A Comparison of Emotional Neural Network (ENN) and Artificial Neural Network (ANN) Approach for Rainfall-Runoff Modelling

Suraj Kumar a, Thendiyath Roshni b*, Dar Himayoun a

a Research Scholar, Department of Civil Engineering, National Institute of Technology, Bihar, Patna, 800005, India.
b Assistant Professor, Department of Civil Engineering, National Institute of Technology, Bihar, Patna, 800005, India.

Received 14 June 2019; Accepted 06 September 2019

Abstract

Reliable method of rainfall-runoff modeling is a prerequisite for proper management and mitigation of extreme events such as floods. The objective of this paper is to contrasts the hydrological execution of Emotional Neural Network (ENN) and Artificial Neural Network (ANN) for modelling rainfall-runoff in the Sone Command, Bihar as this area experiences flood due to heavy rainfall. ENN is a modified version of ANN as it includes neural parameters which enhance the network learning process. Selection of inputs is a crucial task for rainfall-runoff model. This paper utilizes cross correlation analysis for the selection of potential predictors. Three sets of input data: Set 1, Set 2 and Set 3 have been prepared using weather and discharge data of 2 raingauge stations and 1 discharge station located in the command for the period 1986-2014. Principal Component Analysis (PCA) has then been performed on the selected data sets for selection of data sets showing principal tendencies. The data sets obtained after PCA have then been used in the model development of ENN and ANN models. Performance indices were performed for the developed model for three data sets. The results obtained from Set 2 showed that ENN with $R = 0.933$, $R^2 = 0.870$, Nash Sutcliffe $= 0.8689$, RMSE $= 276.1359$ and Relative Peak Error $= 0.00879$ outperforms ANN in simulating the discharge. Therefore, ENN model is suggested as a better model for rainfall-runoff discharge in the Sone command, Bihar.

Keywords: Emotional Neural Network (ENN); Artificial Neural Network (ANN); Cross Correlation; Principal Component Analysis (PCA); Rainfall-Runoff.

1. Introduction

The rainfall-runoff relationship is one of the most complex hydrological phenomena due to presence of complex non-linear relationships in the transformation of rainfall into runoff. This process is quite difficult to comprehend, owing to the presence of huge number of variables involved in the demonstration of physical process [1-4]. Therefore its precise modelling is important for water resources management and development and the prediction of natural calamities like droughts and floods. Based on the involvement of physical aspects, rainfall-runoff models are classified as either physical-based models or system theoretic models [5-7]. The physical-based models also called data driven models require the considerable information about the system mechanism as well as its parameters. However, the system theoretic models do not concern much about the physical processes of the problem. These models are primarily based on rainfall and runoff data and seek to characterize nonlinearity and non-stationary behaviour from those data by the use of transfer functions [8-10]. Among the system theoretic models, Artificial Neural Network (ANN) based models for rainfall-runoff modelling have received global attention because of their capability to capture high degree of non-
linearity and complex nature of relationship between the hydrological variables without fully understanding the processes beneath [11-16].

The ANNs have black box properties and offer a relatively quick and flexible technique of modelling. Owing to their parallel architecture, these models can treat nonlinearity and non-stationary behavior in the data to some extent [17-19]. Several studies could be found in the technical literature that have reported that ANNs outperform the traditional statistical rainfall-runoff models [17, 11]. The ANNs are more promising alternatives for conceptual rainfall-runoff models.

Ghumman et al. [11] used ANN based rainfall- runoff model and compared the results with statistical conceptual model. The results of their study reported that ANN based approach for rainfall-runoff modelling is more promising alternative to conceptual models and this approach could be used when the dataset is of low standard and/or short range. Demirel et al., [17] investigated the ability of both Soil and Water Assessment Tool (SWAT) and ANN model in flow forecasting. They compared the results of SWAT model with ANN model based on the prediction accuracy and concluded that ANNs can be more powerful tools in daily flow forecasts. Despite these promising results, several other studies reveal inefficiency and drawbacks of ANNs over the other rainfall-runoff models [19-22]. The deficiencies like overtraining of data and underestimation of peak values can be fixed to some extent by several data pre-processing approaches. For instance, the ability of wavelet based data pre-processing approaches in decomposing complex hydrological time series into sub-series can be very effective for interpreting hydrological phenomena [23-25]. This technique extracts the useful information from data series at different scales to enhance the modelling efficiency and to extract the seasonal features of the rainfall-runoff process across most areas of hydrology [25-28].

The flexible and data dependent structure of ANN leaves a huge room for its improvement in the context of rainfall-runoff modelling [29]. Wu and Chau [15] employed ANN coupled with Singular spectrum Analysis (SSA) for rainfall-runoff modelling. The purpose of coupling SSA with ANN was to reduce the lag effect in ANN and the results showed that SSA improves the model performance and can eliminate the lag effect. Fellous [30] was the first study that incorporated the emotions into Artificial Intelligence (AI) systems suggesting that emotions must be dynamically interacted with together. After that other researchers investigated the role artificial emotions in AI models by the integration of artificial emotions into the classic ANN framework as Emotional Neural Network (ENN) model [31, 32]. The neurophysiological response of animals from a biological point of view, could be affected by the emotion and mood of animal due to the hormonal activities so that the animals at different moods may provide different actions for the same task. Employing this concept by merging neural network with artificial emotions, there will be a feedback loop between the hormones and neural systems that could relatively enhanced the learning ability of a neural network. A few studies could be find in the technical literature that have successfully applied ENN in hydrological studies [31]. Sharghi et al. [29] employed EANN and WANN approaches for modelling the rainfall runoff process. They reported the superiority of EANN over ANN as well as the better learning capability of EANN in extraordinary and extreme conditions of the training phase.

In this study we have applied ANN and ENN approaches to model rainfall-runoff process of Sone river command, Bihar. We employ Cross Correlation Function (CCF) and Auto Correlation Function (ACF) for the selection of input parameters at different lags. In order to remove the redundancy in input variable we applied Principal Component Analysis (PCA). The results derived from different input combination were compared in both the models and the best combinations were chosen based on the evaluation criteria. Finally the results of ENN are compared with the results of ANN by graphical indicators as well as by the selected evaluation criteria.

2. Study Area and Research Data

The study area (Figure 1) comprises of the Sone Irrigation Project in Bihar, India. It is a river diversion scheme built across the river Sone, which is a tributary of river Ganga. It lies at latitude 24°48’N and longitude 84°07’E. The regions to be specific Patna, Jehanabad, Aurangabad, Gaya, Buxar, Bhujpur, Rohtas and Bhabua are secured under this task. 17,651 sq. km (25%) of the total catchment area of the river i.e. 71,259 sq. km falls in Bihar. Practically plain in geography, it consistently inclines towards the Ganga River. These factors are ideally suited for the development of irrigated agriculture. The major crop grown in the command in the Kharif (monsoon) season is rice and in Rabi season is wheat. Rice is the major crop and other crops occupy less than 2% of cultivable command area. Sugarcane is the main money crop developed in the territory. Linseed and mustard oilseeds are likewise developed in Rabi season over a little zone. The water system framework has been employable since 1871 yet water system in a sorted out way began in 1879.
The data utilized for carrying out this work consists of rainfall and discharge data for the period 1986-2014 (Figure 2). The rainfall data has been obtained from Indian Meteorological Department (Pune) and the discharge data has been obtained from CWC (Patna) (Table 1). The details of the data are shown in table

### Table 1. Statistics of the monthly time series of the Sone Command

<table>
<thead>
<tr>
<th>Time Series</th>
<th>Statistical Parameter</th>
<th>Station</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Indepuri</td>
</tr>
<tr>
<td>Rainfall (mm)</td>
<td>Mean</td>
<td>78.545</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>674.8</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>118.75</td>
</tr>
<tr>
<td>Discharge (m³/s)</td>
<td>Mean</td>
<td>462.09</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>7557.577</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>762.87</td>
</tr>
</tbody>
</table>

![Figure 1. Location of study area (Sone River Command, Bihar, India)](image)

![Figure 2. Box Plot of the data used](image)
3. Research Methodology

3.1. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a procedure mostly known for extraction of constituents for multivariate analysis. The extracted constituents obtained using PCA are uncorrelated. Its objective is to separate the significant data from the information table and to express this data as a lot of new symmetrical factors called principal components. In this study PCA analysis of the datasets has been done in MATLAB.

3.2. Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) is a powerful soft computational technique composed of interconnected nodes (like neurons in the human brain) that are linked by weighted synaptic connections [2, 11, 31]. This technique inspired by the biological cerebral activity is widely applied in the forecasting of water resource and hydrology [2, 11, 17, 19, 31, 33]. This approach being faster, flexible, highly adaptive and robust in newer and noisy environments can solve a wide in the range of problems. Extensive research has been successfully carried out regarding the implementation of ANN in engineering related fields like rainfall-runoff modelling, time series prediction and rule-based control. In ANN, Back Propagation (BP) algorithm network models are common to engineers. As it has been proved that the three-layer BP network model is satisfied for all kinds of engineering problems in terms of simulation and forecasting [15]. The three layer Feed Forward Neural Network (FFNN) usually employed in hydrological time-series forecasting, provide a general framework to perform an input–output mapping using an arrangement of interconnected sensorial components. A detailed explanation regarding the properties of ANN could be found in the study of [18, 23, 31].

3.3. Emotional Neural Network (ENN)

An Emotional Neural Network (ENN) is the new generation of the conventional ANN model, which contains an artificial emotional system where the emission of hormones take place in order to adjust the operation of neurons. The hormonal parameters in the ENN model can be modified by input and output value of neurons of the network. It is evident from the Figure 3, that in each neuron of the ENN model, the information repeatedly passes from inputs to output and vice versa. In addition to that, each node (neuron) provide dynamical hormones of \( H_a, H_b, \) and \( H_c \). These hormones according to the input and output values are firstly initialized in the training phase of model and then are modified through the learning process. In the training phase of the network, these hormonal coefficients can impact the other components of the node such as activation function, weights and net function (Figure 3). The solid and dotted lines in the Figure 3, represent the neural and hormonal paths of the information. The output of the \( i^{th} \) neuron in the proposed ENN model with three hormonal glands of \( H_a, H_b, H_c \) is computed as follows [6, 29]:

\[
Y_i = \left( \lambda_i + \sum_{h} \sigma_{i,h} H_h \right) \times f \left( \sum_{j} \left( \beta_i + \sum_{k} \zeta_{i,k} H_k \right) \times \left( \theta_{i,j} + \sum_{h} \varphi_{i,j,h} H_h \right) X_{i,j} \right) + \left( \alpha_i + \sum_{k} \xi_{i,k} H_k \right) + \left( \delta_i + \sum_{h} \rho_{i,h} H_h \right) \tag{1}
\]

Whereas the overall hormone value is computed as follows [6, 29]:

\[
H_h = \sum_{i} H_{i,h} \quad ; \quad H = (a,b,c) \tag{2}
\]

In the Equation 1, the first term represents the applied weight to the activation function \((f)\). It consists of both the dynamic hormonal weight of \( \sum_{h} \sigma_{i,h} H_h \) and the statistic neural weight of \( \lambda_i \). The second term represents the applied weight to the summation function, the third term represents the applied weight to the input value of \( X_{i,j} \) coming from \( j^{th} \) neuron of previous layer and the fourth term represents the bias of net function from both the hormonal and the neural weights of \( \sum_{h} \xi_{i,k} H_k \) and \( \alpha_i \) respectively. Finally, the fifth term contributes to the activation function, where hormonal and neural weights contribute as \( \sum_{h} \rho_{i,h} H_h \) and \( \delta_i \) respectively.

---

2123
The $i^{th}$ node of the network output ($Y_i$) will provide hormonal feedback of $H_{i,h}$ to the network as follows [6, 29]:

$$H_{i,h} = \text{glandity}_{i,h} \times Y_i \quad (3)$$

Where, $\text{glandity}_{i,h}$ is a parameter representing the production factory of all hormones in the gland. This parameter should be calibrated to produce desired level of hormone in each gland. By considering Equations 1 and 2 and the network output ($Y_i$), the hormone value is updated through the training process to get the reliable agreement between the observed and computed time series of the target.

In this study the time series data of rainfall and runoff have been utilised to develop the best rainfall-runoff model utilising cross correlation and PCA. The flowchart of the methodology adopted in this study is shown in Figure 4.
3.3. Efficiency Parameters

In this work the author intended to use coefficient of determination ($R^2$), Nash Sutcliffe Efficiency (ENS), Root mean square error as efficiency parameters for determining the accuracy of the model developed.

Coefficient of determination ($R^2$) measures the dispersion between observed value and simulated value. It is expressed as:

- **Coefficient of determination:**
  \[
  R^2 = \frac{n(\sum Q_{obs} Q_{sim}) - (\sum Q_{obs}) (\sum Q_{sim})}{\sqrt{n(\sum Q_{obs}^2) - (\sum Q_{obs})^2} \sqrt{n(\sum Q_{sim}^2) - (\sum Q_{sim})^2}}
  \] (4)

Where $Q_{obs}$ = discharge observed, $Q_{sim}$ = discharge simulated, and $n$ = number of observations.

The range of $R^2$ lies between 0 and 1 which represents no correlation and perfect correlation between observed and simulated value.

- **Nash-Sutcliffe efficiency (N):**
  \[
  N = 1 - \frac{\sum_{t=1}^{T} (Q_{obs} - Q_{sim})^2}{\sum_{t=1}^{T} (Q_{obs} - Q_{mo})^2}
  \] (5)

Where, $Q_{mo}$ = average of observed discharges

- **Root mean square error (RMSE):**
  \[
  RMSE = \sqrt{\frac{\sum_{t=1}^{T} (Q_{obs} - Q_{sim})^2}{N}}
  \] (6)

- **Relative peak error (RPE):**
  \[
  RPE = \frac{|Q_p - Q_{pm}|}{Q_p}
  \] (7)

Where, $Q_p$ = observed peak, and $Q_{pm}$ = simulated peak discharge.

4. Results and Discussions

In this research study, time series data of monthly precipitation from two stations (Inderpuri and Buxar) and discharge (Koelwar) from one station were employed for the model development. The foremost step in the process of model development is the selection of significant input variables for the model. In order to select the significant input variables for the model, statistical analysis consisting of Cross Correlation Functions (CCF) and partial Auto Correlation Functions (ACF) were used.

The CCF curves between discharge at Koelwar station and precipitation at different lags for Inderpuri and Buxar stations are presented in Figure 5 (a and b) and the ACF curve for Koelwar station at different lags is presented in Figure 5(c). As can be seen from CCF curves Figure 5 (a and b), the correlation coefficient increases with increase in lag time and reaches maximum at (t-1), then decreases continuously for both Inderpuri and Buxar stations. Therefore, P(t) and P(t-1) are selected as dominant input variables from Inderpuri and Buxar precipitation stations. It is evident from the Figure 5(c), that correlation coefficient decreases continuously with the increase in lag time. Thus, Q(t) and Q(t-1) are selected as dominant input variables from Koelwar discharge gauging stations. Based on the results of CCF and ACF curves three sets of input variables were chosen (Table 2).
For lowering the dimensionality of the input variables, there is a requirement of a practical and efficient method which is capable of changing the correlated discharge-affecting factors into uncorrelated ones. PCA of the selected sets has then been carried out. With the PCA the eigen-vector based multivariate analysis has been done. It helps in establishing sets of variables which exhibits the observed principal tendencies. These new linear combinations are called principal components.

Monthly observed data from 2 rainfall and 1 discharge stations for the period January 1986 to December 2014 of the Sone command have been utilised for the model development. Two models (ANN and ENN) have been used to check the forecasting abilities. Using the top three combinations of data 3 pairs of sets has been developed for rainfall-runoff modelling. The foremost task in rainfall-runoff (r-r) modelling using ANN and ENN approaches is the identification of optimal network geometry. It should be noted that, besides the selection of significant input variables for the model, the optimal adjusting of the network parameters like training iteration epoch, the number of hidden neurons and transfer functions of layers also plays an important role. In this study, the training iteration epoch and the number of hidden neurons were selected based on the trial and error method for each set of input variables in both the models. Both the models were trained on Levenberg–Marquardt Back-propagation algorithm using tangent sigmoid as activation function in ANN model. Based on the study of Sharghi et al. [29] the models were checked for hidden neurons up to the fourfold of the input number.

The comparisons of the efficiency of models have been done using different indices and the results are shown in table. In the table R, R², Nash Sutcliffe, RMSE and Relative Peak Error of the two models are shown (Table 3). The summary of the models have been shown using table. Both the models are showing good results but ENN is showing better results and outperforms ANN for the river runoff forecasting. The performance of the datasets are indicated in Table 3.
Table 3. Performance indices of models

<table>
<thead>
<tr>
<th></th>
<th>ANN</th>
<th>R</th>
<th>R²</th>
<th>NSE</th>
<th>RPE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set 1</td>
<td>0.7685</td>
<td>0.590592</td>
<td>0.5885</td>
<td>0.431</td>
<td>489.322</td>
<td></td>
</tr>
<tr>
<td>Set 2</td>
<td>0.9149</td>
<td>0.837072</td>
<td>0.0197</td>
<td>0.01974</td>
<td>309.8577</td>
<td></td>
</tr>
<tr>
<td>Set 3</td>
<td>0.8072</td>
<td>0.651572</td>
<td>0.6476</td>
<td>0.2766</td>
<td>452.8103</td>
<td></td>
</tr>
</tbody>
</table>

From the table and figure it can be easily deduced that the best performing sets for rainfall-runoff modelling is Set 2 for both ANN and ENN. Therefore, for further analysis the results obtained from set 2 for both the models has been utilised.

The observed versus simulated discharge time series using ENN and ANN for Sone Command has been shown in Figure 6. The observed runoff values are in good accord with the simulated values but it can be seen that simulation by ANN is slightly deviated from the observed discharge. This can be due to the fact that ENN encompasses neural parameters which enhance the network learning process.

Figure 6. Observed versus simulated discharge time series using ENN and ANN for Sone Command

The scatter plot of the simulated discharge using ENN and ANN has been shown in Figure 7. The R² value using ENN is highly significant with value 0.8706 as compared to ANN which shows lesser value of 0.8371.

Figure 7. Verification scatter plot of simulated runoff using ANN and ENN
Figure 8 details the Taylor diagram for the two models ANN and ENN for the period 1986-2014. The Taylor diagram shows the closeness of the observed and model output. Three statistics namely correlation, root-mean-square error and standard deviation are used to quantify the similarity between observed and model stimulated output. RMSE values are indicated by the brown contours. This figure clearly implies that ENN is showing minimum standard deviation, minimum RMSE and maximum correlation.

Figure 8. Taylor diagram of the two developed models

Figure 9 shows the results of the random walk test of the simulated discharge using ENN and ANN at the discharge station for three data sets for the period 1986-2014. The basis of random walk test is sign test which is free of distributional assumptions associated with the errors related to simulation. The results associated with this site are contrasted in each and every step. The upward movement of all the 3 sets shows that ENN model is more skilful approach for discharge simulation. The 95% confidence interval suggests that both the approaches are equally skilful in discharge simulation.

Figure 9. Random Walk Test

Figure 9 shows the results of the random walk test of the simulated discharge using ENN and ANN at the discharge station for three data sets for the period 1986-2014. The basis of random walk test is sign test which is free of distributional assumptions associated with the errors related to simulation. The results associated with this site are contrasted in each and every step. The upward movement of all the 3 sets shows that ENN model is more skilful approach for discharge simulation. The 95% confidence interval suggests that both the approaches are equally skilful in discharge simulation.

5. Conclusion

The impacts of flood can be considerably reduced if the relationship between rainfall and runoff can be properly established. This study has been done to test the applicability of soft computing technique based rainfall-runoff models namely ENN and ANN to simulate runoff in the Sone Command, Bihar. Runoff and antecedent runoff, precipitation, antecedent precipitation over the basin, at three gauging stations in the basin were first identified as appropriate input variables, and then CCF curves at different time lags have been plotted to select the potential input variables. Monthly rainfall data of two stations and discharge data of one station for the period 1986-2014 have been utilised as data sets for the development of ENN and ANN models. Three datasets have been selected based on cross correlation as potential inputs for rainfall-runoff modelling. PCA of the selected scaled input has then been carried out to reduce the dimensionality.

Many checks have been done to estimate the reliability and performance of the models. Based on the statistical indices it has been established that ENN outperformed ANN and is more accurate as compared to the traditional ANN
method for rainfall-runoff modelling. Also, other graphical indicators like Taylor diagram, Random walk test and one to one correlation signifies the outperformance of ENN over ANN in the Sone River Command, Bihar. The results of this study will be helpful in selecting the appropriate model for the discharge simulation in the Sone command, Bihar and thereby helping planners for effective flood mitigation.

6. Conflicts of Interest
The authors declare no conflict of interest.

7. References


Latinez Sotomayor, Karen A. "Comparison of adaptive methods using multivariate regression splines (MARS) and artificial neural networks backpropagation (ANNB) for the forecast of rain and temperatures in the Mantaro river basin." (March 2010).