

Streamflow Modelling at Rampur, Sutlej Basin, Punjab

A. R. Senthil Kumar, S. D. Khobragade, Manohar Arora, R. D. Singh and R. K. Nema

National Institute of Hydrology, Roorkee

E-mail: arsk@nih.ernet.in

Abstract : Snow melt runoff is one of the main sources of streamflow in many of Himalayan Rivers. Conceptual models to simulate the snow melt runoff such as Snowmelt Runoff Model (SRM) and Snowmelt Model (SNOWMOD) require a large quantity of data which are generally not available for most locations in India. Applications of Artificial Neural Networks (ANN) in many water resources area indicate its better performance over other traditional models such as conceptual models and black box models. This paper discusses the development of ANN models for the simulation of streamflow at Rampur in Sutlej river basin. Rainfall, snowfall, temperature and discharge data of stations located at the upstream of Rampur were used as input to the models. Different combinations of significant lagged series of rainfall, snowfall, temperature and discharge data, determined from statistical parameters such as auto correlation function (ACF), partial auto correlation function (PACF) and cross correlation function (CCF), were used as input to the model. The performance of the model was evaluated using statistical criteria such as coefficient of correlation, root mean squared error (RMSE) and model efficiency. The results of the best ANN model during the calibration indicate that the all range of discharge values were simulated fairly well. However, the medium and high range values of discharge slightly deviated from the observed values during the validation of the model. The overall performance of the model, as exhibited by the various statistical criteria, indicates the suitability of ANN modelling technique to reasonably simulate the streamflow at Rampur in Sutlej river basin. Further, the development of two separate ANN models for simulating the low, medium and high did not yield better performance than the generalized ANN model with continuous data.

Keywords: Streamflow modelling, ANN models, Input vector, Statistical parameters, Sutlej Basin

INTRODUCTION

Snowmelt runoff estimates are of high interest for flood warning and management of reservoirs for hydropower generation, irrigation and drinking water supply in drainage basins with significant snowmelt contribution. The rivers originating from Himalayas receive a significant contribution from snow melt. The estimation of snow melt runoff in Sutlej River at Rampur is important for operating the Bhakra reservoir as well as for operating many other small hydropower power projects downstream of Bhakra reservoir. Conceptual models such as Snowmelt Runoff Model (SRM) (Martinec et al, 2007) and SNOWMOD (Arora, 2008) have been developed

to simulate the snow melt runoff using elevation, rainfall, aspect, temperature and snow cover area as inputs. Development and application of conceptual models for the simulation of snow melt runoff require large quantity of data. Recently, neural networks approach has been applied in many areas of water resources due to its capability in representing any nonlinear processes by given sufficient complexity of the trained networks (Maier and Dandy, 2000). ANNs are proven to produce better performance over other traditional models such as conceptual models and black box models, in numerous hydrological studies (Hsu et al., 1995). The main advantage of the ANN models over traditional models is that they do not require

information about the complex nature of the underlying process under consideration to be explicitly described in mathematical form. ANNs have found applications in various fields such as pattern recognition, non-linear modelling, classification, association, control. Some of the applications in hydrological studies are rainfall-runoff modeling, rainfall prediction, flood forecasting, water quality modeling, ground water modeling, development of water management policy, suspended sediment concentration, snow melt runoff modelling and reservoir operation studies. Tokar and Johnson (1999) developed ANN model for simulating the snowmelt runoff with observed temperature, precipitation (rain plus snow), snowmelt runoff as inputs. They compared the results of ANN with conceptual and regression model and found that the ANN model performed better than both the traditional models. Parent et al. (2008) simulated the snow melt runoff using ANN model with the inputs as considered by Tokar and Johnson (1999) in addition to the snow covered area. They also found that ANN model performed better than other models considered in the study. This paper discusses the development

and evaluation of ANN models for the simulation of streamflow at Rampur in Sutlej River.

ANN – AN OVERVIEW

ANNs are a form of computing inspired by the functioning of the brain and nervous system and are discussed in detail in a number of hydrologic papers, for example, ASCE, 2000a,b; Maier and Dandy, 2000. The architecture of a feed forward ANN can have many layers where a layer represents a set of parallel neurons. The basic structure of ANN usually consists of three layers: the input layer, where the data are introduced to the network; the hidden layer or layers, where data are processed; and the output layer, where the results of given outputs are produced. The neurons in the layers are interconnected by strength called weights. A typical three-layered feed forward ANN is shown in Fig. 1

In general, a neuron can have n inputs, labeled from 1 through n . For example neuron 3 in the hidden layer shown in Fig. 1, $n = 2$. In addition, each neuron has an input that is equal to 1.0, called *bias*. Each neuron j receives

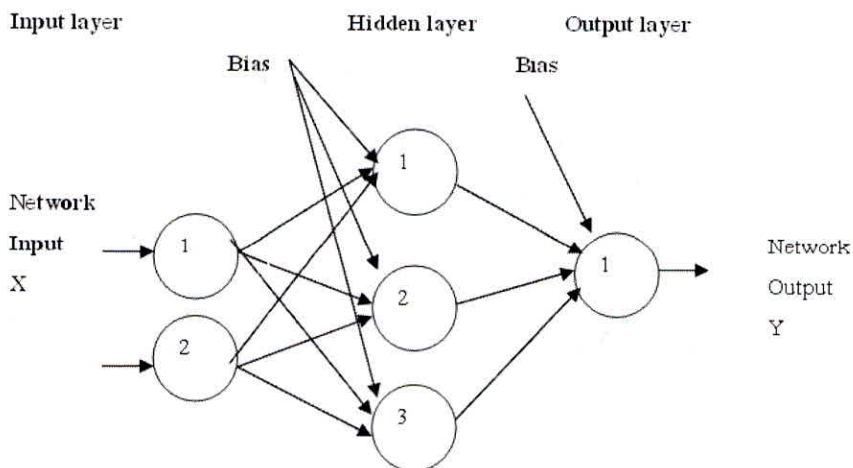


Fig. 1 : A Typical Three-Layer Feed Forward ANN (ASCE, 2000a)

information from every node i in the pervious layer. A weight (w_{ji}) is associated with each input (x_i) to node j . The effective incoming information (NET_j) to node j is the weighted sum of all incoming information, otherwise known as the net input, and is computed as:

$$NET_j = \sum_{i=0}^n w_{ji} x_i \quad (1)$$

where x_0 and w_{j0} are called as the bias term ($x_0 = 1.0$) and the bias respectively. Equation 1 applies to the nodes in the output layer and hidden layer(s). The weighted sum of input information is passed through an activation function, called transfer function, to produce the output from the neuron. The transfer function introduces some nonlinearity in the network, which helps in capturing the nonlinearity present in the function being mapped. The commonly employed transfer function is the sigmoid function (ASCE, 2000a) and is given as follows:

$$OUT_j = \frac{1}{1 + e^{-NET_j}} \quad (2)$$

The interconnected weights are adjusted using a learning algorithm such that the output from the ANN model is very close to the observed values by minimizing the error through a mathematically formulated procedure. This procedure is called training of network.

Using a set of examples from a given problem domain, comprising inputs and their corresponding outputs, an ANN model can be trained to learn the relationship between the input-output pairs. The feed forward ANN is generally adapted in all studies because of its applicability to a variety of different problems (Hsu et al., 1995). However, there are no guidelines in developing an effective ANN architecture, though some researchers have reported suggestions that can be implemented while developing an ANN model.

For instance, Maier and Dandy (2000) report that not more than one hidden layer is required in feed forward networks because a three-layer network can generate arbitrarily complex decision regions. Also, the appropriate input vector to the ANN model can be identified according to the procedure of Sudheer et al. (2002).

The input values should be normalised to the range between 0 and 1 before passing into a neural network since the output of sigmoidal function is bound between 0 and 1. The output from the ANN should be denormalised to provide meaningful results. In this study, following equation is used to normalize the data set:

$$N_i = \frac{R_i - Min_i}{Max_i - Min_i} \quad (3)$$

where R_i is the real value applied to neuron i ; N_i is the subsequent normalized value calculated for neuron i ; Min_i is the minimum value of all values applied to neuron i ; Max_i is the maximum value of all values applied to neuron i .

Training a network is a procedure during which an ANN processes training set (input-output data pairs) repeatedly, changing the values of its weights, according to a predetermined algorithm and the environment in which the network is embedded. The main objective of training (calibrating) a neural network is to produce an output vector $Y = (y_1, y_2, \dots, y_p)$ that is as close as possible to the target vector (variable of interest or forecast variable)

$$T = (t_1, t_2, \dots, t_p)$$

when an input vector

$X = (x_1, x_2, \dots, x_p)$ is fed to the ANN. In this process, weight matrices W and bias vectors V are determined by minimizing a predetermined

error function as explained as follows:

$$E = \sum_P \sum_p (y_i - t_i)^2 \quad (4)$$

where t_i is a component of the desired output T ; y_i is the corresponding ANN output; p is the number of output nodes; and P is the number of training patterns.

Back propagation is the most popular algorithm used for the training of the feed forward ANNs (Maier and Dandy, 2000). Each input pattern of the training data set is passed through the network from the input layer to output layer. The network output is compared with the desired target output, and an error is computed based the equation 4. This error is propagated backward through the network to each neuron, and the connection weights are adjusted based on the equation

$$\Delta W_{ij}(n) = -\epsilon * \frac{\partial E}{\partial W_{ij}} + \alpha * \Delta W_{ij}(n-1) \quad (5)$$

where $\Delta W_{ij}(n)$ and $\Delta W_{ij}(n-1)$ are weight increments between node i and j during n th and $(n-1)$ th pass, or epoch (ASCE, 2000a). A similar equation is written for correction of bias values. In the equation 5, ϵ and α are called learning rate and momentum respectively. The momentum factor can speed up training in very flat regions of the error surface and help prevent oscillations in the weights. A learning rate is used to increase the chance of avoiding the training process being trapped in local minima instead of global minima. The literature by Rumelhart et al, 1986 can be referred for the details of the algorithm.

Criteria for Performance Evaluation of ANN model

The whole data length is divided into two,

one for calibration (training) and another for validation of artificial neural network model. The performance during calibration and validation is evaluated by performance indices such as root mean square error (RMSE), model efficiency (Nash and Sutcliffe, 1970) and coefficient of correlation (R). They are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{k=1}^K (t-y)^2}{K}} \quad (6)$$

$$Efficiency = 1 - \frac{\sum (t-y)^2}{\sum (t-\bar{t})^2} \quad (7)$$

$$Coefficient\ of\ Correlation = \frac{\sum TY}{\sqrt{\sum T^2 \sum Y^2}}$$

where K is the number of observations; t is the observed data; y is computed data; $T = t - \bar{t}$ in which \bar{t} is the mean of the observed data; and $Y = y - \bar{y}$ in which \bar{y} is the mean of the computed data.

STUDY AREA

The catchment of Sutlej river up to Rampur was considered for the analysis. The catchment area up to Rampur is 50298 sq.km. The location of the study area is presented in Figure 2. For the development of the model, the daily rainfall values at Rampur, Kalpa, Rakccham, Kaza and Namagai, snowfall values at Kalpa, Rakccham, Kaza and Namgia, maximum temperature values at Rampur, Kalpa, Rakccham, Kaza and Namagia, minimum temperature values at Rampur, Kalpa, Rakccham, Kaza and Namagia were available from 1987 to 2000.

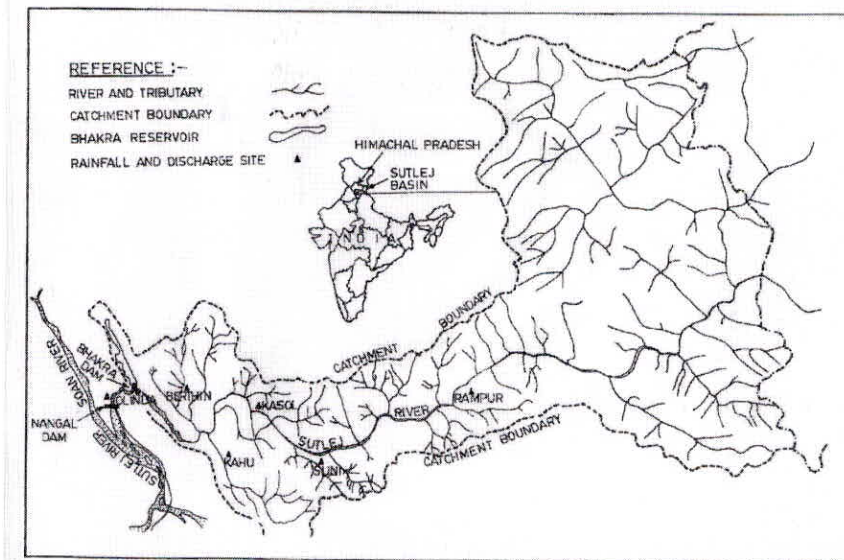


Fig. 2 : Index map of Sutlej basin

The discharge values at Rampur for the same period were also available.

MODEL DEVELOPMENT

The ANN models have been developed for simulating the streamflow in Sutlej at Rampur using the available data. The details of the model development are described in the following sections.

Selection of input

The ANN model for the prediction of streamflow generally uses the antecedent rainfall, snowfall, temperature and discharge values as input vector. Determining the number of antecedent rainfall, snowfall, temperature and discharge values involves finding the lags of rainfall, snowfall, temperature and discharge values that have significant influence on the predicted streamflow. These influencing values corresponding to different lags can be very well established through statistical analysis of the data

series. The input vector is selected generally by trial and error method. However, Sudheer et al. (2002) have presented a statistical procedure that avoids the trial and error procedure. They reported that the statistical parameters such as auto correlation function (ACF), partial auto correlation function (PACF) and cross correlation function (CCF) can be used for this purpose. The PACF of the discharge series for Sutlej at Rampur with 95 % confidence levels and CCF of discharge series at Rampur between the daily rainfall values at Rampur, Kalpa, Rakccham, Kaza and Namagial, snowfall values at Kalpa, Rakccham, Kaza and Namgial, maximum temperature values at Rampur, Kalpa, Rakccham, Kaza and Namagial, minimum temperature values at Rampur, Kalpa, Rakccham, Kaza and Namagial suggest the input vector to the ANN model. The ACF and PACF of discharge series at Rampur and CCF of discharge series at Rampur with rainfall, snowfall, maximum temperature and maximum temperature at the stations mentioned are represented in figures 3 to 23 respectively.

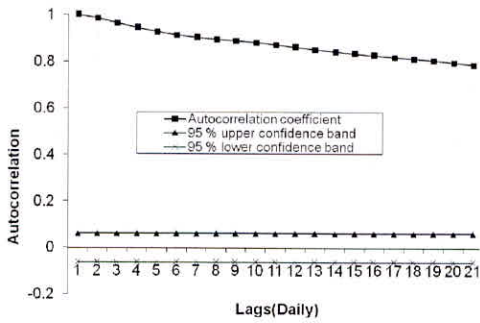


Fig. 3 : The autocorrelation of the discharge series at Rampur

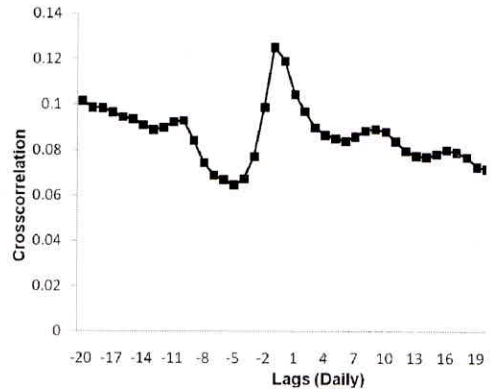


Fig. 6 : The cross correlation between discharge at Rampur and rainfall at Kalpa

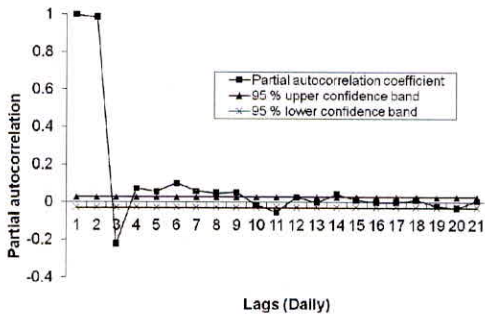


Fig. 4 : The partial autocorrelation of the discharge series at Rampur

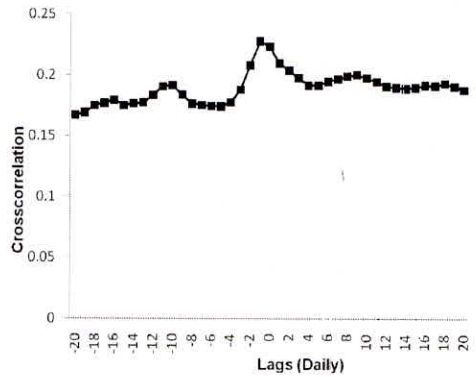


Fig. 7 : The cross correlation between discharge at Rampur and rainfall at Rakccham

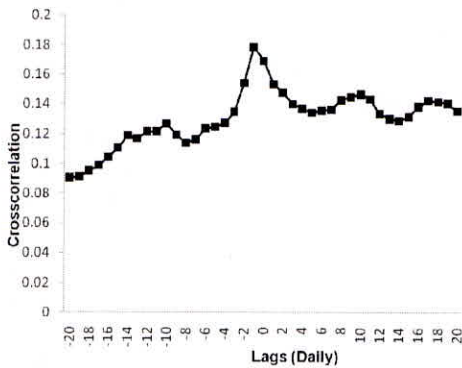


Fig. 5 : The cross correlation between discharge at Rampur and rainfall at Rampur

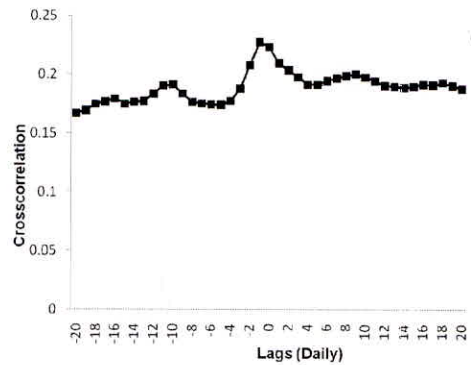


Fig. 8 : The cross correlation between discharge at Rampur and rainfall at Kaza

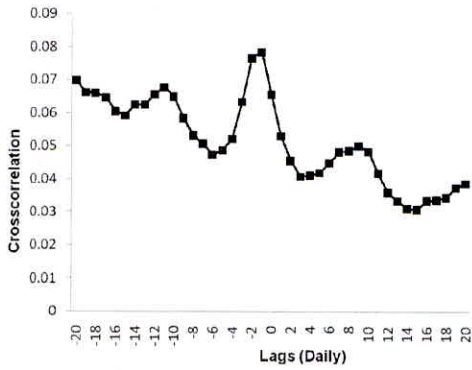


Fig. 9 : The cross correlation between discharge at Rampur and rainfall at Namagia

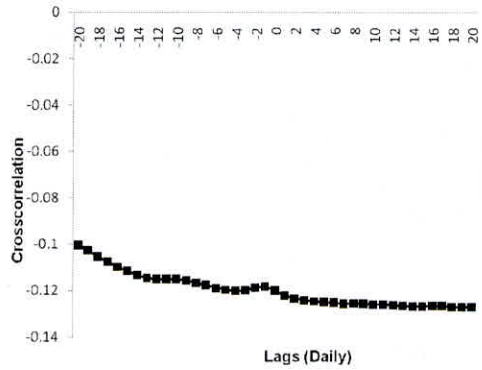


Fig. 12 : The cross correlation between discharge at Rampur and snowfall at Kaza

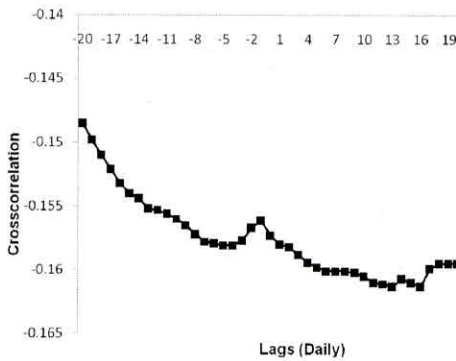


Fig. 10 : The cross correlation between discharge at Rampur and snowfall at Kalpa

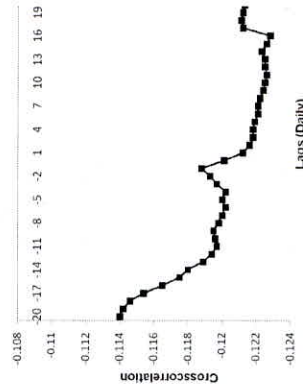


Fig. 13 : The cross correlation between discharge at Rampur and snowfall at Namagia

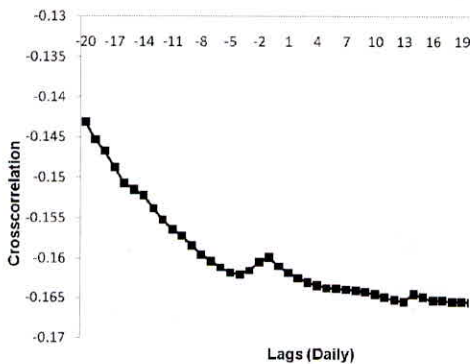


Fig. 11 : The cross correlation between discharge at Rampur and snowfall at Rakccham

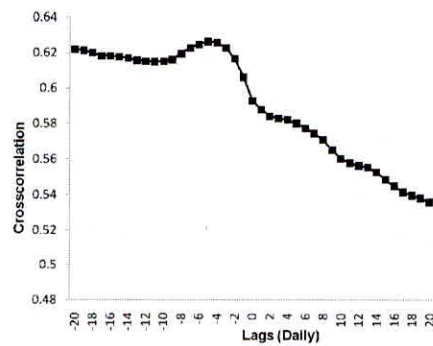


Fig. 14 : The cross correlation between discharge at Rampur and maximum temperature at Rampur

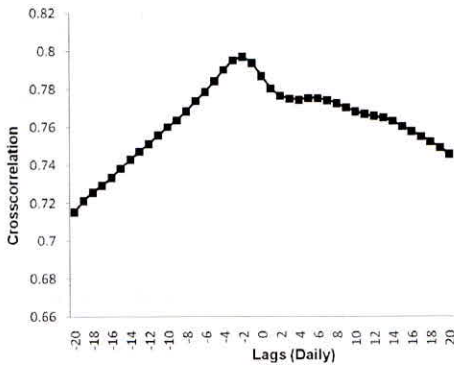


Fig. 15 The cross correlation between discharge at Rampur and minimum temperature at Rampur

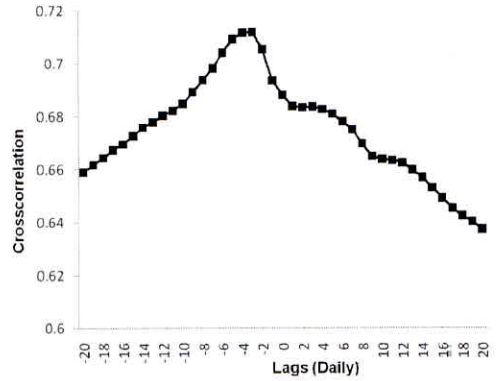


Fig. 18 : The cross correlation between discharge at Rampur and Maximum temperature at Rakccham

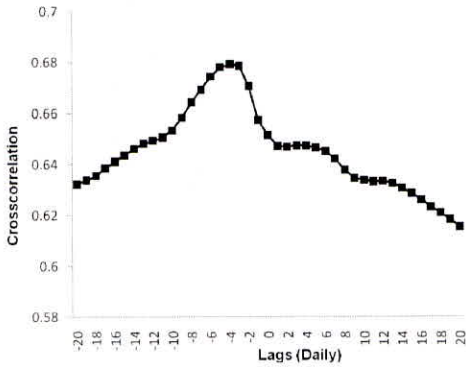


Fig. 16 : The cross correlation between discharge at Rampur and maximum temperature at Kalpa

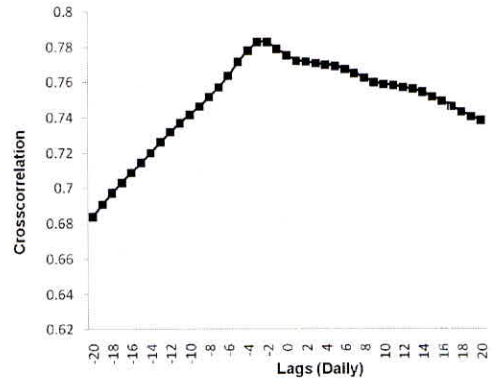


Fig. 19 The cross correlation between discharge at Rampur and minimum temperature at Rakccham

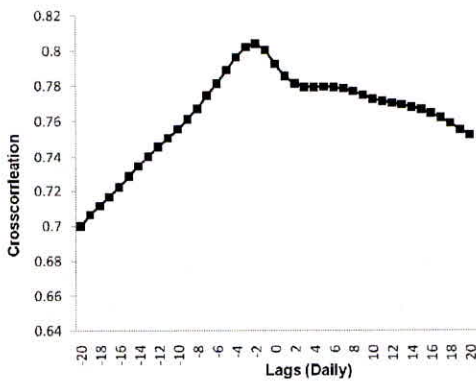


Fig. 17 : The cross correlation between discharge at Rampur and minimum temperature at Kalpa

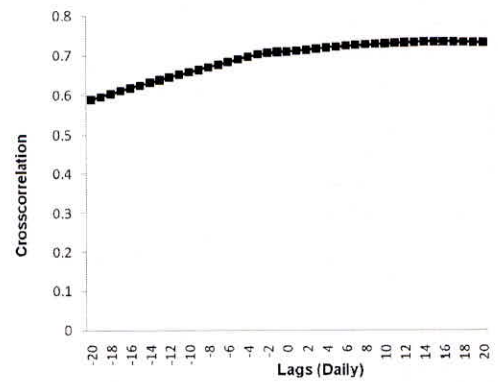


Fig. 20 : The cross correlation between discharge at Rampur and maximum temperature at Kaza

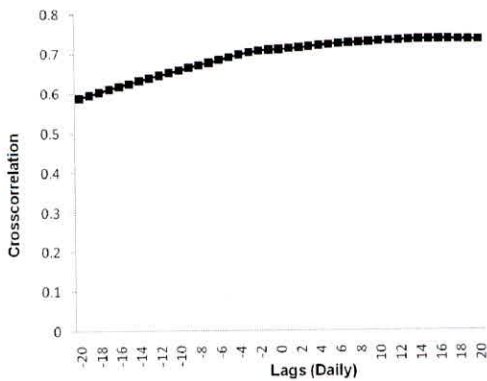


Fig. 21 : The cross correlation between discharge at Rampur and minimum temperature at Kaza

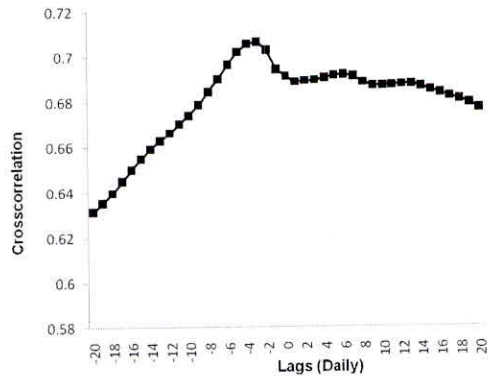


Fig. 22 : The cross correlation between discharge at Rampur and maximum temperature at Namagia

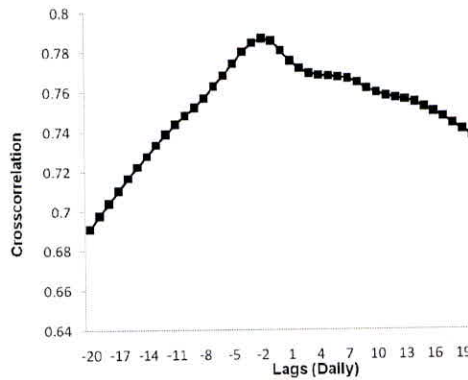


Fig. 23 : The cross correlation between discharge at Rampur and minimum temperature at Namagia

Based on PACF and CCF of the data series, the following input vector was selected for neural network training.

$$\begin{aligned}
 Q_{\text{rampur},t} = f\{ & R_{\text{rampur},t-1}, R_{\text{kalpa},t-1}, R_{\text{rakccham},t-1}, R_{\text{kaza},t-1}, R_{\text{namagia},t-1}, S_{\text{kalpa},t-5}, S_{\text{rakccham},t-4}, S_{\text{kaza},t-4}, \\
 S_{\text{namagia},t-4}, & \text{TMX}_{\text{rampur},t-5}, \text{TMN}_{\text{rampur},t-2}, \text{TMX}_{\text{kalpa},t-4}, \text{TMN}_{\text{kalpa},t-2}, \text{TMX}_{\text{rakccham},t-3}, \\
 \text{TMN}_{\text{rakccham},t-2}, & \text{TMX}_{\text{kaza},t}, \text{TMN}_{\text{kaza},t}, \text{TMX}_{\text{namagia},t-3}, \text{TMN}_{\text{namagia},t-2}, Q_{\text{rampur},t-1} \} \quad (9)
 \end{aligned}$$

In which Q, R, S, TMX and TMN are discharge, rainfall, snowfall, maximum and minimum temperature values respectively.

MODEL TRAINING

The ANN models have been trained using back propagation algorithm. The whole data set were divided into two sets for the training and validation purpose of the ANN model. The data

from 1991 to 2000 were considered for the training of the model since it contained the extreme values of the model. The data of 1987 to 1990 were considered for the validation of the model. The software used for the training of the model was MATLAB (The Mathworks, Inc., 2001). The

From the above table, the model 2 is found to be the best model based on the performance indices during calibration as well as validation of the model. The results of the best ANN model during the calibration indicate that the all range of discharge values was simulated fairly well. However, the medium and high range values of discharge slightly deviated from the observed values during the validation of the model. The overall performance of the model, as exhibited by the various statistical criteria, indicates the suitability of ANN modelling technique to reasonably simulate the stream-flow at Rampur in Sutlej river basin.

Since there is a significant variation in discharge rates during the low discharge and medium to high discharge seasons, so an attempt was also made to evaluate the performance of ANN in modelling these processes separately. For this, two separate ANN models were developed, one for low discharges and other for medium and high discharges using the same input variables, with optimum model structure obtained from the continuous data. The data from November to March were considered as low flow values and those from April to October were considered as medium and high flow values. The results of ANN

models for low, medium and high discharge values during calibration and validation are presented in Figures 26-29 respectively. No improvement in the performance of the ANN models was observed. As a matter of fact, it was found that the generalized ANN model with continuous data, performed better than the ANN models for low discharge and medium and high discharges. The comparison of the results of ANN models for low flow values, medium and high flow values is given in Table 2. It is observed from the results that the generalized ANN can be successfully used to reasonably simulate the streamflow discharges of a river in general and of Sutlej river at Rampur in particular.

CONCLUSIONS

In this paper, ANN Models have been developed for simulating the streamflow at Rampur in Sutlej river using daily discharge, rainfall snowfall and temperature at Rampur and the stations located upstream of Rampur. The statistical parameters ACF, PACF and CCF have been used for selection of Input vector. The performance of the best ANN model for simulating streamflow at Rampur indicated that medium and high flow values were simulated with less accuracy

Model	ANN Structure	Calibration				Validation			
		CORR	EFF	RMSE	Percent error in peak flow estimation	CORR	EFF	RMSE	Percent error in peak flow estimation
General ANN model	11-5-1	0.99	98.56	38.73	-0.79	0.99	97.86	49.56	-0.43
ANN model with low flow values	11-5-1	0.99	98.55	2.21	-0.87	0.96	93.39	4.96	-24.86
ANN model with medium and high Low flow values	11-5-1	0.98	96.79	58.54	-37.41	0.98	95.95	71.07	-17.99

than the low flow values during validation of the model. However, the performance of two separate ANN models for low, medium and high flow values was not better than ANN model with continuous data. It is concluded from the overall performance that ANN model with continuous data is well suitable for simulating the streamflow at Rampur.

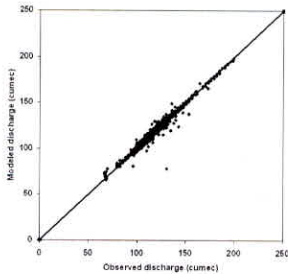


Fig. 26 : Scatter plot of observed Vs modelled discharge at Rampur for ANN model during calibration (Low discharge)

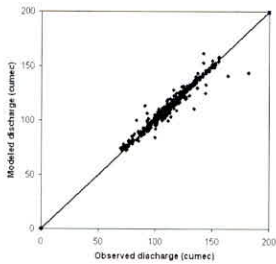


Fig. 27 : Scatter plot of observed Vs modelled discharge at Rampur for ANN model during validation (Low Discharge)

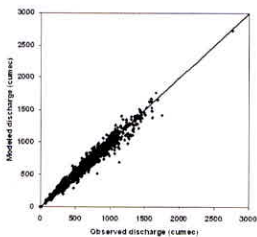


Fig. 28 : Scatter plot of observed Vs modelled discharge at Rampur for ANN model during calibration (Medium and high discharge)

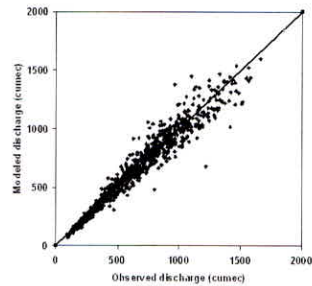


Fig. 29 : Scatter plot of observed Vs modelled discharge at Rampur for ANN model during validation (Medium and high discharge)

REFERENCES

ASCE (2000 a). Artificial neural networks in hydrology-I: Preliminary concepts, *Journal of Hydrologic Engineering*, ASCE, 5(2), pp. 115-123.

ASCE (2000b). Artificial neural networks in hydrology-II: Hydrologic applications, *Journal of Hydrologic Engineering*, ASCE, 5(2), pp. 124-137.

Arora, M. (2008). Seasonal characterization of ablation, storage and drainage of melt runoff and simulation of streamflow for the Gangotri glacier, Project report, D.S.T., GOI.

Hsu, K-L., H.V. Gupta. and S. Sorooshian. (1995). Artificial neural network modeling of the rainfall-runoff process, *Water Resources Research*, 31(10), pp. 2517-2530.

Maier, H.R. and G.C. Dandy. (2000). Neural networks for the prediction and forecasting of water resources variables: A review of modelling Issues and applications, *Environmental Modelling & Software*, 15, pp. 101-124.

Martinez, J., A. Rango. and R. Roberts. (2007). Snowmelt Runoff Model (SRM) – User’s Manual, USDA, 1400 Independence Ave., S.W. Washington, DC 20250.

Nash, J.E. and J.V. Sutcliffe (1970). River flow forecasting through conceptual models:1. A discussion of principles, *Journal of Hydrology*, 10, pp. 282-290.

Parent, A. C., F. Anctil, V. Cantin. and M. A. Boucher (2008). Neural network input selection for hydrological forecasting affected by snowmelt, *Journal of American Water Resources Association*, 44(4), pp. 679-688.

Rumelhart, D. E., E. Hinton. and J. Williams. (1986). Learning internal representation by error propagation, *Parallel Distributed Processing*, Vol. 1, MIT Press, Cambridge, Mass., pp. 318-362.

Sudheer, K. P., A.K. Gosain. and K.S. Ramasastri (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models, *Hydrological Processes*, 16, pp. 1325-1330.

The MathWorks, Inc. (2001). ANN Toolbox User's Guide, 3 Apple Hill Drive, Natick, MA 01760-2098.

Tokar, A.S. and A. Johnson (1999). Rainfall-runoff modeling using artificial neural networks, *Journal of Hydrologic Engineering*, ASCE, 4(3), pp. 232-239.

