Scientific Research and Essays Vol. 8(3), pp. 120-127, 18 January, 2013
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE12.516
ISSN 1992-2248 ©2013 Academic Journals

Full Length Research Paper

Seepage evaluation of an earth dam using Group Method of Data Handling (GMDH) type neural network: A case study

Saeed Poorkarimi Kokaneh[1], Shahram Maghsoodian[2], Hossein MolaAbasi[3] and Afshin Kordnaeij[4]*

[1]Sama Technical and Vocational Training College, Islamic Azad University, Rasht Branch, Rasht, Iran.
[2]Islamic Azad University Noor, Noor, Iran.
[3]Babol University of Technology, Babol, Iran.
[4]Institute of Higher Education Allameh Mohaddes Noori, Noor, Iran.

Accepted 17 December, 2012

The seepage critical impact and significant destructive nature during and after construction of earth dams have been increasingly important topics during the past decades. In this paper, a new approach is presented for determination of seepage induced flow under and through an earth dam based on Group Method of Data Handling (GMDH) algorithm. After careful (detailed) studying of an earth dam called Fileh Khase dam located in Zanjan province of Iran, the permeability of soils was estimated by back analysis method using a Finite Element Method (FEM) software called SEEP\W. Then, a number of 96 data sets were provided using SEEP\W to use as a database according to allowable range of effective parameters such as permeability of clay core foundation of dam and water head in reservoir without any changes in geometry properties of the dam. This study addresses the question of whether GMDH type of artificial neural networks (ANN) optimized with genetic algorithms (GAs) could be used to estimate flow discharge through and under Fealeh Khase Dam. Results showed that GMDH type of ANN, provides an effective means of efficiently recognizing the patterns in data and accurately predicts the flow discharge through the Fileh Khase dam.

Key words: Seepage, earth dam, back analysis, Group Method of Data Handling (GMDH), artificial neural networks.

INTRODUCTION

Among hydro-geological, geological, and geotechnical problems in earth dams, seepage is a major concern associated with dams. Because it threatens dam stability and may cause unforeseen failure (Close et al., 2002). Many methods such as using of cut off wall and injection to avoid seepage are proposed. Otherwise the prediction of seepage discharge through earth dam can be useful for seepage controls. Fealeh Khaseh dam located in Zanjan province, west of Iran, is used as a case study in this paper. Although there are many established theoretical relation between seepage and soil properties, their association and evaluation of seepage, requires system identification techniques. The interdependency of factors involved in such problems prevents the use of regression analysis and demands a more extensive and sophisticated method. The Group Method of Data Handling (GMDH) type of artificial neural networks (ANN) optimized by Genetic Algorithms (GAs) can be used for complex systems modeling, where unknown relationships exist between the variables, without having a specific knowledge of the processes. In recent years, the use of such self-organizing neural networks has led to successful application of the GMDH-type of algorithm in geotechnical sciences (Ardalan et al., 2009; Kalantary et al., 2009).

This paper aims to develop a GMDH-type of ANN for the prediction of seepage discharge (Q) based on various

*Corresponding author. E-mail: afshin_geotec@yahoo.com. Tel: +989111133765.

Figure 1. View of FealehKhase earth Dam.

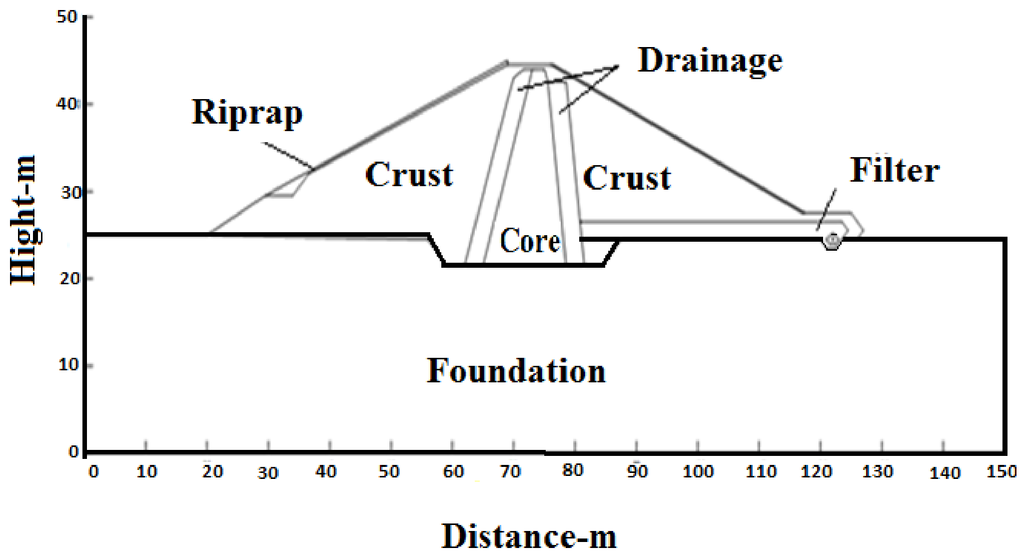

Figure 2. Cross section of FealehKhaseh earth Dam.

parameters such as soil permeability, water level in reservoir and etc.

This paper reviews FealehKhaseh dam at first, then a brief explanation of the databases under consideration and the process of modeling with GMDH are presented. Finally the developed GMDH model is described and its accuracy is assessed.

SEEPAGE ANALYSIS METHODS

The literature presents a portfolio of research regarding application of seepage discharge for geotechnical characterization (earth dams). All of researchers have proposed based on Darcy law. Nowdays numerical methods such as finite element method (FEM) and finite difference method (FDM) and softwares base on these methods such as Geo Office, Plaxis and Flac are commonly used. Therefore in this paper the seepage discharge through FealehKhaseh earth dam based on FEM (Geo Office software: SEEP\W) and a new approach is investigated (William Lambe and Whitman, 1969; Das, 1997).

OVERVIEW OF FEALEHKHASEH EARTH DAM

FealehKhase Dama typical earth dam with a clay core, filter zone, and sandy gravel shell, is located in the Zanjan province, north west of Iran (Figures 1 and 2).

FealehKhaseh Dam, one of the four largest dams in the Zanjan province, is used exclusively for domestic, agricultural, and industrial water supply of FealehKhaseh village. It is 17 m in height and 107 m in length. The full reservoir area is 2.42 km^2 with a storage capacity of 1.77 million cubic meters. Other information can be found in Report of FealehKhaseh dam design (2006).

PRINCIPLES OF MODELING USING GROUP METHOD OF DATA HANDLING (GMDH) TYPE OF ARTIFICIAL NEURAL NETWORK

The GMDH algorithm is a self-organizing approach by which gradually complicated models are generated based on the evaluation of their performances on a set of multi-input single-output data pairs (x_i, y_i) (i=1, 2,…, m). The GMDH was first developed by Ivakhnenko as a multivariate analysis method for complex system modeling and identification (Ivakhnenko, 1971). The main idea of GMDH is to build an analytical function in a feed forward network based on a quadratic node transfer function whose coefficients are obtained using regression technique (Farlow, 1984).

A model can be represented as a set of neurons in which different pairs of them in each layer are connected through a quadratic polynomial by GMDH algorithm, and thus, produce new neurons in the next layer. Such representation can be used in modeling to map inputs to outputs. The formal definition of the identification problem is to find a function $\hat{f}$ that can be approximately used instead of the observed one, f in order to predict output $\hat{y}$ for a given input vector $X = (x_1, x_2, x_3, \dots, x_n)$ as close as possible to its observed output y. Therefore, given M observations of multi-input, single output data pairs so that:

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) \qquad (i = 1, 2, 3, \dots, M) \tag{1}$$

It is now possible to train a GMDH type of artificial neural network to predict the output values $\hat{y}_i$ for any given input vector $X = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$, that is:

$$\hat{y}_i = \hat{f}(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) \qquad (i = 1, 2, 3, \dots, M) \tag{2}$$

The problem is now to determine a GMDH type of artificial neural network so that the square of the differences between the observed and predicted output is minimized as follow:

$$\sum_{i=1}^{M} [\hat{f}(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) - y_i]^2 \to min \tag{3}$$

The general connection between input and output variables can be expressed by a complicated discrete form of the Volterra functional series, known as the Kolmogorov-Gabor polynomial. Hence:

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots \tag{4}$$

This full form mathematical description can be represented by a system of partial quadratic polynomials consisting of only two variables (neurons) in the form of:

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{5}$$

By this means, the partial quadratic description is recursively used in a network of connected neurons to build the general mathematical relation between inputs and output given in equation (4). The coefficients a_i in equation (5) are calculated using regression techniques, so that the difference between the observed output, …, and the calculated one, $\hat{y}$, for each pair of $\mathbf{x_i}$, y_i as input variables is minimized. Apparently, a tree of polynomials is constructed using the quadratic form given in equation (5) whose coefficients are obtained in a least squares scheme. In this way, the coefficients of each quadratic function $\mathbf{G_i}$ are obtained to fit optimally the output in the whole set of input–output data pairs, that is:

$$E = \frac{\sum_{i=1}^{M} (y_i - G_i(\ ))^2}{M} \to min \tag{6}$$

In the basic form of the GMDH algorithm, all the possibilities of two independent variables out of the total n input variables are taken in order to construct the regression polynomial in the form of equation (5) that best fits the dependent observations $(y_i, i = 1, 2, \dots, M)$ in a least squares sense. Consequently, $\binom{n}{2} = \frac{n(n-2)}{2}$ neurons will be built up in the first hidden layer of the feed forward network from the observations

$\{(y_i, x_{ip}, x_{iq}); (i = 1, 2, 3, \dots, M)\}$ for different $p, q \in \{1, 2, 3, \dots, n\}$.

In other words, it is now possible to construct M data triples $\{(y_i, x_{ip}, x_{iq}); (i = 1, 2, 3, \dots, M)\}$ from observations using $p, q \in \{1, 2, 3, \dots, n\}$ in the form of:

$$\begin{bmatrix} x_{1p} & x_{1q} & \vdots & y_1 \\ x_{2p} & x_{2q} & \vdots & y_2 \\ \dots & \dots & \vdots & \dots \\ x_{Mp} & x_{Mq} & \vdots & y_M \end{bmatrix}$$

Using the quadratic sub-expression in the form of equation (5) for each row of M data triples, the following matrix equation can be readily obtained as

$$Aa = Y \tag{7}$$

$$a = \{a_0, a_1, a_2, a_3, a_4, a_5\} \tag{8}$$

$$Y = \{y_1, y_2, y_3, \dots, y_m\}^T \tag{9}$$

Where; a is the vector of unknown coefficients for the quadratic polynomial in equation (5), and Y is the vector of output values from observation. It can be readily seen that:

$$A = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & x_{1p}^2 & x_{1q}^2 \\ 1 & x_{2p} & x_{2q} & x_{2p}x_{2q} & x_{2p}^2 & x_{2q}^2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{Mp} & x_{Mq} & x_{Mq}x_{Mq} & x_{Mp}^2 & x_{Mq}^2 \end{bmatrix} \tag{10}$$

The least squares technique from multiple regression analysis leads to solution of the normal equations, in the form of:

$$a = (A^T A)^{-1} A^T \tag{11}$$

This determines the vector of the best coefficients of the quadratic equation (5) for the whole set of M data triples. Notice that this procedure is repeated for each neuron of the next hidden layer according to the connectivity topology of the network. However, such a solution directly from normal equations is rather susceptible to round off errors and, more importantly, to the singularity of these equations (Nariman-Zadeh et al., 2002).

There are two main concepts involved within GMDH type of artificial neural networks design, namely, the parametric and the structural identification problems. Nariman-Zadeh et al.present hybrid genetic algorithm (GA) and singular value decomposition (SVD) method to optimally design such polynomial neural networks (Nariman-Zadeh et al., 2002).

APPLICATION OF GA IN THE TOPOLOGY DESIGN OF GROUP METHOD OF DATA HANDLING (GMDH) TYPE NEURAL NETWORKS

Stochastic methods are commonly used in the training of neural networks in terms of associated weights or coefficients and have successfully performed better than traditional gradient-based techniques. The literature shows a wide range of evolutionary design approaches either for architectures or connection weights separately, in addition to efforts for them simultaneously. In the most GMDH type neural network, neurons in each layer are only connected to neuron in its adjacent layer as it was the case in Methods I and II previously reported in. Taking this advantage, it was possible to present a simple encoding scheme for the genotype of each individual in the population as already proposed by Nariman-Zadeh et al. (2005). The encoding schemes in generalized GMDH neural networks (GS-GMDH) must, however, demonstrate the ability of representing different length and size of such neural networks.

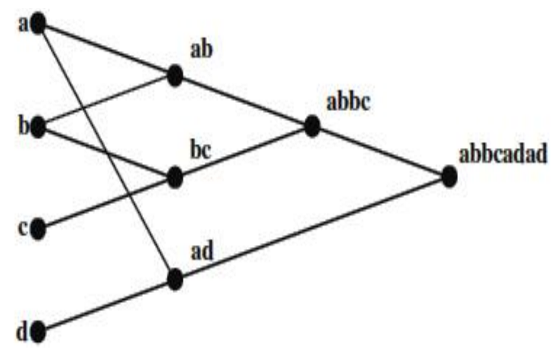

Figure 3. A generalized GMDH network structure of chromosome.

In a GS-GMDH neural network, Figure 3, neuron ad in the first hidden layer is connected to the output layer by directly going through the second hidden layer. Therefore, it is now very easy to notice that the name of output neuron (network's output) includes ad twice as abbcadad. In other words, a virtual neuron named adad has been constructed in the second hidden layer and used with abbc in the same layer to make the output neuron abbcadad as shown in the Figure 3.

Such repetition occurs whenever a neuron passes some adjacent hidden layers and connects to another neuron in the next 2nd, or 3rd, or 4th, or . . . following hidden layer. In this encoding scheme, the number of repetition of that neuron depends on the number of passed hidden layers, $\tilde{n}$, and is calculated as $2^{\tilde{n}}$ It is easy to realize that a chromosome such as abab bcbc, unlike chromosome abab acbc for example, is not a valid one in GS-GMDH networks and has to be simply re-written as abbc.

The genetic operators of crossover and mutation can now be implemented to produce two offspring from two parents. The natural roulette wheel selection method is used for choosing two parents producing two offspring (Nariman-Zadeh et al., 2005).

The incorporation of genetic algorithm into the design of such GMDH type neural networks starts by representing each network as a string of concatenated sub-strings of alphabetical digits. The fitness, φ, of each entire string of symbolic digits which represents a GMDH type neural network model is evaluated in the following form:

$$\varphi = \frac{1}{E} \tag{12}$$

where E is the mean square of error given by equation

Table 1. A sample of databases.

No.	Water level in reservoir (cm)	Foundation permeability (cm/s)	Core permeability (cm/s)	Seepage discharge (cm^3/s)
1	1000	0.000099	1.23E-06	2.7863
2	500	0.000099	1.93E-07	1.3364
3	1500	0.000033	5E-07	2.1326
4	500	0.000033	5E-07	0.4677
5	500	0.000033	6.07E-07	0.46774
6	1000	0.00001	5E-07	0.38535
7	1500	0.000023	1.93E-07	1.6116
8	1000	0.000099	6.07E-07	2.7858
9	200	0.000023	1.93E-07	0.12894
10	200	0.000099	6.07E-07	0.41371
11	200	0.00001	1.93E-07	0.044967
12	1500	0.000023	6.07E-07	1.61172
13	1500	0.00001	5E-07	0.83566
14	1500	0.000033	1.23E-06	2.1395

Table 2. Permeability of dam elements.

Dam element	Permeability (m/s)
Core	10^{-8}
Filter	10^{-3}
Crust	10^{-4}
Riprap	10^{-4}
Drainage	10^{-3}

(11), which is minimized through the evolutionary process by maximizing the fitness, φ. The evolutionary process starts by randomly generating an initial population of symbolic strings, each as a candidate solution. Then, using the genetic operations of roulette wheel selection, crossover, and mutation the entire population of symbolic strings improve gradually. In this way, GMDH type neural network models with progressively increasing fitness, φ, are produced until no further significant improvement is achievable.

Other neural network and machine learning methods can also be used in prediction of parameters such as chen et al., 2008 and lin et al., 2006. This study method suggested an specific formula with more generality. Therefore this study method proposed.

EVALUATION OF SEEPAGE THROUGH FEALEHKHASEH DAM

Evaluation of foundation permeability by back analyses

Information from FealehKhaseh dam shows that the most soil classification used as dam materials are ML, CL and GW. Soils permeability that is an important factor in evaluation of seepage is extracted in Table 2 (Report).

Back analyses method was used in order to evolve the permeability of dam foundation. By modeling of dam sections in Geo Office software (SEEP\W) within the allowable range of foundation permeability and comparing of piezometers head which measured in field, the foundation permeability estimated as 2.37×10^{-7} m/s. Results of back analyses at critical section of dam, the largest section of dam, is shown in Figure 4. Another comparison by field Logan test in geotechnical reports shows the accuracy of foundation permeability.

ESTIMATION OF SEEPAGE DISCHARGE

After determining the foundation permeability which has a great influence on seepage analyses by back analyses, seepage discharge through the total width of dam estimated as much as $2.0231 \times 10^{-6}\ \mathrm{m}^2/\mathrm{s}$. This discharge was estimated by modeling of dam cross sections (7 sections) with the maximum water head in reservoir using SEEP/W software (Figure 5).

At the end the total seepage through the dam was estimated as $114.17\,\mathrm{m}^3/\mathrm{day}$ by considering of seven section of dam. According to monthly hydrology information of FealehKhaseh River, the seepage hazard is not significant by the probability of 85% (IRCOLD[1]).

For similar dams with the same geometry seepage discharge variation was investigated. A total of 96 different conditions included material and foundation permeability and water level in reservoir modeled using SEEP\W software to identify probable conditions in order to establish a databases. A part of generated data is shown in Table 1.

[1] Iranian committee on large dams.

PZ - 3	
H (m)	K *10E-6 (m/s)
1360.275	1
1360.089	1.3
1359.897	1.8
1359.780	2.3
1359.701	2.8
1359.643	3.3

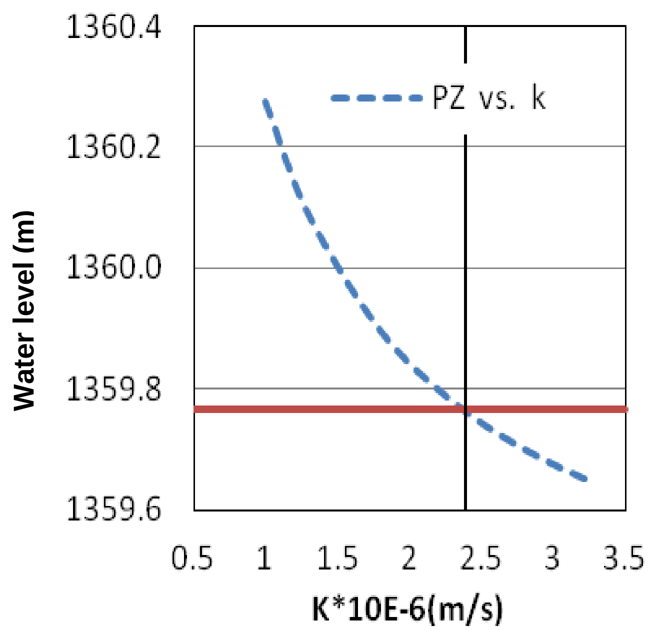

Figure 4. Back analyses of foundation permeability at critical section of dam.

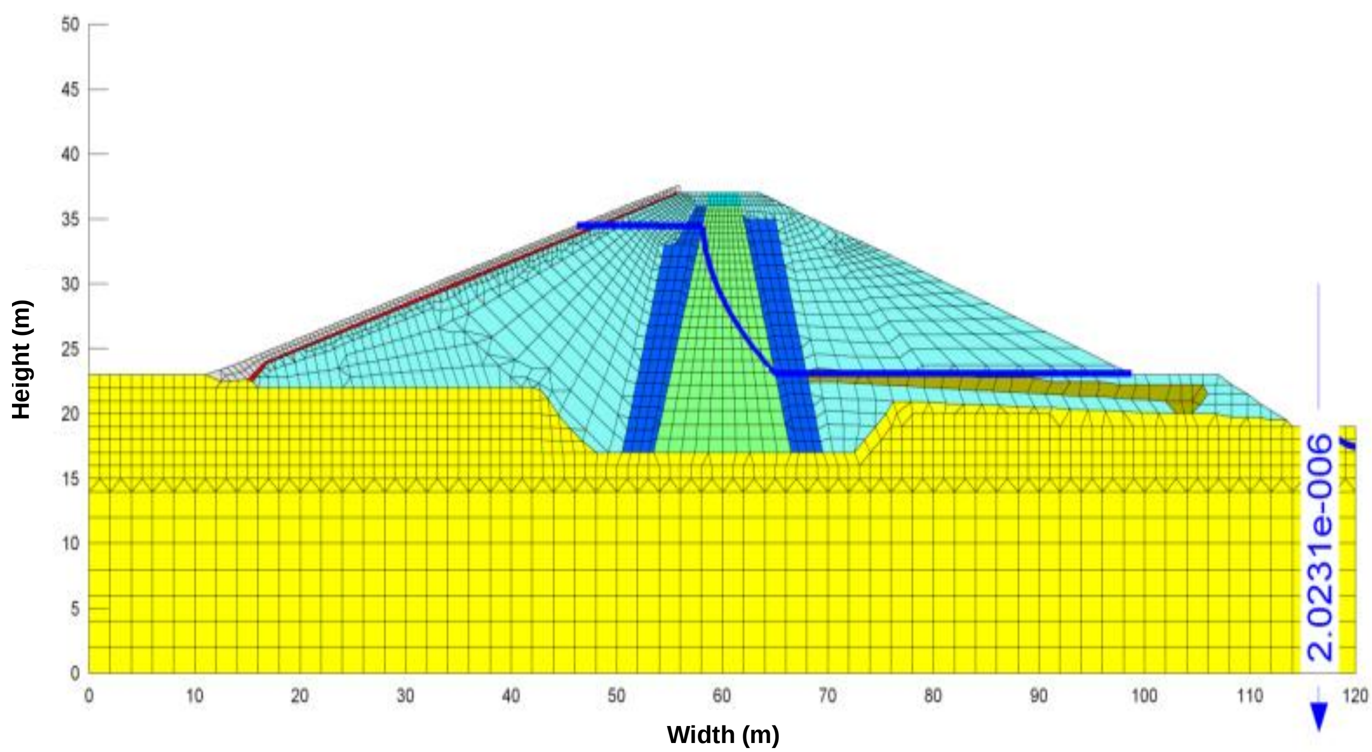

Figure 5. Seepage model in critical section by SEEP\W software.

EVALUATION OF SEEPAGE BY ARTIFICIAL NEURAL NETWORK

In order to demonstrate the prediction ability of evolved GMDH-type neural networks, databases have been divided into two different sets, namely, training and testing sets. The GMDH type neural networks are now used for such inputs-output data to find the polynomial model of flow discharge. The structure of the 2-hidden layer GMDH type of neural networks is shown in Figure 7

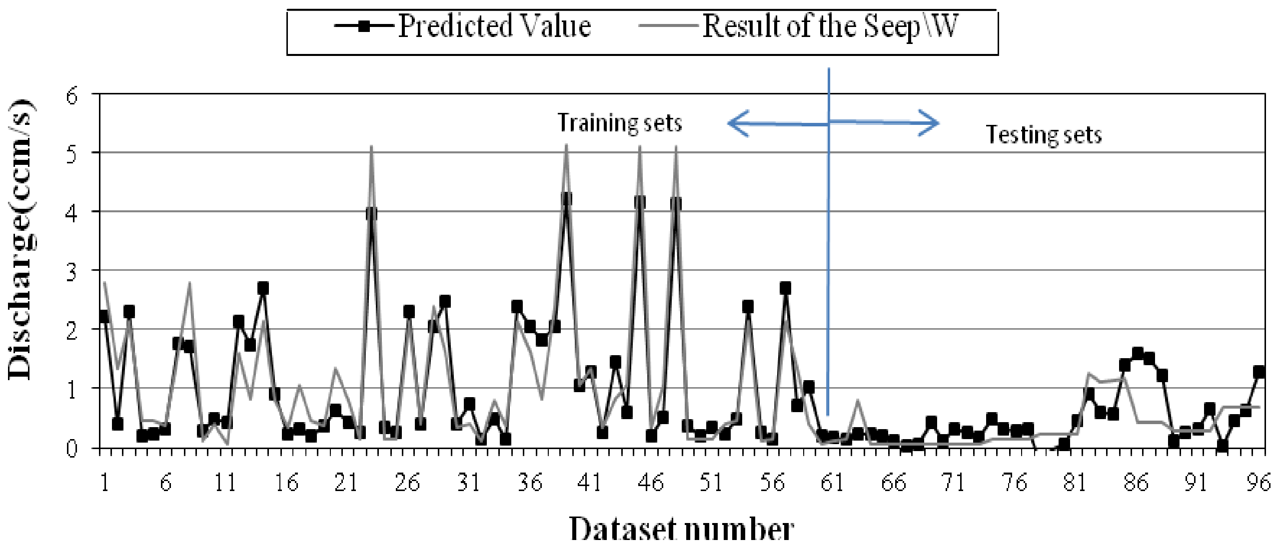

Figure 6. Neural network model predicted performance in comparison with SEEP\W predictions.

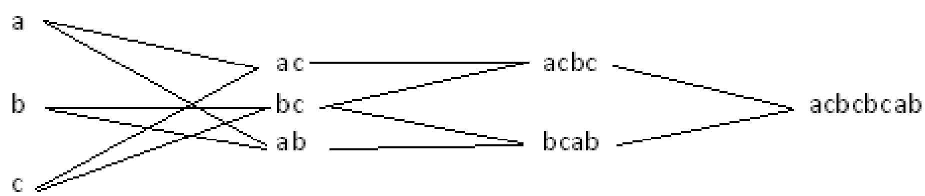

Figure 7. Evolved structure of GMDH neural network for prediction of seepage discharge.

corresponding to the genome representations of acbcbcab for seepage discharge in which a , b and c stands for water level in reservoir (cm), foundation permeability (cm/s) and core permeability (cm/s) of dam, respectively.

In order to demonstrate the prediction ability of the evolved GMDH type of neural networks, the generated database was divided into two different sets, called training and testing data. The training set consists of 60 inputs–output data pairs. The testing set, consisting of 31 inputs–output data pairs unforeseen during the training process, was merely used for testing the trained GMDH type of neural network models. Two hidden layers were considered for each model. To genetically design the neural networks, a population of 100 individuals with a cross over probability of 0.9, mutation probability of 0.1 and 400 generations was used. The evolved GMDH type neural networks parameters have been used to obtain the best model for the prediction.The model's predictive performance in comparison with SEEP\W modeling, are shown in Figure 6. As shown in this figure the predicted and measured values are fairly close.

The corresponding polynomial relationship resulted from neural network for seepage discharge is as follows:

$$y_1 = 0.053 + 0.0004x_1 + 0.00000003x_3 + 0.0000005x_1^2 + 0.00000000000003x_3^2 + 0.0000000003x_1x_3$$

$$y_2 = 1.01 - 349.29x_2 - 27819.89x_3 - 22.38x_2^2 - 0.04x_3^2 - 5.584x_2x_3$$

$$y_3 = 0.053 + 0.00049x_1 - 0.00027x_2 + 0.0000005x_1^2 - 0.0000003x_2^2 - 0.32x_1x_2$$

$$y_4 = -66.2 + 0.38y_1 + 169.83y_2 - 0.02y_1^2 - 104.35y_2^2 + 0.74y_1y_2$$

$$y_5 = -65.63 + 168.45y_2 + 1.095y_3 - 103.53y_2^2 - 0.02\,y_3^2 - 0.05y_3y_2$$

$$Q = 0.086 - 14.19y_5 + 14.19y_4 + 191.36y_5^2 + 168.96\,y_4^2 - 359.88y_4y_5$$

As presented in Table 3, the statistically assessed accuracy of the model is determined via R^2 (absolute fraction of variance), RMSE (root-mean squared error), MSE (mean squared error), and MAD (mean absolute

Table 3. Model statistics and information for the GMDH type of neural network model for predicting Shear wave.

Statistic RMSE	R^2	MSE	MAD
Neural training 43	0.93	1870	31
Neural testing 41	0.92	1718	25

deviation) defined as follows:

$$R^2 = 1 - \left[\frac{\sum_{i=1}^{M}(Y_{i(model)} - Y_{i(Actual)})^2}{\sum_{i=1}^{M}(Y_{i(Actual)})^2}\right]$$

$$RMSE = \left[\frac{\sum_{i=1}^{M}(Y_{i(model)} - Y_{i(Actual)})^2}{M}\right]^{\frac{1}{2}}$$

$$MSE = \frac{\sum_{i=1}^{M}(Y_{i(model)} - Y_{i(Actual)})^2}{M}$$

$$MAD = \frac{\sum_{i=1}^{M}\left|Y_{i(model)} - Y_{i(Actual)}\right|}{M}$$

CONCLUSION

It has been attempted in this study to deploy a system identification technique to develop the discharge correlation over an earth dam seepage properties. The evolved GMDH type of neural networks was used to obtain a model for the prediction of seepage discharge.

The validation and performance of the SEEP\W software was assessed and contrasted with results of pizometers in order to determination of foundation permeability and evaluation of total discharge through the dam.

According to ICOLD, seepage hazard is not significant by the probability of 85% is concluded.

Produced database consist of 94 data resulted from modeling of FealehKhaseh earth dam by SEEP\W software was compiled with the seepage discharge. A polynomial model was developed for discharge based on water level in reservoir, foundation and earth dam core permeability. Using evolved GMDH type of neural networks it was indicated that the water level in reservoir have less meaningful influence on the correlation relationships.

Also obtained results revealed that empirical correlations derived from this dataset should not be used to approximate seepage discharge for other sites. Therefore, these proposed relationships should be used with caution in geotechnical engineering and should be checked against measured seepage discharge.

REFERENCES

Ardalan H, Eslami A, Nariman-Zadeh N (2009). Piles shaft capacity from CPT and CPTu data by polynomial neural networks and genetic algorithms. Comput. Geotechn. 36: 616-625.

Chen CT, Wang WC, Xu DM, Chau KW (2008). Optimizing hydropower reservoir operation using hybrid genetic algorithm and chaos. Water Resour Manag. 22(7):895-909.

Close ME, Stanton GJ, Pang L (2002). Use of rhodamine WT with XAD-7 resin for determining groundwater flow paths. J. Hydrogeol. J. 10: 368-376.

Das BM (1997). Advanced Soil Mechanics. Second Edition. Taylor & Francis.

Farlow SJ (1984). Self-organizing method in modelling: GMDH type algorithm. M Dekker Inc.

Ivakhnenko AG (1971). Polynomial theory of complex systems. IEEE Trans. Syst. Man Cybern. SMC-1:364-378.

Kalantary F, Ardalan H, Nariman-Zadeh N(2009). An investigation on the Su-N_{SPT} correlation using GMDH type neural networks and genetic algorithms. Eng. Geol. 104:144-155.

Lin JIE, Cheng CT, Chau KW (2006). Using support vector machines for long-term discharge prediction. Hydrol. Sci. J. 51(4):599-612.

Nariman-Zadeh N, Darvizeh A, Darvizeh M, Gharababaei H (2002). Modelling of explosive cutting process of plates using GMDH type neural network and singular value decomposition. J. Mater. Process. Tech. 128:80-87.

Nariman-Zadeh N, Darvizeh A, Jamali A, Moeini A (2005). Evolutionary design of generalized polynomial neural networks for modeling and prediction of explosive forming process. J. Mater. Process. Tech. 164-165. 1561-1571.

Report of FealehKhaseh dam desin (2006). FarazAb Engineering consultant.

SEEP/W DEFINE Version 5.11 (2002). GEO-SLOPE International Ltd.

William Lambe T, Whitman RV (1969). Soil Mechanics. SI Version.John Wiley & Sons Inc.