Earthfill dam seepage analysis using ensemble artificial intelligence based modeling

Elnaz Sharghi, Vahid Nourani and Nazanin Behfar

ABSTRACT

In this paper, an ensemble artificial intelligence (AI) based model is proposed for seepage modeling. For this purpose, firstly several AI models (i.e. Feed Forward Neural Network, Support Vector Regression and Adaptive Neural Fuzzy Inference System) were employed to model seepage through the Sattarkhan earthfill dam located in northwest Iran. Three different scenarios were considered where each scenario employs a specific input combination suitable for different real world conditions. Afterwards, an ensemble method as a post-processing approach was used to improve predicting performance of the water head through the dam and the results of the models were compared and evaluated. For this purpose, three methods of model ensemble (simple linear averaging, weighted linear averaging and non-linear neural ensemble) were employed and compared. The obtained results indicated that the model ensemble could lead to a promising improvement in seepage modeling. The results indicated that the ensembling method could increase the performance of AI modeling by up to 20% in the verification step.

Key words | artificial intelligence, black box models, earthfill dam seepage, ensemble, Sattarkhan earthfill dam

Elnaz Sharghi (corresponding author)
Vahid Nourani
Nazanin Behfar
Department of Water Resources Engineering,
  Faculty of Civil Engineering,
University of Tabriz,
Tabriz,
Iran
E-mail: *sharghi@tabrizu.ac.ir*

Vahid Nourani
Department of Civil Engineering,
Near East University,
P.O. Box 99138, Nicosia, North Cyprus, Mersin 10,
Turkey

INTRODUCTION

Monitoring of dam parameters by instrumentations is extremely important in the operation of dams. Such monitoring items include seepage, pore pressure, deformations, earthquake motion, temperature variations, etc. Physical explanation of considerable variations in the indicators is the most important part in the operation of the dams. A seepage problem is one of the most important challenges in the design, construction and operation of the earthfill dams and many earthfill dams are vulnerable to failure because of seepage problems in the dam core. Several studies have been performed to simulate seepage through earthfill dams by various methods (e.g. see Bardet & Tobita 2002; Nourani *et al.* 2014). There are different approaches to develop models for prediction of the non-linear structural behavior of the dam, which are usually classified into three main groups, including physical-based, black box and conceptual

models (Nourani 2017). Although physical-based and conceptual models are reliable tools for investigating the actual physic of the phenomenon, they have practical limitations, and when accurate predictions are more important than the physical understanding, employing black box models can be more useful. Auto Regressive Integrated Moving Average (ARIMA) is a classic black box tool for modeling aqua-related time series (Salas *et al.* 1990). These types of models, which are basically linear, lose their merit toward modeling hydraulic processes that are embedded with high complexity, dynamism and nonlinearity in both spatial and temporal scales. Recently, artificial intelligence (AI) approaches have been widely used for modeling hydraulic processes. AI models such as Feed Forward Neural Network (FFNN), Support Vector Regression (SVR), and Adaptive Neural Fuzzy Inference System

(ANFIS) are relatively new black box methods which have been used in various aspects of hydraulic engineering (e.g. ASCE 2000; Shahin *et al.* 2001). In the field of earthfill dam seepage modeling, Tayfur *et al.* (2005) employed Artificial Neural Networks (ANN) for predicting piezometric heads of an earthfill dam in Poland. They used upstream and downstream water levels as inputs of ANN and the outputs of ANN were compared with the results of the Finite Element Method. An ANN model was linked to the Radial Basis Function (RBF) interpolator by Nourani & Babakhani (2012) for spatio-temporal modeling of water heads in earthfill dams. Nourani *et al.* (2012) developed sole and integrated ANN modeling approaches to simulate the piezometric heads of an earthen dam so that in sole ANN modeling, a single ANN was developed for each piezometer, whereas in the integrated ANN modeling only a unique ANN was trained for all piezometers at various sections of the dam. For the neuro-fuzzy and SVR (based on the Support Vector Machine, SVM, concept) models, as other types of AI methods, only one study can be found in technical literature for each. Novaković *et al.* (2014) developed a neuro-fuzzy model which serves both ANN and fuzzy concepts within a unique framework to predict the water levels in piezometers of the Iron Gate 2 dam using the tail water levels as input data and the water levels in the piezometers as the targets; and finally Yongbiao (2012) employed an SVR model to analyse the seepage and piping in the dam body.

Although such black box models (e.g. ARIMA, ANN, ANFIS, and SVR) may lead to quite reliable results, it is a well-known issue that for solving a unique problem, using different models may lead to slightly different results. For example, in a time series prediction, one model may appropriately simulate the maximum values whereas the other may represent well the lower values. Therefore, by combining different models via an ensemble modeling framework as a post-processing method, different aspects of the underlying patterns may be captured more accurately. The concept of such a model combination has been already discussed and used in different engineering fields (e.g. Krogh & Vedelsby 1995; Shamseldin *et al.* 1997; Zhang 2003) but not, as far as the present authors are aware, in the context of seepage modeling in general and earthfill dam analysis in particular. The objective of the present paper is to apply such a concept to the earthfill dam seepage analysis. With

this regard first FFNN (a commonly used AI model), SVR (more recently used AI model), ANFIS (an AI model which can handle uncertainty of process via fuzzy concept) and ARIMA (a linear conventional method) models are used for modeling an earthfill dam (Sattarkhan dam) seepage via three scenarios with different input combinations. Then, an ensemble model is formed using the outputs of the mentioned models for each scenario to improve the modeling performance. In this way, three ensembling methods of simple linear averaging, weighted linear averaging and non-linear neural ensemble are used for each scenario.

MATERIALS AND METHODS

Sattarkhan earth dam and data

Sattarkhan earthfill dam is a reservoir dam that was constructed in Iran's East Azerbaijan Province on Ahar Chai River. The height of the dam is 78 m above the bed rock and its crest length is 340 m and volume of reservoir is 131.5 million m^3 when the water level is at normal level. Figure 1 shows a view of the dam. Different electrical piezometers have been installed in four sections of the dam. Water levels in the piezometers have been monitored every month for the period from 1999/4/20 to 2013/1/19; also daily water levels in the upstream of the reservoir have been recorded for the dam but the variations in water levels downstream of the reservoir are almost negligible.

Figure 1 │ Sattarkhan earthfill dam.

In this study data of piezometers numbers 207, 212, 216 and 217 at cross section number 2 at different levels were used for the modeling purpose (see Figure 2). Table 1 summarizes the statistics of the used data and Figure 3 presents the observed water levels in piezometers numbers 207 and 217 and the water levels of upstream of the dam for the considered period as examples. Due to the calibration and verif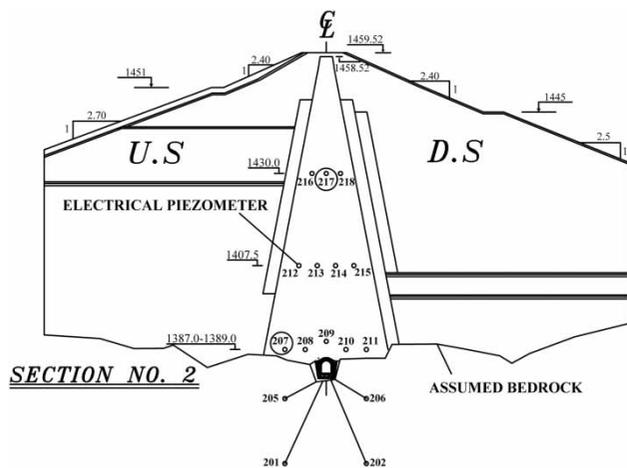ication goals, data set was divided into two parts. The first 75% of the data were employed for the training of the models and the remaining data were used for verification purposes.

Proposed methodology

In the proposed method in this study first FFNN, SVR, ANFIS and ARIMA models were separately created and trained, based on three scenarios. Then, an ensemble model was formed via outputs of the single models. Figure 4 shows the schematic of the proposed methodology. In the proposed ensembling approach for piezometric head time series prediction, linear ARIMA model was also used as well as AI models. The main idea of model ensembling is based on the following perspectives. Firstly, it is often difficult in practice to determine whether a time series under study is generated from a linear or non-linear underlying process or whether one particular method or model is better than others. Therefore, choosing the appropriate method for a unique issue is a difficult task for the predictors. By ensembling various models, the problem of choosing a suitable model can be handled. Although linear models sometimes could not provide accurate results based on their limitations to handle nonstationary and non-linearity, these models are still used because: (a) linear models are low-cost and simple, and the superposition principle can be applied in such linear models; and (b) the noise (or error) included in the used data (or employed computational scheme) increases linearly in a linear model, but such a noise (or error) may non-linearly be magnified over further time/space steps. Hence, it is recommended to apply a linear model for linear portions of a natural process.

Figure 2 │ Piezometers' positions of cross section number 2.

Table 1 │ Statistics of the observed heads in piezometers at section number 2

	Reservoir	Piezometer number			
		207	212	216	217
Maximum (m)	1,447	1,439	1,439	1,442	1,442
Minimum (m)	1,438	1,424	1,424	1,429	1,425
Average (m)	1,432	1,431	1,431	1,434	1,434
Standard deviation (m)	2	3	3	3	4

Figure 3 │ The observed heads in piezometers numbers 207 and 217 and the upstream water levels.
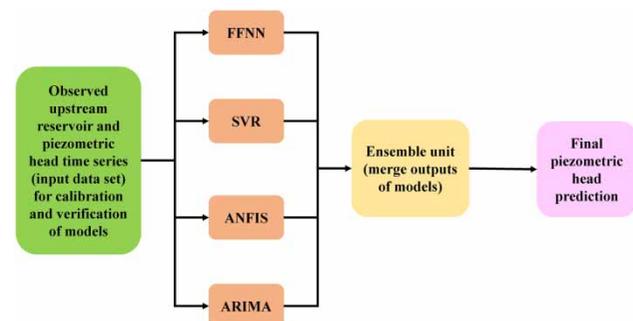

Figure 4 │ Schematic of the proposed methodology.

Secondly, the time series of the real world processes may include both linear and non-linear characteristics. If this is the case, then neither ARIMA nor AIs can be adequate in modeling and forecasting the time series since the ARIMA model cannot deal with non-linear relationships while an AI model may magnify the noise of a linear pattern. Hence, by combining ARIMA and AI models, complex auto-correlation structures in the data can be detected more accurately. Thirdly, it has been proved in most previous studies that no unique method is able to investigate the process perfectly (Zhang 2003). This is largely due to the fact that a real-world problem is often complex in nature and any unique model may not be able to capture different patterns of the process.

The aim of the proposed methodology in this study was to predict piezometric time series of an earthfill dam using upstream and related piezometer's data as inputs of the models. For this purpose, four different black box models were created and then, outputs of single models combined via three ensembling methods, based on three different input scenarios in which each scenario can be applied in a specific operation situation as discussed below.

Scenario 1

In scenario 1, it is attempted to predict the piezometric head of each piezometer using its previous time steps data, as well as upstream reservoir water levels. So the prediction of the ith piezometer's head could be patterned as:

$$P_t^i = f(P_{t-1}^i, P_{t-2}^i, \ldots, P_{t-n}^i, h_t, h_{t-1}, \ldots, h_{t-m}) \tag{1}$$

where the head of ith piezometer at time step t is considered as a function (f) of ith piezometric heads at previous time steps (t–1, t–2, …) up to lag time n and upstream level in time t and previous time steps (t–1, t–2, …) up to lag time m. The dominant lags of m and n could be determined through a trial and error procedure.

Scenario 2

In scenario 2, as another modeling scenario, two other adjacent piezometers' data and reservoir level time series are employed to predict the head of each piezometer in a way

that the head of the piezometer i is related to the reservoir level and two other piezometers as:

$$P_t^i = f(P_{t-1}^j, P_{t-2}^j, \ldots, P_{t-o}^j, P_{t-1}^k, P_{t-2}^k, \ldots, P_{t-r}^k, h_t, \\ h_{t-1}, \ldots, h_{t-m}) \tag{2}$$

where P_{t-o}^j and P_{t-r}^k are sub-series of the jth and kth piezo-meters up to lag times n and r, respectively. A correlation coefficient may be employed here for identification of proper piezometers. In this scenario just like scenario 1, the dominant lag times could be determined through trial and error procedure. Comparing Equations (1) and (2) shows that in scenario 2, for prediction of each piezometer's water head, the data of other piezometers which have reliable correlation with the target piezometer are used without the need for previous observed values of that piezo-meter. Therefore, this scenario can be a helpful approach when a piezometer is failed during the dam operation and the data of other neighboring piezometers can be used to predict that piezometer's water head.

Scenario 3

The third scenario is a combination of scenarios 1 and 2. In scenario 3, it is attempted to model the piezometric heads using each piezometer's data, two other piezometers data and also the upstream reservoir water level time series. Therefore, the general mathematical formulation of this scenario can be expressed as:

$$P_t^i = f(P_{t-1}^i, P_{t-2}^i, \ldots, P_{t-n}^i, P_{t-1}^j, P_{t-2}^j, \ldots, P_{t-o}^j, P_{t-1}^k, \\ P_{t-2}^k, \ldots, P_{t-r}^k, h_t, h_{t-1}, \ldots, h_{t-m}) \tag{3}$$

Dominant piezometers and lag times that are deter-mined in scenarios 1 and 2 can be used in this scenario. Although this scenario with a more complicated structure uses more input data, it is expected that it leads to more accurate outcomes with regard to scenarios 1 and 2.

Feed Forward Neural Network

It has been already shown that an FFNN as the most com-monly used ANN, with only three layers of input, hidden

and target layers, can be satisfactorily used at different fields of water resources engineering (ASCE 2000; Nourani 2017). The output value of a FFNN with three layers can be obtained through the following equation (Nourani & Parhizkar 2013):

$$\hat{y}_k = f_0 \left[\sum_{j=1}^{M_N} w_{kj} . f_h \left(\sum_{i=1}^{N_N} w_{ji} x_i + w_{j0} \right) + w_{k0} \right] \tag{4}$$

where W_{ji} is applied to a neuron in the hidden layer which connects the ith neuron in the input layer to the jth neuron in the hidden layer; W_{jo} is the applied bias to the jth neuron of the hidden layer; f_h denotes the activation function of the related hidden layer neuron; W_{kj} indicates the applied weight to a target neuron which connects the jth hidden neuron to the kth target neuron; W_{k0} is the applied bias to the kth target neuron; f_0 stands for the activation function of the target neuron; x_i is the ith input neuron, and $\hat{y}_k$, y are respectively the network output and observed values. N_N and M_N respectively show the number of input and hidden neurons. Hidden and target layers' weights are different from each other and should be estimated within the training phase.

Adaptive Neural Fuzzy Inference System

ANFIS as a neuro-fuzzy model combines the neural network and fuzzy logic concepts to enjoy the benefits of both within a unique framework (Farhoudi *et al.* 2010). Any fuzzy system contains three main parts, fuzzification, fuzzy data base and de-fuzzification, whereas the fuzzy data base part includes a fuzzy rule base and an inference engine. Among different fuzzy inference engines which can be used for fuzzy operation, the Sugeno engine (Jang *et al.* 1997) was employed in the current research. To show the typical mechanism of ANFIS to create a target function of f, for instance with two input vectors of x and y, the first order Sugeno inference engine may be applied to two fuzzy if–then rules as (Aqil *et al.* 2007):

Rule (1): if $\mu(x)$ is A_1 and $\mu(y)$ is B_1; then $f_1 = p_1x + q_1y + r_1$

(5)

Rule (2): if $\mu(x)$ is A_2 and $\mu(y)$ is B_2; then $f_2 = p_2x + q_2y + r_2$

(6)

in which A_1, A_2 and B_1, B_2 show respectively the MFs of inputs (i.e. x and y), p_1, q_1, r_1 and p_2, q_2, r_2 are the target function parameters. The operation of an ANFIS can be briefly described as follows.

Layer 1: Each node in this layer produces membership grades of an input variable. The output of the ith node in layer k is denoted as Q_i^k. Assuming a generalized bell function (gbellmf) as the membership function (MF), the output Q_i^1 can be computed as (Jang *et al.* 1997):

$$Q_i^1 = \mu_{A_i}(x) = \frac{1}{1 + ((x - c_i)/a_i)^{2b_i}} \tag{7}$$

where $\{a_i, b_i, c_i\}$ are adaptable premise variables.

Layer 2: The imposed signal to this layer is multiplied by each node of this layer as:

$$Q_i^2 = w_i = \mu_{A_i}(x) . \mu_{B_i}(y) \quad i = 1, 2 \tag{8}$$

Layer 3: Node i in this layer computes the normalized firing strength:

$$Q_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1, 2 \tag{9}$$

Layer 4: In this layer, the contribution of the ith rule towards the target is determined as (Jang *et al.* 1997):

$$Q_i^4 = \bar{w}_i(p_ix + q_iy + r_i) = \bar{w}_i f_i \tag{10}$$

where $\bar{w}$ is the output of layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set.

Layer 5: Finally, the output of the ANFIS model is computed by the sole node of this layer as (Jang *et al.* 1997):

$$Q_i^5 = \bar{w}_i(p_ix + q_iy + r_i) = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{11}$$

To calibrate the premise parameters set $\{a_i, b_i, c_i\}$ and consequent parameters set $\{p_i, q_i, r_i\}$ of the ANFIS,

the conjunction of least squared and gradient descent methods are used as a hybrid calibration algorithm (Aqil *et al.* 2007).

Support Vector Regression

SVR developed on the basis of the SVM concept is used for non-linear regression issues. Unlike many other black box forecasting methods, SVM based methods such as SVR consider operational risk as the objective function to be minimized instead of minimizing the error between the observed and computed values. In SVR, first a linear regression is fitted on the data and then the outputs go through a non-linear kernel to catch the non-linear pattern of the data. Given a set of training data $\{(x_i, d_i)\}_i^N$ (x_i is the input vector, d_i is the actual value and N is the total number of data patterns), the general SVR function is (Wang *et al.* 2013):

$$y = f(x) = w\varphi(x_i) + b \tag{12}$$

where $\varphi(x_i)$ indicates feature spaces, non-linearly mapped from input vector x (Vapnik 1998). Regression parameters of b and w may be determined by assigning positive values for the slack parameters of ξ and ξ^* and minimization of the objective function (Equation (13)) (Wang *et al.* 2013):

$$\text{Minimize:} \frac{1}{2}\|w\|^2 + C\left(\sum_i^N (\xi_i + \xi_i^*)\right) \tag{13}$$

$$\text{Subject to:}\begin{cases} w_i\varphi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^* \\ d_i - w_i\varphi(x_i) + b_i \leq \varepsilon + \xi_i^* \quad i = 1, 2, \ldots, N \\ \xi_i, \xi_i^* \end{cases}$$

where $(1/2)\|w\|^2$ is the weights vector norm and C is referred to the regularized constant determining the tradeoff between the empirical error and the regularized term. ε is called the tube size and is equivalent to the approximation accuracy placed within the training data points. Mentioned optimization problems can be changed to the dual quadratic optimization problem by defining Lagrange multipliers α_i and α_i^*. Vector w in Equation (12) can be computed after solving the quadratic optimization problem as (Wang *et al.*

2013):

$$w^* = \sum_{i=1}^N (\alpha_i - \alpha_i^*)\varphi(x_i) \tag{14}$$

So, the final form of SVR can be expressed as (Wang *et al.* 2013):

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^N (\alpha_i - \alpha_i^*)K(x, x_i) + b \tag{15}$$

where α_i^+ and α_j^- are Lagrange multipliers, $k(x_i, x_j)$ is the kernel function performing the non-linear mapping into feature space and b is the bias term. One commonly used kernel function is the Gaussian RBF kernel as (Haghiabi *et al.* 2016):

$$k(x_1, x_2) = \exp\left(-\gamma\|x_1 - x_2\|^2\right) \tag{16}$$

where γ is the kernel parameter.

Auto Regressive Integrated Moving Average

ARIMA is a classic time series-prediction model, commonly used for practical issues. The function of ARIMA is indicated by (p, d, q) where p indicates the order of autoregressive terms, d refers to degree of differencing and q is the order of moving average term in the predicting equation (for more details see Salas *et al.* 1990).

Ensembling unit

It is well known that a combination of different predictors as a post-processing approach can improve overall predictions for a time series. An advantage of combining predictions is that when several methods are used, the results do not seem to be highly sensitive to the specific choice of methods. In this sense, using combined predictors is safer and less risky than relying on a single method. It is recommended by experimental and theoretical researches that ensembling the outputs of various models can be an effective method to enhance the overall efficiency of the time series prediction. It has been shown that it is less risky to use a combination

of relatively simple methods than to use a single method which is more complex and expensive (Makridakis & Winkler 1983).

In this study, three methods were considered for combination of the used models outputs to improve predicting performance as: (a) the simple linear averaging method, (b) the linear weighted averaging method and (c) the non-linear neural ensemble method. In contrast to linear combination methods 1 (Equation (17)) and 2 (Equation (18)), in the neural ensemble method, another FFNN is trained to non-linearly ensemble the outputs of the single models of ARIMA, SVR, FFNN and ANFIS.

Simple averaging is done as:

$$\bar{f}(t) = \frac{1}{N}\sum_{i=1}^{N} f_i(t) \tag{17}$$

where $\bar{f}(t)$ is the output of the simple ensemble model, $f_i(t)$ is the output of the ith single model (here outputs of FFNN, ANFIS, SVR and ARIMA) and N is the number of single models (here, N = 4).

The weighted averaging model is expressed as:

$$\bar{f}(t) = \sum_{i=1}^{N} w_i f_i(t) \tag{18}$$

where w_i is the applied weight on the ith model which can be determined based on the model performance as:

$$w_i = \frac{DC_i}{\sum_{i=1}^{N} DC_i} \tag{19}$$

DC_i is the performance efficiency (e.g. determination coefficient) of the ith single model.

In the neural ensemble method, non-linear averaging is performed by training a neural network. The input layer of the neural ensemble model is fed by the outputs of four considered models, each of which is assigned to one neuron in the input layer. A schematic of the proposed neural ensemble method is shown in Figure 5.

Efficiency criteria

The Determination Coefficient (DC) and Root Mean Square Error (RMSE) efficiency criteria were employed in this study
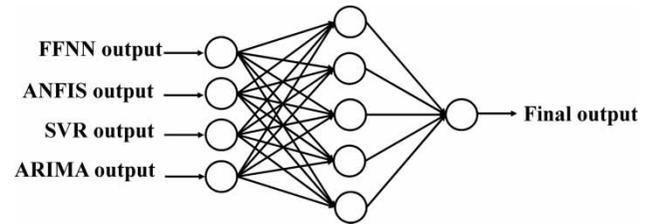

Figure 5 | Schematic of proposed neural ensemble model.

to evaluate the performance of the models as (Nourani 2017):

$$DC = 1 - \frac{\sum_{i=1}^{n} \left(o_{obs_i} - o_{com_i}\right)^2}{\sum_{i=1}^{n} \left(o_{obs_i} - \bar{o}_{obs}\right)^2} \tag{20}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(o_{obs_i} - o_{com_i}\right)^2}{n}} \tag{21}$$

where n, o_{obs_i}, $\bar{o}_{obs}$ and o_{com_i} are data number, observed data, averaged value of the observed data and calculated values, respectively. DC ranges between $-\infty$ and 1, with perfect score of 1.

RESULTS AND DISCUSSION

As the first step, piezometers numbers 207 and 217 of section 2 were modelled via proposed scenario 1 using FFNN, ANFIS, SVR and ARIMA models, separately (see Table 2 for the obtained results).

In scenario 1, correlation coefficient was employed for identification of the proper input set. It was found that each piezometer's head at time step t is mostly correlated with the piezometer's head at time step t–1 and upstream reservoirs levels at time steps t and t–1. Selection of appropriate architecture of the network, i.e. number of neurons in the hidden layer and optimal iteration epoch number are important items in training of each FFNN to prevent the overfitting of the FFNN. So the range of 10–200 for epoch number and the range of 1–10 for number of neurons in the hidden layer were considered and the best results are mentioned in the tables. Considering tangent sigmoid as activation functions of hidden and output layers, the

Table 2 │ Results of single models for scenario 1

Piezometer number	Model	Model structure[a]	DC		RMSE[b]	
			Calibration	Verification	Calibration	Verification
207	FFNN	3-8-1	0.929	0.887	0.061	0.036
	ANFIS	Gaussian-3	0.946	0.876	0.053	0.037
	SVR	0.333, 0.01, 15	0.915	0.883	0.067	0.038
	ARIMA	(5,2,4)	0.880	0.764	0.074	0.056
217	FFNN	3-5-1	0.827	0.756	0.110	0.104
	ANFIS	Trapezoidal-2	0.856	0.690	0.100	0.118
	SVR	0.333, 0.1, 30	0.873	0.667	0.094	0.121
	ARIMA	(5,2,4)	0.800	0.571	0.117	0.136

[a]Numbering of a-b-c in structure of neural network represents the number of input layer, hidden layer and output layer neurons. In ANFIS, structure MF-a refers to used membership function and number of membership functions. Numbering of a, b, c in SVR structure denotes to γ, ε, c and numbering of (a, b, c) in ARIMA structure refers to orders of autoregressive and moving average components and the order of differencing, respectively.
[b]Since all data are normalized, the RMSE has no dimension.

FFNNs were trained using a scaled conjugate gradient scheme of the BP algorithm (Haykin 1994) and the best structure and epoch number of each network were determined through a trial and error procedure to model both piezometers 207 and 217. The obtained results for the best structure are presented for piezometers 207 and 217 in Table 2.

A Sugeno fuzzy inference system was employed for ANFIS models. An ANFIS model includes several rules and membership functions. In the current study, Gaussian and Trapezoidal-shaped membership functions were found to be appropriate for simulation of piezometers number 207 and 217, respectively. Since time series of piezometers numbers 207 and 217 obey normal and semi-normal probability density function (directly or after applying a transformation), respectively, so Gaussian and Trapezoidal-shaped MFs have shown better performance in modeling the mentioned piezometers. Furthermore, constant MFs were employed for the outputs of ANFIS models. In addition to examining the number of membership functions, the alteration of the training epoch was also investigated to obtain the most optimum ANFIS models. The range of 10–100 for epoch number and the range of 2–4 for the number of membership functions were considered and the best results are shown in Table 2.

Also, SVR models were created by RBF kernel for both mentioned piezometers. The RBF kernel's tuning parameters are fewer than two sigmoid and polynomial kernels. Also, this kernel shows better performance by considering smoothness assumptions (Noori *et al.* 2011). For each piezometer, the RBF-kernel's parameters in SVR were tuned to achieve the best performance efficiency (see Table 2).

Finally, ARIMA models were created for both mentioned piezometers. For this purpose, firstly an ARIMA model was fitted to the calibration data to calibrate the model. Then the calibrated model was used to find the values of the verification data, month by month. The best structures of the ARIMA models were obtained through trial and error procedure for both piezometers numbers 207 and 217 (see Table 2).

Among the various predicting models that were developed, AI-based models were more accurate than the ARIMA model. It can be seen that although five lag times were used for calibrating of ARIMA model but results of AI models by three inputs were better than ARIMA models. The lower accuracy of the conventional ARIMA model compared to the AI-based models can be attributed to the linearity of ARIMA and its shortcomings in modeling non-linear processes such as seepage. In addition, ARIMA cannot use water levels of upstream reservoir as inputs. In other words, it can only use previous piezometer heads. Among AI models, the FFNN performance in the verification step was slightly better than the other two AI models.

Piezometer 207 is the closest piezometer to the upstream which is affected mostly by the variations of the reservoir water levels (even capillary), whereas piezometer number 217 is placed at the uppermost level of the core

and the middle of the cross-section and it is far enough from the upstream to parallel with the fluctuations of the upstream water level immediately; on the other hand, due to soil friction these fluctuations would disappear through the dam, and the fluctuations of upstream head do not have a considerable effect on the head of this piezometer. Because of employing upstream water level time series for training the AI models and due to the mentioned reasons, the performance of AI models for piezometer 217 was not as accurate as piezometer number 207 (see Tables 2–4).

In scenario 2, correlation coefficient was applied to determine the dominant piezometers between piezometers of section 2 and piezometer numbers 207 and 217. Consequently, piezometers number 212 and 216 were employed for modeling the piezometer number 207 and piezometers numbers 212 and 216 were used for modeling piezometer number 217. Therefore, upstream reservoir level at time

steps t and $t-1$ and 2 other piezometers' data at time step t were used as four inputs to train the AI models.

Scenario 3 is a combination of scenarios 1 and 2. Thus in scenario 3, data of piezometers numbers 212, 216 and 207 were employed for modeling piezometer number 207, and data of piezometers numbers 212, 216 and 217 were used for modeling piezometer number 217. In this scenario for modeling piezometer i at time step t, piezometer i at time step $t-1$, upstream reservoir level at time steps t and $t-1$ and two other piezometers' data at time steps t and $t-1$ were employed to create the AI models. The results of FFNN, ANFIS, SVR and ARIMA models for both piezometers 207 and 217 via scenarios 2 and 3 are summarized in Tables 3 and 4, respectively. For instance, Figure 6 shows the observed and computed water head time series of piezometer number 217 by ARIMA and AI models for verification step via scenario 2.

Table 3 │ Results of single models for scenario 2

Piezometer number	Model	Model structure[a]	DC		RMSE[b]	
			Calibration	Verification	Calibration	Verification
207	FFNN	4-5-1	0.971	0.927	0.039	0.029
	ANFIS	Gaussian-2	0.966	0.871	0.042	0.038
	SVR	0.333, 0.01, 15	0.957	0.823	0.048	0.047
217	FFNN	4-2-1	0.911	0.791	0.079	0.097
	ANFIS	Trapezoidal-2	0.938	0.696	0.066	0.117
	SVR	0.25, 0.2, 5	0.940	0.737	0.065	0.108

[a]Numbering of a-b-c in structure of neural network represents the number of input layer, hidden layer and output layer neurons. In ANFIS, structure MF-a refers to used membership function and number of membership functions. Numbering of a, b, c in SVR structure denotes γ, ε, c, respectively.
[b]Since all data are normalized, the RMSE has no dimension.

Table 4 │ Results of single models for scenario 3

Piezometer number	Model	Model structure[a]	DC		RMSE[b]	
			Calibration	Verification	Calibration	Verification
207	FFNN	7-3-1	0.977	0.914	0.035	0.031
	ANFIS	Gaussian-2	0.986	0.763	0.027	0.052
	SVR	0.143, 0.1, 20	0.980	0.892	0.033	0.037
217	FFNN	7-2-1	0.946	0.801	0.061	0.094
	ANFIS	Trapezoidal-2	0.972	0.689	0.044	0.118
	SVR	0.5, 0.2, 5	0.957	0.761	0.055	0.103

[a]Numbering of a-b-c in structure of neural network represents the number of input layer, hidden layer and output layer neurons. In ANFIS, structure MF-a refers to used membership function and number of membership functions. Numbering of a, b, c in SVR structure denotes γ, ε, c, respectively.
[b]Since all data are normalized, the RMSE has no dimension.
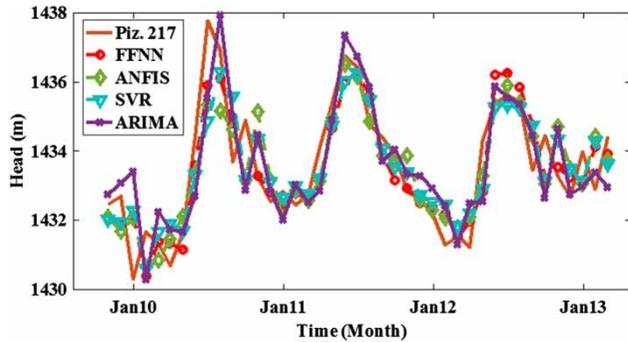
Figure 6 | The results of single AI models for verification step for piezometer 217 via scenario 2.

According to the obtained results via three proposed scenarios, in scenario 2 although the data of each piezometer have not been used in the input layer, the performance of scenario 2 due to the use of synchronous data with targets and using data of two other piezometers for modeling is slightly better than scenario 1. In scenario 3, because of employing each piezometer's data and two other piezometers, the performance is much better than both scenarios 1 and 2 but albeit with more complex structure of modeling. Therefore, in general cases of seepage modeling, scenario 1 could be used and in case of failure of some piezometers, scenario 2 could be employed. Furthermore, scenario 3 could be employed to catch more accurate results with a more complicated modeling framework.

By comparing the results of single models (see Figure 6), it is clear that the FFNN model could not simulate the peaks as well as other AI models. On the other hand, ARIMA

predicted upper and ANFIS and SVR predicted lower values of the observed time series much better. So it could be concluded that each model has some limitations and advantages in modeling of different scenarios and different piezometers. Thus by combining different models, predicting performance could be improved over the single models. In the next step, the three ensembling methods described previously were used for combining outputs of single models to improve predicting performance for each scenario. In this step, ensemble methods were formed for each scenario.

Only the calibration data set was utilized to estimate the parameters of both weighted averaging and neural ensemble methods. In a neural ensemble model like single FFNN, considering tangent sigmoid as activation functions of hidden and output layers, the network was trained using a scaled conjugate gradient scheme of BP algorithm and the best structure and epoch number of the ensemble network were determined through a trial and error procedure.

Obtained results of ensemble models for piezometers 207 and 217 via scenarios 1, 2 and 3 are tabulated in Tables 5–7, respectively. Furthermore, peak DCs and RMSEs of neural ensemble models for the verification step are compared with the individual AI models in Table 8 for scenarios 1, 2 and 3. For example, Figures 7 and 8 show the observed versus computed water head time series (computed by both neural ensemble method and a single model) respectively for the piezometers numbers 207 and 217 at the verification step. Also the scatter plots of the verification step of modeling by scenarios 1–3 for piezometers 207 and 217 are shown in Figures 9 and 10, respectively.

Table 5 | Results of ensemble models for scenario 1

Piezometer number	Ensemble model	Model structure[a]	DC		RMSE[b]	
			Calibration	Verification	Calibration	Verification
207	Simple averaging	–	0.942	0.892	0.055	0.035
	Weighted averaging	0.2530, 0.2577, 0.2494, 0.2309	0.942	0.892	0.055	0.035
	Neural averaging	4-7-1	0.943	0.892	0.054	0.035
217	Simple averaging	–	0.882	0.716	0.091	0.113
	Weighted averaging	0.2464, 0.2550, 0.2602, 0.2385	0.882	0.715	0.090	0.113
	Neural averaging	4-5-1	0.841	0.757	0.105	0.104

[a]Numbering of a,b,c,d in weighted averaging structure denotes to weights of FFNN, ANFIS, SVR, ARIMA models. Numbering of a-b-c in structure of neural averaging represents the number of input layer, hidden layer and output layer neurons.
[b]Since all data are normalized, the RMSE has no dimension.

Table 6 │ Results of ensemble models for scenario 2

Piezometer number	Ensemble model	Model structure[a]	DC		RMSE[b]	
			Calibration	Verification	Calibration	Verification
207	Simple averaging	–	0.973	0.935	0.037	0.027
	Weighted averaging	0.2572, 0.2559, 0.2536, 0.2333	0.973	0.935	0.037	0.027
	Neural averaging	4-2-1	0.974	0.949	0.036	0.024
217	Simple averaging	–	0.955	0.788	0.056	0.097
	Weighted averaging	0.2538, 0.2613, 0.2620, 0.2230	0.955	0.788	0.056	0.098
	Neural averaging	4-7-1	0.948	0.809	0.060	0.093

[a]Numbering of a,b,c,d in weighted averaging structure denotes to weights of FFNN, ANFIS, SVR, ARIMA models. Numbering of a-b-c in structure of neural averaging represents the number of input layer, hidden layer and output layer neurons.
[b]Since all data are normalized, the RMSE has no dimension.

Table 7 │ Results of ensemble models for scenario 3

Piezometer number	Ensemble model	Model structure[a]	DC		RMSE[b]	
			Calibration	Verification	Calibration	Verification
207	Simple averaging	–	0.983	0.885	0.030	0.036
	Weighted averaging	0.2555, 0.2578, 0.2563, 0.2303	0.983	0.886	0.030	0.036
	Neural averaging	4-3-1	0.985	0.900	0.028	0.033
217	Simple averaging	–	0.961	0.767	0.052	0.102
	Weighted averaging	0.2575, 0.2644, 0.2603, 0.2177	0.962	0.768	0.051	0.102
	Neural averaging	4-4-1	0.955	0.821	0.056	0.089

[a]Numbering of a,b,c,d in weighted averaging structure denotes to weights of FFNN, ANFIS, SVR, ARIMA models. Numbering of a-b-c in structure of neural averaging represents the number of input layer, hidden layer and output layer neurons.
[b]Since all data are normalized, the RMSE has no dimension.

Table 8 │ Results of neural ensemble models in comparison of individual AI models for verification step

Piezometer number	Model	DC			RMSE[a]		
		Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
207	FFNN	0.698	0.803	0.621	0.036	0.029	0.042
	ANFIS	0.702	0.521	0.266	0.036	0.045	0.056
	SVR	0.613	0.682	0.561	0.041	0.037	0.043
	Neural averaging	0.716	0.866	0.688	0.035	0.024	0.042
217	FFNN	0.392	0.632	0.601	0.116	0.090	0.105
	ANFIS	0.403	0.339	0.354	0.115	0.121	0.120
	SVR	0.420	0.312	0.435	0.135	0.124	0.122
	Neural averaging	0.516	0.687	0.692	0.104	0.084	0.088

[a]Since all data are normalized, the RMSE has no dimension.

The results of the ensemble methods indicated that almost all of the ensemble models produced better outcomes with regard to the single models. So, simple averaging, weighted averaging and neural averaging methods increased the performance of AI based modeling (according to the obtained DC values) for piezometer No. 207 respectively up to 3, 3 and 3% in the training and up to 7, 7 and 4% in the verification step. While for the piezometer No. 217, these improvements were respectively up to 16, 16, and 18% for the training and up to 13, 13 and 19% for the
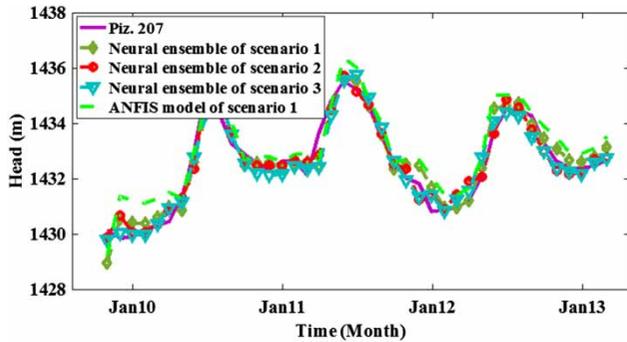
Figure 7 | The results of neural ensemble models and ANFIS model of scenario1 for verification step for piezometer 207.
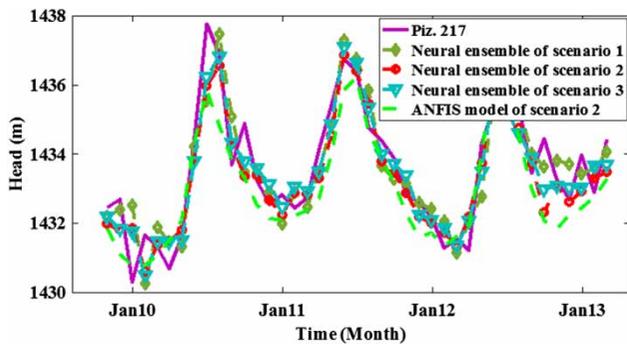

Figure 8 | The results of neural ensemble models and ANFIS model of scenario 2 for verification steps for piezometer 217.

increased the performance of modeling for piezometers 207 and 217 respectively up to 11 and 23% for the verification step in comparison with the AI model with the best performance. As discussed previously, some of the models provided higher and others provided lower estimates, each model having its own advantages and disadvantages. However, the ensemble methods, because of using each model's unique capability, simulate the phenomenon better than single models. As can be seen in Tables 2–7, the results of single models are very close together, and since the performance of simple and weighted averaging methods are in direct relation with individual models, thus the results of simple and weighted averaging models are almost equal.

The performance of the FFNN-based non-linear ensemble was better than the two linear ensemble methods. In non-linear ensemble due to employing FFNN, simulation of the non-linear behavior of the phenomenon would be more accurate than two other ensemble methods. On the other hand, since the results of both simple and weighted ensemble models are in direct relation to the individual models, if the performance of one is poor, the obtained results of ensembling will also be poor; in such conditions the neural ensembling would be very helpful.

verification steps. It should be noted that the DCs in the calibration phase for all methods did not show a significant difference, but the performance improvements are remarkable in the verification phase, which is the ultimate aim of this study. Also, it can be concluded from the results of Table 8 that ensemble models can predict peaks much better than single models. So, the neural averaging method

CONCLUSIONS

In this paper, three methods of model ensembling, including simple, weighted and neural ensembling methods, were employed to combine the models to increase the modeling efficiency. In this way, Sattarkhan earthfill dam piezometric
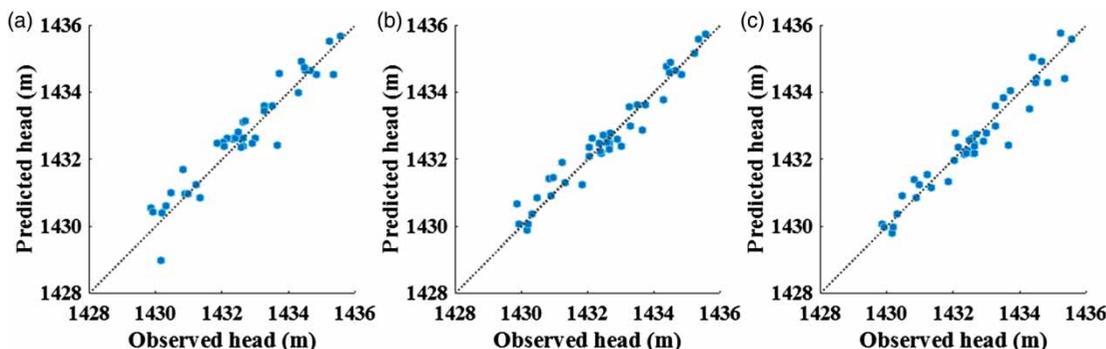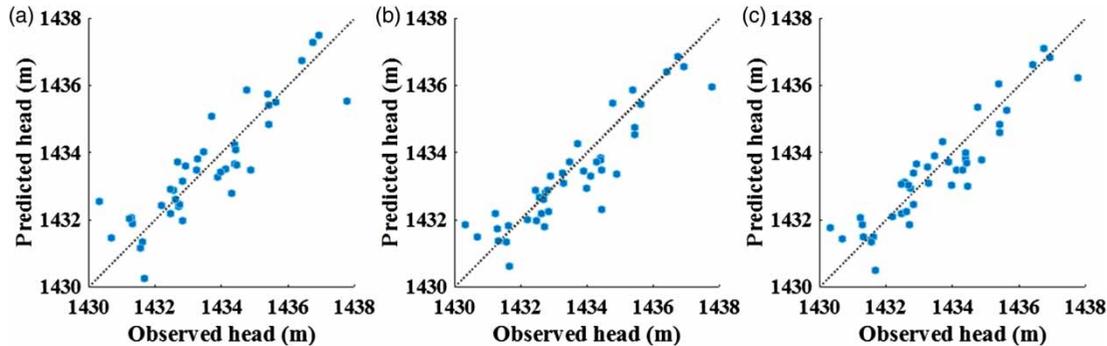

Figure 9 | Verification step scatter plot for neural ensemble models for piezometer 207. (a) Scenario 1 (DC = 0.8923), (b) scenario 2 (DC = 0.9488), (c) scenario 3 (DC = 0.9002).

Figure 10 | Verification step scatter plot for neural ensemble models for piezometer 217. (a) Scenario 1 (DC = 0.7565), (b) scenario 2 (DC = 0.8087), (c) scenario 3 (DC = 0.8214).

heads were simulated via different AI models of FFNN, SVR, ANFIS as well as classic ARIMA. Thereafter ensemble methods were employed as post-processing methods to combine the outputs of the models. For this purpose, three scenarios were considered, each containing different input combinations for different conditions. The comparison of the obtained results of single models showed that an FFNN result in the verification step is slightly better than other single models. According to the results of AI-based and ARIMA models in the verification step, it was clear that AI-based models are more reliable than the linear ARIMA model, since the ARIMA model could not cope with nonlinearity of the seepage phenomenon.

The comparison of considered scenarios showed that employing synchronous data with targets can improve predicting performance by up to 10%. Furthermore, it was investigated that in case of a failure of a piezometer, other piezometers may be used in modeling based on scenario 2.

Ensemble models produced better approximation than the single models and model combination improved the modeling performance by up to 20%.

Not surprisingly, the neural ensemble model was generally found to be a more robust and efficient method of combination and this method could improve the performance of AI modeling for the verification step by up to 20%. Also, the neural averaging method increased the performance of modeling for prediction of peaks by up to 23% for the verification step in comparison with the AI model with the best performance. The success of the neural ensemble was due to the fact that by employing the FFNN model in ensemble methods, the non-linear kernel could simulate the non-linear behavior of the

phenomenon more accurately than other linear ensemble methods.

Overall, the results of this study provide promising evidence for combining models. The results of the application of the three combination methods, especially the neural ensemble method, indicated that the estimated combined outputs can be more accurate than the individual models.

For future studies, other black box models such as Genetic Programming and also conceptual models may be employed to ensemble and compare the obtained results with the individual models. In addition, it is suggested to verify other non-linear ensembling methods such as ANFIS and SVR instead of FFNN. Furthermore, it is suggested to simulate other important aspects of earthfill dams like settlement and its variation at different points of the earthfill dams.

REFERENCES

Aqil, M., Kita, I., Yano, A. & Nishiyama, S. 2007 Analysis and prediction of flow from local source in a river basin using a Neuro-fuzzy modeling tool. *J. Environ. Manage.* **85**, 215–223.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000 Artificial neural networks in hydrology. 2: Hydrology applications. *J. Hydrol. Eng.* **5** (2), 124–137.

Bardet, J. P. & Tobita, T. 2002 A practical method for solving free-surface seepage problems. *Comput. Geotech.* **29**, 451–475.

Farhoudi, J., Hosseini, S. M. & Sedghi-Asl, M. 2010 Application of neuro-fuzzy model to estimate the characteristics of local scour downstream of stilling basins. *J. Hydroinform.* **12** (2), 201–211.

Haghiabi, A. H., Azamathulla, H. M. & Parsaie, A. 2016 Prediction of head loss on cascade weir using ANN and SVM. *ISH J. Hydr. Eng.* **23**, 102–110.

Haykin, S. 1994 *Neural Networks: A Comprehensive Foundation*. Macmillan, New York.

Jang, J. S. R., Sun, C. T. & Mizutani, E. 1997 *Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice-Hall, New Jersey.

Krogh, A. & Vedelsby, J. 1995 Neural network ensembles, cross validation, and active learning. *Adv. Neur. Inform. Process. Syst.* **7**, 231–238.

Makridakis, S. & Winkler, R. L. 1983 Average of forecasts: some empirical results. *Manage. Sci.* **29** (9), 987–996.

Noori, R., Karbassi, A. R. & Moghaddamnia, A. 2011 Assessment of input variables determination on the SVM model performance using PCA, gamma test and forward selection techniques for monthly stream flow prediction. *J. Hydrol.* **401**, 177–189.

Nourani, V. 2017 An emotional ANN (EANN) approach to modeling rainfall-runoff process. *J. Hydrol.* **544**, 267–277.

Nourani, V. & Babakhani, A. 2012 Integration of artificial neural networks with radial basis function interpolation in earthfill dam seepage modeling. *J. Comput. Civil Eng.* **27** (2), 183–195.

Nourani, V. & Parhizkar, M. 2013 Conjunction of SOM-based feature extraction method and hybrid wavelet-ANN approach for rainfall–runoff modeling. *J. Hydroinform.* **15** (3), 829–848.

Nourani, V., Sharghi, E. & Aminfar, M. H. 2012 Integrated ANN model for earthfill dams seepage analysis: Sattarkhan Dam in Iran. *Artif. Intell. Res.* **1** (2), 22–37.

Nourani, V., Aminfar, M. H., Alami, M. T., Sharghi, E. & Singh, V. P. 2014 Unsteady 2-D seepage simulation using physical analog, case of Sattarkhan embankment dam. *J. Hydrol.* **519**, 177–189.

Novaković, A., Ranković, V., Grujović, N., Divac, D. & Milivojević, N. 2014 Development of neuro-fuzzy model for dam seepage analysis. *Ann. Faculty Eng. Hunedoara* **12** (2), 133–136.

Salas, J. D., Delleur, J. W., Yevjevich, V. & Lane, W. L. 1990 *Applied Modeling of Hydrological Time Series*. Water Resources Publications, Denver.

Shahin, M. A., Jaksa, M. B. & Maier, H. R. 2001 Artificial neural network applications in geotechnical engineering. *Aust. Geomech.* **36** (1), 49–62.

Shamseldin, A. Y., O'Connor, K. M. & Liang, G. C. 1997 Methods for combining the outputs of different rainfall-runoff models. *J. Hydrol.* **197**, 203–229.

Tayfur, G., Swiatek, D., Wita, A. & Singh, V. P. 2005 Case study: finite element method and artificial neural network models for flow through Jeziorsko earthfill dam in Poland. *J. Hydr. Eng.* **131** (6), 431–440.

Vapnik, V. 1998 *Statistical Learning Theory*. Wiley, New York.

Wang, W. C., Xu, D. M., Chau, K. W. & Chen, S. 2013 Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. *J. Hydroinform.* **15**, 1377–1390.

Yongbiao, L. 2012 Prediction methods to determine stability of dam if there is piping. *Inform. Eng. Res. Inst. Proc.* **1**, 131–137.

Zhang, G. P. 2003 Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **50**, 159–175.