

Dual Artificial Neural Network for Rainfall-Runoff Forecasting

Pallavi Mittal, Swaptik Chowdhury, Sangeeta Roy, Nikhil Bhatia, Roshan Srivastav

MBS, VIT University, Vellore, India

Email: roshan1979@gmail.com

Received October 1, 2012; revised November 3, 2012; accepted November 14, 2012

ABSTRACT

One of the principal issues related to hydrologic models for prediction of runoff is the estimation of extreme values (floods). It is well understood that unless the models capture the dynamics of rainfall-runoff process, the improvement in prediction of such extremes is far from reality. In this paper, it is proposed to develop a dual (combined and parallel) artificial neural network (D-ANN), which aims to improve the models performance, especially in terms of extreme values. The performance of the proposed dual-ANN model is compared with that of feed forward ANN (FF-ANN) model, the later being the most common ANN model used in hydrologic literature. The forecasting exercise is carried out for hourly river flow data of Kolar Basin, India. The results of the comparison indicate that the D-ANN model performs better than the FF-ANN model.

Keywords: Forecasting; Hybrid Model; ANN; Floods; Non Linear

1. Introduction

One of the most important topics in water resources development and management is rainfall-runoff forecasting. A future aspect of this modelling is to reduce flood risks by providing a flood warning system which includes a complex relationship between precipitation and runoff. This complexity is occurred due to inconsistency of watershed characteristics, non uniformity in precipitation, as well several other factors involved in runoff generation where dominant ones are evaporation, infiltration, soil moisture, overland flow and channel flow [1].

To enhance the understanding of rainfall-runoff process a large number of studies conducted till now used models developed either on physical deliberation (physical considerations) of the process or on the basis of a theoretic approach (systems operation). In spite of providing appropriate accuracy, the implementation of such models can generally result in different complications [2]; hence requiring ambiguous statistical implements, and some extent of proficiency and experience with the model. Usual theoretic models like autoregressive models and their variations [3] experiences from being based on the linear systems theory and may only be slightly appropriate in capturing the highly complex, vibrant, and nonlinear rainfall-runoff process [4,5]. Hence due to the complexity associated with parameter optimization in nonlinear systems, the progress of nonlinear system theoretic models are very restricted [6] and are not very

popular in terms of flood forecasting.

Recently the application of artificial neural networks (ANNs) has marked an impact in the area of hydrological modelling. ANNs are fundamentally semi-parametric regression estimators which are well-matched for hydrological modelling, as they can predict virtually any (measurable) function up to a random degree of precision [7]. Major benefit of this approach over previous methods is the lack of complexity in the statistical form representation *i.e.* no precise process for algorithmically converting an input to an output is required. The only requirement of this network is a collection of representative examples for the required mapping. The ANN then adapts itself to reproduce the desired output when accessible with training model input. The demonstration of neural network technology has provided many remarking results in the area of hydrology and water resources model.

Drawback of this vast amount of network theory has been indicated as their incapability of predicting extreme values in the river flow [8-10] which has given rise to record-breaking downpour and famine conditions. Imrie *et al.* [11] argue that there may be a number of reasons why ANN models are incapable of predicting extreme values, and a range of remedies have been planned [6,12].

This paper addresses this drawback of extreme value forecast in ANN-based runoff flow modelling through

discussion of the probable causes, and thereby developing a new dual ANN (D-ANN) based rainfall-runoff modelling. The performance of this proposed model is illustrated by a real case study of Kolar basin, India. The performance of the proposed D-ANN is compared with a feed forward neural network (FF-ANN) model developed for the same basin and is discussed in the following sections.

The subsequent paper is organised as follows. In Section 2, proposed modelling framework is presented and also a brief introduction on ANN. Following this, in Section 3 the case study on Kolar basin is presented. Section 4 outlines the results and discussions of the present study. Section 5 includes summary and conclusions of the present study and scope for future work.

2. Model Development

In this section basic ANN framework has been discussed which is followed by proposed methodology of D-ANN model.

2.1. Artificial Neural Networks

ANNs are highly simplified mathematical models and computing techniques inspired by biological neural networks. It can be categorized as interconnected groups of simple neurons that function as a combined system for processing information and model complex relationships between inputs and outputs by finding patterns in data. The FF-ANN trained with the back propagation algorithm is perhaps the most popular network for hydrologic modelling [13,14]. This network topology which acts an adaptive system consists of simple artificial nodes (neurons) connected together by links to form a network of nodes usually organized in a number of layers hence the term artificial neural network. Weighted input from previous layer is received and processed output is transmitted to following layer through links. Mostly ANNs have three or more layers: an input layer for presenting data to network, an output layer for producing an appropriate response and intermediate (hidden) layer for collecting feature detectors. Present study highlights on a model back propagation algorithm for training, and the number of hidden neurons is optimized by a trial and error process. The basic structure of the ANN model is shown in **Figure 1**.

Let y and \hat{y} be the actual and the predicted value of ANN model respectively and are related by,

$$y = \hat{y} + \varepsilon \tag{1}$$

where, ε is the residual error in the forecast of the runoff value.

The predicted value of runoff y can be obtained from the following general form of the ANN equation

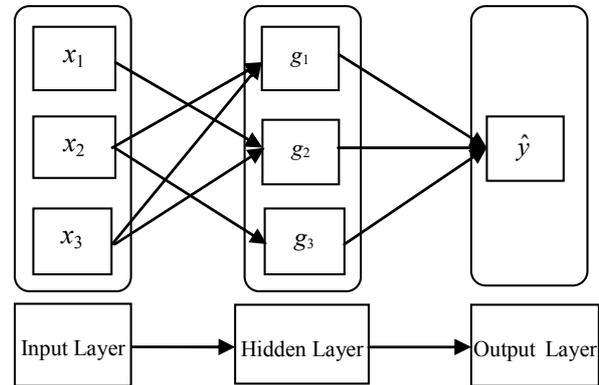


Figure 1. Structure of the feed forward ANN model.

$$\hat{y} = g \left[\beta_i h \left(\sum_{i=1}^n x_i \alpha_i + \phi \right) \right] \tag{2}$$

where,

x_i is the input variables;

α is a weight connecting input node to hidden node;

β is a weight connecting hidden node to output node;

ϕ, φ are the biases at hidden and output nodes respectively;

and $g(), h()$ are the activation functions at hidden and output layers respectively.

2.2. Proposed Dual-ANN Model

The main aim of a D-ANN model is to estimate the error along with the predicted value. The general form of the predicted value is given by

$$\hat{Y} = f(X) \tag{3}$$

where

X is an n -dimensional input vector consisting of variables $x_1 \dots x_i, \dots, x_n$;

Y is a m -dimensional output vector consisting of resultant variables $y_1 \dots y_i \dots y_m$.

In the current modelling vector X comprises of both rainfall and runoff values at recurrent priority time lags and the vector Y is usually the flow for a consecutive period or at a different particular site.

Information is processed in D-ANN on the basis of learning method which is a nonlinear alteration of link weights so that the network can produce an approximate output. In general, in this process the network changes its structure and the strength of the existing matrix of nodal weights is increased. Hence, the probability of achieving similar outputs for same inputs increases. In addition to develop a relation between the input vector and output vector, it is suggested to use another subsequent relationship between the input variables and the errors from the earlier network. The detail steps involved in dual-ANN model are as follows.

Step 1: Compilation of the statistics of rainfall (R) and the corresponding runoff (Q). The subsequent relation can be derived,

$$Q_t = f(R_{t-9}, R_{t-8}, R_{t-7}, Q_{t-1}, Q_{t-2}) \quad (4)$$

Step 2: Allotment of the patterns in the calibration data set and the validation data set. Evaluation and estimation of the predicted values and errors of the runoff values (of calibration data set),

$$\varepsilon = y - \hat{y} \quad (5)$$

where, ε -Value of Error, y -Observed value of runoff and \hat{y} -redicted value of runoff.

Step 3: Development of the relation,

$$\varepsilon_t = f(R_{t-9}, R_{t-8}, R_{t-7}, Q_{t-1}, Q_{t-2}) \quad (6)$$

Now an additional model is trained, to estimate the value of error corresponding to the predicted runoff, the value of $\hat{\varepsilon}$. After validating the model, D-ANN can be used for the forecast of the runoff value (\hat{y}) associated to the specified inputs using the relation,

$$\hat{y} = \hat{y} + \hat{\varepsilon} \quad (7)$$

Figure 2 illustrates the methodology of the D-ANN pictorially.

3. Case Example

The application of the proposed D-ANN model is carried out on a real case study on Kolar river basin, in India (**Figure 3**). The Kolar basin is a descendant of the river Narmada. The basin has a total drainage area of about 1350 km² which constitutes an area of 903.87 km² lies between north latitude 21°90' - 23°17' and east longitude 77°10' - 77°29'. The climate of the basin is humid and landscape of the Kolar basin is hilly consisting of mainly black soil. The basin can be divided into three distinct zones: low land areas, hilly slopes or semi hilly areas, and upland or hilly areas.

Data are collected during monsoon season during years 1987 to 1989. This available data is divided into two sets, calibration set (data during years 1987-1988) and validation set (data during year 1989). Parameters of the model are obtained using calibration data set where

as the model is tested using validation set. The consequential hydrographs from the model is analyzed statistically using an assortment of assessment measures. In this study areal average value of rainfall data for three upstream gauging stations have been used.

4. Results and Discussions

As discussed previously, the performance of the proposed D-ANN model is compared with a FF-ANN model for forecasting the runoff of Kolar River at a lead time of 1 hour. The results of the study are discussed in detail in the following paragraphs.

One of the most important steps in the ANN hydrological model development is determination of significant input variables which requires prior knowledge and generating an analytical approach of cross correlation to find the dependence (linear) between these variables [6,17,18]. The foremost drawback related with this method is that the correlation can be nonlinear but it is only capable of identifying linear dependency between two variables. The present study uses a statistical approach of data series which is based on the scrutiny that the input variables analogous to different time lags can be acknowledged using cross correlations, autocorrelations and partial autocorrelations. To certify good overview by ANN model, many associations between weighted inputs and output samples have been recommended in the literature [19]. The input variables selected in this study are $R(t - 9)$, $R(t - 8)$, $R(t - 7)$, $Q(t - 2)$ and $Q(t - 1)$, where R and Q represent the rainfall and runoff values, respectively at time “ t ”. The hidden nodes are identified by various trials.

The performance of the proposed D-ANN model is compared with that of the feed-forward neural network by means of a variety of statistical criteria coefficient of correlation (R), coefficient of efficiency (E), Root-mean-square error (RMSE) between the calculated and computed flow values. The statistics of the above criteria for D-ANN and FF-ANN model is presented in **Table 1**.

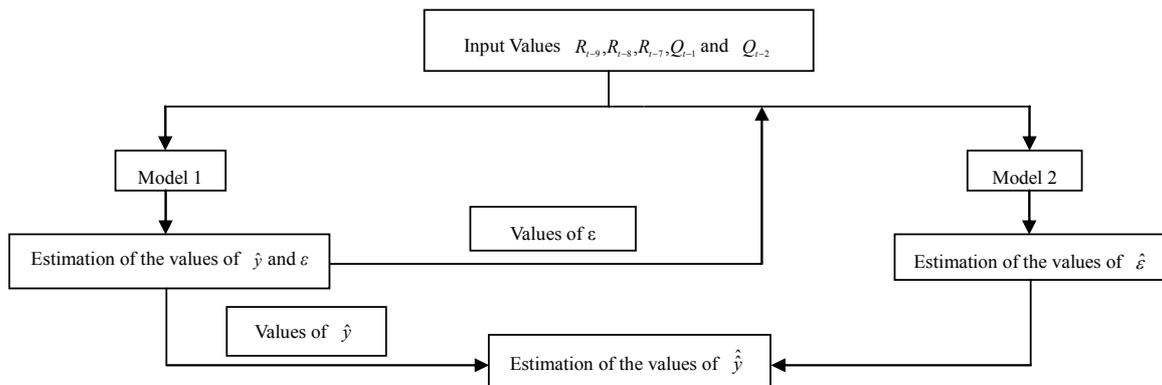


Figure 2. Methodology of D-ANN.

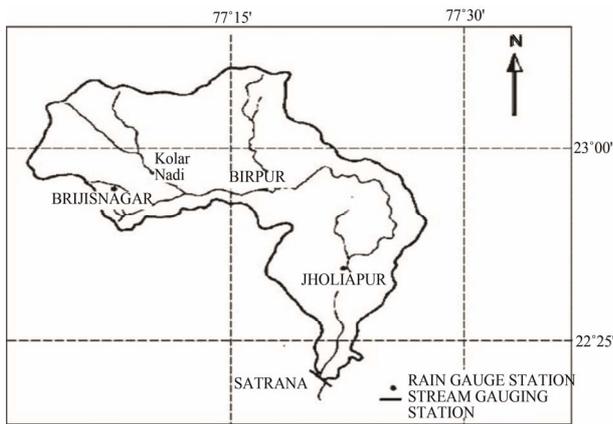


Figure 3. Map of Kolar Basin [15].

Table1. Statistical indices—comparison between D-ANN and FF-ANN model.

	D-ANN	FF-ANN
Coefficient of correlation (<i>R</i>)	0.99	0.99
Coefficient of efficiency (<i>E</i>)	0.98	0.98
Root-Mean-Square Error (RMSE)	27.16	23.24

$$R = \left[\frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \right] \quad (8)$$

$$E = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (10)$$

where y and \hat{y} be the actual and the predicted value of ANN model respectively.

It is observed from the **Table 1** that performance of both the models in terms of statistical indices is very similar and satisfactory. The correlation statistics, for evaluating the linear correlation between the observed and predicted runoff, is persistent for all models during calibration as well as validation period. While evaluating capability of the model for predicting runoff values away from the mean, efficiency of both the models is found to be greater than 90%, which according to Shamseldin [20] is very reasonable. Similarly RMSE statistic for indicating quantitative measure of the model error in units of the variable was also found good for all models as is

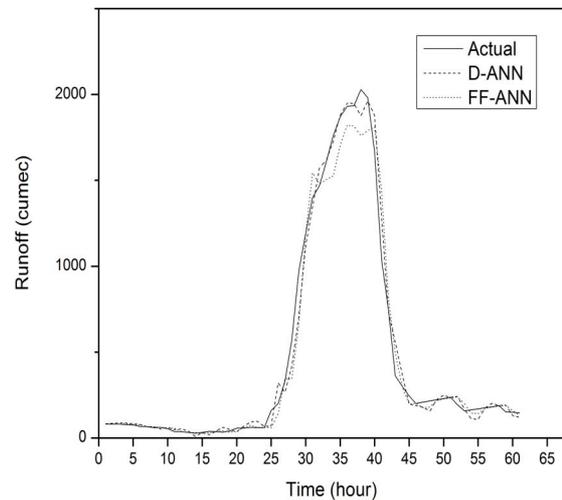


Figure 4. Model computed flows for a typical event during validation showing historical flows and predicted flows from D-ANN and FF-ANN.

evidenced by the low values. Further, it can be observed from **Figure 4** that both the models are able to predict the flows. However, it is observed (**Figure 4**) that the D-ANN model is able to predict the peak flows better than the FF-ANN model. In general it is observed that the D-ANN model although has a similar statistical performance in comparison to FF-ANN, it outperforms the later in terms of prediction of high flows.

5. Summary and Conclusion

This paper presents a dual-ANN model to improve the performance of the model in terms of prediction of high flows. The performance of the model is compared with that of the feed-forward ANN model in terms of statistical indices such as coefficient of correlation, coefficient of efficiency and root means square error. The exercise was carried out for the hourly data in Kolar river basin, India. It is observed that the proposed D-ANN model and the FF-ANN model show similar performances in terms of statistical indices. However, the D-ANN model outperforms the FF-ANN model in prediction of high flows (extremes). The performance of the D-ANN models has to be tested on various time scales. Further extensions of this model can be examined to improve the forecasting accuracy.

6. Acknowledgements

The authors thank the Vellore Institute of Technology Vellore, India, for providing the necessary facilities to carry out this research work.

REFERENCES

[1] K. J. Beven, "Rainfall-Runoff Modelling: The Primer,"

- John Wiley, Hoboken, 2000.
- [2] Q. Duan, S. Sorooshian and V. K. Gupta, "Effective and Efficient Global Optimization for Conceptual Rainfall Runoff Models," *Water Resources Research*, Vol. 28, No. 4, 1992, pp. 1015-1031.
- [3] G. E. P. Box and G. M. Jenkins, "Time Series Analysis: Forecasting and Control," Holden Day Inc., San Francisco, 1976.
- [4] A. Jain, and S. Srinivasulu, "Development of Effective and Efficient Rainfall-Runoff Models Using Integration of Deterministic, Real-Coded Genetic Algorithms and Artificial Neural Network Techniques," *Water Resources Research*, Vol. 40, No. 4, 2004, Article ID: W04302. [doi:10.1029/2003WR002355](https://doi.org/10.1029/2003WR002355)
- [5] R. K. Srivastav, K. P. Sudheer and I. Chaubey, "A Simplified Approach to Quantifying Predictive and Parametric Uncertainty in Artificial Neural Network Hydrologic Models," *Water Resources Research*, Vol. 31, No. 10, 2007, pp. 2517-2530.
- [6] K. Hsu, V. H. Gupta and S. Sorooshian, "Artificial Neural Network Modelling of the Rainfall-Runoff Process," *Water Resources Research*, Vol. 31, No. 10, 1995, pp. 2517-2530. [doi:10.1029/95WR01955](https://doi.org/10.1029/95WR01955)
- [7] K. Hornik, M. Stichcombe and H. White, "Multi Layer Feed forward Networks Are Universal Approximators," *Neural Networks*, Vol. 2, 1989, pp. 359-366.
- [8] A. W. Minns and M. J. Hall, "Artificial Neural Networks as Rainfall-Runoff Models," *Journal of Hydrology Science*, Vol. 41, No. 1, 1996, pp. 399-417.
- [9] C. W. Dawson and R. Wilby, "An Artificial Neural Network Approach to Rainfall Runoff Modelling," *Hydrological Science*, Vol. 43, No. 1, 1998, pp. 47-66.
- [10] M. Campolo, P. Andreussi and A. Soldati, "River Flood Forecasting with a Neural Network Model," *Water Resources Research*, Vol. 35, No. 4, 1999, pp. 1191-1197. [doi:10.1029/1998WR900086](https://doi.org/10.1029/1998WR900086)
- [11] C. E. Imrie, S. Durucan and A. Korre, "River Flow Prediction Using Artificial Neural Networks: Generalization beyond the Calibration Range," *Journal of Hydrology*, Vol. 233, 2000, pp. 138-153. [doi:10.1016/S0022-1694\(00\)00228-6](https://doi.org/10.1016/S0022-1694(00)00228-6)
- [12] N. Karunanithi, W. J. Grenney, D. Whitley and K. Bovee, "Neural Networks for River Flow Prediction," *Journal of Computing in Civil Engineering*, Vol. 8, No. 2, 1994, pp. 201-220. [doi:10.1061/\(ASCE\)0887-3801\(1994\)8:2\(201\)](https://doi.org/10.1061/(ASCE)0887-3801(1994)8:2(201))
- [13] ASCE Task Committee, "Artificial Neural Networks in Hydrology-I: Preliminary Concepts," *Journal of Hydrologic Engineering*, Vol. 5, No. 2, 2000, pp. 115-123. [doi:10.1061/\(ASCE\)1084-0699\(2000\)5:2\(115\)](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(115))
- [14] ASCE Task Committee, "Artificial Neural Networks in Hydrology-II: Hydrologic Applications," *Journal of Hydrologic Engineering*, Vol. 5, No. 2, 2000, pp. 124-137. [doi:10.1061/\(ASCE\)1084-0699\(2000\)5:2\(124\)](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(124))
- [15] P. C. Nayak, K. P. Sudheer, D. M. Rangan and K. S. Ramasastri, "Short-Term Flood Forecasting with a Neurofuzzy Model," *Water Resources Research*, Vol. 41, 2005, Article ID: W04004. [doi:10.1029/2004WR003562](https://doi.org/10.1029/2004WR003562)
- [16] G. J. Bowden, G. C. Dandy and H. R. Maier, "Input Determination for Neural Network Models in Water Resources Applications: 1. Background and Methodology," *Journal of Hydrology*, Vol. 301, No. 1-4, 2004, pp. 75-92. [doi:10.1016/j.jhydrol.2004.06.021](https://doi.org/10.1016/j.jhydrol.2004.06.021)
- [17] G. J. Bowden, G. C. Dandy and H. R. Maier, "Input determination for Neural Network Models in Water Resources Applications: 2. Background and Methodology," *Journal of Hydrology*, Vol. 301, No. 1-4, 2004, pp. 93-107. [doi:10.1016/j.jhydrol.2004.06.020](https://doi.org/10.1016/j.jhydrol.2004.06.020)
- [18] K. P. Sudheer, A. K. Gosain and K. S. Ramasastri, "A Data-Driven Algorithm for Constructing Artificial Neural Network Rainfall-Runoff Models," *Hydrological Processes*, Vol. 16, No. 6, 2002, 1325-1330. [doi:10.1002/hyp.554](https://doi.org/10.1002/hyp.554)
- [19] H. R. Maier and G. C. Dandy, "Neural Networks for the Prediction and Forecasting of Water Resources Variables: A Review of Modelling Issues and Applications," *Environmental Modelling & Software*, Vol. 15, No. 1, 2000, pp. 101-124. [doi:10.1016/S1364-8152\(99\)00007-9](https://doi.org/10.1016/S1364-8152(99)00007-9)
- [20] A. Y. Shamseldin, "Application of a Neural Network Technique to Rainfall-Runoff Modelling," *Journal of Hydrology*, Vol. 199, No. 3-4, 1997, pp. 272-294. [doi:10.1016/S0022-1694\(96\)03330-6](https://doi.org/10.1016/S0022-1694(96)03330-6)