

MODELLING OF INDIAN SUMMER MONSOON RAINFALL USING GLOBAL CLIMATIC TELE- CONNECTION DATA

Debsunder DUTTA¹ & Ashu JAIN²

- (1) M. Tech Student, Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur 208 016, India, e-mail: ddutta@iitk.ac.in
(2) Professor, Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur 208 016, India, e-mail: ashujain@iitk.ac.in

ABSTRACT

Development of the models for forecasting the Indian Summer Monsoon Rainfall (ISMR) has been the subject of research among various researchers in the past. An extensive literature review in this area suggests a strong need to consider the global tele-connection data in and around Indian sub-continent in a systematic manner in order to develop meaningful models to forecast ISMR. Indian monsoon dynamics is extremely complex, non-linear, and dynamic in nature and needs to be modeled using tools capable of capturing such dynamics inherent in the data involved. A key step in the mathematical model development is the input variable selection. Most of the models proposed so far have employed the use of linear dependence measures of correlation coefficients, which are not suitable for assessing the strength of relationship inherent in the inputs and outputs of a physical system involving high degree of non-linearity and complexity.

The objective of the present study are to (a) investigate the use of global climatic tele-connection data in developing models to forecast ISMR in a systematic manner and (b) employing both regression and artificial neural network (ANN) models to develop the monsoon rainfall in India. The monthly rainfall data for All India, global tele-connection data of monthly average surface temperatures, and monthly average of the zonal and meridional wind components at a resolution of $2.5^\circ \times 2.5^\circ$ in a window covering the entire Indian Sub-continent on the North and Indian Ocean up to Australia to South are employed in this study. The results from naive models, linear and non-linear regression models, and ANN models are presented here. The performance of various models is evaluated using a wide variety of standard statistical parameters and graphical aides. The preliminary results obtained in this study for forecasting monsoon rainfalls for the months of June, July, August, September and JJAS for All India indicate that the ISMR can be effectively modeled using the global tele-connection data in the window considered in this study. The systematic approach adopted for input variable selection in this study has helped in developing accurate models for ISMR prediction.

1 INTRODUCTION

The Indian economy is mainly dependent on agriculture, irregularities in monsoon affects agriculture and also cause droughts and floods which greatly affects the Indian

economy. Therefore, accurate prediction of rainfall is beneficial to a billion people and helps directly or indirectly in the early planning and execution of agricultural strategies. The Indian Summer Monsoon Rainfall (ISMR) is highly dynamic in nature and its prediction on monthly and seasonal time scales is a scientifically challenging problem which is presently an important subject of research for the hydrologists and meteorologists.

The ISMR forecasting over the country started more than a century ago (Blanford 1884; Walker 1908, 1933) and generally requires the knowledge of spatial and temporal variability. The relation between ISMR and El Nino Southern Oscillation (ENSO) was discovered in the 1980s and a strong link of ISMR with ENSO was manifested as an increased propensity of droughts during El Nino or warm phase and of excess rainfall during the opposite phase i.e., La Nina (Sikka, 1980; Pant and Parthasarathy, 1981). Singh, Bhadram and Mandal (1995) proposed a regression model for All India seasonal summer monsoon rainfall and identified the average lowest Mean Sea level (MSL) pressure of heat low over central Pakistan and adjoining North West (NW) India for the month of May as a potential predictor and it had a stable, inverse and significant correlation with ISMR. Dube, Luther and O'Brien (1990) showed that interannual variability of the monthly mean upper layer thickness for central Arabian Sea is significantly correlated with ISMR. A significant negative correlation was shown between the Eurasian snow cover for the months of December to March and the monsoon rainfall of western, central and all India regions (Parthasarathy and Yang, 1995). Gadgil et al. (2004) suggested that a strong relation existed between large deficits or excess in ISMR and a composite index based on ENSO and EQUINO indices and further that the two modes are a part of a coupled system. Sahai et al., (2003) identified the Geo-potential height over the Asian land mass and over the Caribbean region in May and the tendency of Southern Oscillation Index (SOI) over the tropical Pacific ocean as the three main Teleconnections associated with ISMR and in combination can give better predictability. In another study, Sahai et al. (2003) identified 14 regions of 10° latitude $\times$ 10° longitude (hot spots of Global Oceans) that were able to capture biennial oscillation and reasonably predicted ISMR and showed the relationship between global sea surface temperature (SST) and ISMR particularly North Atlantic and South Pacific. A strong and significant dipole structure in the correlation pattern between ISMR and OLR over the tropical Pacific and Atlantic was proposed by Bansod (2004). Sadhram and Murthy (2008) identified that SSTA over southeastern Arabian Sea and northwest of Australia along with East Asia sea level pressure is significantly correlated with ISMR.

Development of models for the prediction of monsoon rainfall has also been investigated by many researchers. The prediction models can be broadly classified into two broad categories statistical and dynamical models. The dynamical models include general circulation models (GCMs). The advantage of current dynamical models is that they can provide time evolution and spatial distribution of rainfall at the models spatial and temporal resolution and the cause effect relationship between the various processes represented can also be analysed. The accuracy in long range prediction of ISMR with GCMs is still not of the desired level and so statistical models are still preferred. The statistical models are generally based on empirical relationships between ISMR and various predictors (teleconnection parameters). They can be generally classified as linear regression, power regression, time series analysis or other statistical methods.

Sahai et al. (2003) developed a multiple linear regression (MLR) model with 14 parameters from the Global Oceans for predicting All India ISMR. Iyengar and R Kant (2003) attempted to model the all India as well as regional summer monsoon rainfall as a non linear time series and observed that the model could account for 50% of the variability of the rainfall. Bansod (2003) made an attempt for predicting the ISMR by a MLR model and used OLR data. Sadharam et al. (2008) attempted to increase model efficiency by using three predictors only in a MLR model.

Artificial Neural Network (ANN) has been used widely in recent times for mapping or approximating non linear functions and modelling of complex, non linear, dynamical phenomenon (Muller and Reinhardt 1991). Navone and Ceccatto (1994) used ANN for ISMR prediction and used mainly two approaches, a stochastic approach based on predictors and a dynamical based on time series and also proposed a hybrid approach. Sahai et al. (2000) used ANNs in a time series approach for predicting all India summer monsoon rainfall. Sahai et al. (2003) employed ANNs for modelling the all India summer monsoon rainfall using three predictors only consisting of SSTA and SOI. An attempt towards regional rainfall forecasting was made by Kumar et al. (2007) and they used ANNs and evolutionary algorithms with SSTA, EQWIN and OLTC as predictors.

The review of literature does not give us a well defined insight into what variables should be selected as predictors or inputs for modelling the ISMR. Moreover, different researchers have used different explanatory variables and the monsoon dynamics is a constantly changing phenomenon as a result the factors governing the ISMR or which are prevalent in a particular decade may not be so in some other times. Thus, a further study of the explanatory variables is necessary. In recent times, the researchers have mainly focussed on the associations of ISMR with ENSO and some other indices like Equatorial Indian Ocean Oscillation (EQUINOO), Ocean Land Temperature Contrast (OLTC) etc. It can be seen that apart from ENSO other predictors are a form of the variables Wind or Temperature for regions in and around the Indian Subcontinent. Therefore, in this study, a window was selected covering the oceans around the Indian subcontinent and the variables wind and temperature were studied for that region. ANN models which are very good in capturing non linear relationships were developed with the selected predictors and some encouraging results were obtained.

2 STUDY AREA AND DATA

This study was carried out for the rainfall data of whole of India. The rainfall data is obtained from the Indian Institute of Tropical Meteorology(IITM), Pune. The Homogeneous Monthly Rainfall Datasets were used for this study. The Teleconnection data required for this study have been obtained from The National Oceanic & Atmospheric Administration (NOAA). The Physical Sciences Division (PSD) at the Earth Science Research Laboratory (ESRL) maintains a large number of climate datasets for use in climate diagnostics. In this study we have used the Reanalysis-I data sets, these data sets are in a gridded form and has data from 1948-present. These gridded data has spatial coverage of the entire surface of the earth (90°N to 90°S, 0°E to 357.5°E) and a spatial resolution of 2.5° latitude X 2.5° longitude global grid. The required data for a spatial and temporal range was extracted for use. Data of ENSO index such as the Sea Surface Temperature Anomaly (SSTA) from the NINO 3.4 region is obtained from the website of the National Weather Service, Climate Prediction Centre

of NOAA (<http://www.cpc.ncep.noaa.gov/data/indices/>). The monthly data of NINO 3.4 along with the anomaly are available from 1950 till the present day.

3 STUDY OF TELECONNECTION VARIABLES IN AND AROUND THE INDIAN SUBCONTINENT

A study of literature shows that there are no established guidelines for the selection of the explanatory variables for modeling the Indian Summer Monsoon Rainfall. The most commonly studied physical variables are the wind and temperature in and around the subcontinent. In this study we have studied the Air Temperature and Wind (Zonal and Meridional components) in and around the Indian subcontinent. A rectangular window of 60° latitudinal extent (30°N to 30°S) and 80 ° longitudinal extents (40°E to 120°E) was chosen as the study area covering the Arabian Sea, Bay of Bengal and Indian Oceans around the Indian Subcontinent. The study area also covered part of landmasses of Africa, Australia and South East Asia. A representation of the selected window is shown in Figure 1.

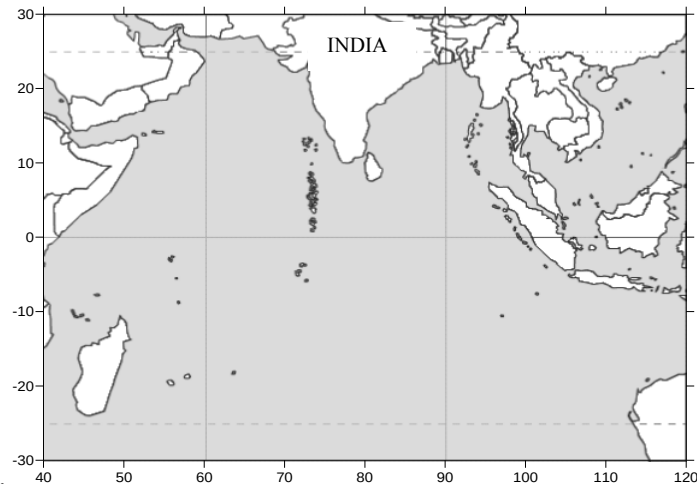


Figure 1: Spatial Window for the Study of Temperature and Wind around the Indian Subcontinent

The data is a gridded data of 2.5° X 2.5° resolution so we have a total of $33 \times 25 = 825$ spatial grid values in the entire window. The period from 1950 to 2008 was selected as the study duration and the data is obtained from the sources mentioned.

Cross Correlation (CC) coefficients were calculated between the All India Rainfall values and the Temperature or Wind values at each spatial grid point for the duration 1950-2008. This CC values were calculated between June, July, August, September and JJAS (sum of June, July, August and September) rainfall values and the variables with lags upto January. The most efficient way to represent the results is to present them in the form of contour plots. Thus, a number of contour plots were prepared. Figure 2 represents the plots for June rainfall with June Air Temperature, Zonal Wind and Meridional Wind.

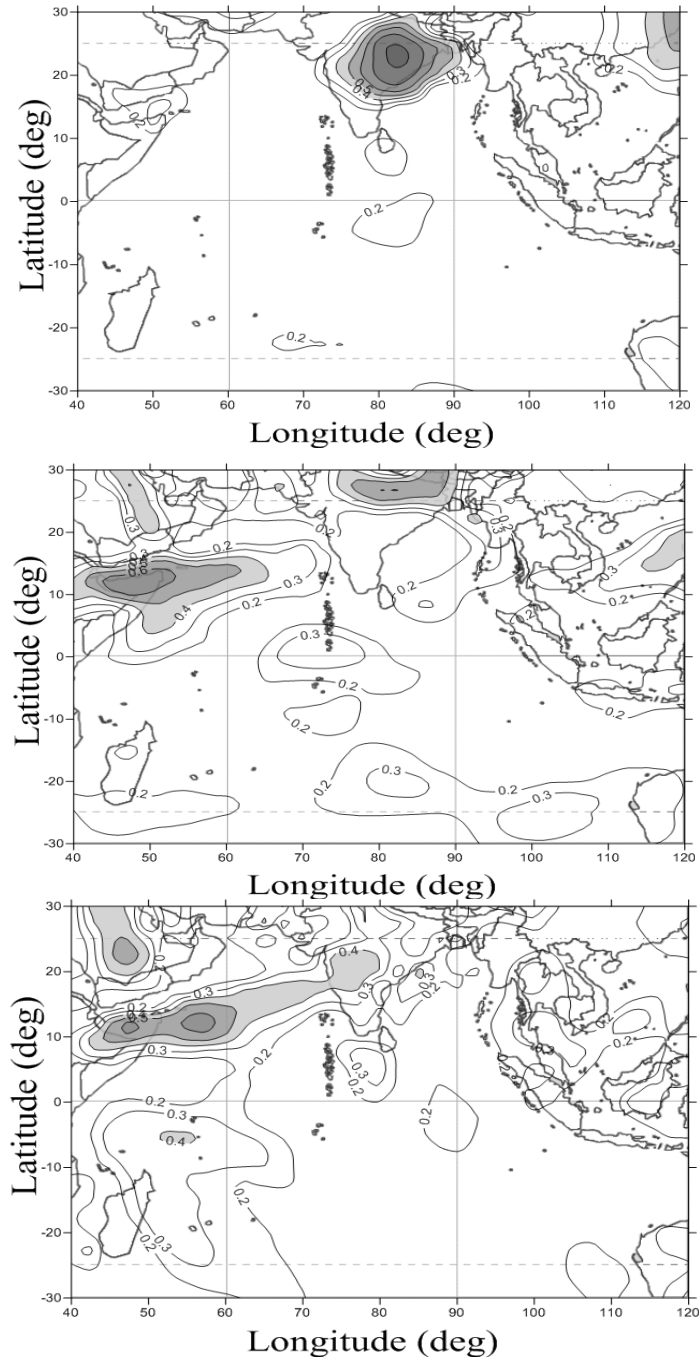


Figure 2: Cross correlation contour plots of all India June rainfall with June surface air Temperature, Zonal wind and Meridional wind from top to bottom respectively.

The observations showed that maximum values of CC were around 0.6, 0.7 and 0.8 for different cases so to include these values and to ignore the low values such as 0.3, 0.2 and 0.1, 0.4 was chosen as the cut-off value, the shaded regions are CC values of 0.4 and above. Since a lot of plots were available with different spatial extents having CC values of 0.4 and above a computer program was written to select all the areas with reasonable extents and averaging them out to get a single input vector for that window. The CC values were also computed for the NINO 3.4 anomalies and finally the Input set was prepared for all the five different cases. Table 1 shows the input variables which were finally selected for model development. Following the conventional modeling technique the entire data set of 59 years were divided into two approximate equal sets of 30 years (1950-1979) for calibration and 29 years (1980-2008) for testing.

Table 1. Input variables for the Various Models (T=Air Temperature, ZW=Zonal Wind, MW=Meridional Wind)

Months	No. Of Variables	Predictor Variables
June	9	T(May), T(June), ZW(Mar), ZW(May), ZW(June), MW(Jan), MW(May), MW(Jun) and NINO 3.4(Mar).
July	11	T(Jan), T(Jul), ZW(Jan), ZW(Feb), ZW(Jun), ZW(Jul), MW(Feb), MW(Mar), MW(Jun), MW(Jul), NINO 3.4(Jul).
August	9	T(Aug), ZW(Feb), ZW(Jun), ZW(Aug), MW(Mar), MW(Jun), MW(Aug), NINO 3.4(Jan), NINO 3.4(Feb).
September	14	T(Aug), T(Sept), ZW(Feb), ZW(Mar), ZW(Sept), MW(Mar), MW(Apr), MW(May), MW(Jun), MW(Jul), MW(Sept), NINO 3.4(Jul), NINO 3.4(Aug), NINO 3.4(Sept).
JJAS	14	T(Jan), T(Jul), T(Sept), ZW(Apr), ZW(Jun), ZW(Jul), ZW(Aug), ZW(Sept), MW(May), MW(Jun), MW(Sept), NINO 3.4(Jul), NINO 3.4(Aug), NINO 3.4(Sept).

4 REGRESSION AND ARTIFICIAL NEURAL NETWORK MODEL DEVELOPMENT

Four types of Regression models were developed, all the models are multiple regression models. One model is a linear model abbreviated as Multiple Linear Regression Model (MLRM Equation (1)) and the remaining three non-linear models abbreviated as Non-Linear Multiple Regression Models (NLMRM1, NLMRM2 and NLMRM3 Equation (2), (3) and (4)). The Equations representing the various models are as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \quad (1)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_4^2 + \dots \quad (2)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + \beta_4 X_2 + \beta_5 X_2^2 + \beta_6 X_2^3 \dots \quad (3)$$

$$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \dots \quad (4)$$

Where, Y is the rainfall, X_i 's are the input variables and β_i 's are the regression model parameters to be determined.

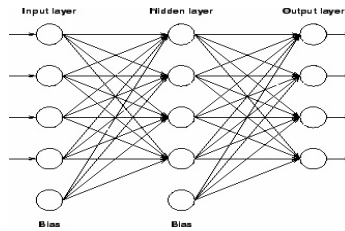


Figure 3: A feed forward Artificial Neural Network Model with a single hidden layer.

The ANNs are mathematical or computational models that tries to simulate the structural and functional aspects of a biological neuron. Figure 3 shows the structure of a single hidden layer neural network model. It is generally composed of units known as neurons which are interconnected by weights. There is an input layer, a number of hidden layers and an output layer in a feed forward ANN model. Activation function are used at the hidden and output layers to transfer the inputs received to the next layer. ANN is an adaptive system that changes its connection strengths (weights) based on the information that flows through the network during its learning or training phase. The neural network models are able to capture the non linear dynamics from a set of Input patterns during training. The activation function used for the present study is sigmoidal function. The networks have a single layer of hidden neurons. The error back propagation training algorithm is used for the training of the network. For this algorithm the value of learning rate (η) and momentum factor (α) were chosen to be 0.01 and 0.075 respectively after some trial and error. The number of iterations were fixed at 20,000 for optimum training of the networks. The determination of optimum number of hidden neurons was done by a trial and error procedure. The optimum numbers of hidden neurons for the June, July, August, September and JJAS models were found to be 3, 7, 3, 4 and 20 respectively.

5 RESULTS AND DISCUSSIONS

A number of performance evaluation statistics were calculated both for training and testing for all the models which are developed. They are correlation coefficient (R), Nash-Sutcliffe coefficient of Efficiency (E), Average Absolute Relative Error (AARE), Threshold Statistics (TS1, TS25, TS50 and TS100), Normalized Mean Bias Error (NMBE) and Normalized Root Mean Square Error (NRMSE)

Table 2. Performance Statistics of ANN models during Training

Model	R	E	AARE	TS1	TS25	TS50	TS100
June (9-3-1)	0.910	0.828	7.7	10.0	96.7	100.0	100.0
July (11-7-1)	0.901	0.812	4.8	13.3	100.0	100.0	100.0
August (9-3-1)	0.819	0.670	6.3	13.3	100.0	100.0	100.0
September (14-4-1)	0.948	0.898	5.5	6.7	100.0	100.0	100.0
JJAS (14-20-1)	0.940	0.881	2.9	30.0	100.0	100.0	100.0

Table 3. Performance Statistics of ANN models during Testing

Model	R	E	AARE	TS1	TS25	TS50	TS100
June (9-3-1)	0.753	0.454	9.9	10.3	96.6	100.0	100.0
July (11-7-1)	0.838	0.675	7.0	10.3	96.6	100.0	100.0
August (9-3-1)	0.759	0.479	6.8	0.0	100.0	100.0	100.0
September (14-4-1)	0.802	0.524	12.9	3.4	93.1	100.0	100.0
JJAS (14-20-1)	0.773	0.590	5.0	13.8	100.0	100.0	100.0

Table 4. Performance Statistics of Best Regression models during Training

Model	R	E	AARE	TS1	TS25	TS50	TS100
June (NLMRM3)	0.895	0.817	8.9	0.0	96.7	100.0	100.0
July (MLRM)	0.868	0.753	5.4	13.3	100.0	100.0	100.0
August (MLRM)	0.805	0.648	6.5	10.0	100.0	100.0	100.0
September (MLRM)	0.934	0.873	6.3	13.3	100.0	100.0	100.0
JJAS (MLRM)	0.977	0.955	2.0	26.7	100.0	100.0	100.0

Table 5. Performance Statistics of Best Regression Models during Testing

Model	R	E	AARE	TS1	TS25	TS50	TS100
June (NLMRM3)	0.769	0.597	10.2	3.4	100.0	100.0	100.0
July (MLRM)	0.844	0.606	8.5	10.3	96.6	100.0	100.0
August (MLRM)	0.720	0.006	9.3	3.4	96.6	100.0	100.0
September (MLRM)	0.805	0.644	10.8	10.3	93.1	100.0	100.0
JJAS (MLRM)	0.386	-0.281	8.6	0.0	93.1	100.0	100.0

Performance statistics in terms of R, E, AARE and TS for the ANN models during training and testing are shown in Table 2 and Table 3. The performance statistics of the best Regression models among the four models during calibration and testing are shown in Tables 4 and Table 5. The naive models like the Persistence model and the Mean model were also prepared but as their performances were very poor compared to these models so they are not reported here. For the month of June, the NLMRM3 performs better than the ANN model and for the September models MLRM performs better than ANN model. For all the other remaining models the ANN models perform consistently well. Even for June and September the performance of ANN model is good. The Regression models show a good performance measure terms of CC value but its E values are not that good compared to those from the ANN models. The regression models are not fully able to capture the dynamics as such for some months their performance is good but for others it is poor for JJAS the ANN model performs much better than the regression models. The scatter plots of the ANN models during testing are shown in Figure 4. The scatter plots of the Regression models for testing are shown in Figure 5. From the scatter plots it can be seen that for July and August a more number of points are below the 45° line indicating a slight under prediction from the models and for September the number of points above the 45° line are slightly more indicating overprediction. For the other models approximately an equal number of points lies both above and below the 45° line.

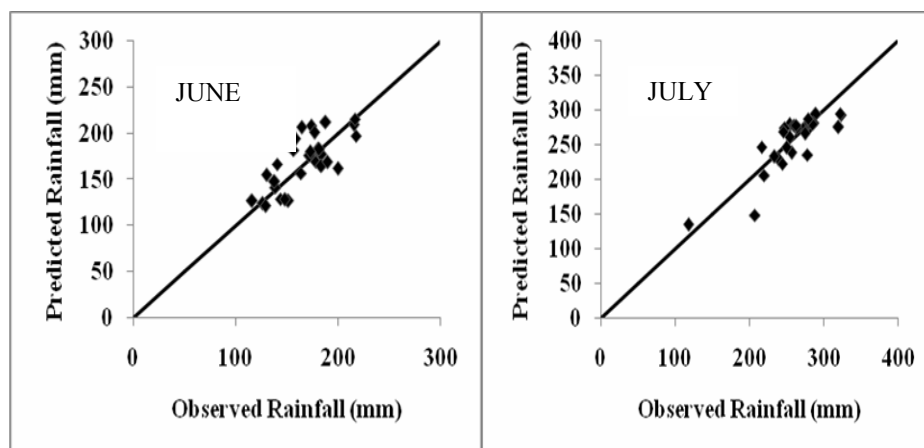


Figure 4(a): Scatter Plot for ANN model for JUNE and JULY Rainfall during Testing

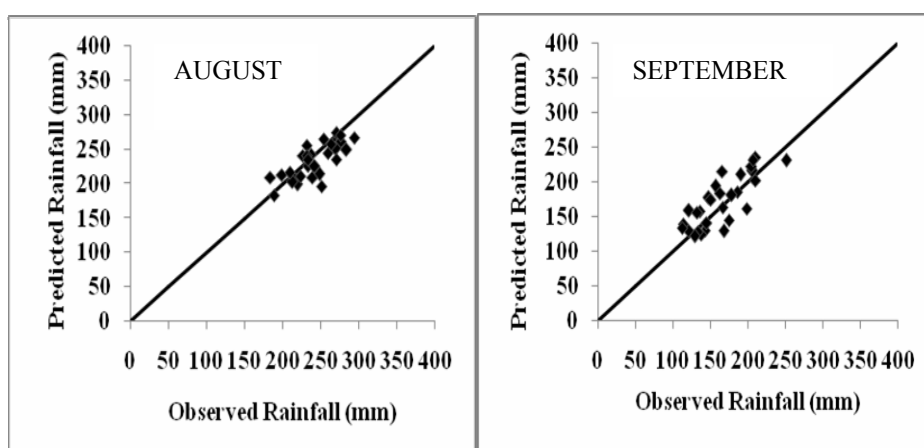


Figure 4(b): Scatter Plot for ANN model for AUGUST and SEPTEMBER Rainfall during Testing

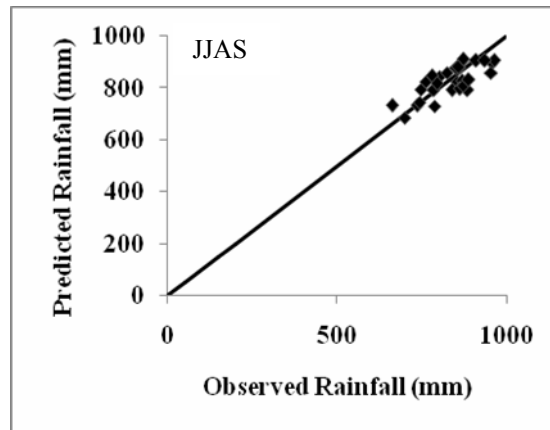


Figure 4(c): Scatter Plot for ANN model for JJAS Rainfall during Testing

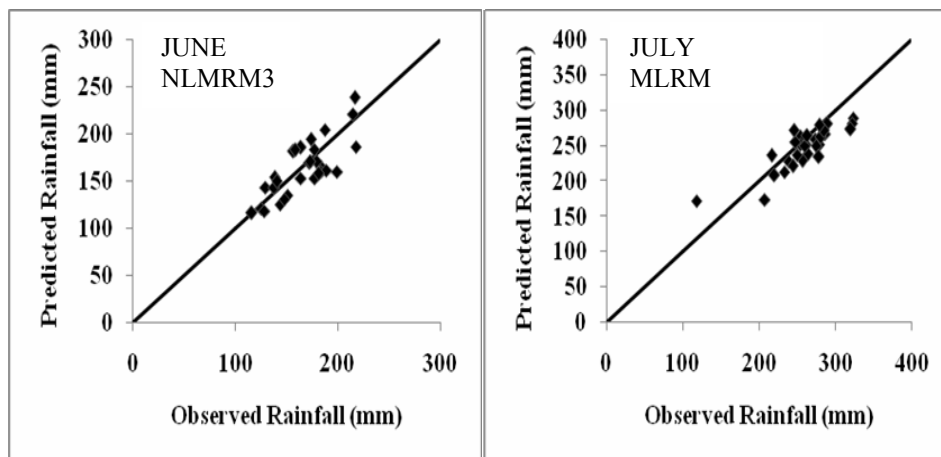


Figure 5(a): Scatter Plot for best regression models for JUNE and JULY Rainfall during Testing.

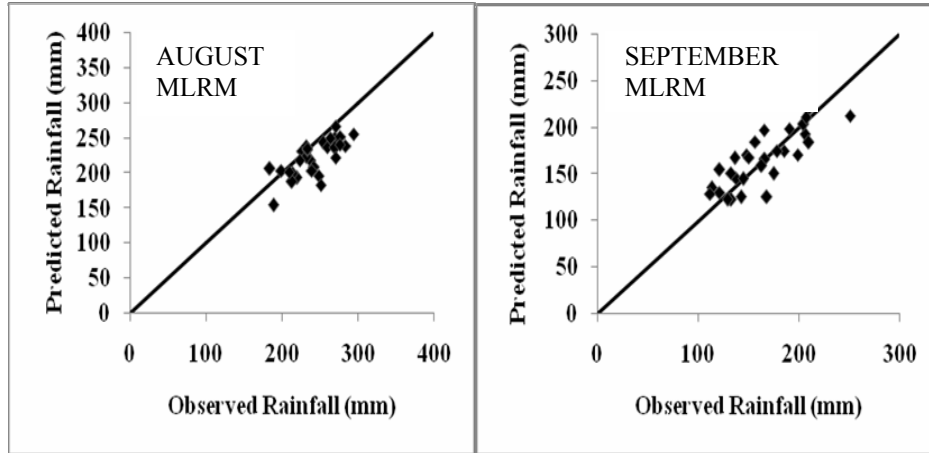


Figure 5(b): Scatter Plot for best regression models for AUGUST and SEPTEMBER Rainfall during Testing.

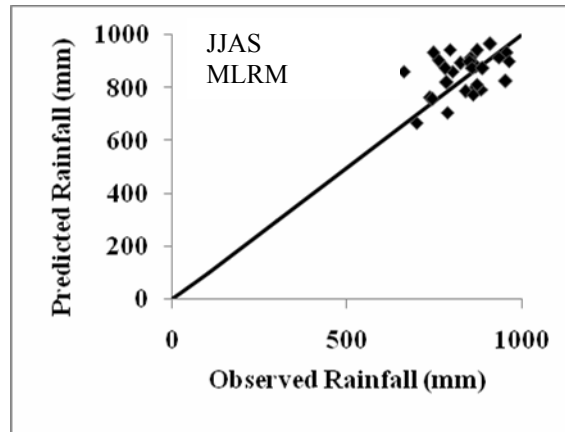


Figure 5(c): Scatter Plot for best regression models for JJAS Rainfall during Testing.

6 CONCLUSION

In this study the climate indices of ENSO (NINO 3.4 SSTA), Air Temperature, Zonal Wind and Meridional Wind in and around the Indian Subcontinent were studied and used as predictors for monthly as well as seasonal monsoon rainfall forecasting in India. The most influential variables from a larger pool were selected on the basis of linear correlation for monsoon prediction and ANNs were used to model the highly non linear dynamics. The results obtained are quite encouraging and prove the appropriateness of ANNs and further motivates for more insights into monsoon prediction.

Acknowledgments. The authors wish to acknowledge The Holocene (2008) paper and IITM web site (<http://www.tropmet.res.in>) for the rainfall data. For the Reanalysis datasets the authors wish to thank the NOAA for making the data available freely and Kalnay et al. (1996). are acknowledged for the datasets.

REFERENCES

- Blanford, H.F. (1884). On the connection of the Himalayan snowfall with dry winds and seasons of draughts in India. *Proc R Soc London* 37:3-22.
- Walker, G.T.(1908). Correlation in Seasonal variation of Climate (Introduction). *Mem Ind Meteorological Dept* 20:117-124.
- Walker, G.T. (1933). Seasonal weather and its prediction. *Brit Assoc Adv Sci*, 103:25-44
- Gadgil et al. (2004). Extremes of the Indian summer monsoon rainfall, ENSO and equatorial Indian ocean oscillation. *Gepophysical Research Letters*, Vol., 31, L12213.
- Sikka, D.R. and Gadgil, S. (1980). On the maximum cloud zone and the ITCZ over India longitude during the southwest monsoon. *Mon. Weather Rev.*, 108,1122-1135.
- Pant, G.B. and Parthasarathy, B. (1981). Some aspects of an association between the southern oscillation and Indian summer monsoon. *Arch. Meterol. Gepphys. Bioklemtol.*, 29, 245-251.
- Singh et al. (1995). New Regression Model for Indian Summer Monsoon Rainfall. *Meteorology and Atmospheric Physics* 55,77-86.
- Dube et al. (1990). Relationships Between Interannual Variability in the Arabian Sea and Indian Summer Monsoon Rainfall. *Meteorology and Atmospheric Physics*, 44, 153-165.
- Parthasarathy, B. And Yang, S. (1995). Realationships between Regional Indian Summer Monsoon Rainfall and Eurasian Snow Cover. *Advances in Atmospheric Sciences*, Vol. 2, No.2.
- Sahai et.al. (2003). Teleconnections in recent times and prediction of Indian Summer monsoon rainfall. *Meteorol Atmos Phys* 84, 217-227.
- Sahai et al. (2003). Long-lead prediction of Indian summer monsoon rainfall from global SST evolution. *Climate Dynamics* 20:855-863.
- Sadhuram, Y. And Ramana Murthy T.V. (2008). Simple Multiple Regression Model for long range forecasting of Indian Summer Monsoon Rainfall. *Meteorol Atmos Phys* 99, 17-24.
- Bansod, S.D. (2004). Outgoing long-wave radiation over the Tropical and Atlantic Ocean and Indian summer monsoon rainfall. *Theor. Appl. Climatol.* 77, 185-193.
- Iyengar, R.N. and Raghu Kanth, S.T.G. (2003). Empirical modelling and forecasting of Indian monsoon rainfall. *Current Science*, Vol. 85, No. 8, 25.
- Sahai et al. (2000). All India summer monsoon rainfall prediction using an artificial neural network. *Climate Dynamics* 16:291-302.
- Muller, B. And Reinhardt, J. (1991). Neural Networks: an introduction, Vol 2, The physics of neural networks series. Springer, Berlin, Heidelberg New York.
- Navone, H.D. and Ceccatto, H.A. (1994). Predicting Indian monsoon rainfall: a neural network approach. *Climate Dynamics* 10:305-3212.
- Kumar et al. (2007). Regional Rainfall Forecasting using large Scale Climate Teleconnections and Evolutionary Algorithms. *Journal of Intelligent Systems*, Vol. 16, No. 4.
- Kalnay et al. (1996).The NCEP/NCAR 40-year reanalysis project, Bull. Amer. Meteor. Soc., 77, 437-470.
- Haykin, S. (1999) Neural Networks. New York, MacMillan Publishing Company.