Regional Rainfall Forecasting using Large Scale Climate Teleconnections and Evolutionary Algorithms

D. Nagesh Kumar, M. Janga Reddy and Rajib Maity
Department of Civil Engineering, Indian Institute of Science, Bangalore-560 012, India
{nagesh, mjreddy, rajib}@civil.iisc.ernet.in

Abstract. The continuous occurrence of changes in the global climate causes significant variability in the seasonal and intra-seasonal rainfall pattern, which often leads to frequent floods and droughts in India. To reduce the magnitude of effect of such natural calamities and for better management of water resources, it is essential to predict the rainfall, well in advance. In this study the possible relationship between the large scale climate indices like, El-Niño Southern Oscillation (ENSO), EQUitorial INdian Ocean Oscillation (EQUINOO) and a local climate index of Ocean-Land Temperature Contrast (OLTC) are studied and used to forecast regional monsoon rainfall of Orissa state in India. To handle the highly non-linear and complex behavior of the climatic variables in forecasting the rainfall, this study employs Artificial Neural Networks (ANNs) methodology. In this study, to optimize the ANN architecture, Genetic Optimizer is used. After identifying the lagged relationship between climate indices and monthly rainfall, the rainfall values are forecasted for summer monsoon months of June, July, August and September (JJAS) individually as well as total monsoon rainfall. The models are trained individually for monthly and for seasonal rainfall forecasting. Then the trained models are tested to evaluate the performance of the model. The results show reasonably good accuracy for monthly and seasonal rainfall forecasting. The study demonstrates the possible teleconnection between large scale climate indices and regional rainfall over Orissa and also shows the usefulness of ANNs in rainfall forecasting.

1 Introduction

Reliable prediction of Indian Summer Monsoon Rainfall (ISMR) on monthly and seasonal time scales is not only scientifically challenging, but is also important for planning and devising agricultural strategies in the country and consequently helps the development of the country’s economy. The recent changes in global climate and the uneven spatial and temporal distribution of rainfall are causes for more severe problems like floods and droughts. Most of the rainfall in the region occurs during monsoon period. The rainfall received during the four months, i.e., June, July, August, September (JJAS) is considered as summer monsoon, and is very crucial for Indian farming community. Researchers have used various approaches to study and predict the seasonal and intra-seasonal rainfall. To forecast the rainfall at different spatial and temporal scales, the models used in the past can be broadly classified as
empirical or dynamical [1]. This study is concerned with empirical models only. Empirical modeling strategies include identification of reliable precursors as well as an optimal utilization of the information contained in the data on precursors. The reasonable success achieved by the empirical approach has motivated persistent exploration of regional/global teleconnections of the Indian summer monsoon season since Walker’s time, which resulted in a large number of predictors as well as a variety of statistical techniques. Since any modeling effort will have to be based on an understanding of the variability of the past data; ANNs have some special characteristics in this regard to be used. In contrast to conventional modeling approaches, ANNs do not require an in depth knowledge of driving processes, nor do they require the form of the model to be specified a priori [2]. This is the case, when modeling various climate variables for prediction (forecasting) of hydrological variables like rainfall, streamflows etc. In view of this, ANNs are found to be suitable approach for prediction of Indian monsoon rainfall using large scale climate variables as input to the network. The next section provides a brief overview of climate variables and their influence on Indian monsoon rainfall.

2 Influence of Climate Variables on Indian Monsoon

At present, assessment of the nature and causes of seasonal climate variability is still uncertain. There are still uncertainties associated with local and global climatic variables. For any rainfall prediction model, these are sources of variance in predictability [3]. Recently, researchers have studied the influence and the possible relationships between various global climate variables and Indian monsoon rainfall. Also they brought out several regional parameters based on sea-level pressure, temperature and wind fields over India and sea surface temperature (SST) data from the adjoining Indian seas. Although their performance in seasonal forecasting has been encouraging, there is still a large variance in the monsoon rainfall unaccounted by the predictors identified so far [4].

Several observational and modelling studies have indicated that the slowly varying surface boundary conditions, particularly in the winter and pre-monsoon seasons, constitute a major forcing factor on the interannual variability of the monsoon rainfall. Parameters representing these conditions, global as well as regional, provide the handle for seasonal prediction. On interannual timescales, the Indian monsoon rainfall has a strong and positive correlation with the pre-monsoon spring tropospheric temperature anomaly ([5], [6], [7]). Factors that influence the Indian summer monsoon include the sea surface temperature in the Pacific and Indian oceans [8], the Indian Ocean Dipole Mode [9], Eurasian snow cover ([10], [11]), the Atlantic circulation variation [12], global warming and human activities, among others. In the subsequent sections some of the highly influencing large scale climate indices, like El-Niño Southern Oscillation (ENSO), EQUitorial INdian Ocean Oscillation (EQUINOO) and Ocean-Land Temperature Contrast (OLTC) are discussed.
2.1 El Niño Southern Oscillation (ENSO)

The planetary-scale tropical Sea Level Pressure (SLP) anomalies associated with the Southern Oscillation (SO) occur in conjunction with the episodes of large-scale sea surface temperature (SST) anomalies (El-Niño/La Niña) over the tropical Pacific ([13], [14]). The intensities of El-Niño events are generally assessed on the basis of the average SSTs over the Niño regions in the Pacific Ocean, widely known as NINO1+2 (0°-10° S, 80° W-90° W), NINO3 (5°- 5° N, 90° W - 150° W), Nino3.4 (5°S-5° N, 120° W -170° W) and NINO4 (5° S -10° N, 160° E -150° W). After the discovery of strong links between the El-Niño Southern Oscillation (ENSO) and the ISMR ([15], [16], [14], [17]), the empirical models for monsoon prediction have developed rapidly. Many studies have used the relationship of ENSO to forecast Rainfall ([16], [14], [18]) over the Indian sub-continent. Parthasarathy et al. [18] found that NINO 4 SSTs are having significant relationships with the ISMR, while others observed that NINO3 SSTs [19], and NINO3.4 SSTs ([20], [21]) present better relationships with ISMR.

2.2 Equatorial Indian Ocean Oscillation (EQUINOO)

More recently Gadgil et al. ([20], [21]) established a possible link of the Indian monsoon to events over the equatorial Indian ocean. According to ([20], [21]), over the equatorial Indian Ocean, enhancement of deep convection in the atmosphere over the western part is generally associated with suppression over the eastern part and vice versa. The oscillation between these two states, which is reflected in the pressure gradients and the wind along the equator, is known as Equatorial Indian Ocean Oscillation (EQUINOO). The index of EQUINOO is based on the negative of the anomaly of the zonal component of surface wind in the equatorial Indian Ocean region (60° E - 90° E, 2.5° S – 2.5° N). This index is known as Equatorial zonal Wind Index (EQWIN) ([20], [21]). They also showed that every season with excess rainfall/drought during 1979–2002 can be ‘explained’ in terms of favorable / unfavorable phase of either the EQUINOO or the ENSO or both. For example in 1994, ENSO was unfavorable, but EQUINOO was favorable and India received above normal rainfall. On the other hand in 1979 and 1985, ENSO was favorable, but EQUINOO is unfavorable and India received a below normal rainfall. Thus they suggested that by using those two climate indices together, the predictability of Indian monsoon rainfall can be improved.

2.3 Ocean-Land Temperature Contrast (OLTC)

The generation of the Indian summer monsoon can be attributed to the land-sea thermal contrast and studies reveal that the thermal conditions during the pre-monsoon season over India play a significant role in the performance of the ensuing monsoon ([22], [6]). Various studies have used this phenomenon in predicting summer monsoon and found that regional temperature data can be a useful predictor in summer monsoon rainfall prediction ([22], [6], [4]). Sahai et al. [23] have used the
links of global SST data with Indian monsoon seasonal data and produced successful results in forecasting ISMR. Recently, it is found that OLTC, the difference of SSTA over the region of (10° S-10° N, 60° E - 85° E) and the anomalous surface air temperature over east coast of India is having better relationship with Indian monsoon rainfall.

All these factors have some influence, but it is not clear as to what amount of variance is caused by each of these climate variables on rainfall predictability. Hence in this study we have chosen the highly influencing global climate variables ENSO, EQUINOX, and OLTC to predict the Indian summer monsoon rainfall. To represent the ENSO event, Nino3.4 Sea Surface Temperature Anomaly (NINO 3.4 SSTA) and for EQUINOX, equatorial zonal wind index (EQWIN) at equator (2.5° S - 2.5° N, 60° E -90° E); for OLTC, the difference of SSTA over the region of (10° S-10° N, 60° E -85° E) and the surface air temperature over east coast of India are used in this study.

3 Forecasting Indian Summer Monsoon Rainfall

Most of the models that were used to forecast ISMR come under empirical modeling approach. A general overview of forecasting models for Indian monsoon rainfall can be found in [24]. Excellent reviews of the empirical models used for prediction of ISMR are presented in ([3], [23]). In this study we considered ANNs as the forecasting tool. A brief description of application of ANNs for rainfall forecasting is given below.

Artificial Neural Networks (ANNs) have been used in various fields for the prediction and forecasting of complex nonlinear time series, including forecasting of Indian monsoon rainfall. The neural network technique is able to learn the dynamics within the time series data [25]. In the past ANNs have been successfully used to predict Indian monsoon rainfall ([26], [27], [28], [21], [29]). Goswami and Srividya [26] have used time series approach, in which previous values from the time series were used to predict future values. Venkatesan et al. [27] have used neural network technique to predict monsoon rainfall of India using few predictors and compared the results with linear regression techniques and showed that the model based on neural network technique performed better. Guhathakurta et al. [28] have used hybrid principal component and Neural Network approach for long range forecasting of the Indian summer monsoon rainfall. They observed improved accuracy in prediction. The neural network technique contains the advantages of both, the regression analysis and nonlinear dynamics that need to be incorporated in order to predict the dynamic rainfall values. Sahai et al. [1] applied the ANN technique to five time series of June, July, August, September monthly and seasonal rainfall. The previous five years values from all the five time-series were used to train the ANN to predict for the next year. They found good performance in predicting rainfall. Other studies, which have used ANNs for summer monsoon rainfall forecasting over India, include [29]. They decomposed the Indian monsoon rainfall data into six empirical time series (intrinsic mode functions). Then they have identified the first empirical mode as a nonlinear part and the remaining as the linear part of the data. The nonlinear part is handled by ANN techniques, whereas the linear part is modeled through simple regression. They
showed that their model can explain 75 to 80% of the interannual variability (IAV) of eight regional rainfall series, considered in their study.

4 Artificial Neural Networks Methodology

Neural Networks are used to detect hidden relations in the set of patterns given during training period. In this study Feed-forward ANN procedure is used. A typical ANN will have an input layer, an output layer and one or more hidden layers. Neurons in the input layer simply act as a buffer to next layer. The neurons in different layers are connected by means of weights. The neurons in the hidden and output layers use activation function to transfer the received input to the next layer neurons. The activation function adopted is a Fermi function, which is given as follows.

\[
f(x) = \frac{1}{1 + e^{-4(x-0.5)}}
\]  

The input to each neuron in the hidden layer is the sum of weighted input signals from the neurons of the previous layer. The neuron will use the transfer function and outputs the signal to next layer which will become input to the neurons in the next layer and so on. The same procedure will be followed for all neurons of different layers. Once the output signals are found at output layer, they will be compared with the actual observed values and the error is back propagated into the network. The most commonly used ANN training algorithm, Back Propagation (BP) algorithm, is adopted to train the ANNs. The Back Propagation Neural Network (BPNN) uses steepest gradient descent procedure. Description of working of ANNs can be found in any standard reference [2].

The ANNs learning is the process of finding the optimal weight matrices in a systematic manner, in order to achieve the desired value of target outputs. The performance measure used to train ANNs is minimizing the network error function, which is given as

\[
E = \sum_{i=1}^{p} \sum_{j=1}^{q} (y_{i,j} - y_{i,j}')^2
\]  

where, \( y_{i,j}' \) is desired (target) value for \( j^{th} \) output node and \( i^{th} \) pattern; \( y_{i,j} \) is computed value for \( j^{th} \) output node and \( i^{th} \) pattern; \( q \) is number of output nodes; and \( p \) is number of training patterns.

The selection of suitable ANN architecture will play a significant role in performance of the model. If the architecture is too small, the network may not have sufficient degree of freedom to learn the process correctly. On the other hand if the network is too large, it may not converge during training or it may overfit the data [2]. In general, the architecture of the network is finalized after a number of trials. The features of network architecture most commonly optimized are the number of hidden layers and the number of nodes in the hidden layer. In general, initially the architecture begins with a very small network and several parameters are varied including the number of hidden layers and the number of nodes in the hidden layer to
obtain an appropriate architecture for each data set. In this study to optimize the ANNs architecture, Genetic Optimizer is used. A brief description on working procedure of the genetic optimizer with neural network is given below.

4.1 Description of the genetic optimizer

The main steps involved in Genetic optimizer with ANN are as follows.

1. Initialize parameters: First all the parameters like population size (N), number of maximum generations, probability of crossover (Pc) and probability of mutation (Pm) are to be set to specific values.

2. Generate initial population: At initial generation, the Genetic Optimizer randomly creates networks for the initial population, which is equal to the size of population (N).

3. Train the network and evaluate fitness: Each network within the current generation is to be trained with BPANN and their fitness values are determined according to the goals to be achieved.

4. Propagation of networks: a new generation of networks will be created from the old one according to the following procedure:
   a. Two “parent” networks will be chosen out of the old generation. The selection algorithm will choose networks with a high fitness by a higher probability.
   b. Two “children” networks will be created from the two “parent” networks. Using the cross over probability, Pc, the two “parent” networks will be crossed over, i.e., they will swap a portion of the network with each other.
   c. The “children” will be mutated with a mutation probability, Pm. Here, mutation means insertion or deletion of a layer and/or insertion or deletion of a neuron into a layer.
   d. A few elitist members of the population in current generation are carried to next generation. The selection continues until the new generation has N members too. After completion the new generation will be evaluated.

5. Check for Termination Criteria: Stop the evolution, if the target level achieved or the pre-set maximum number of generations is reached; else increment the generation counter and go to step (3), repeat the evolution.

6. Output the best solution obtained so far during the evolution.

The input and output patterns are scaled to 0-1 range through mapping with the help of minimum and maximum values of the patterns, whereby it can be modeled using Fermi function (Eq.1). This scaling helps the algorithms towards faster convergence. For ANNs training, the termination criteria used to stop the learning process are, either the epoch counter reaches 1000 or the maximum value squared deviation of neural network output from observed value among all training patterns is less than 0.001. To avoid over fitting, first the model is trained with BPANN using training data set, then the model is cross validated, by testing its performance with
validation data set, which is not used during training period. In this process the model makes better generalization for the new data set.

5 Data

The study uses data on various parameters, viz., monthly NINO3.4 SSTA, EQWIN index, OLTC index and monthly rainfall anomaly over Orissa subdivision. Monthly Nino3.4 SSTA and SSTA for region (10° S-10° N, 60° E - 85° E) (1958-1990) data have been collected from the web site of Climate Analysis Centre, National Centre for Environmental Prediction (NCEP), USA (www.cpc.ncep.noaa.gov/); Wind data (1958-1990) have been collected from NCEP, USA (www.cpc.ncep.noaa.gov/) to obtain EQWIN index for EQUINNOO ([20], [21]). Monthly rainfall and temperature data (1901-1990) have been collected from the web site of Indian Institute of Tropical Meteorology (http://www.tropmet.res.in/).

6 Model Selection

To identify the months which can be used as input to the ANNs model, cross correlation analysis is carried out for predictor variables. The inputs driven to the ANNs are monthly Nino 3.4 SSTA, EQWIN and the OLTC indices. The output variable is rainfall. Table 1 gives a complete list of inputs that have been used for forecasting the rainfall for June, July, August, September months and for the monsoon season (JJAS).

The predictor variable data set is available for 33 years (1958-1990). To train the ANNs, 23 years data set is selected and then to test the performance of the trained model, the remaining 10 years data set is used. The models are trained with different combinations of network architectures by using genetic optimizer. For genetic optimizer the parameters adopted are as follows: population size (number of networks created per generation) =50; maximum number of generations =100; probability of crossover (probability for a child network to be crossed over with another child network), $P_c = 0.6$ and probability of mutation (probability for a network to be modified during rollover to a new generation), $P_m=0.04$. For BPANN the parameters adopted are, learning parameter =0.2; momentum parameter=0.1; maximum number of epochs=1000. The best suited network architecture obtained for different monthly and seasonal models is given in Table 2.
Table 1. Climate indices of predictor variables considered for rainfall prediction of June, July, August, September months and for summer monsoon (JJAS)

<table>
<thead>
<tr>
<th>Rainfall</th>
<th>Predictor variables considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>June</td>
<td>Nino 3.4 (February, March, April)</td>
</tr>
<tr>
<td></td>
<td>EQWIN (May, June)</td>
</tr>
<tr>
<td></td>
<td>OLTC (May, June)</td>
</tr>
<tr>
<td>July</td>
<td>Nino 3.4 (February, March, April)</td>
</tr>
<tr>
<td></td>
<td>EQWIN (June, July)</td>
</tr>
<tr>
<td></td>
<td>OLTC (June, July)</td>
</tr>
<tr>
<td>August</td>
<td>Nino 3.4 (February, March, April)</td>
</tr>
<tr>
<td></td>
<td>EQWIN (July, August)</td>
</tr>
<tr>
<td></td>
<td>OLTC (July, August)</td>
</tr>
<tr>
<td>September</td>
<td>Nino 3.4 (February, March, April)</td>
</tr>
<tr>
<td></td>
<td>EQWIN (August, September)</td>
</tr>
<tr>
<td></td>
<td>OLTC (August, September)</td>
</tr>
<tr>
<td>JJAS</td>
<td>Nino 3.4 (February, March, April)</td>
</tr>
<tr>
<td></td>
<td>EQWIN (May, June)</td>
</tr>
<tr>
<td></td>
<td>OLTC (May, June)</td>
</tr>
</tbody>
</table>

Table 2. Architecture of ANNs selected and the performance of the models for rainfall forecasting in the monsoon season for Orissa sub-division

<table>
<thead>
<tr>
<th>Month/Season/</th>
<th>Network Architecture</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training</td>
</tr>
<tr>
<td>June</td>
<td>7,7,7,1</td>
<td>0.9941</td>
</tr>
<tr>
<td>July</td>
<td>7,8,9,1</td>
<td>0.9994</td>
</tr>
<tr>
<td>August</td>
<td>7,10,1</td>
<td>0.9969</td>
</tr>
<tr>
<td>September</td>
<td>7,8,1</td>
<td>0.9998</td>
</tr>
<tr>
<td>JJAS</td>
<td>7,8,1</td>
<td>0.9975</td>
</tr>
</tbody>
</table>

7 Results and Discussion

For June month ANN model, the correlation coefficient (CC) values obtained are 0.9941 and 0.8349 during training and testing periods respectively. Fig. 1 shows the comparison of ANN forecasted rainfall with observed rainfall for June month. It can be seen that, except for one year (1986) the model results are within reasonable accuracy and it predicts well the low rainfall (1981, 1982, 1983, 1987) and high rainfall (1984, 1989) values during the testing period. For July month, the CC values obtained are 0.9994 and 0.8002 during training and testing periods respectively. Fig. 2 shows, the comparison of ANN model forecasted rainfall with observed rainfall for
July month during the testing period. For July month, even though there are a few deviations from observed rainfall (1981, 1982, 1985, 1989, 1990), the model shows reasonable accuracy. For August month, the CC values obtained are 0.9969 and 0.8102 during training and testing periods respectively. Fig. 3 shows, the comparison of ANN model forecasted rainfall with observed rainfall for August month during the testing period. The model results show that for two years ANN over forecasted (1984, 1985), for two years under forecasted (1983, 1989) and for the remaining years forecasts are in reasonable agreement. For September month model, the CC values obtained are 0.9998 and 0.5775 during training and testing periods respectively. Fig. 4 shows the comparison of ANN forecasted rainfall with observed rainfall for September month during the testing period. It can be observed that the model has under forecasted for one year (1981), over forecasted for four years (1986, 1987, 1988, 1989) and for the remaining years the results are in reasonable accuracy.

Fig. 1. Comparison of rainfall predicted using ANN model and observed rainfall for June month (1981-1990).
Fig. 2. Comparison of rainfall predicted using ANN model and observed rainfall for July month (1981-1990).

Fig. 3. Comparison of rainfall predicted using ANN model and observed rainfall for the month of August (1981-1990).
Similarly the ANN model is trained for summer monsoon season, June-September (JJAS) rainfall forecasting. Fig. 5 shows, the comparison of observed rainfall with ANN model forecasted rainfall for monsoon season (JJAS) during testing period.

The CC values obtained for JJAS seasonal model are 0.9975 and 0.8951 during training and testing periods respectively. The model results are within reasonable accuracy with observed rainfall for most of the seasonal rainfall values, except minor deviation for one year (1982). When the monthly rainfall forecasting is compared
with seasonal rainfall forecasting performance, it can be observed that seasonal rainfall model is performing better than monthly models. This may be due to the fact that the dynamic nature of climate variables causes more uncertainty in monthly time scale than seasonal scale, thus larger variability can be observed in intra-seasonal rainfall prediction.

8 Conclusions

The climate indices of ENSO (Nino 3.4 SSTA), EQUINOO (EQWIN), and Ocean-Land Temperature Contrast (OLTC) have been used as predictor variables to predict the monthly as well as seasonal rainfall. After investigating the global climate variables that are highly influencing the regional rainfall, correlation analysis is carried out to select the monthly climate indices which have more influence on the individual monthly and seasonal rainfall. In a highly nonlinear time-series, to handle multiple variable influences on the rainfall prediction, ANNs have been adopted. Genetic optimizer is used to optimize the network architecture of ANNs. The models are trained and tested for individual months and for seasonal rainfall as a whole. The obtained results are encouraging and show improvement in rainfall forecasting. Incorporation of global climate information into rainfall prediction is very much useful and ANNs are found to be suitable technique for this purpose.

Acknowledgement

First author wishes to thank Department of Science Technology, Govt. of India, for supporting this work through a sponsored project with reference number ES/48/010/2003.

References