Chiang Mai J. Sci. 2016; 43(6) : 1358-1367
http://epg.science.cmu.ac.th/ejournal/
Contributed Paper

Application of Radial Basis Function Neural Networks for Modeling Rainfall-runoff Processes: A Case Study of Semenyih River Catchment, Malaysia

Nadeem Nawaz* [a,c], Sobri Harun [a], Rawshan Othman [b] and Arien Heryansyah [a]

[a] Department of Hydraulics and Hydrology, Faculty of Civil Engineering, Universiti Teknologi Malaysia,Skudai, 81310 Johor Bahru, Malaysia.

[b] Petroleum Department, Koya Technical Institute, Erbil Polytechnic University, 44001 Erbil, Kurdistan Regional Government, Iraq.

[c] Faculty of Water Resources Management, Lasbela University of Agriculture, Water and Marine Sciences, 90150 Uthal, Balochistan, Pakistan.

* Author for correspondence; e-mail: nn.engr96@yahoo.com

Received: 20 February 2016

Accepted: 8 July 2016

ABSTRACT

The gradual transformation of arable lands into urbanized environments in built-up areas is common in fast developing countries like Malaysia. Such changes have a large effect on hydrologic processes in the catchment area, which eventually results in an increase of both the magnitude and frequency of floods in urban areas. Therefore there is a great need of reliable rainfall-runoff models that are able to accurately estimate the discharge for a catchment. So far various physically-based models have been developed to capture the rainfall-runoff process, but the drawback has been the estimation the several numbers of parameters which is quite difficult and time consuming. Recently, artificial intelligence tools are being used because of their capability of modeling complex nonlinear relationships. These tools have been widely used in hydrological time series modeling and prediction. Radial basis function neural network (RBFNN) is a popular artificial intelligence technique that is well used in hydrological modeling. In this study, 30 extreme rainfall-runoff events were extracted from twelve years of hourly rainfall and runoff data. An input selection method based on correlation analysis and mutual information was developed to identify the proper input combinations of rainfall and discharge antecedents. The results obtained by RBFNN model were then compared with a traditionally used statistical model known as auto-regressive moving average with exogenous inputs (ARMAX), as a bench mark. Results showed that RBFNN performance is superior then the traditional statistical model and has good potential to be used as a reliable rainfall-runoff modeling tool.

Keywords: rainfall-runoff modeling, RBFNN, ARMAX

1. INTRODUCTION

Rapid population growth, urbanization, and industrialization in many parts of the world have increased the demand for water. The increase in water demand has resulted in altered watersheds and river systems and it has become critical to plan and manage water resources systems intelligently. Understanding the dynamics of rainfall-runoff process is one of the most important problems in hydrology in order to predict or forecast streamflow for purposes such as water supply, flood control, irrigation, drainage, recreation, power generation and so on. This rainfall-runoff relationship is known to be highly non-linear and complex due to the involvement of many parameters such as watershed geomorphology, initial soil moisture, infiltration, evapotranspiration, distribution and duration of rainfall, temperature and land use. Reliable estimations of streamflow generated from catchments are known to be the essential information sets which help policy makers to make decisions on planning and water resources management. The characteristics of the streamflow time series that have direct impact on planning and water resources management can include the spatial and temporal variability of flows, sequencing of flows on different time steps, seasonal distribution and characteristics of high and low flows [1]. Rainfall-Runoff models are commonly used to describe the hydrological behavior of the catchment. The processes and relationships between the rainfall and runoff are required to have good understanding for the satisfactory drainage and river management purposes [2]. Researchers have so far developed many rainfall-runoff models. Some of them are physical and mathematically based which required significant efforts in handling and data input while on the other hand the black box models attempt to describe the Rainfall-Runoff relationship without explicit consideration of internal hydrological processes [3]. The physical models provide reasonable accuracy but the implementation and calibration of such models can present various problems such as requiring sophisticated mathematical tools, a significant amount of calibration data and some degree of expertise and experience with the model [4]. In last few decades, artificial neural networks (ANN) have shown promising capability in modeling and forecasting non-linear hydrological time series. ANN offers an effective approach for handling large amounts of dynamic, non-linear and noisy data, especially when the underlying physical relationships are not fully understood. This makes them well suited to time series modeling problems of a data-driven nature. In general, the ANN modeling technique does not require a priori knowledge of the underlying physical process involved in transformation of rainfall into runoff [5]. ANN was introduced by [6], inspired by an aspiration to understand human brain and imitate it's working. Over the last few decades ANN has undergone an enormous renaissance because of the advancement in sophisticated algorithms. A remarkable progress occurred with the discovery of a scientifically demanding theoretical context of ANN i.e., back-propagation algorithm [7]. Since the early nineties, ANN applications can be found in various hydrological studies such as rainfall-runoff modeling [8, 9], ground water modeling [10, 11], stream flow forecasting [12, 13], suspended sediment load [14, 15], rainfall forecasting [16, 17], reservoir operations [18, 19] and so on. The goal of this study is to check the capabilities of Radial Basis Function Neural Network (RBFNN) as an alternative rainfall-runoff modeling tool for the Semenyih River catchment.

2. MATERIALS AND METHODS

2.1 Artificial Neural Network

ANN is an immensely parallel-distributed information processing system that has certain performance features like biological neural networks of the human brain [20]. Approximately 100 billion neurons are present in the brain, which interconnects over electro-chemical signals. The construction of ANN was an attempt to represent the computational reflection of brain behavior since it is not comparable due to the presence of larger number of neurons with multiple functions in a human brain than those in ANN. An ANN structure is a network that consists of artificial neurons, called “nodes” which are interconnected to one another. Based on their connection strength a value is assigned known as inhibition or excitation which is maximally -1.0 and +1.0 respectively. Higher value reflects strong connection and the lower value indicate the weakness of the connection. A transfer function is built after each node’s design. The transfer functions commonly used are Sigmoid, Gaussian, Piecewise Linear and Unit step (threshold). ANN consists of three types of nodes known to be as input nodes, hidden nodes and output nodes as can be seen in Figure 1. The function of input nodes is to collect the information which is expressed numerically and represented as activation values. Every node is assigned a value where the higher value characterized as higher activation. All collected information is further conveyed to hidden nodes. Activation value is then forwarded by one node to another on the basis of inhibition or excitation, transfer functions and connection strengths (weights). Every node sums the collected values upon receiving and modifies them on the basis of activation function. The same process moves throughout the network until it approaches to the output nodes. At the end of process the output nodes redirect collected information in an informative manner to the desired output. ANN is the network of simple computational elements that are able to adapt to an information environment. This adaptation is realized by adjustment of the internal network connections through applying a certain algorithm. Thus, ANN is able to uncover and approximate relationships that are contained in the data that is presented to the network. Since 1993, hydrologists have put their efforts to having clear understandings on the capabilities of ANN to model rainfall-runoff processes and also analyzed the factors influencing ANN performances. Several algorithms have been developed by researchers and employed for different applications. The hydrologists’ interest remained to use those algorithms to simulate complex hydrological time series. This study presents the application of radial basis function based ANN to simulateextreme rainfall-runoff events.

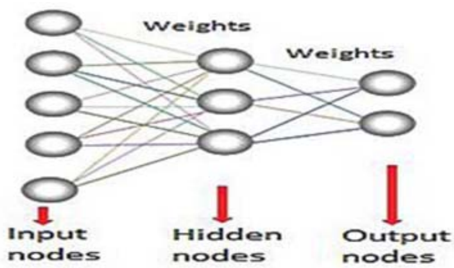

Figure 1. Basic structure of artificial neural networks.

2.2 Radial Basis Function (RBF)

This network is consisting of three layers: 1) input layer, 2) hidden layer and 3) output layer. For each input layer, one neuron resembles to every forecaster variable and concerned with firm variables (n-1) neurons are entertained where n represents the number of categories. The variable numbers of neurons are placed in hidden layer as can be seen Figure 2. Each neuron contains RBF concentrated to the similar dimensions with those forecaster variables. The output layer comprises of a weighted sum of outputs obtained from the hidden layer.

$$f(x) = \sum_{j=1}^{m} w_j h_j(x) \quad (1)$$

$$h(x) = \exp\left[-\frac{(x - c)^2}{r^2}\right] \quad (2)$$

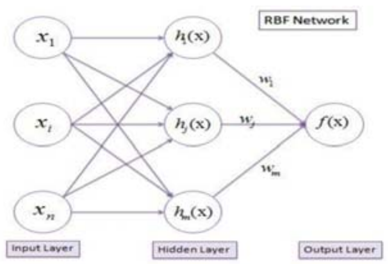

Figure 2. Radial basis function neural network.

where h(x) represents Gaussian activation function and r is the radius or standard deviation and c represents the center or average taken from the input space defined separately at each RBF unit. The calibration process bases on adjustment of the parameters associated with the network to produce a set of input-output patterns. Mainly three parameters are involved in calibration process; (1) weight "w" between the hidden and output nodes, (2) center "c" of each neuron of hidden layer and (3) unit width "r". Many clustering algorithms are being used for determining the RBF unit centers for example, K-means clustering algorithm. The input variables or nodes determine the set of clusters and theirr-dimensional centers which converts the centers into the RBF units. The number of clusters "H" is a design parameter, which determines the number of nodes in the hidden layer. The following procedure is monitored by the K-means clustering algorithm:

a) Adjust the center of all independent clusters to several arbitrarily chosen training patterns.

b) Allocate all the training patterns to the nearby clusters. This is done with calculation of the Euclidean distances among the training patterns and the cluster centers.

c) After assigning all training patterns it determines the average position for each cluster center to get new cluster centers.

d) Steps "a" and "b" are repeated until the cluster centers stop changing after the succeeding iterations.

After the establishment of RBF centers, the unit width "r" of every RBF unit can be found by using the K-nearest neighbor's algorithm. After selecting the number of K in each center, the nearest K center originates, followed by calculated root mean squared spaceamong the current cluster center and its K nearest neighbor. The values are then selected for the unit width, therefore if it is assumed the cluster center is "c_j" and then the r value is:

$$r_j = \sqrt{\frac{\sum_{i=1}^{k} (c_j - c_i)^2}{k}} \quad (3)$$

A standard value of K is 2, where "c_i" is set to be the average distance from the two nearest neighboring cluster centers. By means of the linear mapping, w vector is calculated using the output vector (y) and the design matrix H.

$$y = wH \tag{4}$$

$$w = (H'H)H'y \tag{5}$$

In literature there are few successive applications of RBFNN in rainfall-runoff modeling. The study performed by Mason [21] is one of the earlier successful applications of RBFNN to simulate the complex relationship. Lin [22] employed RBFNN in Fei-Tsui Reservoir Watershed in northern Taiwan and reported the RBFNN can be successfully applied to build the relationship between rainfall and runoff.

2.3 ARMAX Model

The conventional regression method uses linear or piecewise-linear framework for the forecasting function. In this mechanism, linear combination determines the functional relationship that supplies the requested forecast, which assumes linear relationship without adequate reasons. An ARMAX (p-d-q;X) model can be explicitly represented as;

$$Y_t = \mu + p1y_{t-1} + p2y_{t-2} + \ldots + p_p y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \beta_k x_{t-k} + \varepsilon_t - q_1\varepsilon_{t-1} - q_2\varepsilon_{t-2} - \ldots - q_q\varepsilon_{t-q} \tag{6}$$

where μ is the constant term, β parameters are the regressors for lagged distributed x explanatory variables, p parameters are the autoregressive parameters for lagged distributed y exogenous dependent variables, q parameters are the moving average parameters for lagged distributed ε stochastic variables, and d is thedegree of differencing. The same lag structure is not necessarily applied to y_t and ε_t; which is required in the autoregressive distributed lag models. ε_t is the serially undistributed constant variance random variable. Furthermore the input-output functional relationship between observed phenomena and its underlying cause are more often not stationary in conventional regression like in the case of Rainfall Runoff processes. Hence the conventional regression approach produces averaged results as it does not have enough adaptability to identify inherent spatio-temporal variation. ARMAX linear models with their improved efficiency for time series analysis have been developed by Box and Jenkins [23] in 1970. ARMAX model is available in the MATLAB environment. This model is frequently used because of producing acceptable prediction in time series modeling.

2.4 Study Area and Data Used

Semenyih River catchment is located at the central part of Peninsular of Malaysia. Malaysia receives approximately 2500mm rainfall per annum. The northeast monsoon contributes heavy rainfall events in the eastern part of Peninsular Malaysia that occur from November to January. The western part of Peninsular Malaysia receives southwest monsoon from May to September. Semenyih River catchment is in the state that also experiences inter-monsoon period that occurs during the transaction of southeast and southwest monsoon during April and October. Figure 3 shows the location map of Semenyih catchment.The catchment (with an area of 225 km^2) has four rainfall stations and one discharge station located at the outlet. The locations of the rainfall stations can be seen in Figure 4. The Department of Irrigation and Drainage (DID), Malaysia is the responsible institution to maintain the hydrological records. The rainfall data of the presently active stations was arranged from DID. Twelve years

(2002-2013) of hourly rainfall and runoff data was provided from the department of irrigation and drainage (DID), Malaysia.

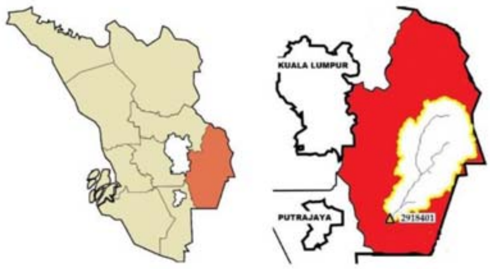

Figure 3. Location map Semenyih River catchment, Hulu Langat District, Selengor state, Malaysia.

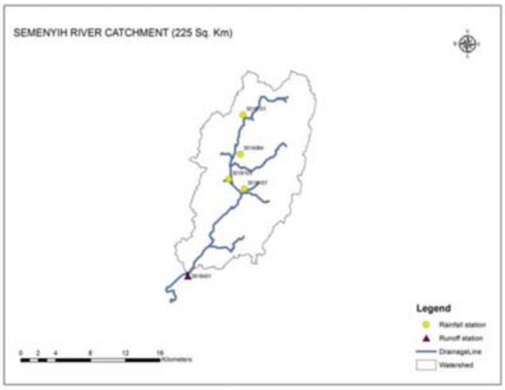

Figure 4. Map showinglocations of rainfall stations.

2.5 Homogeneity Tests

Homogeneity is important to check the variability in time series data as it is always affected by the changes made in measurement techniques and environmental characteristics. Homogeneity tests were performed for hourly rainfall and runoff data from the four stations. Three commonly used approaches were adopted to perform homogeneity tests that included: (1) Pettitt test developed by [24]; (2) standard normal homogeneity test (SNHT) developed by [25]; and (3) Von Neumann Ratio (VNR) test developed by [26]. The performance of the tests was evaluated on two different variables, which are annual mean and annual median.

2.6 Events Selection and Data Processing

This study was performed for event based rainfall-runoff modeling. Thirty (30) of extreme events were extracted from rainfall-runoff time series to be used for this study from which twenty two (22) events were used for training the model and the remaining eight (08) for the testing phase. The testing events were randomly selected rather than selecting the last eight events. All the rainfall and runoff data were normalized before analysis. Normalization concentrates the dispersed data into a defined interval. The normalization method used in this study was adopted from [27] which can be given by:

$$x_n = F_{min} + \left[\frac{x_i - x_{min}}{x_{max} - x_{min}}\right] \times (F_{max} - F_{min}) \quad (6)$$

where FMIN and FMAX are the required minimum and maximum of the new domain (e.g. 0.1-0.9), x_n is the standardized data, x_{min} and x_{max} are the minimum and maximum observed data, respectively; and xi is the observed data.

2.7 Model Performances

The performances of RBFNN model were assessed on the basis of different statistics such as coefficient of efficiency (CE), coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Relative Peak Error (RPE).

$$CE = 1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}(Q_i - \bar{Q}_i)^2} \quad (7)$$

$$R^2 = \frac{\sum_{i=1}^{n}(Q_i - \bar{Q})(\hat{Q}_i - \tilde{Q})}{\sqrt{\sum_{i=1}^{n}(Q_i - \bar{Q})^2} \times \sqrt{\sum_{i=1}^{n}(\hat{Q}_i - \tilde{Q})^2}} \quad (8)$$

$$\text{RMSE} = \frac{\sqrt{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}}{n} \quad (9)$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|Q_i - \hat{Q}_i|}{n} \quad (10)$$

$$\text{RPE} = \frac{|(Q_p) - (\hat{Q}_p)|}{(Qp)} \quad (11)$$

where $\bar{Q}$ is the average observed discharge and n is the total number of the observations, Q_i is observed flow rate and $\hat{Q}_i$ is the simulated flow rate, Q_p and $\hat{Q}_p$ are the observed peak discharge and simulated peak discharge respectively.

3. RESULTS AND DISCUSSION

To assess the homogeneity of rainfall data obtained from DID for Semenyih River catchment; the critical values were adopted from [27] which were 57, 6.95 and 1.30 for Pettitt test, SNHT and VNR respectively. The results of the homogeneity tests showed that the rainfall and runoff data from all stations are homogeneous and found suitable for further analyses. In data driven modeling approaches the input combination selection is of key importance. In most of the studies the different input combinations were evaluated and the best one is selected based on their statistical performances obtained from the model. The selection of random combinations of different stations with different lags is time taking and needed to run the model again and again to get optimum performance. In this study the input selection procedure adopted was based on correlation and mutual information analyses. The understanding of lag time from different rainfall stations to the runoff observed station is necessary to perform these analyses. As the Semenyih catchment has 4 rainfall stations, the analyses were carried out to identify the suitable input combination for the model development. The input selection procedure showed that out of the four available rainfall stations, the rainfall antecedents of only two stations with different lag demonstrate sufficient correlation with output. Moreover, including one discharge antecedent Q(t-1) in inputs, was found to be effective in enhancing the model performance. Out of the different possible combinations between rainfall antecedents, the correlation analysis showed that the input combination of R3(t-2), R4(t-4), and Q(t-1) is the best to develop RBFNN model. The model was then trained with the 22 selected events and then the model was tested by selected 8 testing events. Model performances were obtained in both training and testing stages based on several statistics such as CE, R^2, RMSE and MAE. Moreover RPE was also tested to check the performance of the model for simulation of peak discharges. The performances obtained from RBFNN for 8 testing events can be seen in Table 1.

Table 1. Performances of the RBFNN for testing events.

Events	RBFNN performances in testing phase				
	CE	R^2	RMSE	MAE	RPE
Event 1	0.86	0.87	4.46	2.43	0.04
Event 2	0.83	0.9	7.66	5.02	0.06
Event 3	0.96	0.96	3.66	2.02	0.04
Event 4	0.76	0.8	5	2.37	0.11
Event 5	0.82	0.86	8.01	4.32	0.05
Event 6	0.86	0.9	2.35	1.16	0.03
Event 7	0.81	0.88	3.65	2.16	0.13
Event 8	0.78	0.86	3.09	1.85	0.12

The coefficient of efficiency obtained for each testing event was greater than 0.75, which shows that the RBFNN model was able to simulate discharge very well. Similarly the RPE obtained were also at minimum level for each testing event, which shows that the model is able to simulate the peak discharges. The simulated and observed hydrographs for testing events can be seen in Figure 5.

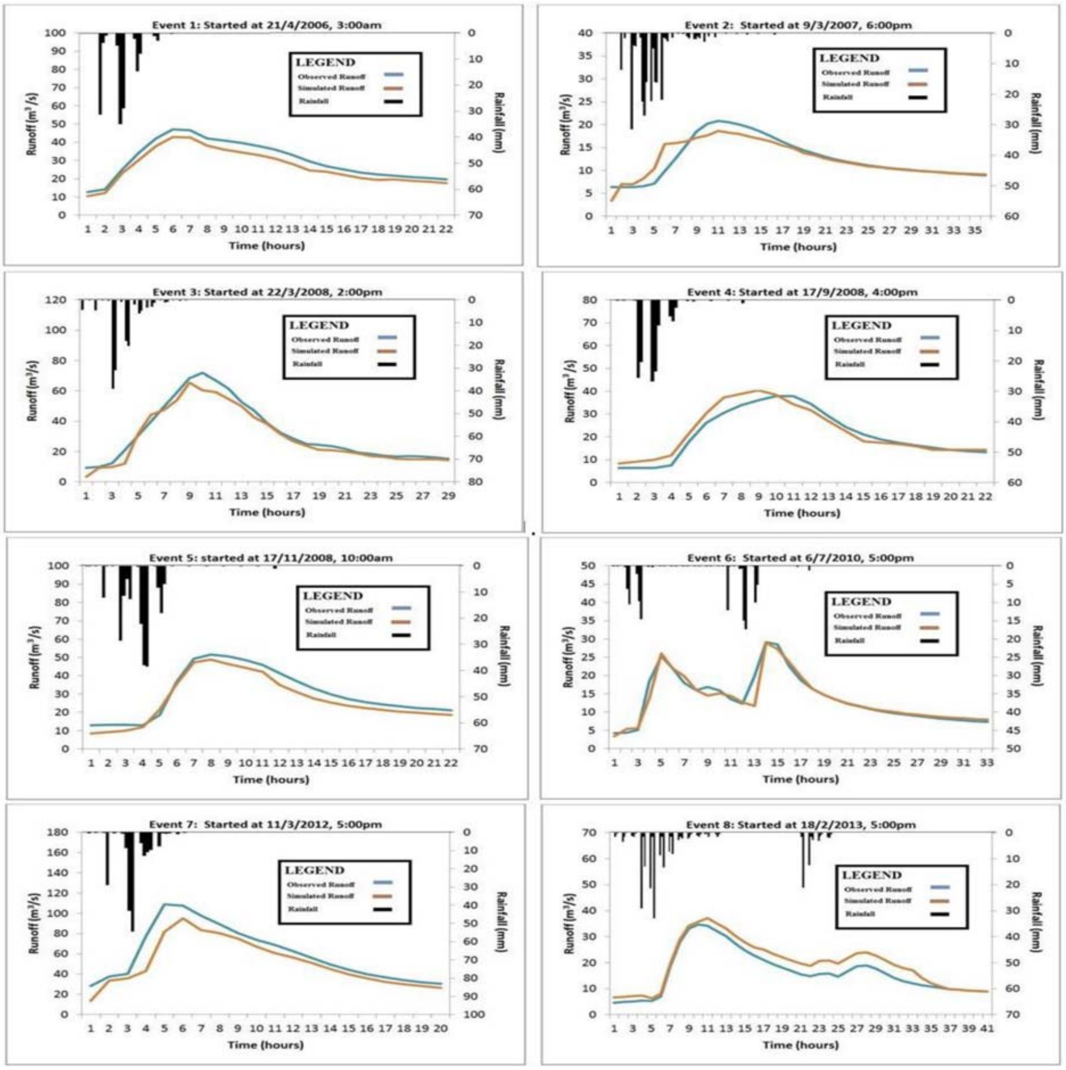

Figure 5. Hyetograph and hydrograph of the observed and simulated discharge by RBFNN model for 08 testing events.

In this study, the ARMAX model was developed using different combinations of rainfall antecedents (up to present time) as well as discharge antecedents (up to t-1) as exogenous inputs to estimate the runoff at present time t. The performance assessment based on several statistics was also performed on ARMAX results for the testing events. To further assess the performance of RBFNN, the results were also compared with ARMAX model. The Results of RBFNN were much better than the ones obtained from the ARMAX model in terms of all statistics. Table 2 presents the comparison of average performances obtained from RBFNN and ARMAX model for 08 testing events.

Table 2. Comparison of Average performances of RBFNN and ARMAX.

Models	Average performances in testing phase				
	CE	R^2	RMSE	MAE	RPE
RBFNN	0.83	0.87	4.73	2.67	0.07
ARMAX	0.62	0.63	5.81	3.03	0.12

The modeling techniques like RBFNN and ARMAX provide a direct mapping between input and the desired output without understanding the internal physical behavior of the catchment. These techniques are useful when data of some parameters like infiltration rate, soil moisture conditions, land use map, temperature, and evapotranspiration are missing. The high quality and small resolution of recorded metrological data can enhance the model performances. Such techniques are also useful when hydrologists dealing with missing record of data. These highlighted problems relating to data record are common in developing countries and such models are preferable in these conditions.

4. Conclusions

This study was performed to simulate rainfall-runoff process for the extreme events of Semenyih River catchment. The Homogeneity tests were performed on the historical rainfall and runoff data. The analyses found that both time series data were homogeneous. The catchment has 4 rainfall stations so an input selection analyses were performed to select best input combination. Thirty extreme events were selected to simulate rainfall runoff process. The results were evaluated on different statistical measures. This study has found that RBFNN is a reliable tool and can be used as alternate model where the physical data is deficient. Moreover RBFNN results were much better then traditionally used statistical model ARMAX.

References

[1] Solomatine D.P. and Wagener T., 2.16 - Hydrological Modeling; in Wilderer P., ed., *Treatise on Water Science*, Elsevier: Oxford, 2011: 435-457.

[2] Kisi O., Shiri J. and Tombul M., *Comput. Geosci.*, 2013; **51(0)**: 108-117. DOI 10.1016/j.cageo.2012.07.001.

[3] Nasr A. and Bruen M., *J. Hydrol.*, 2008; **349(3-4)**: 277-290. DOI 10.1016/j.jhydrol.2007.10.060.

[4] Duan Q., Sorooshian S. and Gupta V., *Water Resour. Res.*, 1992; **28(4)**: 1015-1031. DOI 10.1029/91WR02985.

[5] Nourani V., Kisi O. and Komasi M., *J. Hydrol.*, 2011; **402(1-2)**: 41-59. DOI 10.1016/j.jhydrol.2011.03.002.

[6] McCulloch W.S. and Pitts W., *B. Math. Biophys.*, 1943; **5(4)**: 115-133.

[7] Rumelhart D.E., McClelland J.L. and Group P.R., *Parallel Distributed Processing*, Vols 1 and 2. The MIT Press, Cambridge, 1986.

[8] Minns A. and Hall M., *Hydrol. Sci. J.*, 1996; **41(3)**: 399-417. DOI 10.1080/02626669609491511.

[9] Hsu K.L., Gupta H.V. and Sorooshian S., *Water Resour. Res.*, 1995; **31(10)**:

2517-2530. DOI 10.1029/95WR01955.

[10] Coulibaly P., Anctil F., Aravena R. and Bobee B., *Water Resour. Res.*, 2001; **37(4)**: 885-896. DOI 10.1029/2000WR900368.

[11] Coppola E., Szidarovszky F., Poulton M. and Charles E., *J. Hydrol. Eng.*, 2003; **8(6)**: 348-360. DOI 10.1061/(ASCE)1084-0699(2003)8:6(348).

[12] Nor N.I., Harun S. and Kassim A.H., *J. Hydrol. Eng.*, 2007; **12(1)**: 113-123. DOI 10.1061/(ASCE)1084-0699(2007)12:1(113).

[13] Wu J., Han J., Annambhotla S. and Bryant S., *J. Hydrol. Eng.*, 2005; **10(3)**: 216-222. DOI 10.1061/(ASCE)1084-0699(2005)10:3(216).

[14] Nourani V., Kalantari O. and Baghanam A., *J. Hydrol. Eng.*, 2012; **17(12)**: 1368-1380. DOI 10.1061/(ASCE)HE.1943-5584.0000587.

[15] Senthil K.A., Ojha C., Goyal M., Singh R. and Swamee P., *J. Hydrol. Eng.*, 2011; **17(3)**: 394-404. DOI 10.1061/(ASCE)HE.1943-5584.0000445.

[16] French M.N., Krajewski W.F. and Cuykendall R.R., *J. Hydrol.*, 1992; **137(1)**: 1-31. DOI 10.1016/0022-1694(92)90046-X.

[17] Ramirez M.C.V., de Campos Velho H.F., and Ferreira N.J., *J. Hydrol.*, 2005; **301(1)**: 146-162. DOI 10.1016/j.jhydrol.2004.06.028.

[18] Jain S., Das A. and Srivastava D., *J. Water Resour. Plann. Manage.*, 1999; **125(5)**: 263-271. DOI 10.1061/(ASCE)0733-9496(1999)125:5(263).

[19] Rama H. and Chandramouli V., *J. Water Resour. Plann. Manage.*,1996; **122(5)**: 342-347. DOI 10.1061/(ASCE)0733-9496(1996)122:5(342).

[20] Haykin S., *Neural Networks: A Comprehensive Approach*, IEEE Computer Society Press, 1994.

[21] Mason J.C., Price R.K. and Tem'Me A., *J. Hydrol. Res.*, 1996; **34(4)**: 537-548. DOI 10.1080/00221689609498476

[22] Lin G.F. and Chen L.H., *J. Hydrol.*, 2004; **289(1)**: 1-8. DOI 10.1016/j.jhydrol.2003.10.015.

[23] Box G.E.P. and Jenkin G.M., *Time Series Analysis Forecasting and Control*, San Francisco: Holden-Day, 1970.

[24] Pettitt A.N., *Appl. Stat.*, 1979; **28(2)**: 126-135. DOI 10.2307/2346729.

[25] Alexandersson H., *J. Climatol.*, 1986; **6(6)**: 661-675. DOI 10.1002/joc.3370060607.

[26] Von Neumann J., *Ann. Math. Stat.*, 1941; **12(4)**: 367-395.

[27] Van Ooyen A. and Nienhuis B., *Neural Networks*, 1992; **5(3)**: 465-471. DOI 10.1016/0893-6080(92)90008-7.

[28] Wijngaard J., Tank K.A. and Konnen G., *J.Climatol.*, 2003; **23(6)**: 679-692. DOI 10.1002/joc.906.