns (observations), Yo represents the observed response (target output), and Yp the model response (predicted output). In the back-propagation training, minimization of the error function (see (1)) is attempted using the steepest descent method and computing the gradient of the error function by applying the chain rule on the hidden layers of the feed-forward neural network. Consider a typical multilayer feed-forward neural network whose hidden layer contains M neurons. The network is based on the following equations:

$$net_{PJ} = \sum_{I=1}^{N} W_{ji} x_{pi} + W_{jo} \tag{2}$$

$$g(net_{PJ}) = \frac{1}{1 + e^{-net_{PJ}}} \tag{3}$$

where net_{PJ} is the weighted inputs into the jth hidden unit, n is the total number of input nodes, W_{ji} is the weight from input unit i to the hidden unit j, x_{pi} is a value of the ith input for pattern P, W_{jo} is the threshold (or bias) for neuron j, and $g(net_{PJ})$ is the jth neuron's activation function assuming that g is a logistic function. Note that the input units do not perform operation on the information but simply pass it onto the hidden nodes. The output unit receives a net input of

$$net_{pk} = \sum_{J=1}^{M} W_{kj} \cdot g(net_{PJ}) + W_{ko} \tag{4.a}$$

$$y_{pk} = g(net_{pk}) \tag{4.b}$$

where M is the number of hidden units, W_{kj} represents the weight connecting the hidden node j to the output k, W_{ko} is the threshold value for neuron k, and y_{pk} is the kth predicted output. The ultimate goal of the network training is to find the set of weights W_{ji}, connecting input units i to the hidden units j and W_{kj}, connecting the hidden units j to output k, that minimize the objective function (1, [17]). Since (1) is not an explicit function of the weight in the hidden layer, the first partial derivatives of E in (1) are evaluated with respect to the weights using the chain rule, and the weights are moved in the steepest descent direction. This can be formulated mathematically as follows:

$$\Delta W_{kj} = -\eta \frac{\partial E}{\partial W_{kj}} \tag{5}$$

where η is the learning rate, which scales the step size. The usual approach in back-propagation training consists in choosing η according to the relation $0 < \eta < 1$. The learning rule is a procedure for modifying the weights and biases of the network. This procedure may also be referred to as the training algorithm. The learning rule is provided with a known input/output set of data, and an algorithm is then used to adjust the weights and biases of the network in order to move the network outputs closer to the targets. Therefore, modeling capabilities of networks trained using supervised learning algorithms are limited to the range of the input used in training the network.

3.2 ANN for modeling soil-water profile

Artificial neural network model have been used in a broad range of applications including patterns classification, identification, prediction, optimization, and control systems.

ANNs learn by using some examples, namely patterns. In other words, to train and test a neural network, input data and corresponding target values are necessary [17, 32]. In this study, 49 records for saturated hydraulic conductivity (K_s) and saturated soil-water content (SMC) with respect to soil depth (D), obtained from the experiment location, were used to train and test the model.

Our pilot investigation showed that soil moisture content (SMC) is based on hydraulic conductivity (K_S) at certain depth D. Therefore, in this study, ANN with its non-linear and stochastic modeling capabilities is utilized to develop a soil-water profile model. The soil moisture content SMC_{di} predicted at certain depth d_i with respect to saturated hydraulic conductivity HC_{di} can be expressed as follows:

$$SMC_d(i) = f(HC_d(i), d(i)) \tag{6}$$

The ANN model is established using the above equation. The architecture of the network consists of an input layer of two neurons ($HC_d(i), d(i)$), an output layer of one neuron (corresponding to $SMC_d(i)$), and a number of hidden layers of arbitrary number of neurons at each layer. In order to achieve the desirable prediction accuracy, ANN architectures were developed utilizing 40 records out of 49 experimental records in order to train the network. The performance and the reliability of the ANN model were examined using the rest 9 records.

In order to accelerate the training procedure and to achieve minimum mean square estimation error, the records of data were normalized [10, 11]. All data in input and output layers were normalized in the (−1, 1)–(0, 1). Different MLP-ANN architectures (while keeping two neurons in the input layer and only one neuron in the output layer) were used to examine the best performance. The choice of the number of hidden layers and the number of neurons in each layer is based on two performance indices. The first index is the root mean square (RMS) value of the prediction error, and the second index is the value of the maximum error. The exact ANN architecture used for predicting saturated soil moisture content $SMC_d(i)$ is presented in Fig. 2.

There are several training algorithms for feed-forward networks. All these algorithms use the gradient of the performance function to determine how to adjust the weights to minimize performance. The gradient is determined using a technique called back-propagation, which involves performing computations backward through the network. The basic back-propagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient), the direction in which the performance function decreases most rapidly. It turns out that, although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. In the conjugate gradient algorithms, a search
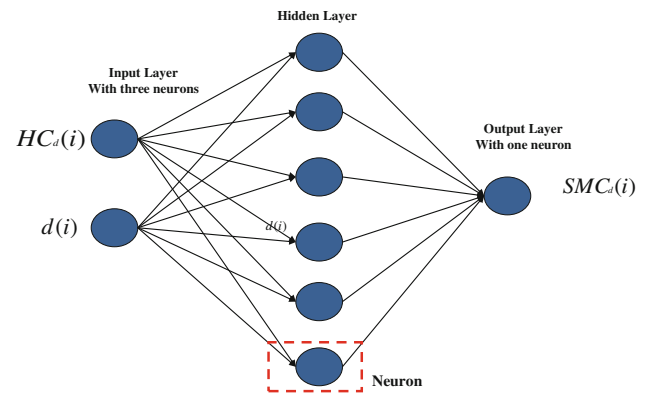

Fig. 2 The exact neural network architecture utilized for saturated soil-water content $SMC_d(i)$

is performed along conjugate directions, which produces generally faster convergence than steepest descent directions. Each of the conjugate gradient algorithms discussed so far requires a line search per iteration. This line search is computationally expensive, because it requires that the network response to all training inputs be computed several times for each search. The scaled conjugate gradient algorithm (SCG) was designed to avoid the time-consuming line search. This algorithm combines the model-trust region approach (used in the Levenberg–Marquardt algorithm), with the conjugate gradient approach [4]. The scaled conjugate gradient criterion was used to update the ANN parameters while training, since it was reported that this method is the most suitable in case of high randomness on the input data, which is the case in this study [10]. This criterion is based on the conjugate gradient method, but with small modification that significantly saves computational time [4].

Training of the network was performed by using Levenberg–Marquardt (LM) feed-forward back-propagation algorithms. A computer program was performed under Matlab 6.5 software. Hyperbolic tangent sigmoid function (HTSF) and purelin function (PF) were used as the transfer function in the hidden layer and output layer, respectively. These are shown by the (7) and (8), respectively.

$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{7}$$

$$y = ax + b \tag{8}$$

Finally, the Levenberg–Marquardt learning algorithm, which is a variant of back-propagation, was chosen as the learning algorithm of the model. The model was constructed, trained, and tested using different numbers of neurons (from 2 to 30) in their hidden layer using software developed by Matlab. For each network, RMSE value of the outputs was calculated.

3.3 Network over-fitting and regularization procedure

Network over-fitting is a classical machine learning problem that has been investigated by many researchers [35, 38]. Network over-fitting usually occurs when the network captures the internal local patterns of the training data set rather than recognizing the global patterns of the data sets. The knowledge rule base that is extracted from the training data set is therefore not general. As a consequence, it is important to recognize that the specification of the training samples is a critical factor in producing a neural network capable of making the correct responses. The problem of over-fitting has also been investigated by researchers with respect to network complexity [32].

Here, to avoid an over-fitting problem, we utilized the regularization technique [26]. This is known as a suitable technique when the scaled conjugate gradient descent method is adopted for training, as is the case in this study. The regularization technique involves modifying the performance function, which is normally chosen to be the sum of squares of the network errors on the training set defined as follows:

$$MSE = \frac{1}{2} \sum_{P=1}^{n} (Y_O - Y_P)^2 \qquad (9)$$

The modified performance function is defined by adding a term that consists of the mean of the sum of squares of the network weights and biases to the original mean square error (MSE) function as follows:

$$MSE_{reg} = \gamma \times MSE + (1 - \gamma) \times MSW \qquad (10)$$

where γ is the performance ratio that takes values between 0 and 1 and MSW is computed as follows:

$$MSW = \frac{1}{M} \sum_{j=1}^{M} w_j^2 \qquad (11)$$

where M is the number of weights utilized inside the network structure and w is weight matrix of the network. Using the performance function of (10), the neural networks to predict the saturated water content were developed with the intention to avoid over-fitting of data.

3.4 Regression model

Linear regression analysis has been used to establish a quantitative relationship between water quality parameters. Moreover, linear regression adopts to construct a simple formula that will predict what value will occur for a quantity of interest when other related variables take given values. Analyses were performed using the Statistica7 software. The general purpose of multiple regressions is to learn more about the relationship between several

independent or predictor variables and a dependent or criterion variable. A linear regression line has an equation of the following form:

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = x_i'\beta + \varepsilon_i \qquad (12)$$

where $'$ denotes the transpose, so that $x_i'\beta$ is the inner product between vectors x_i and β.

Often, these n equations are stacked together and written in vector form as follows:

$$y = x\beta + \varepsilon \qquad (13)$$

The most common method for fitting a regression line is the method of least squares. This method calculates the best-fitting line for the observed data by minimizing the sum of the squares of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0). Because the deviations are first squared and then summed, there are no cancellations between positive and negative values.

4 Results and discussions

The ANN model architecture of Fig. 2 is employed in this study to provide prediction of saturated soil moisture content $SMC_d(i)$. The measured saturated soil moisture content $SMC_d(i)$ over the first 40 records was used to train the ANN model. The proposed ANN model successfully achieved the target MSE of 10^{-4}. Figure 3 demonstrates the training curve and shows that the convergence to the target MSE is achieved after 73 iterations.

Various ANN architectures (while keeping two neurons in the input layer and only one neuron in the output layer) were used to examine the best performance. In fact, there is no
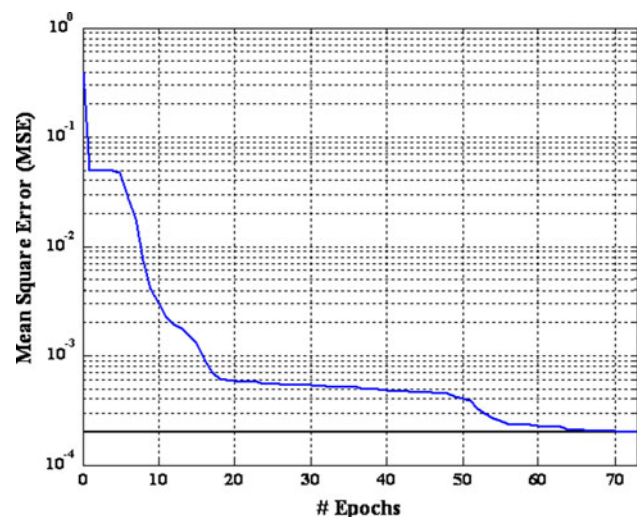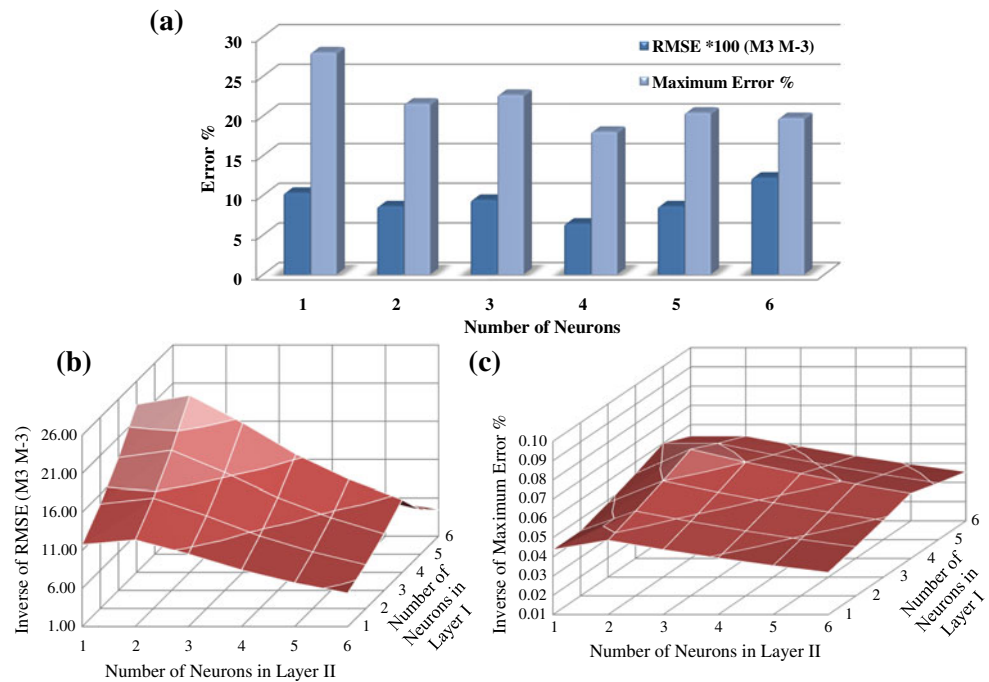

Fig. 3 Training curve for the proposed ANN model utilizing 40 records of the data

Fig. 4 Neural network performance "RMSE and maximum error" utilizing different architectures, **a** one hidden layer, **b** and **c** two hidden layers

formal and/or mathematical method for determining the appropriate "optimal set" set of the key parameters of neural network (number of hidden layers, number of neurons with each hidden layer, and the type of transfer function between two consequence layers). Therefore, it was decided to perform this task utilizing trial-and-error method. Several sets were examined with maximum 2 hidden layers and maximum 6 neurons within each layer. Therefore, the choice of the number of hidden layers and the number of neurons in each layer is based on two performance indices. The first index is the root mean square error (RMSE) value of the prediction error, and the second index is the value of the maximum error. Both indices were obtained while examining the ANN model with the training data. Actually, in developing such prediction model using neural network, the model could perform well during the training period and might provide higher level of error when evaluating during either validation or testing period. In this context, in this study the authors used these performance indices to make sure of that the proposed model could provide consistent level of accuracy for different input patterns. The advantages of utilizing these two statistical indices as a performance indicator of the proposed model are first to make sure that the highest error while evaluating the performance is within the acceptable error for such prediction model. Utilizing the RMSE is to ensure that the summation of the error distribution within the validation period is not high. Consequently, using both indices is guaranteed consistent level of errors, which is providing a great potential for having same level error while examining the model for unseen data in the testing period.

In order to show how the trial-and-error procedure for selecting the best parameter set of certain ANN architecture was performed, an example for month of January is presented in Fig. 4. For better visualization, the inverse value of both RMSE and maximum error was used as seen in Fig. 4b and c instead of the real values, while Fig. 4a shows the real value for both indices. Figure 4 shows the changes in the value of the RMSE and the maximum error versus the number of neurons when the number of hidden layers is one (Fig. 4a) and for two hidden layers in Fig. 4b (RMSE) and Fig. 4c for the maximum error during the validation period between 1930 and 1960. It is interesting to observe the large number of local minima that exist in both domains. The best combination of the proposed statistical indices for evaluating the model can be observed when the ANN architecture has 4 neurons in the first layer and 2 neurons in the second layer, achieving RMSE 0.045BCM and maximum error 15%.

The number of hidden layers (R) and the number of neurons in each layer (N) for the network are presented in Table 1. The transfer functions used in each layer of the

Table 1 The architecture of the neural network model

Layer	Number of neurons	Transfer function
Input layer	2	Tan sigmoid
Hidden layer I	4	
Hidden layer II	2	Pure line
Output layer	1	

networks are also listed in Table 1. The network utilizes the back-propagation algorithm during the training procedure. Once the network weights and biases are initialized, during the training process the weights and biases of the network are iteratively adjusted to minimize the network performance function mean square error MSE—the average squared error between the network outputs a and the target outputs t.

In order to examine the performance of the proposed ANN, a simulation for the model output during the training is performed. Figure 4 illustrates the performance of the model over the first 40 records used for training. The statistical comparisons between predicted and measured soil moisture content $SMC_d(i)$ were performed by estimating the prediction error (PE), which measures the average squared error between the predicted $SMC_d(i)$ obtained from the model and the measured $SMC_d(i)$. The PE is described in (14)

$$PE(\%) = 100 * \left(\frac{\left| SMC_{f(testing)} - SMC_m \right|}{SMC_m} \right) \qquad (14)$$

where $SMC_{f(testing)}$ is the predicted value and SMC_m is the experimentally measured value and m represents the number of samples in each testing group.

4.1 Non-regularized neural network

Prediction errors for the two ANN networks are presented in Fig. 5. It is obvious from Fig. 5 that $SMC_d(i)$ prediction models using ANN have a maximum error of 4% at only the experiments #1 and #8, while a maximum error $\sim 2\%$ at the rest of the 40 records. In addition, almost 0.0% error for 10 experiments can be observed, which is 25% of the whole examined records. As a result, the proposed ANN model successfully provides accurate predictions for $SMC_d(i)$ utilizing the saturated hydraulic conductivity (K_s) at different depth (D).
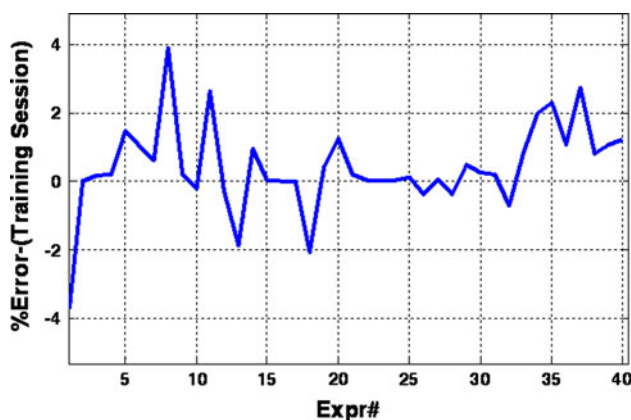

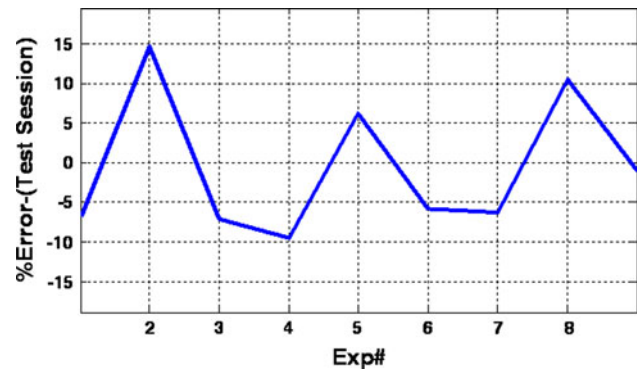Fig. 5 The error distribution for the ANN model during training session

Fig. 6 The error distribution for the ANN model during testing session
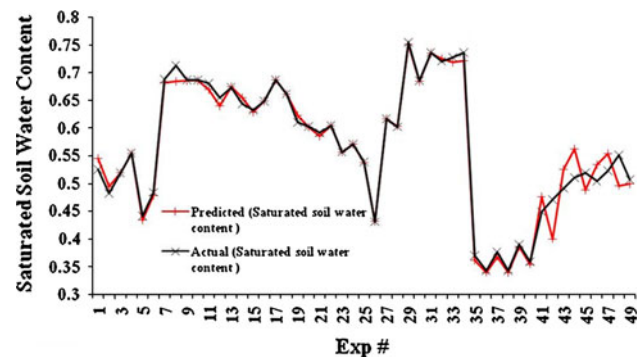

Fig. 7 Observed and predicted saturated soil-water content for the ANN model during training and testing session

To verify the performance of the proposed ANN-based model for saturated soil moisture content, the experiments between #41 and #49 was utilized. Figure 5 shows the error distribution value of the soil moisture content error over these 9 experiments. It can be observed that 6 out of 9 experiments experienced error lower than 5%. On the other hand, relatively higher levels of errors could be observed for experiments 42, 44, and 48, which is above 10% (Fig. 6).

Furthermore, Fig. 7 shows the neural network model output versus the actual saturated soil-water content. It can be observed from Fig. 7 that the proposed neural network model output could mimic the dynamic pattern in the soil-water content during training and testing.

4.2 Regularized neural network

The regularization technique described in Sect. 3.3 was applied to improve the generalization of the training and testing process of the proposed neural network model utilizing the same procedure (40 experiments for training and 9 experiments for testing). A trial-and-error procedure is applied to determine the best γ ratio that overcomes the over-fitting problems. Optimization techniques were not

necessary as value of γ easily converged by simplified trial-and-update procedures [15]. Different values of γ ranging between 0 and 1 are examined for each network. The analysis showed that γ ratio equals to 0.8 provided a considerable reduction in the error distributions of the proposed model.

Figure 8 demonstrates the performance of the regularized neural network model during testing. It can be depicted that the reduction in the distribution of the errors for those experiments experienced relatively poor prediction (Exp# 42, 44, and 48) while utilizing non-regularized neural network. It can be determined that the regularized network significantly improved the distribution of errors for all the experiments compared with non-regularized network.

Furthermore, by comparing the results showed in Figs. 8 and 6, similar level of accuracy for the regularized neural network model during training and testing sessions could be depicted. Such observation proves that the proposed regularized neural network model could provide consistent level of accuracy. In addition, Fig. 9 shows the observation versus the prediction values of the soil-water content. It could be depicted that there is a clear matching between the
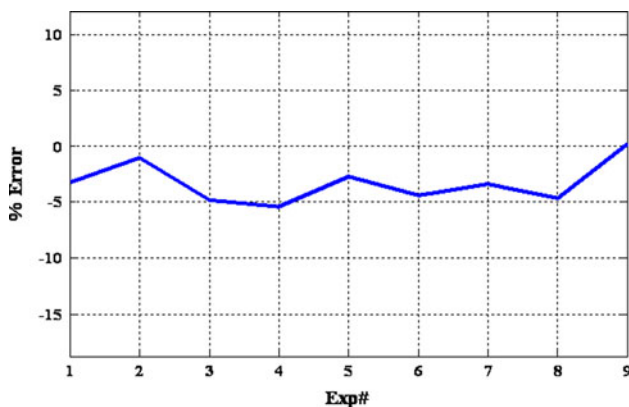

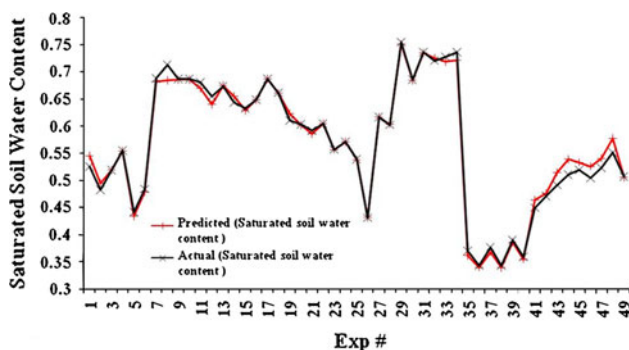**Fig. 8** The error distribution for the non-regularized ANN model during testing session

Fig. 9 Observed and predicted saturated soil-water content for the ANN model during training and testing session

Table 2 Improvement in PE associated with NN model during testing

Exp#	PE before regularization (%)	PE after regularization (%)	Reduction in prediction error (%)
1	−6	−3.3	45
2	15	−1	93
3	−7	−4.8	31
4	−10	−5.4	46
5	−6	−2.7	55
6	−6	−4.3	28
7	−6	−3.4	43
8	10	−4.7	53
9	1.5	0.26	83

observed and the proposed ANN model output, which confirms the ability of the ANN model to provide an adequate accuracy level for soil-water content values.
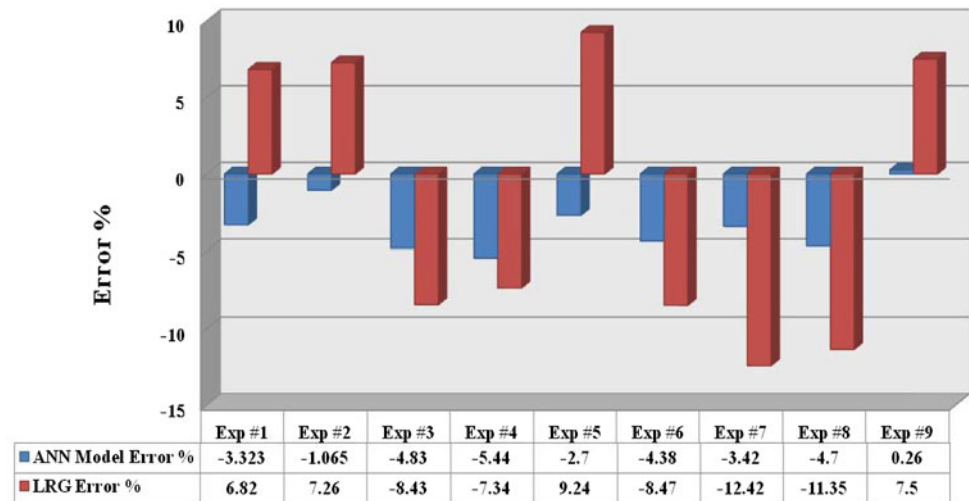
Table 2 shows the PE values of the errors at each experiment for both non-regularized and regularized networks. When compared with non-regularized networks, smaller values of PE errors can be depicted after eliminating the over-fitting problem. For example, comparing the PE value of *SMC* error at the Exp# 42, a reduction from 15% error to 1% error has been achieved. Similar improvement on the performance of almost all experiments can be observed. It can be observed that the results utilizing regularized neural network achieved better level of accuracy over the non-regularized neural network model.

For further assessment, a comparison analysis is carried out between the proposed ANN model and the linear regression model proposed in Sect. 3.3. The same procedure applied while performing the ANN model utilizing 40 records to calibrate the LRM and examined using the rest 9 records. Figure 10 shows the error distribution for those 9 records for both models. It could be observed that the proposed ANN model with the regularized procedure outperformed the LRM for all the records with remarkable improvement in prediction accuracy.

4.3 Recommendation for further research

In fact, it is common in ANN development to train several different networks with different architecture and to select the best one on the basis of performance of the networks with testing/validation sets. A major disadvantage of such an approach is that it assumes that performance of the networks for all other possible testing sets will usually be similar, which is statistically incorrect. Moreover, observing the performance of the developed ANN when tested, it is obvious that no single network has the optimal prediction for all the testing data sets.

	Exp #1	Exp #2	Exp #3	Exp #4	Exp #5	Exp #6	Exp #7	Exp #8	Exp #9
■ ANN Model Error %	-3.323	-1.065	-4.83	-5.44	-2.7	-4.38	-3.42	-4.7	0.26
■ LRG Error %	6.82	7.26	-8.43	-7.34	9.24	-8.47	-12.42	-11.35	7.5

Therefore, a better accuracy than the best reported by any single network can be accomplished if an optimization algorithm that can utilize all these networks is developed.

Another interesting observation is that the effect of the transfer function is as important as the number of layers and neurons in each layer. This can be observed when comparing the performance of two networks with similar number of hidden layers and neurons, but with different transfer functions. Further discussion on the effect of the optimal combination of different transfer function for specific applications is beyond the scope of this study.

5 Conclusion

This article suggests the use of a regularized NN model for developing a prediction model of the saturated soil-water content at different depths. The model was successful to provide an accurate prediction for saturated soil moisture content using only the saturated hydraulic conductivity and the depth as input variables. The proposed ANN model was examined utilizing 49 records of data collected from field experiments. The results showed that the ANN model has the ability to detect and extract the stochastic behavior of the saturated soil-water content with relatively high accuracy. The performance accuracy of the model can be expressed from the training session that a maximum error $\sim 2\%$ at the rest of the 40 records, while the maximum error for the testing session is within $\sim 5\%$ level of accuracy.

References

1. Agyare WA, Park SJ, Vlek PLG (2007) Artificial neural network estimation of saturated hydraulic conductivity. Vadose Zone J 6:423–431

2. Assouline S, Tessier D (1998) A conceptual of the soil water retention curve. Water Res Res 34(2):223–231

3. Baker L, Ellison D (2008) Optimisation of pedotransfer functions using an artificial neural network ensemble method. Geoderma 144:212–224

4. Bishop CM (1996) Neural networks for pattern recognition, 1st edn. Oxford University Press, Oxford

5. Box GH, Jenkins G (1970) Time series analysis: forecasting and control. Holden-Day, San Francisco

6. Bras RL, Rodriguez-Iturbe I (1985) Random functions and hydrology. Addison-Wesley Publishing Company, Reading

7. Brooks RH, Corey AT (1964) Hydraulic properties of porous media. Hydrol Pap 3. Civil Engineering Department, Colo State University, Fort Collins

8. Burdine NT (1953) Relative permeability calculation from size distribution data. Trans Am Inst Min Metall Pet Eng 198:71–78

9. Chigira M (2001) Micro-sheeting of granite and its relationship with landsliding specifically after the heavy rainstorm in June 1999, Hiroshima prefecture. Jpn Eng Geol 59:219–231

10. El-Shafie A, Noureldin AA, Taha MR, Basri H (2008) Neural network model for nile river inflow forecasting based on correlation analysis of historical inflow data. J Appl Sci 8(24):4487–4499

11. El-Shafie A, Abdin AE, Noureldin A, Taha MR (2009) Enhancing inflow forecasting model at aswan high dam utilizing radial basis neural network and upstream monitoring stations measurements. Water Resour Manage 23:2289–2315

12. El-Shafie A, Othman A, boelmagd Noureldin A, Hussien A (2010) Performance evaluation of a nonlinear error model for underwater range computation utilizing GPS sonobuoys. Neural Comp Appl 19(5):272–283

13. El-Shafie A, Abdelazim T, Noureldin A (2010) Neural network modelling of time-dependent creep deformations in masonry structures. Neural Comp Appl Springer 19(4):583–594

14. Fredlund DG, Xing A (1994) Equations for the soil-water characteristic curve. Can Geotech J 31(3):521–532

15. Gelb (1979) Applied optimal estimation. MIT Press, Cambridge

16. Gibson GJ, Cowan CFN (1990) On the decision regions of multilayer perceptrons. Proc IEEE 78(10):1590–1594

17. Haykin S (1999) Neural networks: a comprehensive foundation. Prentice Hall, New York

18. Hendrayanto, Kosugi K, Uchida T, Matsuda S, Mizuyama T (1999) Spatial variability of soil hydraulic properties in a forested hillslope. J Forest Res 4:107–114

19. Kosugi K (1994) Three-parameter lognormal distribution model for soil water retention. Water Resour Res 32:2697–2703

20. Kosugi K (1997) A new model to analyze water retention characteristics of forest soils based on soil pore radius distribution. J Forest Res 2:1–8

21. Kosugi K (1997) New diagrams to evaluate soil pore radius distribution and saturated hydraulic conductivity of forest soil. J Forest Res 2:95–101

22. Livingstone DJ, Manallack DT, Tetko IV (1997) Data modeling with neural networks: advantages and limitations. J Comp Aid Mol Design 11(2):135–142

23. Mualem Y (1976) A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resour Res 12(3):513–523

24. Mukhlisin M, Kosugi K, Satofuka Y, Mizuyama T (2006) Effects of soil porosity on slope stability and debris flow runout at a weathered granitic hillslope. Vadose Zone J 5:283–295

25. Mukhlisin M, Taha MR, Kosugi K (2008) Numerical analysis of effective soil porosity and soil thickness effects on slope stability at a hillslope of weathered granitic soil formation. Geosci J 12(4):401–410

26. Nordström T, Svensson B (1992) Using and designing massively parallel computers for artificial neural networks. J Parallel Distrib Comp 14(3):260–285

27. Ohta T, Tsukamoto Y, Hiruma M (1985) The behavior of rainwater on a forested hillslope. I. The properties of vertical infiltration and the influence of bedrock on it (in Japanese, with English summary.). J Jpn For Soc 67:311–321

28. Ooyen V, Nienhuis B (1992) Improving the convergence of the backpropagation algorithm. Neural Netw 5:465–471

29. Parasuraman K, Elshorbagy A, Si BC (2006) Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles. Soil Sci Soc Am J 70:1851–1859

30. Pinder GF, Jones JF (1969) Determination of the ground-water component of peak discharge from the chemistry of total runoff. Water Resour Res 5:438–445

31. Prechelt L (1998) Early stopping but when? Lecture notes in computer science. Neural Netw Tricks Trade 1524:55–69

32. Ripley BD (1996) Pattern recognition and neural networks. Cambridge University Press, New York

33. Russo D (1988) Determining soil hydraulic properties by parameter estimation: on the selection of a model for the hydraulic properties. Water Resour Res 24(3):453–459

34. Salas JD, Delleur JW, Yevjevich V, Lane WL (1980) Applied modeling hydrological time series. Water Resources Publications, Littleton

35. Schaffer C (1993) Overfitting avoidance as bias. Mach Learn 10(2):153–178

36. Shinomiya Y, Kobiyama M, Kubota J (1998) Influences of soil pore connection properties and soil pore distribution properties on the vertical variation of unsaturated hydraulic properties of forest slopes. J Jpn For Soc 80:105–111

37. Sklash MG, Farvolden RN (1979) The role of groundwater in storm runoff. J Hydrol 43:45–65

38. Stallard BR, Taylor JG (1999) Quantifying multivariate classification performance—the problem of overfitting. CD Proceedings. SPIE Annual Conference, Denver, Co

39. Van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 44:615–628

40. Wood EF (ed) (1980) Workshop on real time forecasting/control of water resource systems. Pergamon Press, New York