The Pharma Innovation Journal 2022; 11(9): 2943-2950

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; 11(9): 2943-2950
© 2022 TPI
www.thepharmajournal.com
Received: 14-07-2022
Accepted: 22-08-2022

Archana Majhi
Department of Soil and Water Conservation Engineering, CAET, Odisha University of Agriculture and Technology, Bhubaneswar, Odisha, India

Sagar Chandra Senapati
Institute of Agricultural Sciences, S'O'A University, Bhubaneswar, Odisha, India

Ambika Prasad Sahu
Department of Soil and Water Conservation Engineering, CAET, Odisha University of Agriculture and Technology, Bhubaneswar, Odisha, India

Janaki Ballav Mohapatra
Department of Agriculture and Farmers' Empowerment, Government of Odisha, Odisha, India

Corresponding Author:
Janaki Ballav Mohapatra
Department of Agriculture and Farmers' Empowerment, Government of Odisha, Odisha, India

Effectuality of neural network in flood forecasting at middle reach of Mahanadi River-Basin

Archana Majhi, Sagar Chandra Senapati, Ambika Prasad Sahu and Janaki Ballav Mohapatra

Abstract
Flood is a severe natural hazard with potentially devastating consequences leading to huge loss of human life, agricultural production and property worldwide. Therefore, the use of flood forecasting and early warning systems is of utmost importance in order to reduce the economic losses and the risk for people. This study presents the application of artificial neural network (ANN) in forecasting lead-time Streamflow discharge at Khairmal station located in the middle reach of Mahanadi River basin using previous discharge data as well as discharge data from three upstream stations (Kantamal, Kesinga and Salebhata) as inputs to the model. Moreover, the study also investigated the effect of length of training data period on network architecture and model performance for attaining the best model efficiency. The findings of the study revealed about reasonable forecast of the one- and two-days ahead Streamflow discharge without relying on other information of the region. The 2003-2008 period trained model predicted the one-day ahead discharge accurately whereas, the 2003-2010 period trained model was found to be superior in predicting the two-days ahead Streamflow discharge. Further, the study inferred that the model trained with shorter training datasets requires complex architecture as compared to the model trained with longer training datasets.

Keywords: Flood, forecasting, artificial neural network, Mahanadi river-basin, Odisha

1. Introduction

Flood is one of the natural disasters that create tremendous havoc and myriad miseries in the affected area which causes loss of life, disruption of human activities, damage to properties, agricultural crops and health hazard. Many districts in Odisha, namely Sambalpur, Subarnapur, Boudh, Nayagarh, Cuttack, Jagatsinghpur and Kendrapara have been affected by flood and heavy rainfall causing damage to property and impacting the normal life of the people in the state of Odisha. On analysis of the past and the present flood scenarios of Odisha, it is found that the Mahanadi River causes the maximum numbers of floods with high magnitude and massive loss factors resulting in devastation of the downstream areas in the coastal tract of Odisha (Beura 2015) [5]. Therefore, there is an urgent need for systems capable of efficiently forecasting water levels or discharge rates in rivers.

Due to the spatial and temporal variation of the rainfall distribution and the inordinately complex and highly non-linear nature of rainfall-runoff relationship, flood forecasting remains one of the most challenging and important tasks of operational hydrology (Chang *et al.* 2007) [7]. Generally, flood forecasting is carried out using physically based, conceptual and black box approaches. Although conceptual and physically based models are reliable in forecasting the important hydrograph features, implementation and calibration of such models can have several difficulties, including development of sophisticated mathematical tools, estimation of many parameters for modeling (Duan *et al.* 1992; Grayson *et al.* 1992) [8, 9]. An approach based on physics is still far from being realized and researchers have therefore, focused attention on the use of data driven techniques in the recent years. In this study, one such data driven technique, the artificial neural networks (ANNs) approach is pursued for predicting daily streamflow discharge using lagged variables. ANN has gained momentum in last few decades for river flow forecasting and has been accepted as a good alternative to physically based models and conceptual models (ASCE, 2000a and 2000b) [3-4]. Recently, many works have been carried out on the use of ANNs for rainfall-runoff modeling (Tokar and Johnson 1999; Rajurkar *et al.* 2004; Kumarasiri and Sonnadara 2008) [17, 14, 10], river flow prediction (Abrahart and See 2000; Agarwal and Singh 2004; Panda *et al.* 2010) [1, 2, 13] and also flood forecasting (Campolo *et al.* 1999; Tiwari and Chatterjee 2010a; Tiwari and Chatterjee 2010b) [6, 15-16].

The neural networks are capable of performing non-linear modeling without prior knowledge about the relationship between input and output variables which makes ANNs a general and flexible alternative modelling tool for hydrological time series.

The study on flood forecasting plays a significant role in saving human lives and helps in organizing timely rescue and flood fighting measures in order to prevent or minimize the damage to flood protection works like embankments. Flood forecasting is vital for developing flood warning systems, flood prevention, flood damage mitigation and soil erosion reduction measures especially for flood prone regions of the state. Keeping the above in view, the present study aims to present an application of ANN for forecasting one-day and two-days ahead streamflow discharge at Khairmal gauging station using its past discharge data along with discharge data from the upstream stations as inputs. Also, the study explores the effect of training data length on model efficiency and architecture for attaining accurate predictions.

2. Materials and Methods

2.1 Study Area

The Mahanadi river-basin is the fourth largest river basin of India. The catchment area of the basin is 141,589 km^2 which accounts for 4.3% of the total geographical area of India (Fig. 1). It extends from 19° 21' to 23° 35' latitude and from 80° 30' to 86° 50' E longitude. About 53% (75, 136 km^2) of the basin is in the state of Chhattisgarh, 46% (65, 580 km^2) is in the coastal state of Odisha, and the remainder of the basin is in the states of Jharkhand and Maharashtra. The middle reach of Mahanadi river-basin located in Odisha between 19° N 82° E and 22° N 86° E is chosen as the study area for this work.

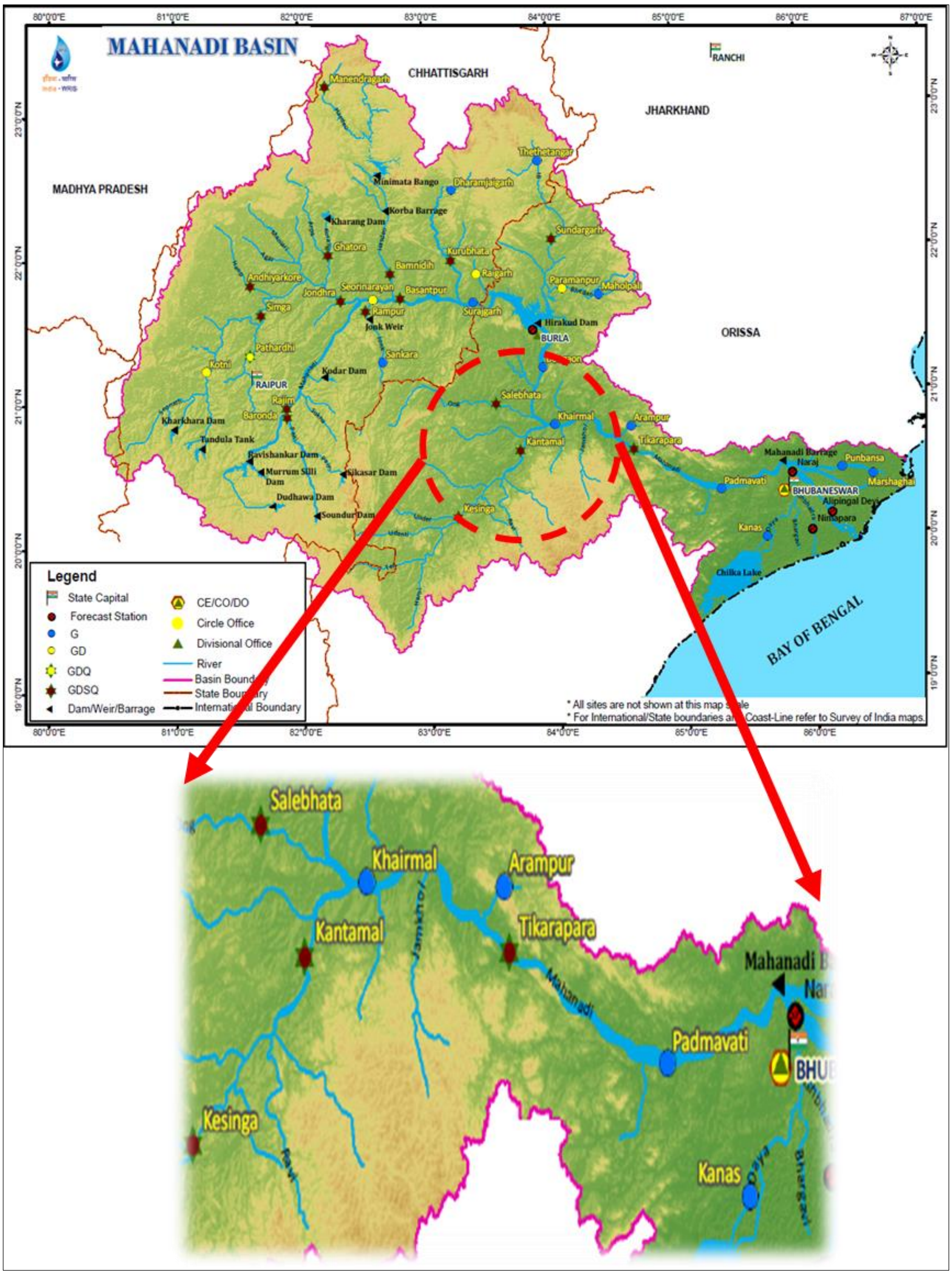

Fig 1: Map of study area

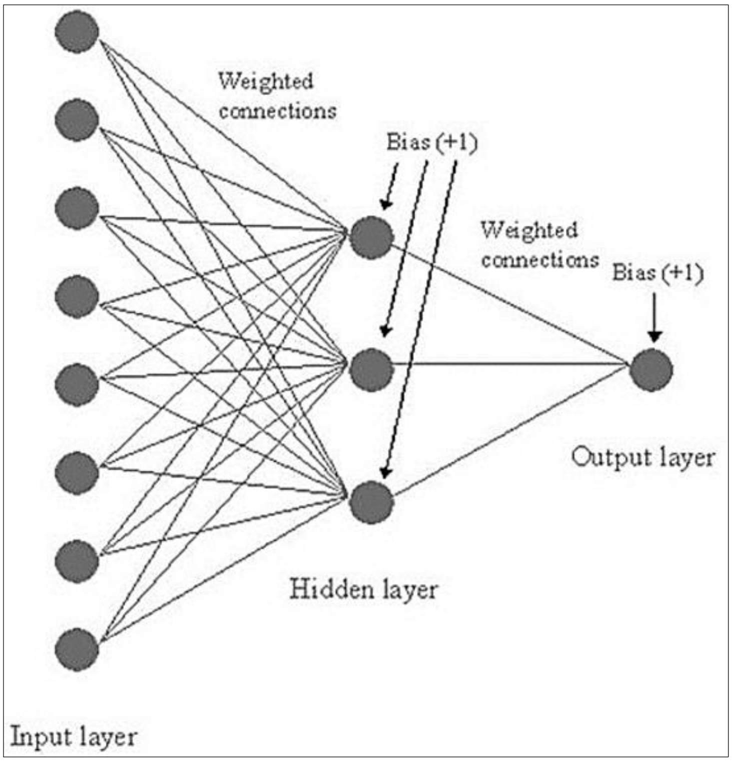

Fig 2: Basic overview of a typical Multi-Layer Perceptron (MLP) topology.

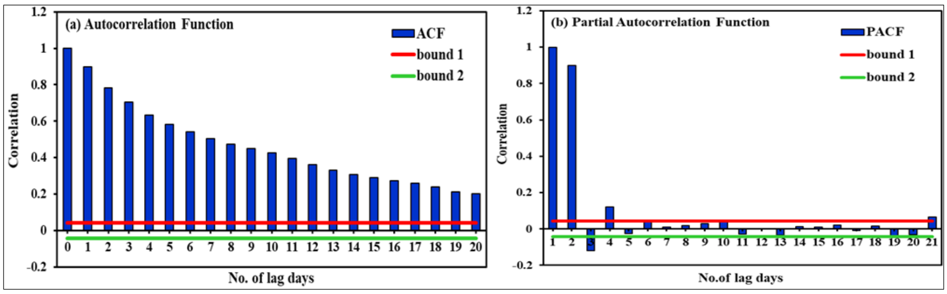

Fig 3: Results of (a) Autocorrelation Function (Left) and (b) Partial Autocorrelation Function (Right) of streamflow discharge at Khairmal gauging station.

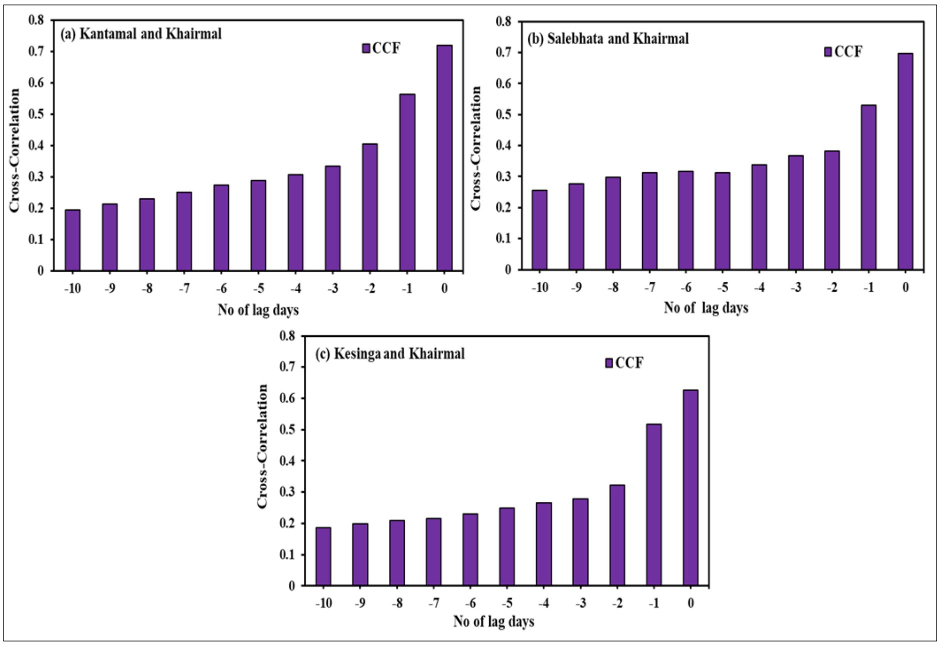

Fig 4: Results of cross-correlation function between Streamflow discharges of (a) Kantamal and Khairmal stations (Top Left); (b) Salebhata and Khairmal stations (Top Right); and (c) Kantamal and Khairmal stations (Bottom).

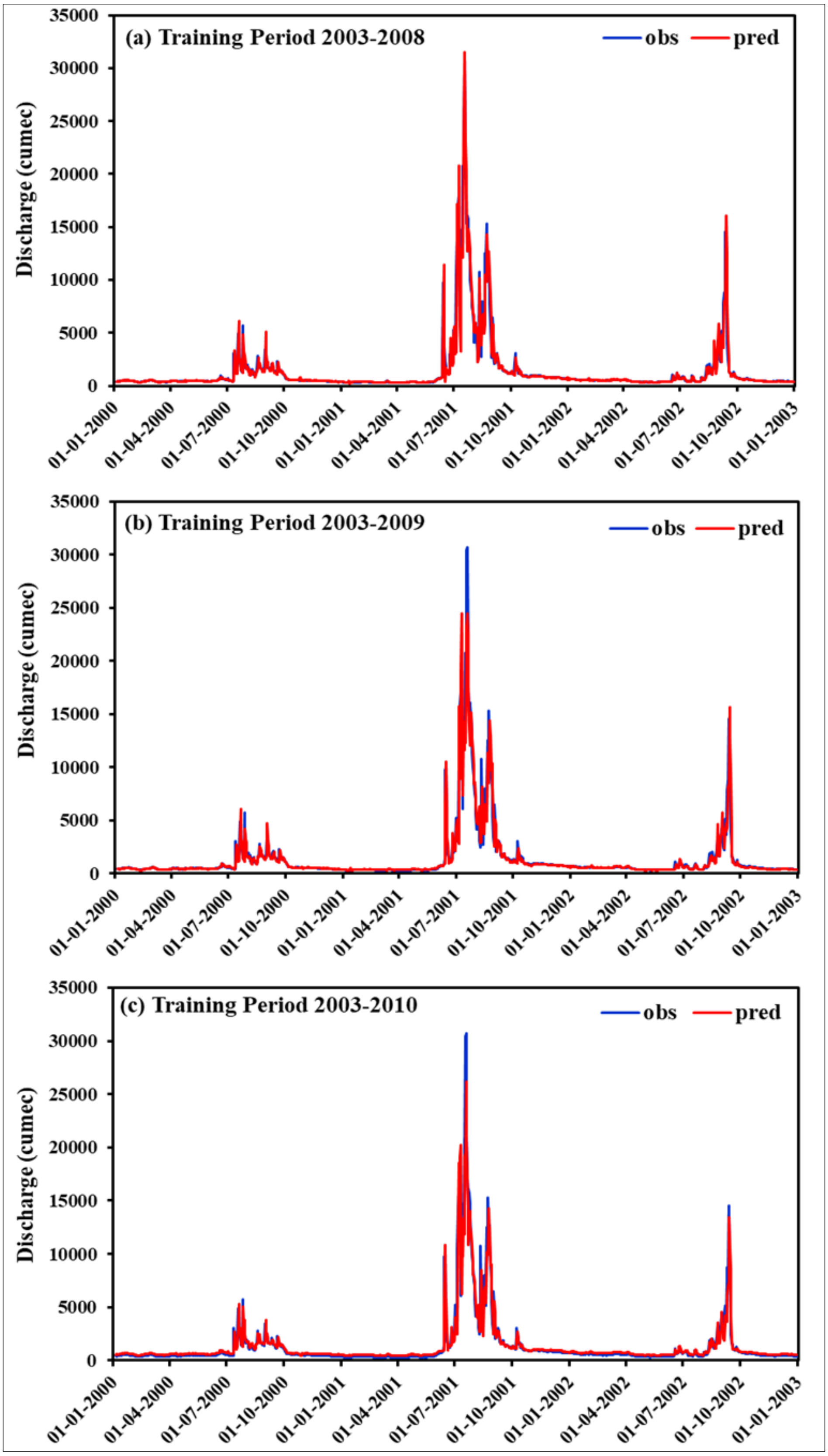

Fig 5: Comparison of observed and model predicted hydrographs for 1-day lead time forecasting during the testing period (2000-2002) by using training period of (a) 2003-2008; (b) 2003-2009; and (c) 2003-2010

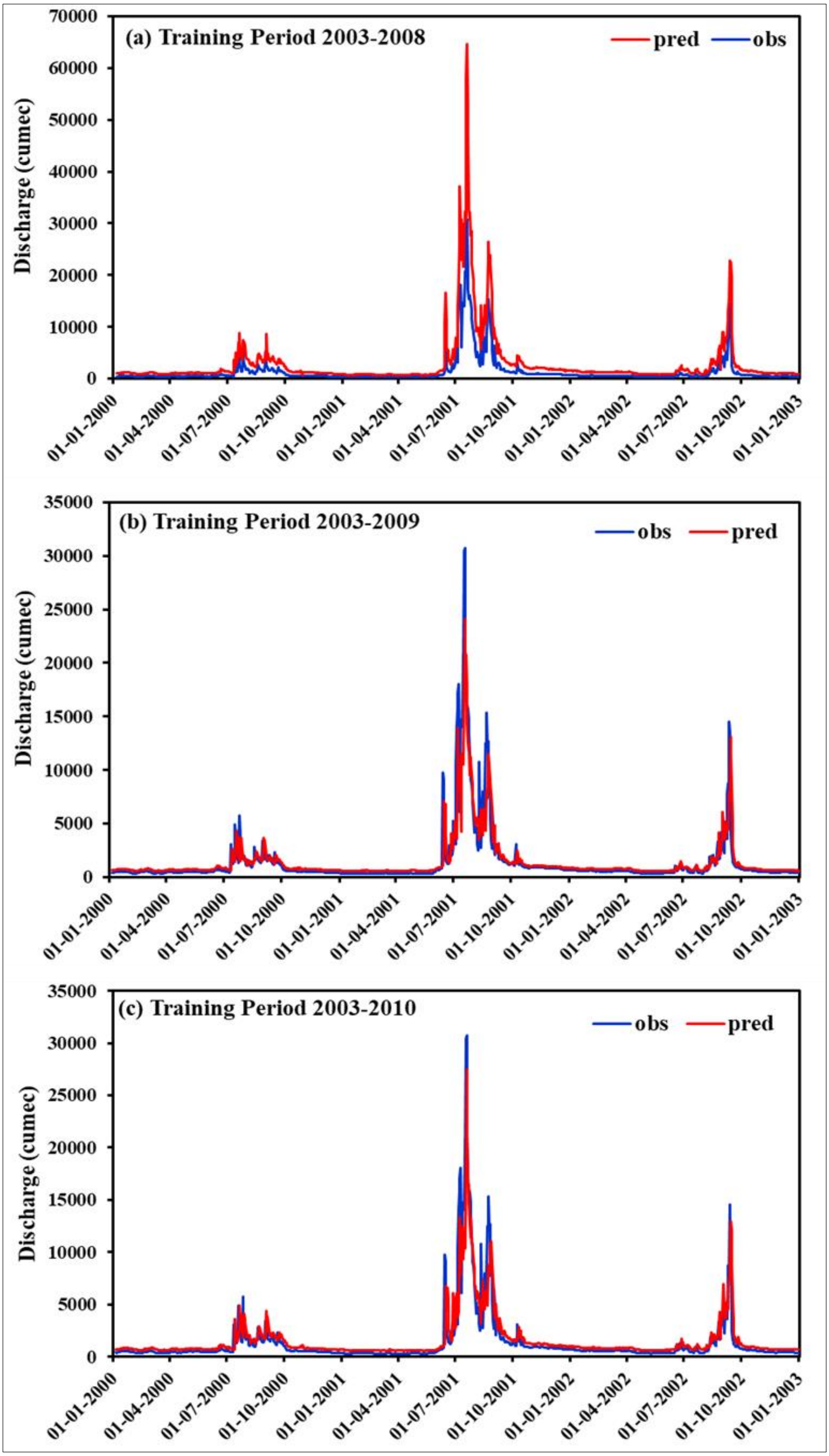

Fig 6: Comparison of observed and model predicted hydrographs for 2-day lead time forecasting during the testing period (2000-2002) by using training period of (a) 2003-2008; (b) 2003-2009; and (c) 2003-2010

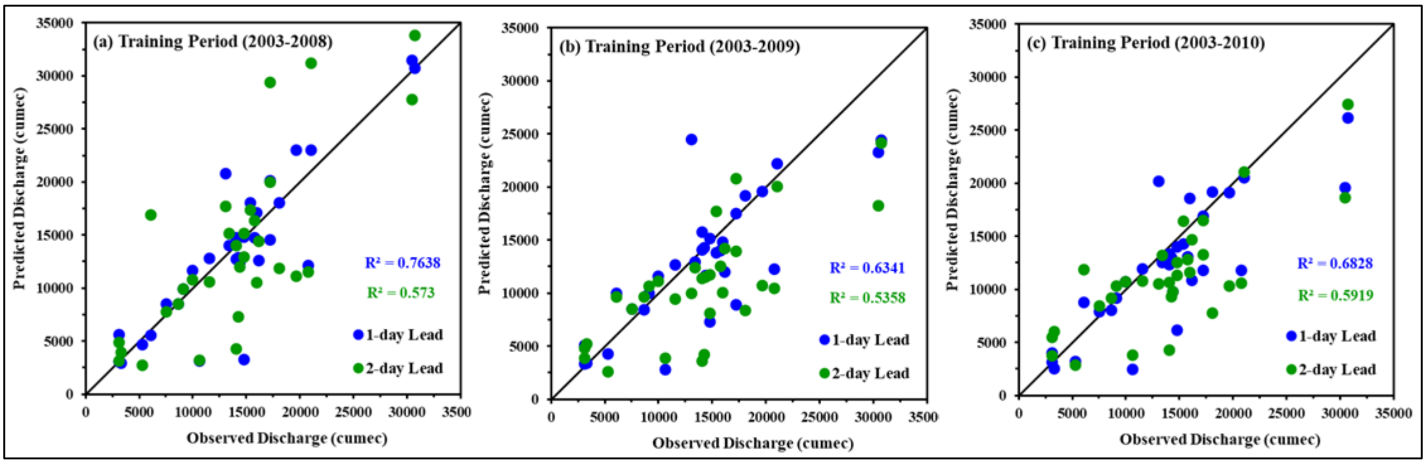

Fig 7: Scatter plots of observed and model predicted discharge during the testing period (2000-2002) by using training period of (a) 2003-2008; (b) 2003-2009; and (c) 2003-2010

2.2 Streamflow Discharge Data

The daily discharge data of four gauging stations [Kesinga (Kalahandi), Kantamal (Boudh), Khairmal (Kalahandi) and Salebhata (Bolangir)] during 2000-2010 in the middle reach of Mahanadi river-basin was obtained from the Central Water Commission, Bhubaneswar. The data was pre-processed in order to handle missing values in the time series using MATLAB R2013a software. Further, the data was transformed between zero and one for nullifying the effect of larger and smaller magnitudes which may confuse the learning algorithm on the importance of each variable leading to rejection of any variable with smaller magnitude (Luk *et al.* 2000) [11]. The data was standardized using the following expression:

$$x_s = \frac{x - \mu}{\sigma} \qquad (1)$$

Where
x_s = Standardized value
x = non-standardized value
μ = mean; and
σ = standard deviation. However, the model outputs were de-standardized for retrieving the predicted discharge data.

2.3 Development of Artificial Neural Network

Artificial Neural Network (ANN) is based on pattern recognition which explores the relationship between sets of inputs and desired outputs without giving any information about the actual processes involved. It is analogous to a biological neuron as it consists of a large number of simple processing units called neurons or nodes linked by weight connections (Mohapatra *et al.* 2021) [12]. The most commonly used multi-layer perceptron (MLP), as shown in Fig. 2, is a feed-forward network with one input layer, one output layer and a minimum of one hidden layer which is trained by static back propagation. The optimal number of hidden neurons was determined by trial-and-error approach. In addition, mean square error was used as the performance function for the model. The activation functions of hidden and output layers were considered as 'hyperbolic tangent' and 'linear' transfer function, respectively for the present study. Moreover, Levenberg-Marquardt algorithm was chosen as the training algorithm as it is regarded as the fastest and highly recommended backpropagation algorithm. The Neural Network Toolbox of MATLAB R2013a was employed for developing the ANN model in this study.

2.3.1 Selection of model inputs

One of the most important steps in ANN modelling is the selection of significant input variables for the model. The input datasets of this study include discharge data of Khairmal station from previous time-step along with the discharge data of neighboring-upstream stations such as Salebhata, Kantamal and Kesinga gauging stations. The most significant inputs for daily discharge forecasting were selected using autocorrelation, partial autocorrelation and cross-correlation techniques. The significant inputs from Khairmal station were determined using autocorrelation and partial autocorrelation statistics, whereas the cross-correlation statistics was used for finding significant inputs from Kantamal, Kesinga and Salebhata gauging stations.

2.3.2 Training and testing of model

A total available data of 11 years (2000-2010) was divided into training and testing sets for the developed ANN model. The weights were adjusted in order to make the model predicted values closer to the target outputs of the network during the training period, whereas the performance of the trained model was evaluated by exposing it to unseen data during the testing period. For the present study, six (2003-2008), seven (2003-2009) and eight years (2003-2010) of the available data were used for training and the remaining three years (2000-2002) were used for testing the model.

2.3.3 Evaluation of model performance

The performance of the developed ANN model was evaluated during training and testing periods in order to examine the model effectiveness in prediction using statistical and graphical indicators. Three statistical indicators, namely root mean square error (RMSE), correlation coefficient (r) and Nash-Sutcliffe efficiency (NSE) were used which are as follows

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \qquad (2)$$

$$r = \frac{\sum_{i=1}^{n}(O_i - \bar{O}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O}_i)^2 \sum_{i=1}^{n}(P_i - \bar{P}_i)^2}} \quad (3)$$

$$NSE = \left[1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O}_i)^2}\right] \quad (4)$$

Where,
O = observed discharge; *P* = predicted discharge; n = number of data points; $\bar{O}$ = mean observed discharge; and $\bar{P}$ = mean predicted discharge. Besides statistical indicators, graphical indicators in form of simultaneous and scatter plots with 1:1 line of observed and predicted discharges were also prepared for examining the model performance.

3. Results and Discussion

3.1 Model Inputs

The autocorrelation function showed a significant correlation for one to more than 20 days, whereas the partial autocorrelation function revealed a significant correlation up to three-days lagged discharge at Khairmal station [Fig. 3(a-b)]. Thus, only three-days lagged discharge data (Q_t, Qt_{-1}, Q_{t-2}, Q_{t-3}) has been considered as inputs from the gauging station. The cross-correlation between Khairmal and Kantamal indicated significant correlation up to two-days lag (Q_t, Qt_{-1}, Q_{t-2}) [Fig. 4(a)]. Similarly, significant correlation up to one-day lag was observed for Salebhata and Kesinga stations from the analysis of cross-correlation function (Q_t, Qt_{-1}) [Fig. 4(b-c)]. The total numbers of significant inputs considered for this study are presented in Table 1 and hence, 11 input nodes were finalized for the model.

Table 1: Most significant inputs for forecasting lead time discharge

Stations	Input variables	Output Variable
Kantamal	Q_{t-2}, Q_{t-1}, Q_t	Q_{t+1} or Q_{t+2} of Khairmal Gauging Station
Kesinga	Q_{t-1}, Q_t	
Salebhata	Q_{t-1}, Q_t	
Khairmal	Q_{t-3}, Q_{t-2}, Q_{t-1}, Q_t	

Note: Q_t indicates the Streamflow discharge of the station at time Step 't'.

3.2 One-Day Lead Time Flow Forecasting using Different Lengths of Training Data

Using the shortest training period (2003-2008), the developed model performed the best with 20 number of hidden nodes in forecasting one-day ahead stream flow discharge with a correlation coefficient (r) of 0.956, Nash-Sutcliffe efficiency (NSE) of 0.912 and root mean square error (RMSE) of 851.15 m^3/s. Likewise, the model, trained with a period of 2003-2009, recorded a value of 0.945, 0.894 and 931.13 m^3/s for the statistical indicators r, NSE and RMSE, respectively. On the other hand, the one-day lead time forecasting was found to be poor for the longest training period (2003-2010) returning indicator values of r, NSE and RMSE as 0.948, 0.894, 932.93 m^3/s, respectively. Further, the models developed with training period 2003-2009 and 2003-2010 underestimated the peak values of the observed data as shown in Fig. 5.

3.3 Two-Day Lead Time Flow Forecasting using Different Lengths of Training Data

The model developed with seven hidden nodes employing the training period 2003-2010 for predicting two-days ahead Streamflow discharge outperformed others models trained with shorter periods which is indicated by correlation coefficient (r) of 0.904, NSE of 0.804 and RMSE of 1267.62 m^3/s. The 2003-2008 trained model returned r value of 0.905, NSE of 0.801 and RMSE of 1282.84 m^3/s, whereas the 2003-2009 trained model recorded figures of 0.896, 0.792 and 1307.24 m^3/s, respectively. It is evident from Fig. 6 that the model developed with 2003-2008 period over-predicts the observed values in most of the cases whereas, the model developed with 2003-2009 under-predicts most of the observed values.

3.4 Overall Performance of the Models

Based on the efficacy of the model in forecasting Streamflow discharge using different lengths of training period, it is clear that the performance of the model developed with shortest training period (2003-2008) is superior than other developed models for forecasting one-day ahead stream flow discharge whereas, the model developed with longest training period (2003-2010) outperforms other developed models in predicting two-days ahead stream flow discharge. This indicates regarding the requirement of long training datasets for forecasting higher lead time stream flow discharge. Moreover, it is worthwhile to note that the ability of the model to simulate the observed data decreases considerably with increase in lead time forecast, i.e., the model predictions in forecasting one-day ahead discharge are more accurate than the model predictions in forecasting two-days ahead discharge. The scatter plots of predicted versus observed discharge values along with 1:1 line for three different training length periods were shown in Fig.7. These figures strengthen the findings of the study as the one-day ahead predicted discharge values are closer to the actual values as compared to the two-day ahead predicted discharge values. Further, it can also be confirmed that with the increase in the length of the training period, the requirement of hidden nodes also decreases in order to attain the best model performance. In other words, the model architecture becomes more complex when the model is developed with short training length data. In contrast to it, simple model architecture attains optimum model efficiency when developed with longer training data.

Table 2: Performance of ANN model in forecasting 1-day and 2-day lead time discharge under different training periods

Sl. No.	Lead Time Forecast	Training Period	Model Architecture (Input-Hidden-Output)	Testing Period (2000-2002)		
				RMSE (Cumec)	r	NSE
1	Q_{t+1}	2003-2008	11-20-1	851.15	0.956	0.912
		2003-2009	11-19-1	931.13	0.945	0.894
		2003-2010	11-8-1	932.93	0.948	0.894
2	Q_{t+2}	2003-2008	11-11-1	1282.84	0.905	0.801
		2003-2009	11-9-1	1307.24	0.896	0.792
		2003-2010	11-7-1	1267.62	0.904	0.804

Note: 1. Q_t indicates the Streamflow discharge of the station at time Step 't'.
The rows highlighted with light grey background shows the best model performance.

4. Conclusions

Artificial neural network (ANN) becomes the most viable option for flood forecasting in the data scarce regions and it is often preferred over other models requiring several variables for producing accurate data. Also, ANNs offer a means of reducing the analytical costs of topographical and hydrological information by reducing the amount of time spent for analyzing the data. The goal of the present study was to forecast one-day and two-days ahead streamflow discharge at Khairmal gauging station using its past discharge values and the discharge values of its neighboring stations (Kantimal, Kesinga and Salebhata) as inputs to the ANN model developed with three training periods (2003-2008, 2003-2009 and 2003-2010). The findings of the study reveal about reasonable forecast of the Streamflow discharge by the model developed using three training length data without relying on other information such as meteorology, hydrology, topography, etc. The model developed with training period of 2003-2008 predicted the one-day ahead discharge data accurately (RMSE = 851.51 m^3/s, r = 0.956, NSE = 0.912) whereas, the developed model using training period of 2003-2010 was found to the best in predicting the two-days ahead stream flow discharge (RMSE = 1267.62 m^3/s, r = 0.904, NSE = 0.804). However, the performance of the model reduces with increase in the lead time forecast. Further, complex model architecture is required while using shorter training length data for attaining the best model efficiency which is not in the case of longer training period. The findings of the study are expected to assist policy makers and water managers in improving the flood forecasting and warning systems for reducing economic losses and human life in the downstream areas of Mahanadi river-basin.

5. Acknowledgments

The authors are thankful to the Central Water Commission, Bhubaneswar for providing Streamflow discharge data at four stations of middle Mahanadi river-basin.

6. References

1. Abrahart RJ, See L. Neural network vs. ARMA modeling: constructing benchmark case studies of river flow prediction. Hydrological Processes. 2000;14:2157-2172.
2. Agarwal A, Singh RD. Runoff modelling through back propagation artificial neural network with variable rainfall-runoff data. Water Resources Management. 2004;18(3):285-300.
3. ASCE Task Committee on application of artificial neural networks in hydrology. Artificial neural networks in hydrology I: Preliminary concepts. Journal of Hydrologic Engineering. 2000a;5(2):115-123.
4. ASCE Task Committee on application of artificial neural networks in hydrology. Artificial neural networks in hydrology II: Hydrologic applications. Journal of Hydrologic Engineering. 2000b;5(2):124-137.
5. Beura D. Floods in Mahanadi River, Odisha, India: Its Causes and Management. International Journal of Engineering and Applied Sciences. 2005;2(2):51-55.
6. Campolo M, Andreussi P, Soldati A. River flood forecasting with a neural network model. Water Resources Research. 1999;35(4):1191-1197.
7. Chang FJ, Chiang YM, Chang LC. Multi-step-ahead neural networks for flood forecasting. Hydrological Sciences. 2007;52(1):114-130.
8. Duan Q, Sorooshian S, Gupta VK. Effective and efficient global optimization for conceptual rainfall–runoff models. Water Resources Research. 1992;28(4):1015-1031.
9. Grayson RB, Moore ID, McMahon TA. Physically based hydrologic modelling: 2. is the concept realistic? Water Resources Research. 1992;28(10):2659-2666.
10. Kumarasiri AD, Sonnadara UJ. Performance of an artificial neural network on forecasting the daily occurrence and annual depth of rainfall at a tropical site. Hydrological Processes. 2008;22(17):3535-3542.
11. Luk KC, Ball JE, Sharma A. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. Journal of Hydrology. 2000;227(1-4):56-65.
12. Mohapatra JB, Jha P, Jha MK, Biswal S. Efficacy of machine learning techniques in predicting groundwater fluctuations in agro-ecological zones of India. Science of the Total Environment. 2021;785:147319.
13. Panda RK, Pramanik N, Bala B. Simulation of river stage using artificial neural network and MIKE 11 hydrodynamic model. Computers & Geosciences. 2010;36(6):735-745.
14. Rajurkar MP, Kothyari UC, Chaube UC. Modeling of the daily rainfall–runoff relationship with artificial neural network. Journal of Hydrology. 2004;285(1-4):96-113.
15. Tiwari MK, Chatterjee C. Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). Journal of Hydrology. 2010a;382(1-4):20-33.
16. Tiwari MK, Chatterjee C. Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. Journal of Hydrology. 2010b;394(3-4):458-470.
17. Tokar AS, Johnson PA. Rainfall-runoff modeling using artificial neural networks. Journal of Hydrologic Engineering. 1999;4(3):232-239.