Vol.4.Issue.4.2016 (October -December)

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

http://www.bomsr.com
Email:editorbomsr@gmail.com

ISSN:2348-0580

RESEARCH ARTICLE

RAINFALL PREDICTION BY SEASONAL NEURAL NETWORKS ENHANCED WITH FIRST ORDER MARKOV MODELS

ARUMUGAM.P[1]* KARTHIK.S.M[2]

[1,2]Department of Statistics, Manonmaniam Sundaranar University
Tirunelveli, India

ABSTRACT

A novel stochastic modeling approach is introduced to monthly rainfall prediction. Amounts of rainfall is a time series with seasonal and cyclic variation. Markov models are used to predict the overall levels of rainfall and seasonal neural networks are utilized to arrive at the exact quantity. The seasonal neural networks take the twelve monthly rainfall figures of the previous year at the input nodes and produces the next year levels at the output nodes. Rainfall patterns are classified into two categories namely wet and dry. First order Markov chains are used to predict the category of the next year. Four different seasonal neural networks are trained for each of the four different transitions. This enhances the prediction accuracy and increases the stability of the prediction for more than one year time frame. Experiments are done on monthly rainfall data in India provided by the Indian Meteorological Department. The results show the superior performance of the proposed method compared to existing seasonal neural networks based method.

Key Words: Seasonal Neural Networks, Markov chain, Rainfall Prediction, Time series, Stochastic modeling

1. INTRODUCTION

The study of rainfall pattern in India has far reaching consequences for the livelihood of more than a billion people. More than half of India's fledgling population is directly involved in the agrarian economy. The process of providing planned irrigation to the agricultural lands is an ongoing process in India. A significant part of India's farming communities have their prosperity intertwined with the rainfall. India gets its rainfall from the monsoon system with two prominent monsoons. The south west monsoon and north east monsoon systems bring moisture from the Arabian sea and

Bay of Bengal and supplies the much needed freshwater to India. Each year the arrival of the monsoon is eagerly looked forward to and the rural economic development is highly dependent on it. Water management is one of the most important issues in India. Proper forecasting of rainfall and adequate desilting of water bodies, proper use of dams, canals are important. Overlooking these issues can lead to strife as evidenced by the several disputes over fresh water. Another aspect of rainfall management is urban planning. Improper draining facilities and the encroachment of lake beds have created difficult scenes during flood in cities. Forecasting of rains enable the proper planning of relief and prevention procedures. Thus the study of rainfall is paramount to the prosperity and well being of India.

The rainfall patterns exhibit a seasonal and cyclical nature. The average annual rainfall over the last century is around 1176 mm. The highest rainfall occurs during June-September months with an average of 280 mm per month. Z` Over the last 114 years, 61 years have above average rainfall while 53 years received lower than average rainfall. The average annual rainfall is shown in figure 1.

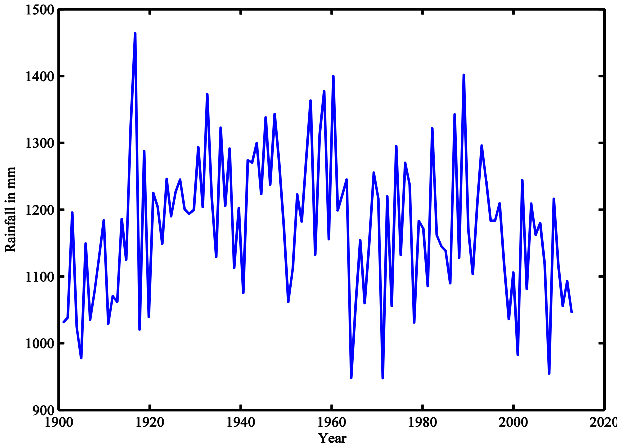

Figure 1. Average Annual Rainfall in India

The average monthly rainfall is shown in figure 2 and the monthly variation between the dry and wet years are shown in figure 3. It can be seen that the wettest months are the ones which have the maximum variation. The highest rainfall of 1463.9 mm occurred in 1917 and the lowest of 947.1 mm in 1972. The data is provided by the Indian Ministry of Earth Sciences and hosted in data.gov.in website [1]. The amounts of rainfall in consecutive years show high correlation. The dry and wet spells occur mostly together and change cyclically.

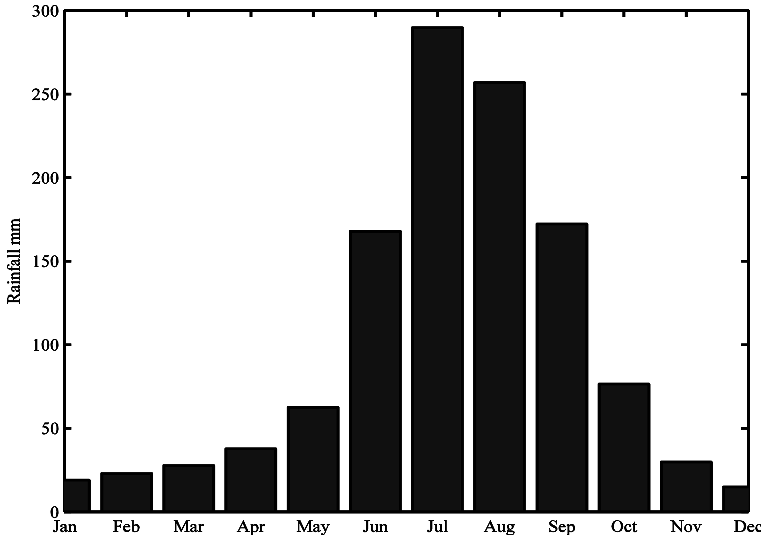

Figure 2. Average Monthly Rainfall in India

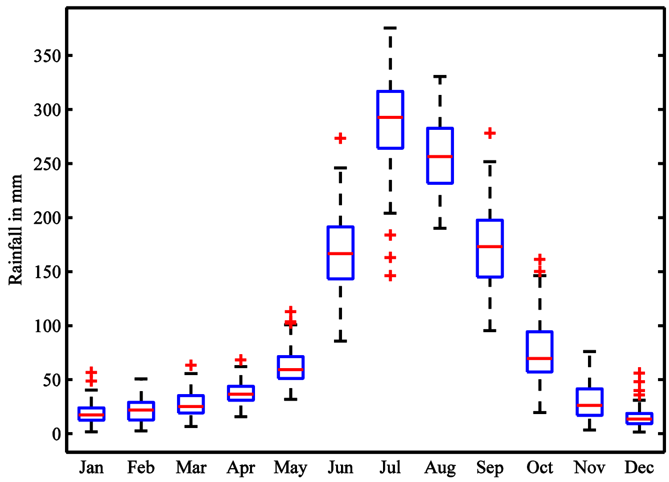

Figure 3. Monthly Variation in Rainfall in India

The attempt to forecast the Indian monsoon started more than a century ago. The MONEX program increased the understanding of the monsoon phenomenon. Many empirical statistical models [2, 3] have been developed with various levels of success. One of the key consideration in forecasting is the accuracy of long range predictions. Sahai et al. [4] used Artificial Neural Networks to predict Indian Summer Monsoon. They found decadal scale variability in predictability. They achieved an absolute prediction error of less than 15%.

2. MATERIALS AND METHOD

The years are classified into two classes namely Dry, Medium and Wet based on the annual rainfall. The monthly rainfall amounts are input to *K*-Means clustering to cluster the years into three categories. The first order transition probability matrix is calculated on the data and is given as

$$T = \begin{bmatrix} 0.2727 & 0.2424 & 0.4848 \\ 0.4000 & 0.3750 & 0.2250 \\ 0.2000 & 0.4250 & 0.3750 \end{bmatrix} \quad (1)$$

Accordingly nine different seasonal neural networks are trained each with the appropriate transitions. The Seasonal neural networks are

$$S = \{S_{11}, S_{12}, S_{13}, S_{21}, S_{22}, S_{23}, S_{31}, S_{32}, S_{33}\} \quad (2)$$

First a seasonal neural network *(SANN)* is trained on the entire dataset and is used as a template. Then the base SANN is trained with the particular transitions to get the specialized SANNs. The functioning of the SANNs are discussed briefly below. Seasonal Artificial Neural Networks *(SANN)* was proposed by Hamzacebi in [5]. It captures nonlinear dependence on the observed and predicted variables. SANN captures the seasonal variations and is more suitable for seasonal data than linear methods like Autoregressive Integrated Moving Average *(ARIMA)* [6] models. It is well suited for predicting time series of quantities that exhibit cyclical patterns. The input layer of SANN [7] consists of s nodes, where s is the period of the seasonal variation. In monthly rainfall data, *s* is taken to be 12 corresponding to the twelve months of the year. The measurements of every month of the year are fed through the input neurons. A hidden layer with fixed number of neurons is taken. Higher the number of neurons in hidden layer leads to better modeling capability but requires more training data. The output layer corresponds to the prediction for the monthly rainfall for the next year. Figure 4 illustrates the SANN used in this work. Let $R_{i,j}$ be the amount of rainfall observed in the j^{th} month of the i^{th} year. The input neurons are fed with $R_{i,j}, j = 1,2,\ldots,12$. The output neurons are trained to give$R_{i+1,j}, j = 1,2,\ldots,12$. The weights $U_{j,k}, j = 1,2,\ldots,12\,; k = 1,2,\ldots,h$ are present in the connections between the 12 neurons in the input layer and the h neurons in the hidden layer.

The weights $V_{k,j}, j = 1,2, \dots ,12\,; k = 1,2, \dots , h$ are present in the connections between the h neurons in the hidden layer and the 12 nodes in the output layer. In addition to these neurons, bias neurons with the weights α_k and $\beta_j$are present. The relationship between the output and input neurons are captured by the equation.

$$\hat{R}_{i+1,j} = \beta_j + \sum_{k=1}^{h} \left(V_{kj} f\left(\alpha_k + \sum_{j=1}^{12} U_{j,k} R_{i,j}\right)\right) \quad (3)$$

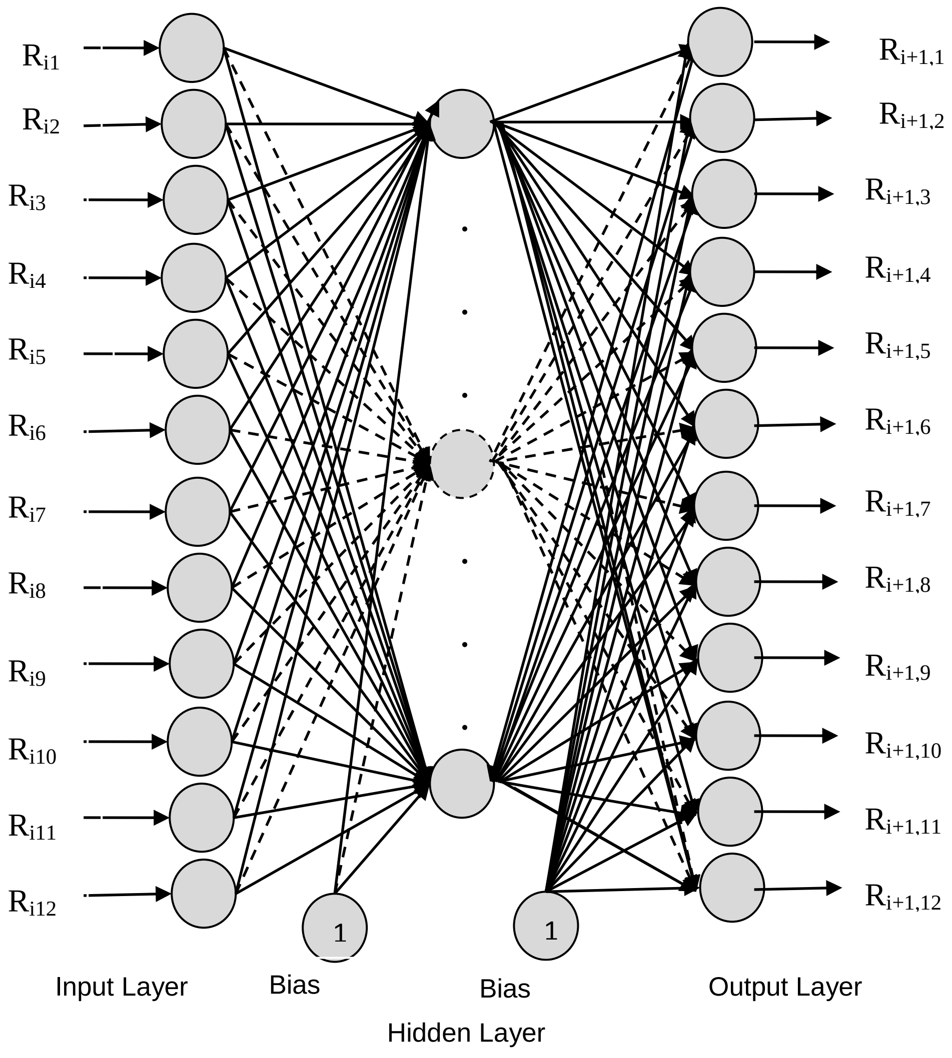

Figure 4. Seasonal Neural Network with s = 12

Here f is the activation function. The choice of a proper architecture is crucial in the accuracy of all types of ANNs. The only parameter here is h which is the number of neurons in the hidden layer. It is decided on an empirical basis. Common activation functions used include logistic function, softmax function and the gaussian function. In this work, the gaussian function is used because the predicted variable is continuous. The stochastic gradient descent algorithm [8] is used to train the SANNs.

When the current year falls into the k^{th} category, the three SANNs $\{S_{k1}, S_{k2}, S_{k3}\}$ are used to predict the next year monthly rainfall patterns $\{M_1, M_2, M_3\}$. The final prediction is taken as

$$M = T(k,1) * M_1 + T(k,2) * M_2 + T(k,3) * M_3 \quad (4)$$

The data consists of 114 years of observations. It is partitioned randomly into training and testing set each consisting of 57 observations. The parameter *h* is varied from 4 to 9. The gradient descent method is used for training. The experimental results are summarized in the next section.

3. EXPERIMENTAL RESULTS AND ANALYSIS

The results of prediction for the average annual rainfall is given in figures 5-8 for different number of hidden layers. The parameter setting h = 7 gives the best result.

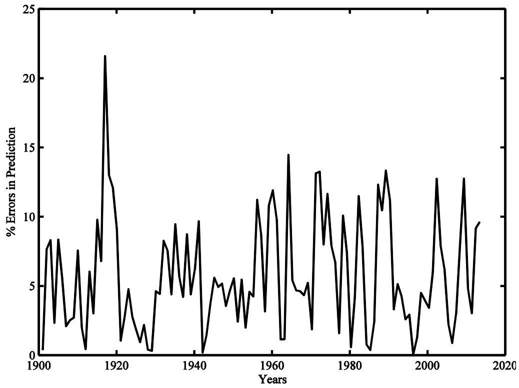

Figure 5. Prediction Errors for number of Hidden Layers = 6

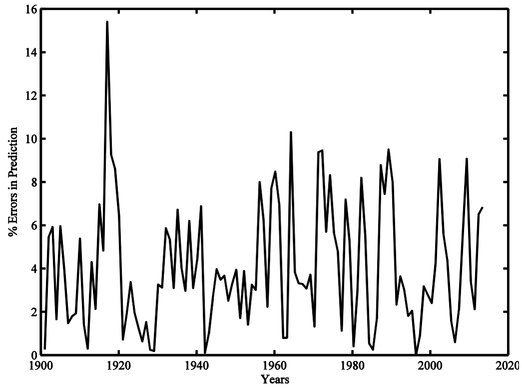

Figure 6. Prediction Errors for number of Hidden Layers = 7

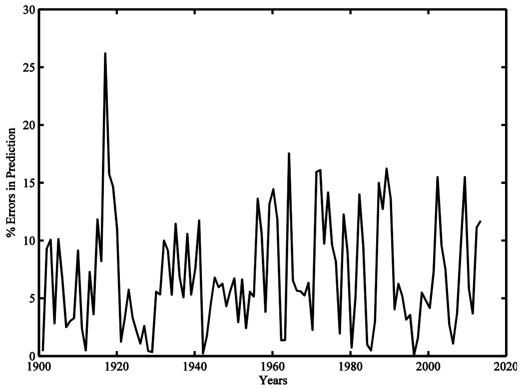

Figure 7. Prediction Errors for number of Hidden Layers = 8

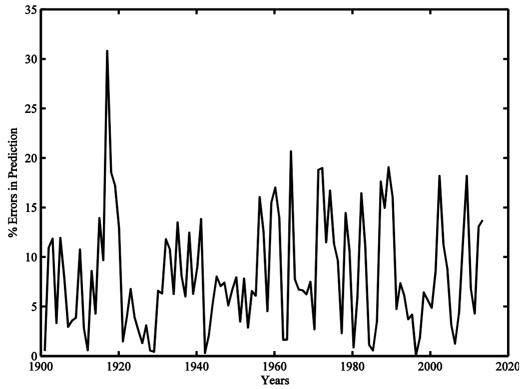

Figure 8. Prediction Errors for number of Hidden Layers = 9

The monthly average prediction errors are displayed in figures 9 -12.

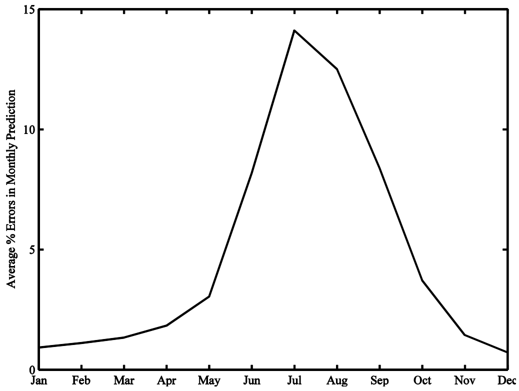

Figure 9. Monthly Average % Errors h = 6

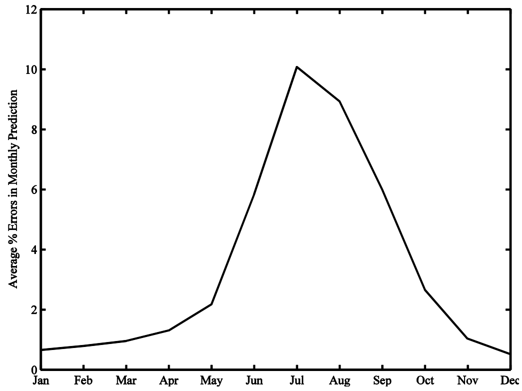

Figure 10. Monthly Average % Errors h = 7

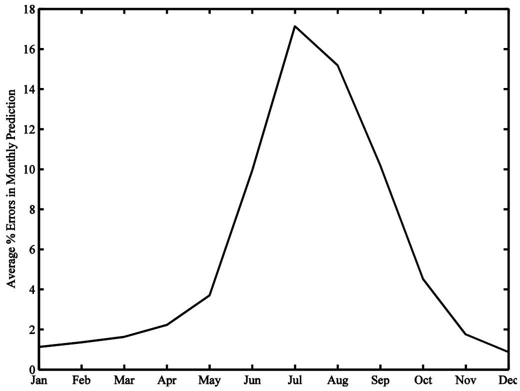

Figure 11. Monthly Average % Errors h = 8

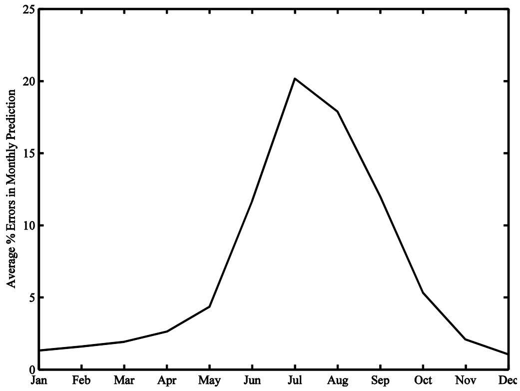

Figure 12. Monthly Average % Errors h = 9

The errors for h = 7 are optimal. The percentage errors in annual rainfall is below 16% and monthly rainfall is below 10%. To analyze the proposed method, the following performance measures are used.

Let $R_{i,j}, i = 1,2,..,n; j = 1,2,\dots,12$ be the actual rainfall recorded in the test set for n years. Let $\hat{R}_{i,j}$ be the prediction for the same period from the SANN. Then the following performance measures are used to assess the accuracy of the prediction.

Mean Squared Error (MSE) is defined as

$$mse = \frac{1}{12n}\sum_{i=1}^{n}\sum_{j=1}^{12}\left(R_{i,j} - \hat{R}_{i,j}\right)^2 \tag{5}$$

Root Mean Squared Error (RMSE) is defined as

$$rmse = \sqrt{mse} \tag{6}$$

Mean Absolute Deviation (MAD) is defined as

$$mad = \frac{1}{12n}\sum_{i=1}^{n}\sum_{j=1}^{12}\left|R_{i,j} - \hat{R}_{i,j}\right| \tag{7}$$

It is the average of all absolute deviations of the predicted from the actual values.

Mean Absolute Prediction Error (MAPE) [9] is also known as Mean Absolute Percentage Deviation (MAPD). It is defined as

$$mape = \frac{1}{12n}\sum_{i=1}^{n}\sum_{j=1}^{12}\left|\frac{R_{i,j}-\hat{R}_{i,j}}{R_{i,j}}\right| \tag{8}$$

MAPE can be used in our current application since there are no zero values in the predicted variable and does not cause division by zero error. When MAPE is multiplied by 100, it is expressed as a percentage.

MSE, RMSE, MAD and MAPE must be low for a good prediction. RMSE, MAD and MAPE are expressed in the same units as the predicted variable i.e., in mm in this case. The performance measures of the proposed method is shown in table 1.

Table 1. Performance measures of Proposed Method

h	MSE	RMSE	MAD	MAPE
4	501.20	22.39	21.60	0.4349
5	263.75	16.24	15.30	0.3213
6	85.31	9.24	9.01	0.2128
7	10.98	3.31	3.80	0.0131
8	32.89	5.74	5.81	0.1027
9	53.32	7.30	7.10	0.1159

The proposed method with h=7 is compared against commonly used time series models [10] SANN, Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA) [11] and Hidden Markov Models [12]. The results are tabulated in table 2. The proposed method performs better than related methods.

Table 2. Comparative Analysis

Method	MSE	RMSE	MAD	MAPE
Hybrid SANN (h=7)	10.98	3.31	3.80	0.0131
SANN (h=8)	32.89	5.73	4.96	0.1027
HMM	38.69	6.22	5.34	0.1131
ARMA	39.38	6.28	5.37	0.1134
ARIMA	40.36	6.35	5.55	0.1194
SARIMA	43.42	6.59	5.82	0.1200

4. CONCLUSION

Rainfall prediction using seasonal neural networks enhanced by first order markov chain is presented. Years are classified into three types based on the monthly rainfall pattern. K-means clustering algorithm was used to classify the years. Based on the average annual rainfall of the cluster centroids, the years were labeled as Dry, Medium and Wet. The transition probability matrix was calculated for the nine types of transitions. Separate seasonal neural networks were trained for each transition to fine tune the networks to different cases. The weighted average of the results of the three applicable networks was taken as the final prediction. Experimental results indicated that this approach gives better results than related methods. This approach can be applied in the future to other chaotic time series models in the future. Also the decadal variability can be incorporated into the stochastic modeling directly.

5. REFERENCES

[1]. Bottou, Léon., 2010, Large-scale machine learning with stochastic gradient descent. Proceedings of COMPSTAT 2010. Physica-Verlag HD, 177-186.

[2]. Hastenrath, Stefan, and Arnold Rosen. 1983, Patterns of India monsoon rainfall anomalies. Tellus, A 35.4, 324-331.

[3]. Hamzaçebi, Coşkun. 2008, Improving artificial neural networks' performance in seasonal time series forecasting. Information Sciences, 178.23, 4550-4559.

[4]. Hagan, Martin T., et al. 1996, Neural network design. Vol. 20. Boston: PWS publishing company.

[5]. Hamilton, James Douglas. 1994 Time series analysis. Vol. 2. Princeton: Princeton university press.

[6]. Kalpakis, Konstantinos, Dhiral Gada, and Vasundhara Puttagunta. 2001, Distance measures for effective clustering of ARIMA time-series. Data Mining, ICDM 2001, Proceedings IEEE International Conference on. IEEE.

[7]. MacDonald, Iain L., and Walter Zucchini. 1997, Hidden Markov and other models for discrete-valued time series. Vol. 110. CRC Press.

[8]. Open Government Data Platform India url: https://data.gov.in/.

[9]. Parthasarathy, B., H. F., Diaz, and J. K. Eischeid., 1998, Prediction of all-India summer monsoon rainfall with regional and large-scale parameters. Journal of Geophysical Research: Atmospheres 93.D5 (1988): 5341-5350.

[10]. Sahai, A. K., M. K. Soman, and V. Satyan. 2000, All India summer monsoon rainfall prediction using an artificial neural network. Climate dynamics, 16.4, 291-302.

[11]. Tseng, Fang-Mei, and Gwo-Hshiung Tzeng. 2002, A fuzzy seasonal ARIMA model for forecasting. Fuzzy Sets and Systems 126.3, 367-376.

[12]. Willmott, Cort J., and Kenji Matsuura. 2010, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research 30.1, 79-82.