Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 74 (2015) 1007 – 1014

Energy Procedia

International Conference on Technologies and Materials for Renewable Energy, Environment and Sustainability, TMREES15

River flow model using artificial neural networks

Imen AICHOURI[a], Azzedine HANI[b], Nabil BOUGHERIRA[b], Larbi DJABRI[b], Hicham CHAFFAI[b], Sami LALLAHEM[c]

[a]Laboratory of Water Ressource and Soustainble Développement, National School of Mines and Metallurgy ENSMM-Annaba ex CEFOS Chaiba BP 233 RP Annaba,
[b]Water Ressource and Soustainble Développement Laboratory,university of Badji Mokhtar, Annaba 23000,Algeria
[c]Laboratory of Civil Engineering & Geo-Environment (LGCgE), University of Lille 1, Villeneuve d'Ascq, 59655, France.

Abstract

The use of artificial neural networks (ANNs) is becoming increasingly common in the analysis of hydrology and water resources problems. In this research, an ANN was developed and used to model the rainfall-runoff relationship, in a catchment located in a semiarid and Mediterranean climate in Algeria. The performance of the developed neural network-based model was compared against multiple linear regression-based models using the same observed data. It was found that the neural network model consistently gives superior predictions. Based on the results of this research, artificial neural network modeling appears to be a promising technique for the prediction of flow for catchments in semi-arid and Mediterranean regions. Accordingly, the neural network method can be applied to various hydrological systems where other models may be inappropriate.

Peer-review under responsibility of the Euro-Mediterranean Institute for Sustainable Development (EUMISD)

Keywords: Artificial neural network; modeling multiple regressions; semi arid climate; Rainfall-runoff; Catchment, MLP, Algeria.

* Corresponding author. Tel.: +213-388-765-62; fax: +213-388-765-62.
E-mail address: haniazzedine@yahoo.fr

1. Introduction

The ANN models are powerful prediction tools for the relation between rainfall and runoff parameters. The results will support decision making in the area of water resources planning and management. Besides, they assist

Peer-review under responsibility of the Euro-Mediterranean Institute for Sustainable Development (EUMISD)
doi:10.1016/j.egypro.2015.07.832

urban planners and managers undertake the necessary measures to face the bad productions. Thus, they help avoid losses in public and private properties, and health and ecological hazards that are likely to occur due to flooding.

In addition, the ANN models have been used increasingly in various aspects of science and engineering because of its ability to model both linear and nonlinear systems without the need to make any assumptions as are implicit in most traditional statistical approaches. In some of the hydrologic problems, ANNs have already been successfully used for river flow prediction (Riad et al.) [1], 2004; Lallahem et al., 2005) [2], for rainfall-runoff process (Smith and Eli,) [3] for the prediction of water quality parameters (Maier and Dandy, 1996) [4]. In addition, ANNs are applied for prediction of evaporation (Sudheer, 2002) [5], for rainfall-runoff forecasting (Minns and Hall, 1996) [6], for prediction of flood disaster (Wei *et al.,* 2002) [7], and for river time series prediction (Hu et al., 2001) [8]. In these hydrological applications, a multilayer feed-forward backpropagation algorithm is use (Lippmann, 1987, Riad et al., 2004) [9]. It usually is composed of a large number of interconnected nodes, arranged in an input layer, an output layer, and one or more hidden layers. The transfer function selected for the network was the sigmoid function. The aim of this paper is to model the rainfall-runoff relationship in the Seybouse catchment located in the northern part of Algeria using a black box type model based on ANN methodology. The Seybouse River Basin is located in northeastern Algeria. With a total area of 6,471 km2, the basin extends over 68 municipalities and 7 prefectures. The Seybouse River and its tributaries are vital for sustaining the majority of economic activities in the region.

2. The artificial neural networks approach

2.1. The basics

An ANN is a computational approach inspired by the human nervous system. It is based on theories of the massive interconnection and parallel processing architecture of biological neural systems. The main theme of ANN research focuses on modeling of the brain as a parallel computational device for various computational tasks that were performed poorly by traditional serial computers.

ANNs have a number of interconnected processing elements (PEs) that usually operate in parallel and are configured in regular architectures. The collective behavior of ANN, like a human brain, demonstrates the ability to learn, recall, and generalize from training patterns or data. The advantage of neural networks is they are capable of modeling linear and nonlinear systems.

In this research, we use an MLP trained with a back propagation algorithm to predict the drainage basin runoff. The MLP consists of an input layer consisting of node(s) representing various input variable(s), the hidden layer consisting of many hidden nodes, and an output layer consisting of output variable(s). The input nodes pass on the input signal values to the nodes in the hidden layer unprocessed. The values are distributed to all the nodes in the hid-den layer depending on the connection weights Wij and Wjk between the input node and the hidden nodes. Connection weights are the interconnecting links between the neurons in successive layers.

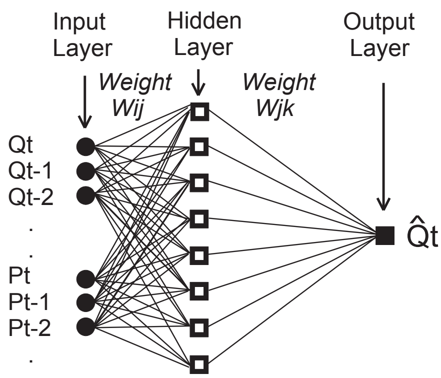


Fig. 1. Architecture of the neural network model in this study.

The architecture of the neural network used in this study and the schematic representation of a neuron are shown in Figure 1. Each node j receives incoming signals from every node i in the previous layer. Associated with each incoming signal (Xi) is a weight (Wij). The effective incoming signal (Sj) to node j is the weighted sum of all the incoming signals and bj is the neuron threshold value.

$$S_j = \sum_{i=1}^{n} X_i W_{ij} + b_j . \quad (1)$$

The effective incoming signal, Sj, is passed through a nonlinear activation function to produce the outgoing signal (yj) of the node. The most commonly used in this type of networks is the logistic sigmoid function. This transfer function is continuously differentiable, monotonic, symmetric, bounded between 0 and 1. It is expressed mathematically as:

$$f(S_j) = \frac{1}{1+e^{-Sj}} . \quad (2)$$

Criteria for model performance

In the present research, both statistical and graphical criteria were adopted to select the desired optimal network model. The statistical criteria consist of average squared of error (ASE), coefficient of determination (R2) and the mean absolute relative error (MARE). They are given by:

$$ASE = \sum_{i-1}^{n} (Qt_i - \hat{Q}t_i)^2 / n \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i-1}^{n} (Qt_i - \hat{Q}t_i)^2}{\sum_{i-1}^{n} (Qt_i - \overline{Q}t_i)^2} \quad (4)$$

$$MARE = \frac{\sum_{i-1}^{n} \left| \left(\frac{\hat{Q}t_i - Qt_i}{Qt_i} \right) \right|}{n} .100 \quad (5)$$

where Qti and $\hat{Q}t_i$ are respectively, the actual and predicted value of flow (normalized between 0 and 1), $\overline{Q}t_i$ is the mean of Qti values and n is the total number of data sets.

The R2 statistic measures the linear correlation between the actual and predicted flows values.

The ASE and MARE statistic measures are used to quantify the error between observed and predicted values. The optimal value for R2 is equal to 1.0 and for ASE and MARE is equal to 0.0.

The graphical performance indicator gives better results when the data pairs are closing to 45° line and the good superposition between the desired and calculated flow values in the training and testing phases.

For the data set considered in this research, the input variables as well as the target variables are first normalized linearly in the range of 0 and 1. This range is selected because of the use of the logistic function (which is bounded between 0.0 and 1.0) as the activation function for the output layer, i.e., equation (2). The normalization is done using the following equation.

$$\overline{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (6)$$

where $\overline{X}$ is the standardized value of the input, Xmin and Xmax are respectively, the minimum and maximum of the actual values, in all observations and X is the original data set.

The main reason for standardizing the data matrix is that the variables are usually measured in different units. By standardizing the variables and recasting them in dimensionless units, the arbitrary effect of similarity between objects is removed.

3. The study catchments and database

In the present research, the flow and rainfall series observed in Seybouse basin at Mirbeck station (Flow) and Pont Bouchet (Rainfall) is analyzed using the ANN model. The Seybouse River Basin extends over an area of 6.471 km2, is located in the northern part of Algeria and has a permanent population of approximately 1.300.000 inhabitants (Figure 2). The River Basin lies within the territories of the wilayas of Guelma, El-Tarf (by Drean) and Annaba. It is limited in the north by the Mediterranean Sea, in the south by the Wilaya of Souk-Ahras, in the west by the Edough Massif, lake Fetzara, Ain Berda, and in the east by oued Mafragh.

The Seybouse River, of 240 km total length, is an important water source, used mainly for the irrigation of large agricultural plains, extending from the Guelma region up to the city of Annaba. Overall the basin extends over the administrative boundaries of 68 municipalities located in 7 wilayas (prefectures). Its water resources are vital for sustaining the majority of economic activities in the region.

The climate of the basin varies from typical Mediterranean along the coast to semi-arid. The mean annual precipitation varies between 700 mm and 400 mm, reaching a monthly in the range of 90- 120 mm in December-January. Minimum and maximum temperatures are observed in December-January (less than 10°C) and in July or August (between 25 °C and 30 °C) respectively. The average annual infiltration is about 162 mm whereas surface run-off accounts for 79 mm/yr., municipalities and 7 wilayas (prefectures).

The Rainfall and Runoff daily data was used for model investigation. The data contains information for a period of eighteen years (1986-2003). The entire database is represented by 5400 daily values of rainfall and runoff pairs. The ANN model was trained using the resulting runoff and rainfall daily data. The database was collected by the National Agency for Water Resources (ANRH).

The input vector is represented by rainfall and runoff values for the preceding seven days, (i.e., t - 1, t - 2, t - 3, t - 4, t - 5, t - 6, t - 7) as well as the rainfall value expected for day t. Accordingly, the output vector represents the expected runoff value for day t ($\hat{Q}t$).

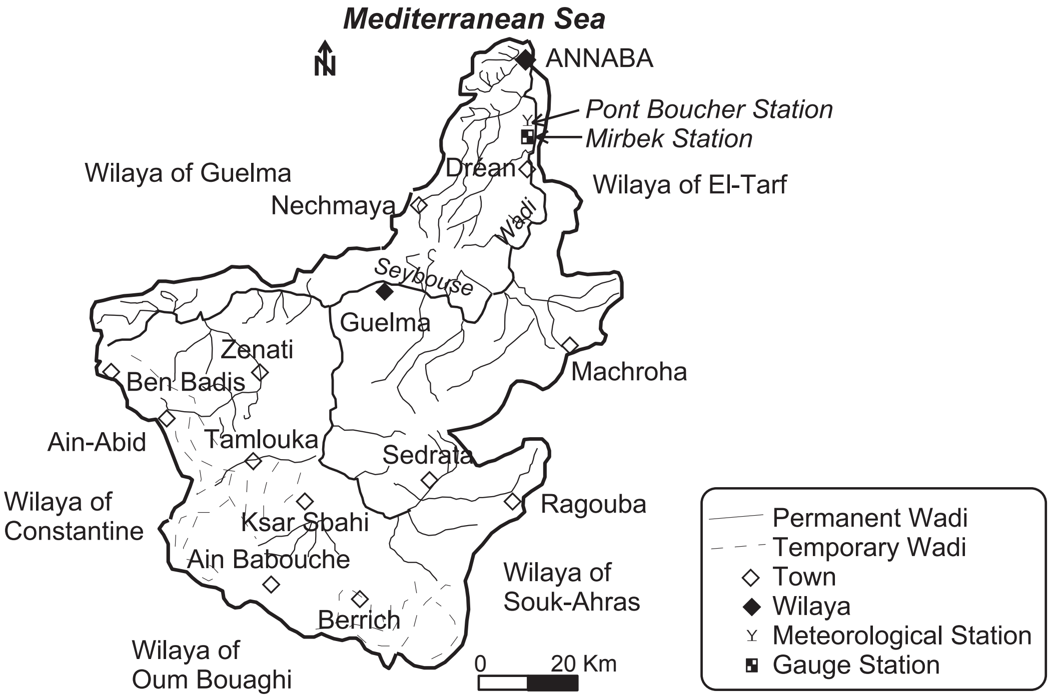


Fig. 2. Location of the Seybouse River Basin

4. Test results and discussions

The database compiled represents eighteen years daily sets of rainfall-runoff values for the Seybouse River basin. In this paper, we used the data for the two last years (2002 and 2003) for model testing, while the other remaining data (1986 to 2001) was used for model training/calibration. The training phase of ANN model was terminated when the average squared error (ASE) on the testing databases was minimal.

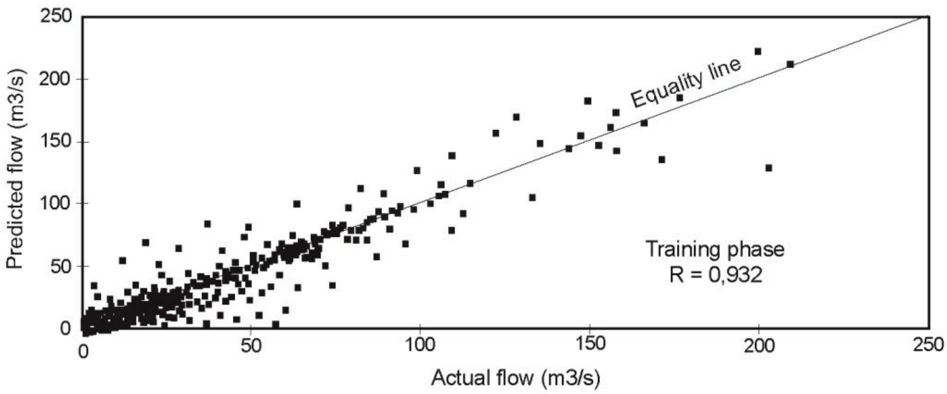
The objective of the training process is to reach an optimal solution based on some per-formance measurements such as ASE, coefficient of determination known as R-square value (R2), and the MARE.

Therefore, required ANN model was developed in two phases: training (calibration) phase, and testing (generalization or validation) phase.

In the training phase, a larger part for database (sixteen years) was used to train the net-work and the remaining part of the database (two years) is used in the testing phase. Testing sets are usually used to select the best performing network model. In this research, the ANN was optimal at 540 iterations with 10 hidden nodes. The corresponding accuracy measures of this network model on testing and training data are given in the following table (Table 1). Generally, accuracy measures on training data are better than those on testing data.

Table 1. Statistical accuracy measures of this network model at testing and training phases.

	ASE	R	MARE
Training phase	0.00009	0.932	1.132%
Testing phase	0.00001	0.902	1.453%



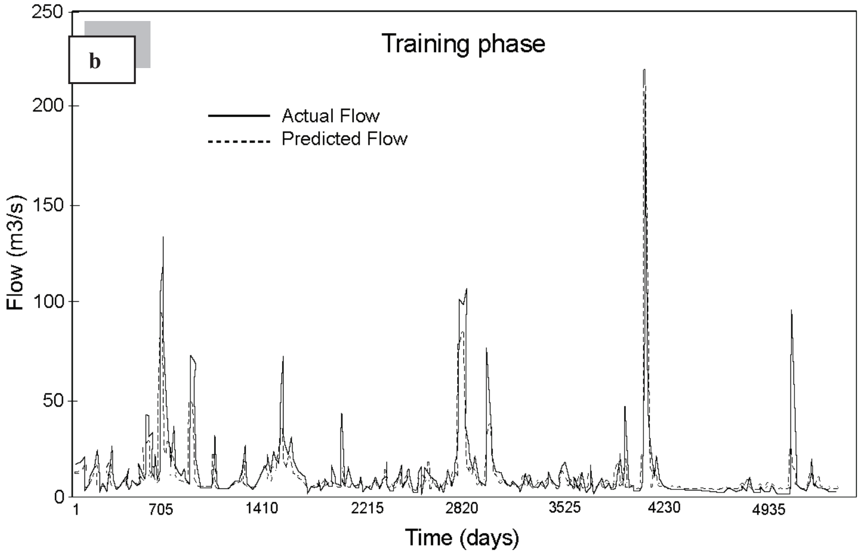


Fig. 3a-b. Comparison between the actual and ANN predicted flow value at training phase.

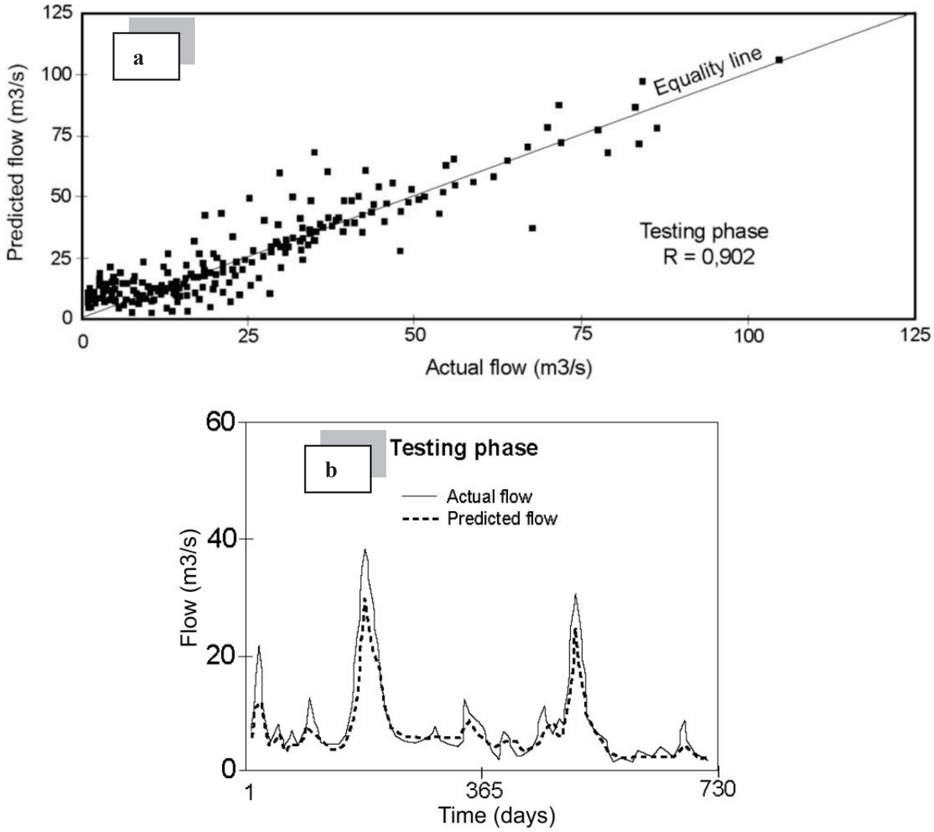


Fig. 4a-b. Comparison between the actual and ANN predicted flow value at testing phase.

Table 2. Statistical parameters of the predicted and actual flow at training and testing phases (a) Training Phase, (b) Testing Phase.

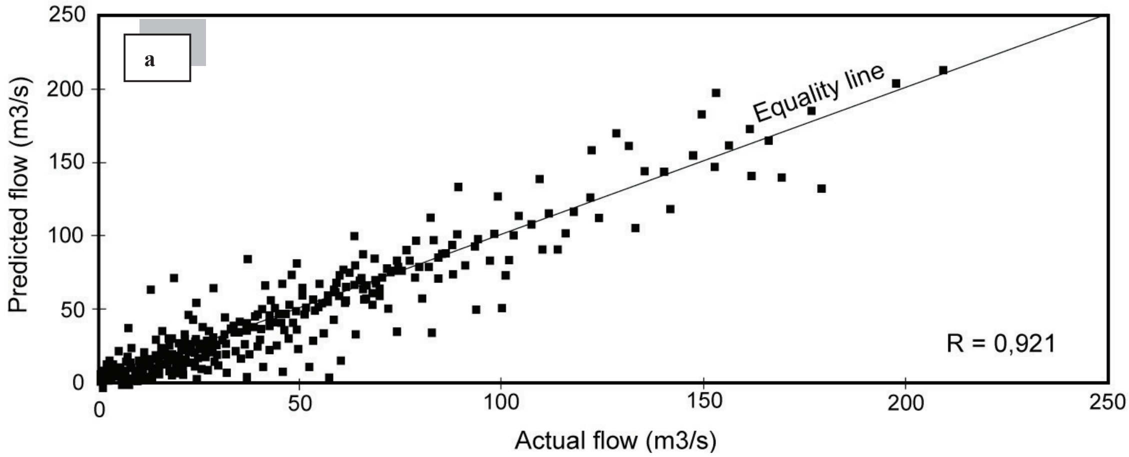
	Training phase	Testing phase		
Statistical parameters	Actual flow (m3.s-1)	Predicted flow (m3.s-1)	Actual flow (m3.s-1)	Predicted flow (m3.s-1)
Average	24.84	22.34	09.89	09.57
Standard deviation	23.2	21.7	16.78	15.67
Minimum	00.00	03.18	04.34	04.30
Maximum	220.38	217.34	39.03	32.25
Coefficient of variation	03.86	05.65	02.31	02.73

The comparison between the predicted and actual flow values at training and testing phases show excellent agreement with the R2 are respectively 0.932 and 0,902. Note that, data pairs closer to the 45° line represent better prediction cases. The good performance and convergence of the model are illustrated in Figure 1.

The statistical parameters of the predicted and actual values of flow for the entire data-base are practically identical (Table 2).

In order to evaluate the performance of the ANN, the multiple linear regression (MLR) technique was applied with the same data sets used in the ANN model.

Figure 2 shows the comparative results obtained by MLR technique. The R2 values for MLR and ANN models are presented in Figure3a-b. In conclusion, the ANN approach gives much better prediction than the traditional method (MLR).



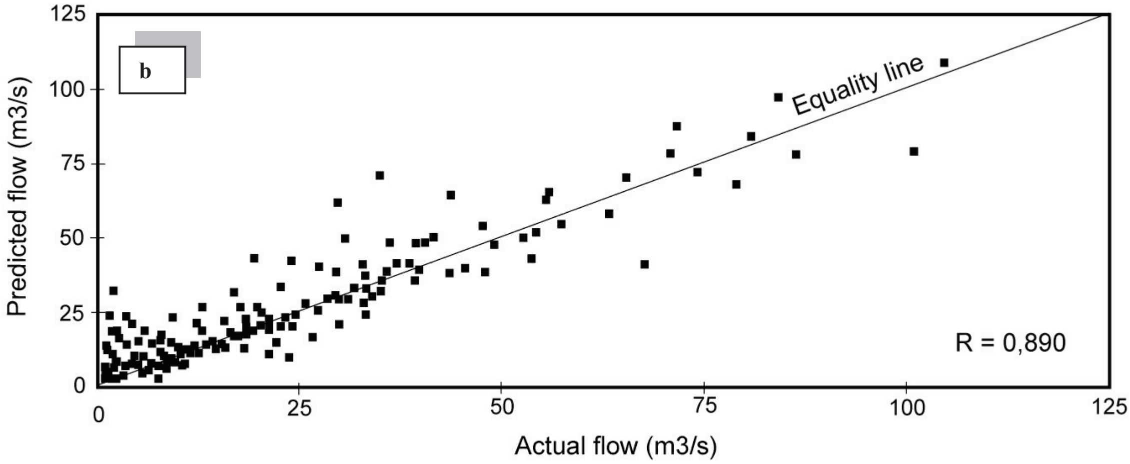


Figure 5a-b. Comparison between the actual and predicted flow values by multiple linear regression (MLR), a) ANN, b) MLR.

5. Conclusion

The artificial neural network models show good ability to model hydrological process. They are useful and powerful tools to handle complex problems compared with the other traditional models. In this research, the results obtained show that the artificial neural net-works are capable of model rainfall-runoff relationship in the semiarid and Mediterranean regions in which the rainfall and runoff are very irregular, thus, confirming the general en-hancement achieved by using neural networks in many other hydrological fields. The results and comparative study indicate that the artificial neural network method is more suitable to predict river runoff than classical regression model. The ANN approach could provide a very useful and accurate tool to solve problems in water resources studies and management.

References

[1] Riad S., Mania J., Bouchaou L., Najjar Y. (2004) Rainfall-Runoff Model Using an Artificial Neural Network Approach. Mathematical and Computer Modelling 40 (2004) 839-846.
[2] Lallahem S., Mania J., Hani A., Najjar Y. On the use of neural networks to evaluate groundwater levels in fractured media. Journal of Hydrology 307; 2005.p. 92-111.
[3] Smith I. and Eli R.N. (1995) Neural-network models of rainfall-runoff process, Journal of Water Resources Planning and Management 121 (6), 499-508.
[4] Maier H.R. and Dandy G.C. (1996) The use of artificial neural network for the prediction of quality water parameters, Water Resources Research 32 (4), 1013-1022.
[5] Sudheer K.P., Gosain A.K., Mohana Rangan D. and Saheb S.M. (2002) Modelling evaporation using an artificial neural network algorithm, Hydrol. Process. 16, 3189-3202.
[6] Minns A.W. and Hall M.J. (1996) Artificial neural networks as rainfall-runoff models, Hydrological Sciences Journal 41 (3), 399-418.
[7] Wei Y., Xu W., Fan Y. and Tasi H.T. (2002) Artificial neural network based predictive method for flood disaster, Computers & Industrial Engineering 42, 383-390.
[8] Hu T.S., Lam K.C. and Ng S.T.River flow time series prediction with a range-dependent neural network. Hydrological Science Journal 46 (5), 2001.p. 729-745.
[9] Lippmann R.P. (1987) An introduction to computing with neural nets, IEEE ASSP Magazine, 4-22.