DOI 10.4430/bgo00347

Pond water level simulation by applying the Hybrid Genetic Evolutionary Artificial Neural Network method

M. Ahmadi[1] and M.A. Riahi[2]

[1] *Department of Computer Engineering, Islamic Azad University, Malãyer, Iran*

[2] *Department of Solid Earth Physics, University of Tehran, Iran*

(Received: 18 June 2020; accepted: 19 November 2020; published online: 20 September 2021)

ABSTRACT The appropriate design of the pond coast structures and the reservoir supplement management entails an accurate pond water surface simulation. A hybrid Genetic Algorithm - Artificial Neural Network (GA-ANN) is presented and applied in this research to estimate the next 3- and 5-day water surfaces. Training and validation of the GA-ANN is performed using the 4-year daily water surface measurements performed on Chahnimeh reservoir located on the eastern side of Iran. Various input combinations are applied to the GA-ANN method. According to the results, for both the next 3- and 5-day estimation models, the input combination, consisting of the past 2-day water surface data, contributes to the optimal yield. Root Mean Squared Error of the optimum next 3- and 5-day prediction GA-ANN models were obtained to be 0.1798 and 0.3102, respectively. This paper found that in modelling the 3- and 5-day ahead pond water level, the best input variables are the information on the one and two previous days.

Key words: artificial neural network, genetic algorithm, pond water surface, simulation, soft computing.

1. Introduction

A simulation is the execution of a model, represented by a computer program that gives information about the system being investigated. The simulation approach of analysing a model is opposed to the analytical approach, where the method of analysing the system is purely theoretical. The simulation modelling is intended to help determine locations where pond-water availability is seasonal and identify possible pond management interventions that can extend the period of water availability. Pond water surface simulation is considered one of the utmost significant processes in designing pond coast construction. Nath and Bolte (1998) developed a water budget simulation model at the detailed individual ponds.

In contrast, continental-scale estimations tend to oversimplify model inputs. Aguilar-Manjarrez and Nath (1998) used simplified estimates of the key variables: surface runoff was estimated as a constant proportion (0.1) of the monthly rainfall, while a constant soil percolation rate was applied for the entire African continent. The annual quantum of water (in mm) required for fish culture was estimated, and critical thresholds set for rating and mapping pond aquaculture suitability. In mapping indicators of water-harvest potential at continental scale for Africa, Senay and Verdin (2004) used a more elaborate procedure to estimate annual surface runoff based on the Curve Number (CN) method of the United States Department of Agriculture Soil Conservation Service (SCS, 1985). Furthermore, accurate data on the pond water surface must be provided to prepare the water supply plane in reservoir management. Different parameters including runoff, water surface evaporation, environmental temperature, wind speed, and rainfall can influence the pond water surface. Therefore, the pond water surface becomes an intricate problem due to such multivariable dependency. The pond water surface simulation entails some data that is considered as input parameters. To choose the input variables, two procedures should be taken into account. Firstly, the model can be evaluated by using the above-mentioned parameters as the input variables. Secondly, the past pond surfaces must be chosen as the input variables in order to simulate the future pond surfaces. Based on the first procedure, application of various input variables obviously can lead to an intricate model. Accordingly, assuming that the environmental variables indirectly affect the pond surfaces, the second procedure can contribute to the most facile and practical simulation models.

Soft computing methods are extensively applied in intricate multivariable engineering problems. One of the powerful soft computing approaches that are commonly applied in the area of hydrology, hydraulics, and engineering is the Artificial Neural Network (ANN). ANN is applied and investigated in a diversity of region containing reservoir inflow prediction (Jothiprakash and Magar, 2012; Huang and Hsu, 2013; Lo *et al.*, 2014), pond and reservoir surface prediction (Biancamaria *et al.*, 2016; Avisse *et al.*, 2017; Busker *et al.*, 2019), wave prediction (Malekmohamadi *et al.*, 2011; Gullu, 2013; Puscasu, 2014), wind prediction (Song and Li, 2014; Nagababu *et al.*, 2016; Barbosa de Alencar *et al.*, 2017), sea surface prediction (Al-Zubaidy and Shambour, 2011; Imani *et al.*, 2014; Nitsure *et al.*, 2014).

The current paper mainly seeks to propose and apply a high-performance hybrid model for forcasting the future pond water surface by using the data relevant to the past two days. The modelling is performed via a novel hybrid Genetic Algorithm-ANN (GA-ANN) procedure. The GA-ANN has the potential to adjust the number of hidden layer neurons. The GA-ANN is applied to model the water surface of the Chahnimeh pond in the eastern Iran. Different input compounds are analysed to explore the utmost proper input variables. Additionally, different activation functions are examined for the ANN models, and the principal goal underlying the models is to simulate the next 3- and 5-day pond water surface.

2. Methodology

2.1. Recorded data of water level of Chahnimeh pond

In the current article, the daily measurement samples relevant to the Chahnimeh pond water surface were utilised for the training and validation of the GA-ANN method in the pond water surface simulation. As seen in Fig. 1, Chahnimeh pond is located on the eastern part of Iran at the latitude of 61° 39′ to 61° 43′ N and longitude of 30° 46′ to 30° 51′ E. The data set was provided via 4-year daily free surface measurements. In the present research, the first 2-year data measurements (700 numbers) are used in the training the GA-ANN models and the remaining 2-year measurements (700 numbers) are taken as the testing samples. The statistical parameters of the Chahnimeh pond level during the considered period are presented in Table 1.

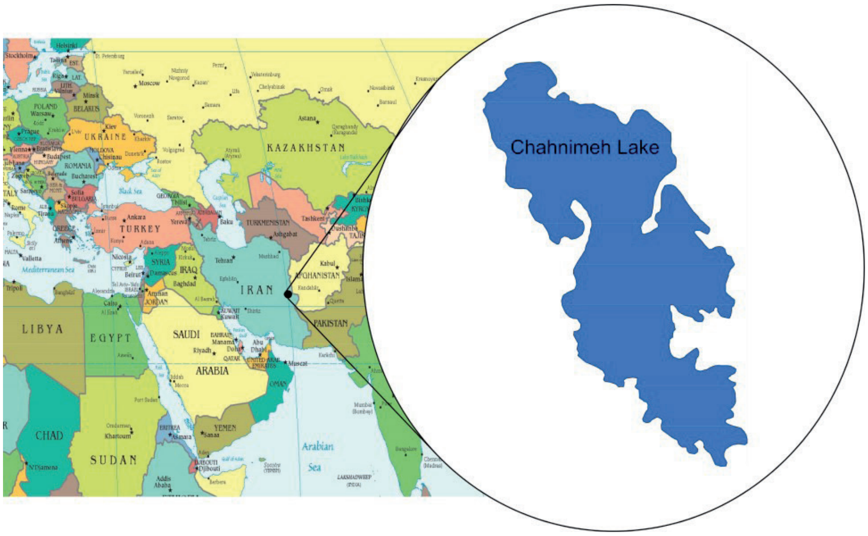

Fig. 1 - Location of the Chahnimeh pond.

Table 1 - Statistical parameters pertinent to the Chahnimeh pond.

Pond name	**Variable**	**Unit**	**Statistical parameters**			
			Min	**Max**	**Mean**	**Standard deviation**
Chahnimeh	*L*	m	484.83	492.50	488.57	1.90

2.2. Genetic Algorithm - Artificial Neural Network (GA-ANN) method

The Genetic Algorithm (GA) is a kind of self-adapting heuristic global search algorithm derived from imitating the thought of natural biological evolution. In nature, it is a cycle process made up of reproduction-crossover-mutation operators. In the process of searching for the global optimum solution, GA needs neither the information of gradient nor the calculus computing; it can find out the global optimum solution or near-optimal solution in the solution space with high probability only by operating the reproduction-crossover-mutation operators, thereby, it could reduce the probability of getting into the local minimum efficiently. The reproduction operator reproduces the individuals to the new colony according to the probability in proportion as their adaptive value. After reproduction, the preponderant individuals are preserved, and the inferior individuals are weed out, and the average fitness degree of the colony is increased. Still, the variety of colonies is a loss at the same time. The action of reproducing operator is to realise the principle of winner priority for preserving predominance and natural selection and make the colony converge on the optimum solution. The crossover operator first selects two individuals stochastically according to the particular exchanging probability Pc, and it can produce two new individuals by exchanging parts of chromogene stochastically. The GA can generate a filial generation colony with higher average fitness and better individuals through the reproduction and crossover operators and make the evolutionary process proceed to the optimum solution. The mutation operator changes several bits of the chromosome string stochastically with a small probability Pm, namely, turn 0 to 1 and 1 to 0. The mutation operator is very important to recoup the loss of colony diversity (Tian and Gao, 2009).

ANN is a highly simplified model of the structure of a biological network (Mandal *et al.*, 2009). The fundamental processing element of ANN is an artificial neuron (or simply a neuron).

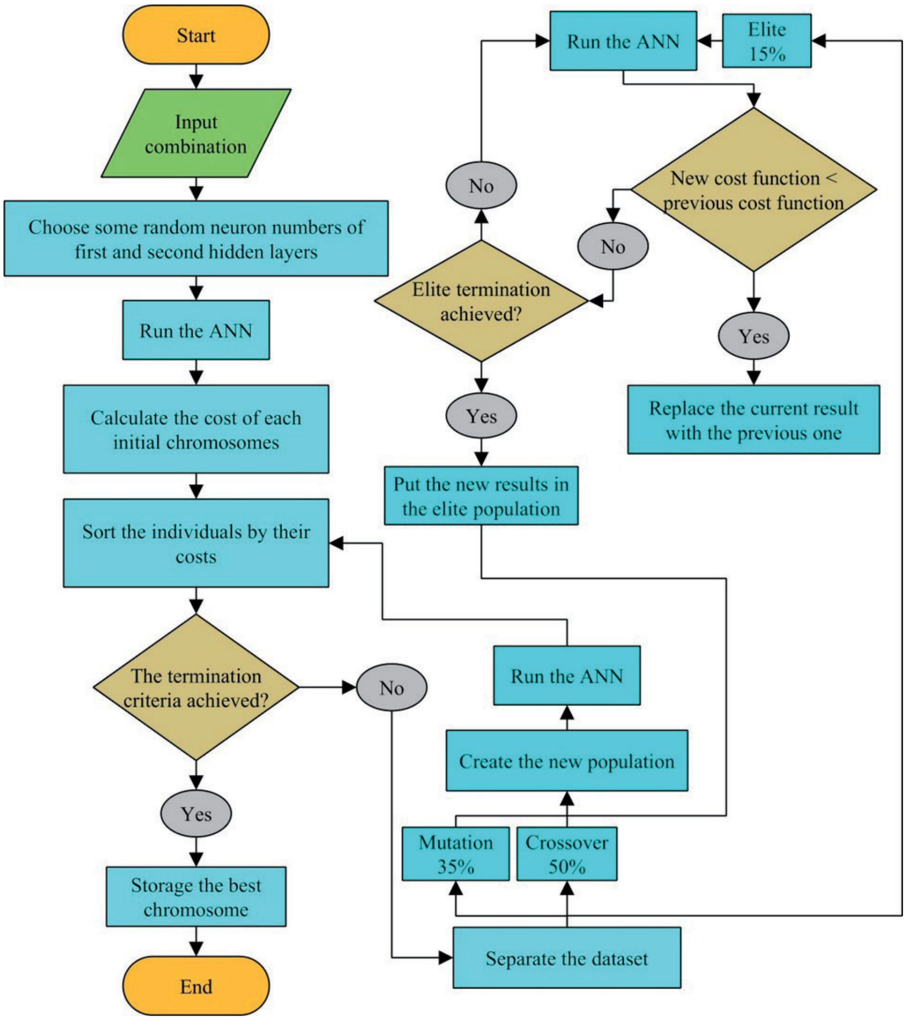

Fig. 2 - GA-ANN flowchart.

A biological neuron receives inputs from other sources, combines them, generally performs a nonlinear operation on the result, and, then, outputs the final result (Bas *et al.*, 2007). The basic advantage of ANN is that it does not need any mathematical model since an ANN learns from examples and recognises patterns in a series of input and output data without any prior assumptions about their nature and interrelations (Mandal *et al.*, 2009). ANN is a good alternative to conventional empirical modelling based on polynomial and linear regressions (Kose, 2008). On the other hand, ANN has several disadvantages such as long training time, unwanted convergence to local instead of optimal global solution, and a large number of parameters; therefore, there have been attempts to remedy some of these disadvantages by combining ANN with another algorithm that can take care of a specific problem. An algorithm that has frequently been hybridised with ANN is GA. A combination GA-ANN methods recently is getting popular among researchers. Therefore, we used GA-ANN to predict pond water surface simulation models.

The GA-ANN procedures are presented in Fig. 2. According to the figure, an initial population of the ANN models, with random hidden layer neuron number, is firstly created. Then, ANNs run as the chromosomes of the GA method, and the ANN model cost is calculated via the fitness function. Afterwards, the ANNs are sorted in terms of their costs. Finally, the GA is applied in order to find the best hidden layer neuron number through constructing different generations.

ANN model training is performed via the Damped Least Squares algorithm (Gavin, 2020). Due to the random weight selection of the Damped Least Squares training algorithm, a proper ANN with reasonable hidden layer neuron number may be regarded as an inapplicable and inaccurate ANN by GA. Thus, based on Fig. 2, a modification was done on the compendium choice of the GA. The compendium crowd was run many times, and the optimal cost of compendium chromosomes was saved. The effect of the local minima and the effect on the Random Weighted Selection of the training process were decreased.

Three kinds of layers including input, hidden and output constitute a GA-ANN. Neurons are the integral parts of each layer. Input layer neurons are taken into account as input variables. In other view, the outputs changeable are the supposed objectives of the neural network model. The upcoming pond water surface is taken as the output variable. The hidden layer neurons are able to perform some essential functions. The tasks of each neuron are: a) calculate the summation of the neurons of the previous layer according to the weight of each neuron, b) add the neuron bias to the result of the first step, c) put the result of the second step into the transfer function, and d) submit the result of the third step to the subsequent layer of neurons. The neuron's number of input and output layers is equal to the number of input and output variables of the considered problem.

Ability to select the proper activation functions plays a significant part in the GA-ANN accuracy. Sigmoid Activation Functions are accurately applied to the different areas of ANN modelling (Kavitha and Naidu, 2011; Khan *et al.*, 2011; Balaji and Baskaran, 2013). By definition, every bounded function with a direct relevance among *x* and *f(x)* is regarded as a sigmoid function (Zhang, 1998). In the current article, the linear activation function (Eq. 1) is applied to the output layer. On the other hand, two functions, namely Logarithmic Activation Function (LAF) (Eq. 2) and Hyperbolic Tangent Activation Function (HTAF) are taken as the hidden layer activation functions:

$$purelin(x) = x \tag{1}$$

$$logsig(x) = \frac{1}{1+e^{-x}} \tag{2}$$

$$tansig = \frac{2}{1+e^{-2x}} - 1. \tag{3}$$

2.3. Applied statistics for efficiency assessment

The statistical approaches, including Root Mean Square Error (*RMSE*), Mean Absolute Error (*MAE*), modulus of designation (R^2), and mean absolute deviance percent ($\delta\%$) were used for GA-ANN method performance evaluation. Obviously, $\delta\%$ is a non-scale statistic. However, *RMSE* and *MAE* are on the identical scale of the outputs. Furthermore, R^2 specifies to what extent the numerical approaches can replicate the behold results. Thus, to obtain an even-handed arbitration on the efficiency of a numerical approach, it is highly better to regard the models altogether. Eqs. 4 to 7 present the equations of *RMSE*, *MAE*, R^2, and $\delta\%$.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(o_i - t_i)^2}{N}} \quad (4)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|o_i - t_i| \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(o_i - t_i)^2}{\sum_{i=1}^{N}(o_i - \bar{o_i})^2} \quad (6)$$

$$\delta\% = \frac{\sum_{i=1}^{N}|(o_i - t_i)|}{\sum_{i=1}^{N} o_i} \times 100. \quad (7)$$

Here, t_i denotes the behold diurnal pond surface amount, and o_i denotes the GA-ANN pond surface estimations. In addition, *N* shows the quantity of examining the specimen.

3. Results

The present research investigated the use of hybrid GA-ANN method of simulating the daily pond surface. The GA-ANN performance was examined in three different input combinations. The input combinations were constituted via the past daily pond surfaces. On the basis of Table 2, the considered input combinations are the last day (L^{-1}), the last 2-days (L^{-1} and L^{-2}) and the last 3-days (L^{-1}, L^{-2} and L^{-3}) data on pond surfaces. The GA-ANN is used to simulate the next 3- and 5-day pond water surfaces. According to Table 2, the input combination formed from the last 2-day pond water surface data (L^{-1}, L^{-2}) contributed to the most accurate GA-ANN model in both the next 3- and 5-day simulations. Table 2 indicates that the GA-ANN model contributes to an acceptable performance in both the next 3- and 5-day predictions. However, when we want to perform the next 3-day predictions and, then, to perform the next 5-day predictions, the *RMSE* increased about two times.

Table 2 - Evaluation of the GA-ANN model performance in the next 3- and 5-day predictions with different input compositions.

Day ahead prediction	Input combinations	GA-ANN		
		RMSE	*MAE*	*δ%*
3	L^{-1}	0.2311	0.1413	0.0289
	L^{-1}, L^{-2}	0.1798	0.0913	0.0187
	L^{-1}, L^{-2}, L^{-3}	0.1873	0.0915	0.0187
5	L^{-1}	0.3598	0.2003	0.0410
	L^{-1}, L^{-2}	0.3102	0.1617	0.0331
	L^{-1}, L^{-2}, L^{-3}	0.3657	0.1640	0.0336

Since obtaining the most proper input combinations, two functions, namely LAF (Eq. 2) and HTAF (Eq. 3) were examined in order to find the optimal hidden layer activation function. The linear activation function (Eq. 1) was applied to the output layer in all models. Table 3 presents the results obtained by investigating the activation function. According to the table, the next 3-day model shows better performance by using the HTAF in the hidden layer. In contrast, the next 5-day model has the better performance by using the LAF.

Table 3 - Performance of GA-ANN models in next 3- and 5-day predictions via different transition functions. Values in bold indicate the best results.

Day-ahead prediction	Transfer function		Selected input combination	GA-ANN		
	Hidden layer	Output layer		**RMSE**	**MAE**	*δ*%
3	logsig	purelin	L^{-1}, L^{-2}	0.1798	0.0913	0.0187
	tansig	**purelin**	L^{-1}, L^{-2}	**0.1751**	**0.0910**	**0.0181**
5	**logsig**	**purelin**	L^{-1}, L^{-2}	**0.3102**	**0.1617**	**0.0331**
	tansig	purelin	L^{-1}, L^{-2}	0.3160	0.1639	0.0335

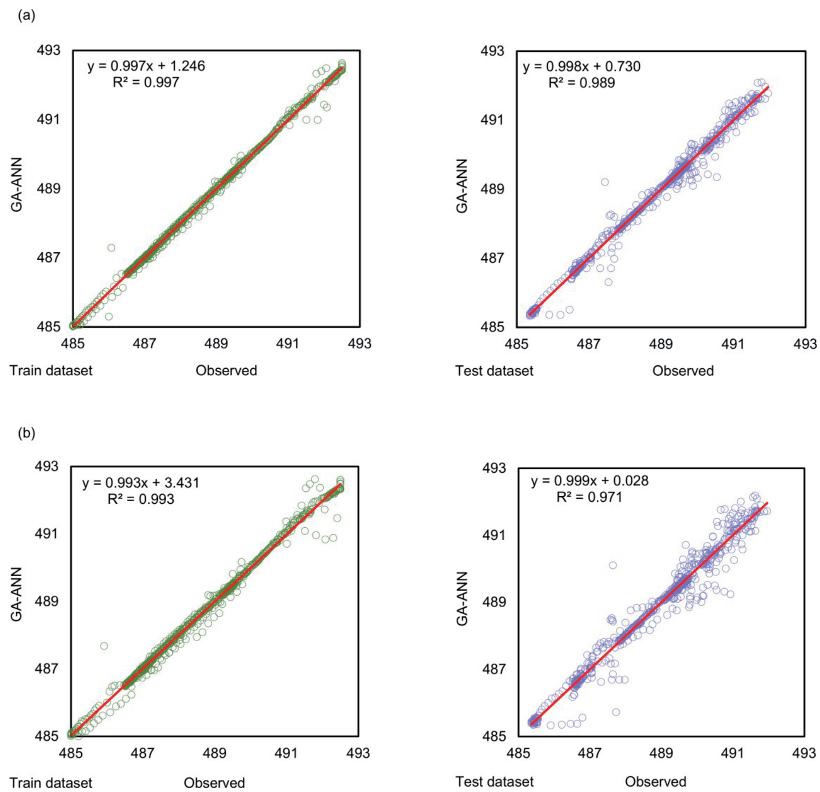

Fig. 3 - Scatter plot of GA-ANN models for the next 3- and 5-day estimation models in train and test data sets. In these figures, the trend line by the equation of $y = ax + b$ and R^2 is shown. Comparison between the results of the 3- and 5-day ahead GA-ANN prediction, indicates that by moving away from the 3-day ahead prediction to the 5-day ahead prediction there is no significant decrease in R^2.

Fig. 3 presents the spread plot of the most appropriate GA-ANN models in the training and testing data sets for the next 3- and 5-days in the hidden layer, and the next 5-day model uses the LAF. Both the next 3- and 5-day prediction models use the last 2-day pond water surface data. In the figure, the horizontal axis represents the observed pond water surfaces, and the vertical axis shows the GA-ANN predicted pond water surfaces. In the current paper, the training samples are presented by the green scatters, and the test samples by the blue scatters. The trend line shown by the red line has the equation of $y = ax + b$. In this equation, the closer a to one and closer b to zero indicate the more homogeneous dispersion of scatters surrounding the exact line. Based on the figure, by moving from the next 3- to 5-day prediction, the amount of R^2 remains almost constant. Furthermore, a comparison of the GA-ANN results in training and test data sets shows that the performance of the two models are close to each other, and there exists no over-training occurred.

A hydrograph is a graph or plot that shows the rate of water flow with time, given a specific point or cross-section. These graphs are often used to evaluate pond water runoff on a particular site considering a development project. Fig. 4 shows the pond water surface hydrographs relevant to the optimum GA-ANN models in the next 3- and 5-day prediction models. The top figures represent the pond water surface in total training and testing periods, and the bottom figures indicate the zoomed-out section of the upper figure rectangle. According to the figure, the GA-ANN can successfully model

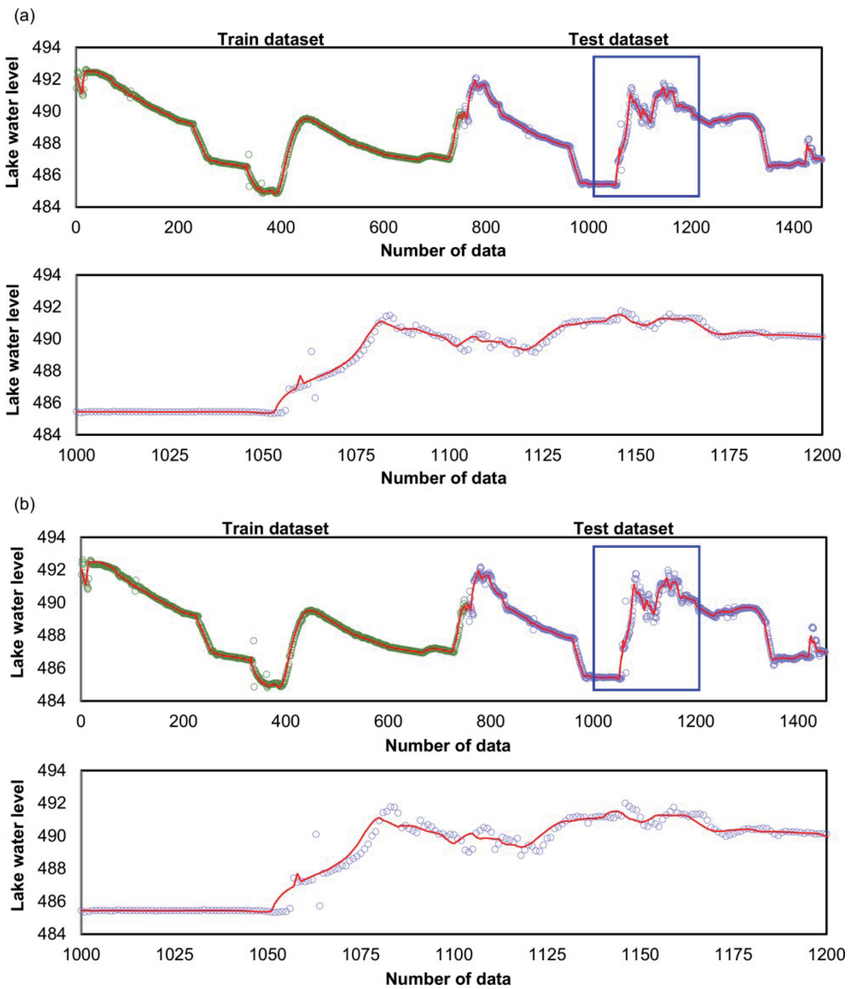

Fig. 4 - Pond water surface hydrographs relevant to the optimum GA-ANN models in 3- and 5-day prediction models:

a) pond water surface in total training and testing periods in the 3-day prediction model;
b) zoomed-out section of the upper figure rectangle;
c) pond water surface in total training and testing periods in the 5-day prediction model;
d) zoomed-out section of the upper figure rectangle.

the pond water surface. However, the next 3-prediction model results are nearby to the corresponding behold pond water surface amounts.

According to Fig. 4, the GA-ANN model could be used successfully in the pond water surface simulation. However, accurate prediction of GA-ANN entails finding the optimal situation of the input variables and the hidden layer activation function. The present paper shows that in modelling the next 3- and 5-day pond water surfaces, the optimal input variables are the data on the past 1- and 2-day predictions. GAs search a population of points in parallel, not just a single point. Therefore, it has the ability to avoid being trapped in local optimal solution like traditional methods, which search from a single point. GAs use probabilistic selection rules, not deterministic ones. The main disadvantages of GAs are that they cannot guarantee an optimal solution.

4. Conclusions

Precise estimation of the pond water surface is one of the utmost critical parameters in pond coast structure designing and water supply management. Past day data was used in the paper in order to simulate the future pond water surface. Modelling was performed via a novel hybrid GA-ANN model, which has the potential to provide an automatic adjustment of its hidden neuron number. The modelling substantially aimed to estimate the next 3- and 5-day pond water surfaces. Examining the different input combinations on the models, the paper shows that the past 2-day data can be regarded as the optimal input variables in modelling the both the next 3- and 5-day pond water surfaces. Activation function examination, performed on the models, indicate that the next 3-day modelling, performance is better when utilising the HTAF, and the next 5-day modelling performance is better when applying the LAF. Based on the results obtained from the optimal GA-ANN models in the next 3- and 5-day models with *RMSE* of 0.175 and 0.310, it could be derived that the GA-ANN model can be successfully applied in the pond water surface prediction problems. The main advantage of GA-ANN is its ability to find good quality solutions in a short time of computation. The main disadvantage is that it cannot guarantee an optimal solution.

Acknowledgements. The second author would like to acknowledge the Research Council of the University of Tehran. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Aguilar-Manjarrez J. and Nath S.S.; 1998: *A strategic reassessment of fish farming potential in Africa*. Committee for Inland Fisheries of Africa (CIFA), FAO, Roma, Italy, Technical Paper n. 32, 170 pp.

Al-Zubaidy R. and Shambour M.K.Y.; 2011: *Prediction of the Dead Sea water level using neural networks.* In: Proc., 4th International Symposium on Innovation in Information and Communication Technology (ISIICT), Amman, Jordan, pp. 147-154.

Avisse N., Tilmant A., Müller M.F. and Zhang H.; 2017: *Monitoring small reservoirs storage from satellite remote sensing in inaccessible areas.* Hydrol. Earth Syst. Sci., 21, 6445-6459, doi: 10.5194/hess-21-6445-2017.

Balaji S.A. and Baskaran K.; 2013: *Design and development of Artificial Neural Networking (ANN) system using sigmoid activation function to predict annual rice production in Tamilnadu.* Int. J. Comput. Sci. Eng. Inf. Technol., 33, 13-31, doi: 10.5121/IJCSEIT.2013.3102.

Barbosa de Alencar D., de Mattos Affonso C., Limão de Oliveira R.C., Moya Rodríguez J.L., Cabral Leite J. and Reston Filho J.C.; 2017: *Different models for forecasting wind power generation: case study*. Energies 2017, 10, 1976, doi: 10.3390/en10121976.

Biancamaria S., Lettenmaier D.P. and Pavelsky T.M.; 2016: *The SWOT mission and its capabilities for land hydrology*. Surv. Geophys., 37, 307-337, doi: 10.1007/s10712-015-9346-y.

Busker T., de Roo A., Gelati E., Schwatke C., Adamovic M., Bisselink B., Pekel J.F. and Cottam A.; 2019: *A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry*. Hydrol. Earth Syst. Sci., 23, 669-690, doi: 10.5194/hess-2018-21.

Gavin H.P.; 2020: *The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems*. Department of Civil and Environmental Engineering, Duke University, Durham, NC, USA, 19 pp.

Gullu H.; 2013: *On the prediction of shear wave velocity at local site of strong ground motion stations: an application using artificial intelligence*. Bull. Earthquake Eng., 11, 969-997.

Huang C.L. and Hsu N.S.; 2013: *Real-time reservoir inflow hydrograph forecast during typhoon warning-tseng-wen reservoir basin as a case study*. Taiwan Water Conservancy, 61, 1-21.

Imani M., You R.J. and Kuo C.Y.; 2014: *Caspian Sea level prediction using satellite altimetry by artificial neural networks*. Int. J. Environ. Sci. Technol., 11, 1035-1042.

Jothiprakash V. and Magar R.B.; 2012: *Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent techniques using lumped and distributed data*. J. Hydrol., 450-451, 293-307.

Kavitha M.M. and Naidu K.B.; 2011: *ANN and fuzzy logic models for the prediction of groundwater level of a watershed.* Int. J. Comput. Sci. Eng., 3, 2523-2530.

Khan Z.H., Tasnim Sharmin Alin T.S. and Hussain M.A.; 2011: *Price prediction of share market using Artificial Neural Network (ANN).* Int. J. Comput. Appl., 22, 42-47, doi: 10.5120/2552-3497.

Lo Y.C., Chang F.J. and Shieh M.C.; 2014: *Real-time multi-step ahead reservoir inflow forecasts by artificial neural networks.* Taiwan Water Conservancy, 62, 26-34.

Malekmohamadi I., Bazargan-Lari M.R., Kerachian R., Nikoo M.R. and Fallahnia M.; 2011: *Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction.* Ocean Eng., 38, 487-497.

Mitchell M.; 1995: *Genetic algorithms: an overview.* Complexity, 1, 31-39, doi: 10.1002/cplx.6130010108.

Nagababu G., Simha R.R., Naidu N.K., Kachhwaha S.S. and Savsani V.; 2016: *Application of OSCAT satellite data for offshore wind power potential assessment of India.* Energy Procedia, 90, 89-98.

Nath S.S. and Bolte J.B.; 1998: *A water budget model for pond aquaculture.* Aquacult. Eng., 18, 175-188.

Nitsure S.P., Londhe S.N. and Khare K.C.; 2014: *Prediction of sea water levels using wind information and soft computing techniques.* Appl. Ocean Res., 47, 344-351.

Puscasu R.M.; 2014: *Integration of artificial neural networks into operational ocean wave prediction models for fast and accurate emulation of exact nonlinear interactions.* Procedia Comput. Sci., 29, 1156-1170.

SCS (Soil Conservation Services); 1985: *National engineering handbook, section 4: hydrology.* U.S. Department of Agriculture, Soil Conservation Service, Washington, DC, USA, 127 pp.

Song S.S. and Li R.; 2014: *Method of wind power prediction based on ANN and improvements.* Appl. Mech. Mater., 521, 143-146, doi: 10.4028/www.scientific.net/AMM.521.143.

Tian J. and Gao M.; 2009: *Network intrusion detection method based on high speed and precise genetic algorithm neural network.* In: Proc., International Conference on Networks Security, Wireless Communications and Trusted Computing, IEEE 2009, Vol. 2, pp. 619-622, doi: 10.1109/NSWCTC.2009.228.

Zhang G., Patuwo B.E. and Hu M.Y.; 1998: *Forecasting with artificial neural networks: the state of the art.* Int. J. Forecasting, 14, 35-62.

Corresponding author: Mohammad Ali Riahi
Department of Solid Earth Physics, University of Tehran
North Kargar Ave., Tehran, Iran
Phone: +98 21 61118219; fax: +98 21 88001115; e-mail: mariahi@ut.ac.ir