SENRA Academic Publishers, Burnaby, British Columbia
Vol. 4, No. 3, pp. 1275-1281, October 2010
Online ISSN: 1920-3853; Print ISSN: 1715-9997

Canadian Journal of pure&applied sciences

RIVER DISCHARGE PREDICTION USING ARTIFICAL NEURAL NETWORK

*Archana Chowdhary[1] and R K Shrivastava[2]
[1]Civil Engg Deptt IES IPS Academy Indore
[2]SGSITS, Indore, India

ABSTRACT

The research described in this article investigates the utility of Artificial Neural Networks (ANNs) for predicting the daily river discharge. The work explores the capabilities of ANNs and compares the performance of Feed Forward Neural Network (FFNNS) and Radial Basis Function (RBF) network. Perceived strengths of ANNS are the capability for representing complex, non linear relationships as well as being able to model interaction effects. The application of the ANN approach is to a portion of Seonath River in Chhattisgarh and forecasting was conducted using daily records. ANN technique shows an enhancement of prediction capabilities & reduces the over fitting problem of neural networks. The results show that the ANN technique can be used to extract information from the data & to describe the non-linearity of river discharge.

Keywords: Artificial neural network, radial basis function, regression analysis.

INTRODUCTION

Many of the activities associated with the planning and operation of the components of a water resource system require forecast of future events. For the hydrologic components, there is a need for both short term and long term forecasts of streamflow events, discharge in order to optimize the system or to plan for future expansion or reduction. Many of these systems are large in spatial extent and have a hydrometric data collection network that is very sparse. These conditions can result in considerable uncertainty in the hydrologic information that is available.

Furthermore, the inherently non-linear relationships between input and output variables complicate attempts to forecast streamflow events. There is thus a need for improvement in forecasting techniques. Many of the techniques currently used in modelling hydrological time-series and generating synthetic streamflow assume linear relationships amongst the variables. The two main group of techniques include physically based conceptual models time-series models. Techniques in the first group are specifically designed to mathematically simulate the sub-processes and physical mechanisms that govern the hydrological cycle. These models usually incorporate simplified forms of physical laws and are generally non-linear, time-invariant, and deterministic, with parameters that are representative of watershed characteristics (Hsu *et al*.,1995) but ignore the spatially distributed, time-varying, and stochastic properties of the rainfall runoff (R-R) process. Kitanidis and Bras (1980[a,b]) state that conceptual watershed model is reliable in forecasting the most important features of the hydrograph. However, the implementation and calibration of such a model can typically present various difficulties (Duan *et al*., 1992), requiring sophisticated mathematical tools (Duan *et al*., 1992, 1994; Sorooshian *et al*.,1993), significant amount of calibration data (Yapo *et al*., 1996) and some degree of expertise and experience with the models (Hsu *et al*.,1995).The problem with the conceptual models is that empirical regularities or periodicities are not always evident and can often be masked by noise.

Currently, environmental prediction and modelling includes a variety of approaches, such as rainfall-runoff modelling or statistical techniques, which entail exogenous input together with a number of assumptions. Conventional numerical modelling addresses the physical problem by solving a highly- coupled, non- linear, partial differential equation set which demands huge computing cost and time. However, physical processes affecting flooding occurrence can be highly complex and uncertain, and can be difficult to capture in some form of deterministic or statistical model.

In time-series analysis, stochastic or time-series model are fitted to one or more of the time-series describing the system for purpose which include forecasting, generating synthetic sequences for use in simulation studies, and investigating and modelling the underlying characteristics of the system under study. Most of the time-series modelling procedures fall within the framework of multivariate autoregressive moving average (ARMA) models (Raman and Sunil Kumar, 1995). Traditionally, the class of ARMA models has been the statistical method

*Corresponding author email: archanakeertichowdhary@yahoo.co.in

most widely used for modelling water resource time-series (Maier and Dandy, 1996). In streamflow forecasting, time-series models are used to describe the stochastic structure of the time sequence of streamflows and precipitation values measured over time. Time-series models are more practical then conceptual models because one is not required to understand the internal structure of the physical processes that are taking place in the system being modeled. The limitation of univariate time-series methods in streamflow forecasting is that the only information they incorporate is that which is present in past flows. Many of the available techniques are deficient in that they do not attempt to represent non-linear dynamics inherent in the transformation of rainfall to runoff.

Therefore a relatively new, improved and efficient soft computing technique such as Artificial Neural Networks (ANNs) is used in this study to overcome the above-mentioned problems. The distinguishing feature of the ANN based hydrological model is the use of the capability of ANNs to approximate virtually any continuous function up to an arbitrary degree of accuracy which is not otherwise true of other conventional hydrological techniques (Hornik *et al*., 1989). Very often, in hydrology, the problems are not clearly understood or are too ill-defined for a meaningful analysis using physically based methods. Even when such models are available, they have to rely on nonlinear rainfall- runoff modelling assumptions that make ANNs seem more attractive. It provides one alternative to hydrological time-series modelling.

This study has the following objectives:

1. To analyze historical flow data for the River Seonath and to test its suitability for streamflow/flood forecasting.
2. To develop an improved and efficient streamflow/flood forecasting model using feed forward neural networks (FFNNs) and radial basis function neural networks (RBFNNs).

A number of flood/ streamflow forecasting studies have been undertaken throughout the world, using various techniques. Among these techniques, conventional auto-regressive integrated moving average (ARIMA) models and ANNs have been extensively applied. The following paragraphs describe major flood/streamflow forecasting studies using these techniques. Bhattacharya and Solomatine (2000) have used an ANN to model the stage-discharge relationship at the Swarupgunj gauging station on the Bhagirathi River in India. They concluded that the percentage error of the ANN output for all ranges of data sets was less compared to that of conventional models like auto-regressive integrated moving average (ARIMA) techniques. Birkundavyi *et al*. (2002) have used feed forward multi-layer perceptron neural networks as predictors for daily streamflow forecasting up to 7 days of lead- time at a Mistassibi River station in Canada. Furthermore they noticed that the reliability value does not depend upon the ANN structure and the input data. Burian *et al*. (2001) have trained artificial neural networks to perform disaggregation of rainfall data from three gauging stations in Alabama, USA. The study concluded that using 20 or 50 hidden neurons would produce accurate results as compared to ANNs with 4 hidden neurons. Coulibaly *et al*. (2001) applied temporal neural networks to predict multivariate time-series, specifically for hydropower reservoir inflow at the Chute-du-Diable watershed, in Canada. It was found that the Elman recurrent neural network (RNN) was more efficient than any of the other models for short term reservoir inflow forecasting. Elshorbagy *et al*. (2001) explored the applicability of chaos theory to find the segments of missing data of streamflows. They concluded that ANN models were superior to linear regression (LR) models. Their results showed that the process of nonlinear noise reduction did not help to improve the accuracy of estimating the missing data. Hsu *et al*. (2002) applied a hybrid ANN model named a self-organizing linear output (SOLO) network with 6 input variables for daily streamflow prediction. They concluded that SOLO could provide not only a quick and effective solution, but also an analysis tool to the modelling system. Islam and Kothari (2000) have used ANNs coupled with remote sensing of hydrological processes and its data at spatial and temporal levels. Self-organization feature maps (SOFMs) provide a lower dimensional representation that preserves the topological structure of the original and higher dimensional data. Liong *et al*. (2000) applied an ANN with feed forward architecture having back propagation algorithms as a flow prediction tool, which yielded a very high degree of water level prediction accuracy at Dhaka Bangladesh up to 7 days leads time. Maier and Dandy (1997) stated that auto-regressive-moving average (ARMA) models have been used conventionally for stochastic modelling of time series data of water resources. Maier and Dandy (1999) also studied that it is important to follow a systematic approach in the development of ANN models, taking into account factors such as data pre-processing, the determination of appropriate model inputs and a selection of appropriate topology, parameter estimation and model validation. Mani and Desai (2005) developed a relationship between stage and discharge using artificial neural networks. This model was applied to three gauging stations, which are located on the downstream side of Godavari River, India. The results showed that the ANN models were able to generalize the relationship between input and output variables. The Levenberg-Marquardt algorithm, which is a standard second-order non-linear least squares technique, based on the back-propagation process was used to increase the speed of training (Masters, 1993).

Thandaveswara and Sajikumar (2000) used adaptive resonance theory (ART) and multi-layer perceptron for pattern clustering to identify the hydrological homogeneity of Indian River basin. Thirumalaih and Deo (2000) developed a neural network model for specific sites in a river basin where adequate meteorological information was not available. The sufficiency of the impending monsoon rainfall was adequately judged by an appropriately trained network, which helped in hydrological forecasting. Tokar and Johnson (1999) stated that the auto-regressive integrated moving average (ARIMA) model does not attempt to represent the non-linearity inherent in the hydrologic processes, and may not always perform well. Tokar and Markus (2000) applied artificial neural networks to rainfall-runoff modelling of the Fraser River, Colorado, USA along with conceptual models. To compare the model results, they used statistical indices such as coefficient of determination between predicted and observed discharge and the ratio between the root mean square error and standard deviation of observed discharge. The results revealed that ANNs could accurately model nonlinear relationships between hydrologic inputs (i.e., rainfall, snowmelt equivalent, temperature) and the output–streamflow.

Study Area

The Seonath River originates near village Panabaras in the Rajnandgaon district. The basin (Fig. 1) is located between latitude 20^0 16' N to 22^0 41' N and longitude 80^0 25' E to 82^0 35' E. The basin area of river up to confluence with the Mahanadi River is 30,860 Sq Km. The river traverses a length of 380 Km. The main tributary of Seonath river are Tandula, Kharun, Arpa, Hamp, Agar and Maniyari rivers. The mean annual rainfall in the basin varies from 1005mm to 1255mm.

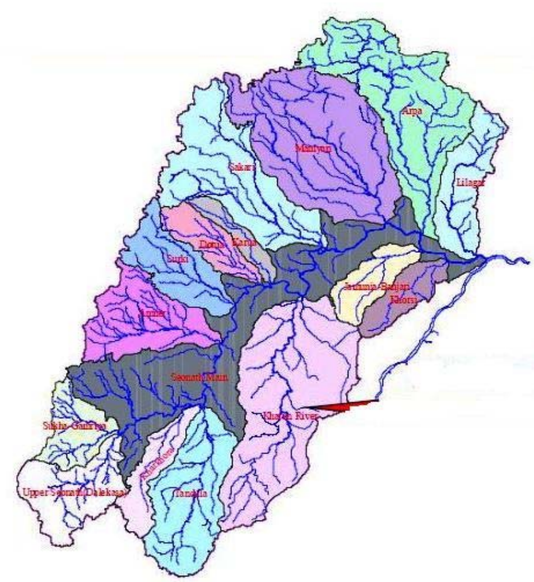

Fig. 1. Location map of Seonath basin.

Artificial Neural Network

A neural network is a computational method inspired by studies of the brain and nervous systems in biological organisms (Haykin, 1999). Typically neural networks consist of a layered processing units and weighted interconnections. Neural networks operate on the principle of learning from a training set. The typical architecture of ANNs is shown in figure 2. The most commonly used training algorithm for feed forward networks is the back propagation algorithm by Rumelhart *et al*. (1986). Earlier studies are limited primarily to feed forward neural networks with logistic activation function, as these are mostly used for the prediction and forecasting of water resource variables. Radial basis function neural network (RBFNN) can be considered as a three-layer network in which the hidden layer performs a fixed nonlinear transformation with no adjustable parameters. The primary difference between the RBF network and back-propagation is in the nature of the nonlinearities associated with the hidden nodes. The nonlinearity in back –propagation algorithm (BP) is implemented by a fixed function such as a sigmoid, whereas in radial basis function (RBF) method applies its nonlinearities on the data in the training set. ANNs offer real merit over traditional modelling, including the ability to handle large amounts of noisy data from dynamic nonlinear systems, especially when the underlying physical relationships are not fully understood (Pan and Wang, 2004).

Selection of input and output variable(s)

For any type of forecasting or estimation problem, it is very important to determine appropriate input variables which will help us for mapping the non-linear relationship between the input and output variables. The goal of an ANN is to generalize a relationship of the form.

$$Y^m = f\,(X^n) \qquad (1)$$

where X^n is an n-dimensional input vector consisting of variables x_1, $x_2 \ldots x_i \ldots\ldots x_n$; and Y^m an m-dimensional output vector consisting of resulting variables of interest $y_1 \ldots\ldots\ldots y_i \ldots\ldots\ldots y_m$. In hydrology, the values of x_i can be causal variables such as rainfall, temperature, previous flows, spatial locations, evaporation, basin area, slope, elevation, meteorological data, and so on. The values of y_i can be hydrological responses such as runoff, streamflow and others. For this present study, the value of X and Y are river stage and discharge respectively. An optimal dataset should be representative of the probable occurrence of an input vector that should facilitate the mapping of the underlying nonlinear process.

Data collection and preprocessing

There is no hard and fast rule for determining the number of input-output data combinations that will be required. An optimal data set should be representative of the probable occurrence of an input vector and should facilitate mapping of the underlying nonlinear process.

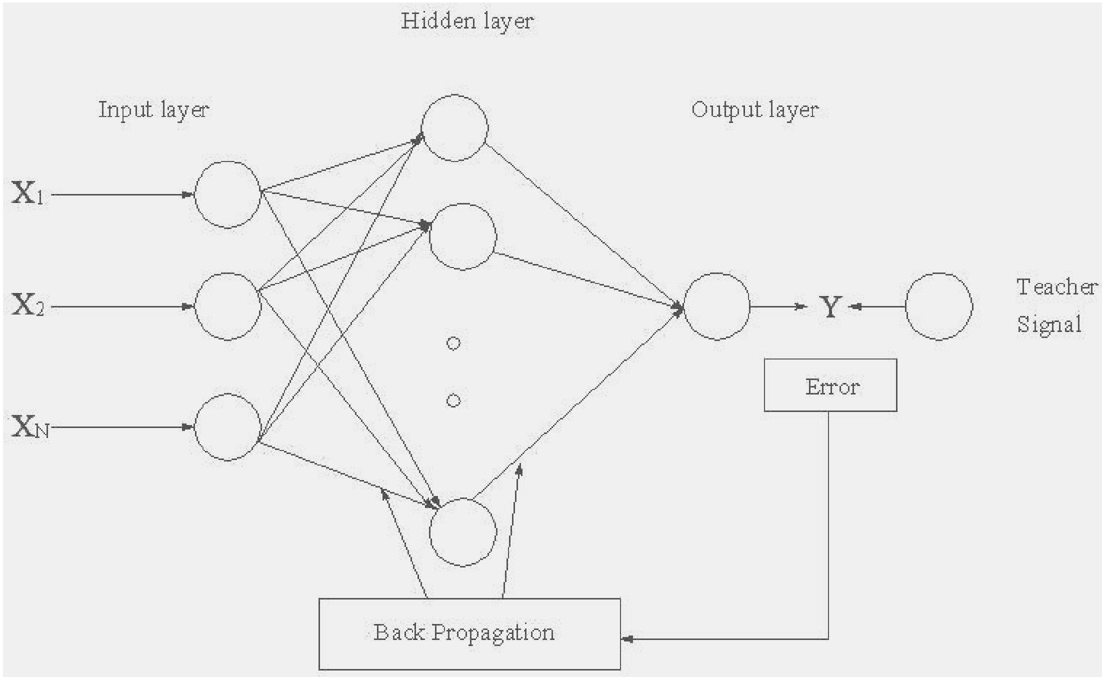

Fig. 2. Typical ANN architecture with three neuron layer.

Inclusion of less important patterns will reduce the network learning speed but an insufficient data set could lead to poor learning. This makes it useful to analyze and pre-process the data before it is used for the artificial neural network application. The following equations are used for scaling the input and output data set.

$$\text{Data}_{\text{scaled}} = \frac{H}{1.24H_{max}} + 0.1 \quad (2)$$

$$\text{Data}_{\text{scaled}} = \frac{Q}{1.24Q_{max}} + 0.1 \quad (3)$$

Where Q and H are the discharge and the stage at any time t and Q_{max} and H_{max} are the maximum discharge and maximum stage within the period of the ANN simulation. The values of Q_{max} and H_{max} used here are 9106cumecs and 249.70m.

RESULTS AND DISCUSSION

The applicability and potential of two different types of ANN in daily streamflow forecasting is explored in this study. Table 1 (a) and (b) gives the preliminary statistical analysis of the daily discharge and the stage. The choice of the neural network architecture, the training algorithm and the definition of error are usually determined based on past experience and preference of the users, rather than the physical aspects of the problem. Note that the ANN simulated (FFNN) daily discharge values are plotted continuously irrespective of missing years up to 726 days (i.e., 2 water years) in figure 4 during testing, 3053 (i.e., 8 water years) in figure 3 during training respectively. In India, the water year generally starts on the first day of June and ends on the last day of May in every year. The applicability and potential of two different types of ANN in daily streamflow forecasting is explored in this study. Performance of the model outputs is compared in terms of correlation coefficient (CC), normalized mean-squared-error (NMSE).

In this case study, daily discharge during the monsoon period was predicted using the daily stage and the previous time-step discharge as inputs. Ten water years of daily discharge from June to May (1996-2006) were used for developing the models. Since neural networks are capable to do well only with interpolation not with extrapolation. Because of that two different training and testing sets were used. In the first case, eight water years of data were used for training and two water years of data were used for testing the models. Therefore, the training and testing of the daily discharge prediction models were performed using a total of 3053 and 726 data sets. The daily discharge model was developed with only one hidden layer. Initially the experimentation was performed by changing the number of hidden layer neurons from 5 to 20. Beyond that of 20 hidden layer neurons, the performance of RBFNN model didn't improve significantly. That's why we used the number of neurons in the hidden layer was kept at a constant of 20. By

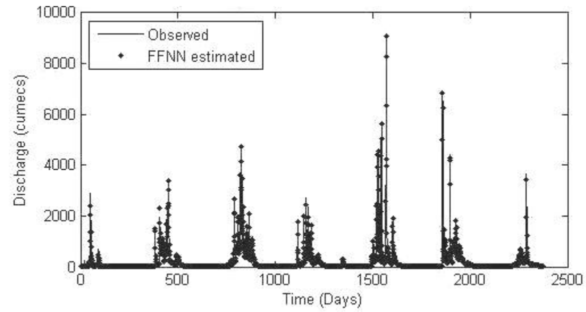

Fig. 3. Comparison of the observed & the FFNN estimated daily discharge during training period.

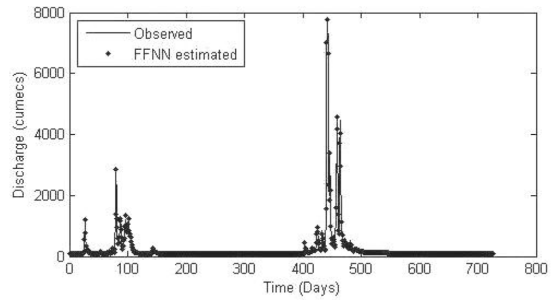

Fig. 4. Comparison of the observed & the FFNN estimated daily discharge during testing period.

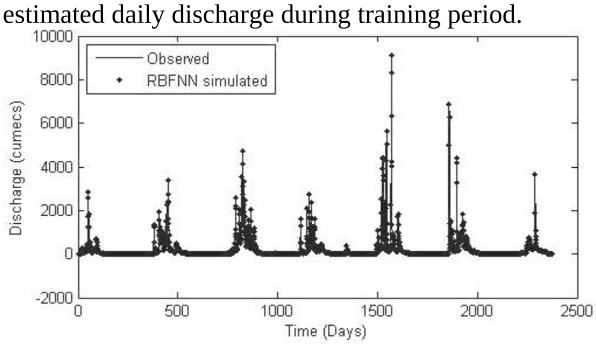

Fig. 5. Comparison of the observed & the RBFNN simulated daily discharge during training period.

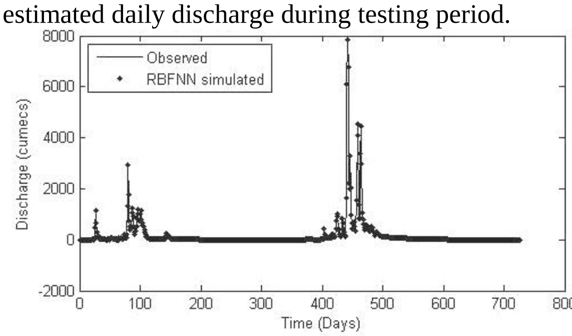

Fig. 6. Comparison of the observed & the RBFNN simulated daily discharge during testing period.

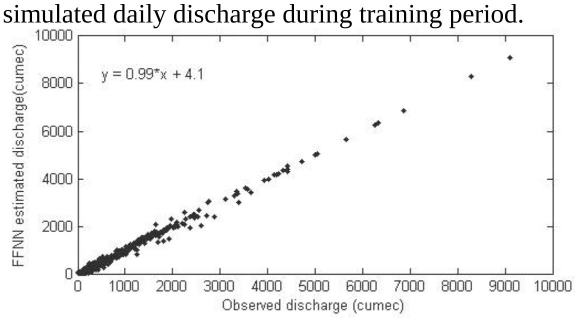

Fig. 7. Scatter plot between observed & the FFNN estimated daily discharge during training period.

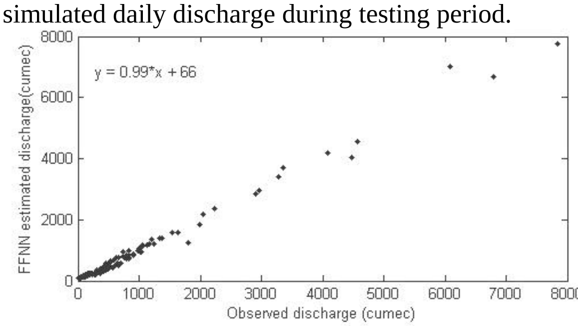

Fig. 8. Scatter plot between observed & the FFNN estimated daily discharge during testing period.

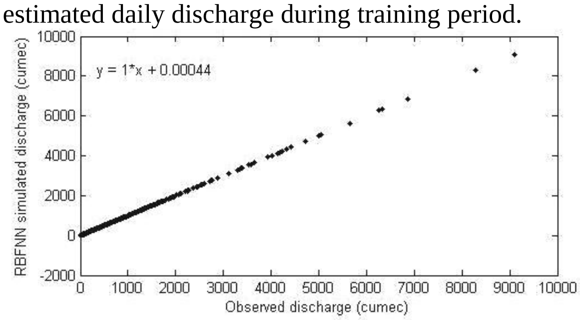

Fig. 9. Scatter plot between observed & the RBFNN simulated daily discharge during training period.

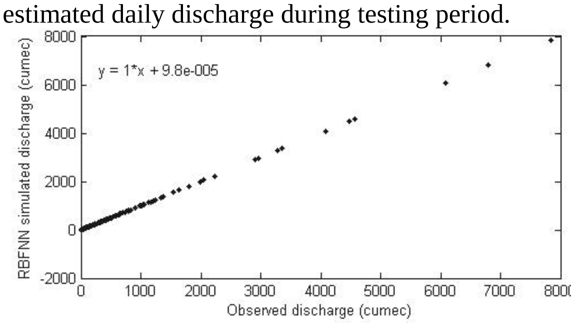

Fig.10. Scatter plot between observed & the RBFNN simulated daily discharge during testing period.

changing the values of spread in RBFNN from 0.05 to 1.00 simulation runs were performed to predict the daily discharge and thereafter the best RBFNN model was selected. The RBFNN simulated daily discharge values are plotted continuously irrespective of missing years up to 726 days (i.e., 2 water years) in figure 6 during testing, 3053 (i.e., 8 water years) in figure 5 during training respectively The daily hydrographs and the linear agreement between the observed and the ANN simulated using FFNN and RBFNN streamflows are depicted in figures 7 & 8 and figures 9 & 10 respectively.

Table1a. SEONATH DISCHARGE.

Year	Mean	Standard Deviation	Skewness	Kurtosis	Covariance	Max	Min
2000-2001	78.4117	268.67	7.262	64.08	72180	2870	0
2001-2002	194.866	430.73	3.711	19.8	185000	3372.63	0.1
2002-2003	93.64	280.144	5.2	39.1362	78400	2914.8	0.04
2003-2004	332.78	657.168	3.209	15.5597	413870	4724.8	0.84
2004-2005	160.1	365.6419	3.8343	19.5274	133690	2722.6	0.65
2005-2006	372.163	1049.9	4.9512	31.6813	1102300	9105.7	0.83
2006-2007	248.0292	830.633	6.0083	43.9904	689950	7848.8	0.76
2007-2008	247.9557	701.9785	6.2978	49.6811	492770	6867.07	2.02
2008-dec2008	185.1562	338.1807	6.8233	64.971	114370	3625.7	2.55

Table1b. SEONATH STAGE.

Year	Mean	Standard Deviation	Skewness	Kurtosis	Covariance	Max	Min
2000-2001	236.07	0.895	3.65	22.31	0.8023	242.86	235
2001-2002	236.55	1.347	2.18	8.14	1.8612	244.35	236.57
2002-2003	237.13	0.933	2.89	11.84	0.871	242.78	236.54
2003-2004	237.93	1.675	2.04	7.53	2.8071	246.63	236.77
2004-2005	237.47	1.416	2.23	8.22	1.3136	242.95	236.55
2005-2006	237.82	2.022	3.04	13.66	4.0909	249.7	236.56
2006-2007	237.54	1.638	3.86	20.46	2.6856	248.25	236.65
2007-2008	237.69	1.523	3.25	17.91	2.3209	248.05	236.67
2008-dec2008	237.56	0.871	3.64	24.69	0.7597	244.3	236.66

CONCLUSIONS

Feed forward neural networks trained with back-propagation usually start with a large network and proceeds by removing weights to which sensitivity of the error is minimal. The optimal architectures of ANN for the present investigation are 5-30-1 for feed forward neural networks and 5-20-1 for radial basis function neural networks with different range of spread values. This is similar to the idea of calibration that is an integral part of most of the time series modelling in the field of hydrology. The results show that the radial basis function neural networks are found to produce an accurate forecast of daily streamflow when compared to feed forward neural networks to the particular gauge/discharge station. The prediction accuracy of the ANN based model is highly dependent on many issues associated with network structure identification and network parameters such as interconnected weights, learning rate, momentum coefficient, and the number of epochs needed for optimization. This study aims to improve the model performance by explicitly incorporating hydrological a priori knowledge, reducing the network sensitivity to input errors and changing the training objective function. The results of this case study, demonstrated in the Seonath nandghat G/D station at Seonath River Basin under experimentation, were encouraging. To improve the performance of the ANN models, the incorporation of other climatic variables such as rainfall and temperature into the input data set and use of recurrent and modular neural networks, may be a good focus for the future direction of research work.

REFERENCES

Bhattacharya, B. and Solomatine, DP. 2000. Application of artificial neural networks in stage discharge relationship, Proc. of 4th International conference on Hydro informatics, lowa City, USA.

Birikundavyi, S., Labib, R., Trung, HT. and J. Rousselle, J. 2002. Performance of neural network in daily streamflow forecasting. J. Hydrologic Engg., ASCE. 7 (5):392-398.

Burian, JS., Durrans, SR., Nix, SJ. and Pitt, RE. 2001. Training artificial neural networks to perform rainfall disaggregation. J. Hydrologic Engg., ASCE. 6(1):43-51.

Coulibaly, P., Anctil, F. and Bobee, B. 2001. Multivariate reservoir inflow forecasting using temporal neural networks. J. Hydrologic Engg., ASCE. 6(5): 367-376.

Duan, Q., Sorooshian, S. and Gupta, VK. 1992. Effective and efficient global optimization for conceptual rainfall runoff models. Water Resources Research. 28 (4):1015-1031.

Duan, Q., Sorooshian, S. and Gupta, VK. 1994. Optimal use of SCE-UA global optimization method for calibrating watershed models, Journal of Hydrology. 158:265-284.

Elshorbagy, A., Panu, US. and Simonovic, SP. 2001. Analysis of cross-correlated chaotic streamflows. J. Hydrological Sciences. 46 (5):781-793.

Haykin, S. 1999. Neural networks- A comprehensive foundation. (2nd ed.) Prentice-Hall of India Private Limited, New Delhi. India.

Hornik, K., Stinchcombe, M. and White, H. 1989. Multilayer feed forward networks are universal approximators. Neural Networks. 2(5):359-366.

Hsu, K., Gupta, HV. and Sorooshian, S. 1995. Artificial neural network modeling of the rainfall runoff process. Water Resources Research. 31:2517-2530.

Hsu, K., Gupta, HV., Sorooshian, XGS. and Imam, B. 2002. SOLO- an artificial neural network suitable for hydrological modeling and analysis. Water Resources Research. 38(12):1302-1310.

Islam, S. and Kothari, R. 2000. Artificial neural networks in remote sensing of hydrologic processes. J. Hydrologic Engg., ASCE. 5(2):138-144.

Kitanidis,PK. and Bras, RL. 1980[a]. Adaptive filtering through detection of isolated transient errors in rainfall-runoff models. Water Resources Research. 16 (4):740-748.

Kitanidis, PK. and Bras RL. 1980[b]. Real time forecasting with a conceptual hydrological model,1,Analysis of uncertainity. Water Resources Research. 16 (6):1025-1033.

Liong, SY., Lim, WH. and Paudyal, GN. 2000. River stage forecasting in Bangladesh Neural networks approach. J. Computing in Civil Engg., ASCE. 14 (1):1-8.

Maier, HR. and G.C Dandy, GC. 1996. The use of artificial neural networks for the prediction of water quality parameters. Water Resources Research. 32:1013-1022.

Maier, HR. and Dandy, GC. 1997. Determining inputs for neural network models of multivariate time series. Microcomputers in Civil Engg. 12:353-368.

Maier, HR. and Dandy, GC. 1999. Neural network based modeling of environmental variables: A systematic approach. Mathematical and Computer Modeling. 33(6-7):669-682.

Mani, A. and Desai, VR. 2005. Artificial neural networks (ANNs) –A state-of-the-art computing technique in stage-discharge relationship, Second Indian International Conference on Artificial Intelligence (IICAI-2005), Dec. 20-22, 2005 Pune, India.

Masters, T. 1993. Practical neural Network Recipes in C++. Academic Press. San Diego, Calif, USA. 173-199.

Pan, TY. and Wang, RH. 2004. State space neural network for short-term rainfall runoff forecasting. J. Hydrol. 297:34-50.

Raman, H. and Sunilkumar, N. 1995. Multivariate modeling of water resources time series using artificial neural networks. J. Hydrol. Sci. 40:145-163.

Rumelhart, DE.**,** Hinton, GE. and Williams, RJ. 1986. Learning Internal Representations by Error Propagation. In: Parallel Distributed Processing Explorations in the Microstructure of Cognition **(**vol**.** 1). Eds. Rumelhart, DE.**,** McClelland, JL. MIT Press, Bradfords Books, Cambridge**,** MA. 318-362**.**

Sorooshian, S., Duan, Q. and Gupta, VK. 1993. Calibration of rainfall-runoff models: Application of global optimization to the Sacramento soil moisture accounting model. Water Resources Research. 29(4):1185-1194.

Thandaveswara, BS. and Sajikumar, N. 2000. Classification of river basins using artificial neural network. J. Hydrologic Engg., ASCE. 5(3):290-298.

Thirumalaiah, K. and Deo, MC. 2000. Hydrological-forecasting using neural networks, J. Hydrologic Engg., ASCE. 5(2):180-189.

Tokar, AS. and Johnson, PA. 1999. Rainfall-runoff modeling using artificial neural networks. J. Hydrologic Engg., ASCE. 4(3):232-239.

Tokar, AS. and Markus, M. 2000. Precipitation-runoff modeling using artificial neural networks and conceptual models. J. Hydrologic Engg. ASCE. 5(2):156-161.

Yapo, P., Gupta, VK. and Sorooshian, S. 1996. Calibration of conceptual rainfall–runoff models: Sensitivity to calibration data. Journal of Hydrology. 181:23-48.

Received: May 1, 2010; Revised: Sept 6, 2010; Accepted: Sept 27, 2010