MANTECH Publications

Journal of Engineering Mathematics & Statistics
Volume 3, Issue 1, January-June 2019

New ANN Model for Forecasting Indian Monsoon Rainfall

Kokila Ramesh[1], Anita Chaturvedi[2], Radha Gupta[3]

Assistant Professor[1], Professor[2, 3]

Department of Mathematics

Jain University, Bangalore, Karnataka, India[1, 2], Dayanand Sagar College of Engineering, Bangalore, Karnataka, India[3]

Corresponding Author's Email: *r.kokila@jainuniversity.ac.in[1]*

Abstract

Variability in monsoon rainfall is a natural hazard that India faces every year. The total seasonal precipitation in the month of June, July, August and September is generally known as southwest monsoon (SWM) rainfall. While the season recurs annually, the variation about the long term expected value can be as high as 40-50% in some parts of the country. The hazard due to droughts and floods caused by extreme variations can be mitigated to some extent if the rainfall time series can be modeled efficiently for simulation and forecasting. Rainfall data is a strongly non-Gaussian time series exhibiting non-stationarity. Artificial neural network (ANN) models are known to be versatile in handling such complex and unstructured data. In this paper a new ANN model general enough to model the available century long data in several sub regions of India is developed. The model is found to be efficient in explaining nearly 80% of the data variance. One year ahead forecast on a set of data, independent from the training period is also found to perform very well.

Keywords*: - SWM, Indian Rainfall, ANN, Modeling, forecasting*

INTRODUCTION

It is estimated that in large parts of India, 75% of the annual rainfall happens during the four months from June to September, popularly known as southwest monsoon (SWM) season. Whereas, the rainfall season is joyfully celebrated in folk and classical literature the vagaries of the phenomenon are dreaded alike by the farmers and the common public. The expected

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

total SWM rainfall is an index that is sought after in advance by the administrators to foreshadow the economic activity of the country. Like other atmospheric parameters the SWM rainfall value shows strong spatial and temporal variability. The spatial average of the country as a whole, known as All India Rainfall (AIRF), has a long-term time average (LTA) of about 89 cm with the standard deviation hovering about 9 cm. On the other hand, rainfall values in smaller regions show greater variability. For example, the coefficient of variation (%) of the subdivision comprising of Saurashtra & Kutch is as high as 44%.

The past literature on Indian rainfall modeling studies can be classified into two broad groups. The first group is dynamical modeling based general circulation theories of the atmosphere and the oceans to simulate summer monsoon circulation and associated rainfall pattern. Several such general circulation models with varying levels of refinement have been proposed in the past (Latif et al ,1994; Goswami, 1998; Krishnamurthy et al ,2000; Wang et al ,2005; Sajani et al, 2007; Rajendran et al, 2008). These models are qualitatively promising but have not yet shown quantitatively verifiable statistical skill to simulate the SWM rainfall quantity and its inter-annual variability (Gadgil and Sajini, 1998; Kang et al, 2002; Gadgil et al, 2005; Krishna Kumar et al, 2005; Wang et al., 2005).

The second approach is empirical modeling based on past rainfall and other atmospheric data. The statistical correlation between rainfall and antecedent climate parameters, whenever significant is attractive in forecasting. Gilbert Walker (1923, 1924) initiated simple linear regression models based on statistical correlations between SWM rainfall, snow accumulation over Himalayas in the month of May and South American atmospheric pressure parameters during spring. The performance of this model was found to be reasonable for 65% of the samples during 1924-1987, but deteriorated slowly over time. Inspired by such models based on precursors several linear and nonlinear regressions using as many as sixteen precursors have been proposed (Gowariker et al 1989 and 1991). Such models tend to have more undetermined parameters than the available data and hence show artificial skill but lead to poor forecasts (Delsole and Shukla 2002). Thapliyal (1990 and 2001) attempted to reduce the number of predictors by a transformation, but still the number of parameters required was 36 for a time series of 30 years, leading to spurious fit. Sahai et al (2003) made use of global sea surface temperature data for long lead prediction of Indian monsoon rainfall, through multivariate correlation analysis. They showed that the performance efficiency of their model

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

fit for 105 years of data was as high as 0.72. The corresponding value in the independent verification period for 22 years was also quite significant at 0.71. Nevertheless, such a model failed to forecast the drought of 2002 and 2004 (Gadgil et al, 2005). New statistical models using two stage forecasting, first stage requiring the precursor predictor data set up to March and the second stage up to May with six predictors to improve the official operational forecasting were developed by Rajeevan et al (2005) of the India Meteorological Department (IMD). Ashok Kumar et al (2012) improved the above with step wise linear regression and nonlinear ANN techniques for three stage forecasting (April, June and July) of SWM rainfall over India. It is generally found that ANN methods work better in comparison with linear methods of forecasting (Eisner and Tsonis, 1992). Since the predictors used are atmospheric parameters such as pressures and sea surface temperatures, which are time varying and non-Gaussian, their correlation with SWM rainfall changes with time and hence selection of the right predictors remains unsolved.

Another approach to the problem is to handle rainfall data as a time series with no other climate parameters in either modeling or forecasting. Sahai et al (2000) proposed ANN techniques to forecast monsoon rainfall using only past data. They considered the All-India spatial average of each of the four monsoon months (June, July, August, September) and also their sum as time series data for the period 1871-1960. Five antecedent values of all the five-time series were used to train the multilayer ANN to model and predict the five-component vector for the subsequent year. This network requires 25 input nodes, 2 hidden layers and 1 output node, leading to a total of 276 model parameters. These are found from five time series each of ninety sample values, but the seasonal value is just the sum of the other four and hence is not an independent sample. Further since five antecedent years are used in training the model the sample size is effectively 335. The modeling efficiency, in terms of variance explained, was found to be 0.8 which is not surprising, since the number of independent parameters is more than half the sample size. The ANN model proposed by Guhathakurta (2008) for the 36 subdivisions of India includes 11 to 12 antecedent rainfall values in the input layer with 3 neurons in the hidden layer and one output. This amounts to 40 to 43 parameters in the model, more than half the length of the sample size of 51 used in the training period (1941-1991). The modeling efficiency in the training period of 51 years for all India time series was shown to be 0.7 and for the subdivisions it was found to be 0.8. Another example of this kind of ANN model was by Pritpal and Bhogeswar (2013) for All

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

India rainfall data. They proposed five different three layered (input, hidden and output) ANN architectures with 43, 57, 73, 91 and 111 number of parameters. The efficiency of this ensemble model in the training period of 84 years was found to be 0.65. It is observed that some of the ANN approaches above are not skilful since by increasing the number of parameters to the sample size a polynomial function can be made to fit the data series exactly. A new approach of studying SWM rainfall by decomposing the data series into finite number of uncorrelated components called intrinsic mode functions (IMFs) was developed by Iyengar and Raghukanth (2004). They demonstrated that the SWM data (1901-2000) can be partitioned as the sum of a strongly non-Gaussian short period component and a nearly Gaussian long period component. The first part was modeled using ANN methods with one input layer having 5 nodes, a hidden layer with 5 nodes and a single output. The remaining part was modeled as a linear five step auto-regression. The total number of unknown parameters to be found in this model would be just 42, for a sample size of 100. The modeling efficiency of the non-Gaussian part through ANN was 0.84 and the overall variance explained by this model was as high as 0.83. This approach is very efficient in modeling and simulation, but for forecasting the empirical mode decomposition (EMD) suffers from end point approximations, because the numerical method is anticipative of at least one future value.

The above review brings out that among the empirical models, ANN technique appears to be the most promising. This should not be surprising since, the data is strongly non-Gaussian and hence linear methods valid for Gaussian time series will prove to be unsuccessful for SWM rainfall data. However, the ANN architecture to be selected remains subjective and hence not unique. Modeling of time series data can be seen as an exercise in curve fitting, with minimum number of parameters in the model not exceeding half the sample size. At present the ANN models reviewed above do not satisfy this condition. Another limitation of the available models is the inherent assumption of stationarity and hence use of the same model parameter values in the forecasting period. The EMD analysis of Iyengar and Raghukanth (2004 and 2006) with high modeling skill indicates that their ANN architecture can perhaps be improved to account for the long period linear component also in the training stage itself. The present study investigates such a possibility to arrive at ANN model that can be used for making one step ahead forecast of SWM rainfall at the All India and at the regional levels.

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

RAINFALL DATA

Six sets of SWM rainfall data are considered here for further work. The first is the All India rainfall value (AIRF) representing the whole country, which is spatial average based on the sub-regions. The other five data chosen are for the four sub-regions and one subdivision along the west coast with high variability. In Figure 1, the six regions are marked for clarity (Ref: Long range forecasting update, IMD, April 2013).

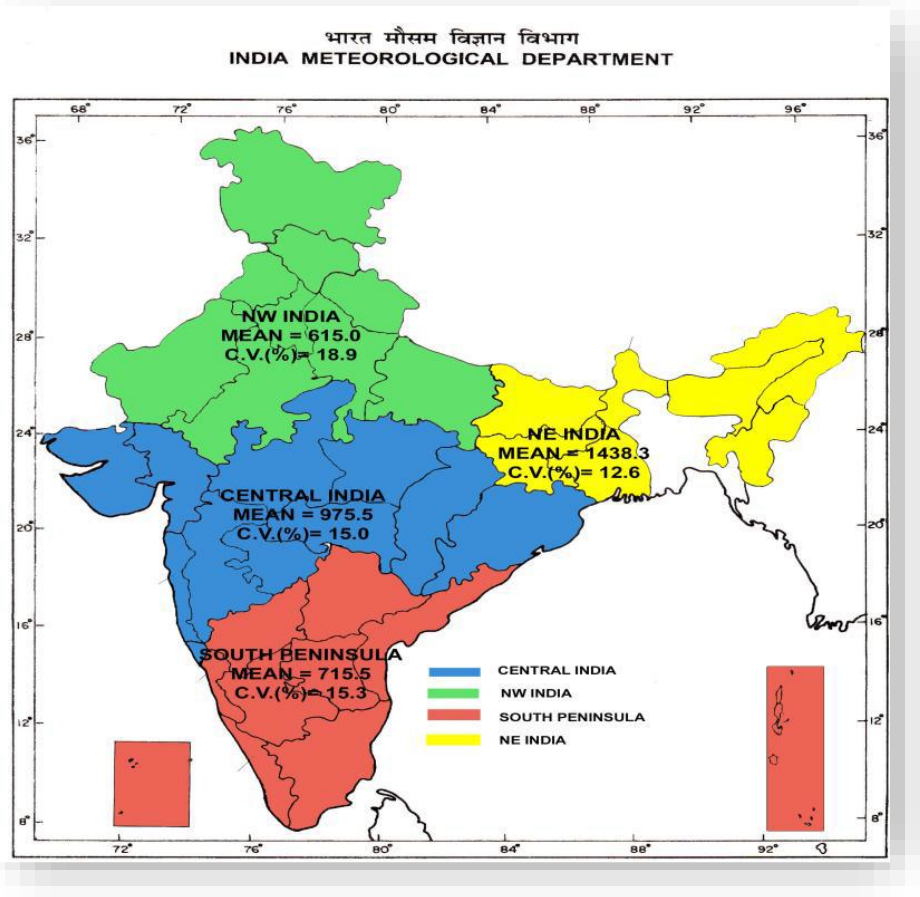


Fig 1 India map with homogeneous regions and subdivisions by IMD

The basic details of the data for the period (1901-2000) are shown in Table 1. The first five data series are from the official reports of the India Meteorology Department (IMD). The data for the Coastal Karnataka subdivision is taken from the data base of the Indian Institute of Tropical Meteorology (http://www.tropmet.res.in). It may be noted here that, since the data series R_i,(i=1,2…N) is nonstationary, the long term average (LTA) and the long term deviation (LTD) cannot be strictly termed the mean and standard deviation of the rainfall.***(See Table:-1)***

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

This can be seen from the variation of m_R and σ_R presented in Figure 2 with increasing sample length for NWIND. The two parameters slowly oscillate over years without converging to a constant value. This slow variation may not be important in modeling exercises, but will influence forecasts made ahead of time using a stationary model. However, LTA for a known length of sample can be used as a standard number with reference to which, inter annual variability (IAV) can be quantified. Hence, sometimes LTA is cited as the normal rainfall value. It also serves as a scaling factor to make the data series non-dimensional for further work. ***(See Figure:-2)***

Table 1: Basic Statistics of Rainfall Data (1901-2000)

Name	Area (Sq Km)	LTA ($m_R$cm)	LTD ($\sigma_R$cm)	Skewness	Kurtosis
AIRF	32, 87,782	88.49	8.97	-0.2168	2.6723
CEIND	11, 06,673	96.41	12.89	0.0264	3.1296
NWIND	10, 10,719	61.60	11.53	0.1395	3.1859
NEIND	5, 24,065	143.97	12.29	-0.2936	3.1422
PEIND	6, 46,325	72.26	9.91	0.2567	3.1186
COKNT	18,717	289.19	46.80	0.6644	4.9179

AIRF-All India Rainfall CEIND–Central India NEIND-Northeast India NWIND-Northwest India PEIND-Peninsular India, COKNT-Coastal Karnataka

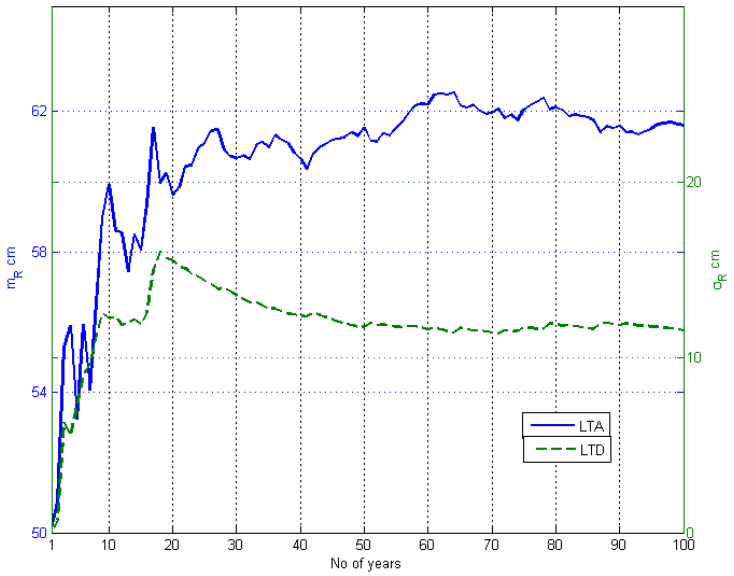


Fig 2: Variation of LTA and LTD with increasing data length

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

MODELING:

A mathematical model for Indian rainfall is of considerable importance in simulation and forecasting exercises. Rainfall as a random variable is non-Gaussian, but can be approximately represented by a log-normal distribution. Hence it is advantageous to transform the data series R_i,(i=1,2…N) for further analysis as

$$r_i = log\left(\frac{R_i}{m_R}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

The LTA (m_r) and long term variance LTV (μ^2) of the transformed time series r_i along with its estimated skewness and kurtosis are presented in **Table 2.**

Gaussianness of the random data r_i is checked with respect to the skewness and kurtosis using Jarque-Bera test (Jarque and Bera 1987). For the null hypothesis H_0 accepting the data as Gaussian, the tabulated statistic at 90% confidence is 3.14. The calculated statistics listed in the above table exceed (3.14) and hence even after the logarithmic transformation the data series cannot be treated as Gaussian.

Iyengar and Raghukanth (2004) decomposed the rainfall data as R_i=x_i+y_i where y_i was a linear equation with six unknowns accounting for about 33% of the data variance. The short period component x_i carrying the bulk of data variance of 67% was modeled by ANN having 36 parameters. The network used by them is shown in the **Figure 3.**

Table 2: Basic Statistics of data 〖r〗_i for the period 1901-2000

Name	LTA (m_r)	LTV (μ^2)	Skewness	Kurtosis
AIRF	-0.0052	0.0107	-0.4661	2.8333
CEIND	-0.0090	0.0186	-0.4312	3.6478
NWIND	-0.0190	0.0405	-0.7803	3.5614
NEIND	-0.0037	0.0074	-0.3338	3.4381
PEIND	-0.0095	0.0193	-0.3215	3.9287
COKNT	-0.0129	0.0262	-0.1393	4.4364

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

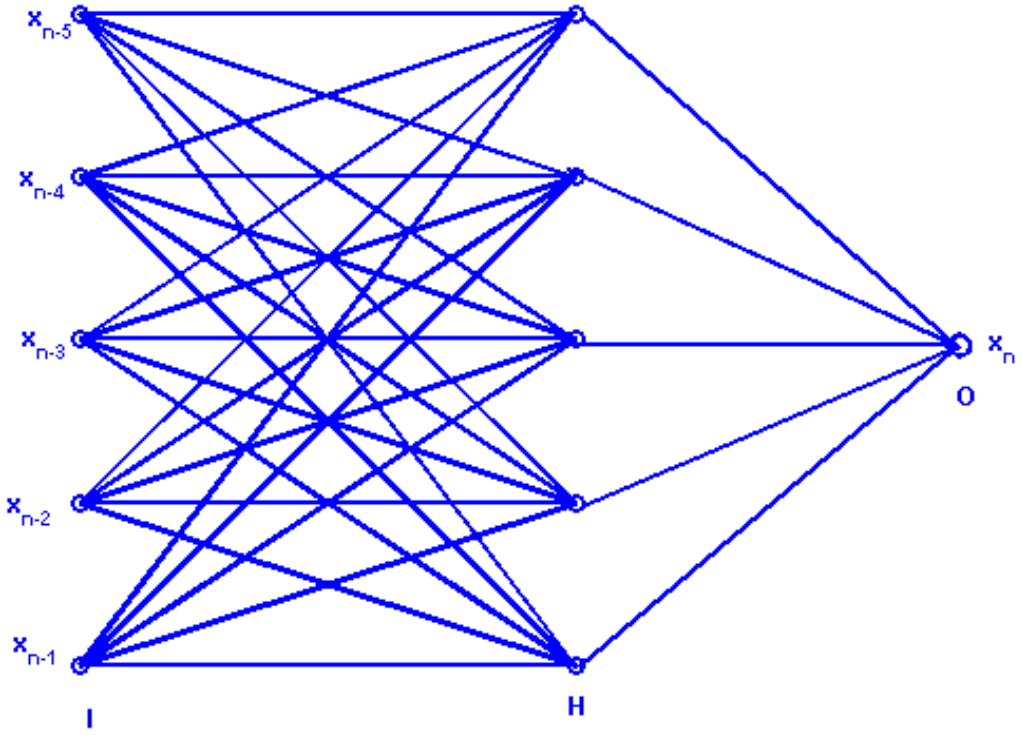


Fig 3: Five input node ANN used by Iyengar and Raghukanth (2004)

At this stage it would be beneficial to know how the above ANN model would perform on the original data without any type of decomposition. This would be a preliminary ANN model for r_i and hence for R_i with only 36 parameters. The statistical skill of such a model for the period (1906:2000) is shown in Table 3.

Table 3: Performance of ANN Model with 5 input nodes (Ref: Fig.2)

Region	Modeling Period (1906:2000)		
	σ_m (cm)	CC_m	PP_m
AIRF	4.52	0.85	0.74
CEIND	7.40	0.82	0.66
NWIND	6.40	0.83	0.69
NEIND	7.17	0.81	0.65
PEIND	5.18	0.85	0.72
COKNT	23.62	0.86	0.74

Here the modeling skill is measured in terms of three different parameters, namely the root mean square error (σ_m) between the data R_i and simulated $\hat{R}_i$, the correlation coefficient (CC_m) between R_i and $\hat{R}_i$, and the performance parameter (PP_m) defined as

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

$$PP_m = 1 - \frac{\sigma_m^2}{\sigma_R^2} \ldots\ldots\ldots\ldots(2)$$

CC_m may not be a very good indicator of the model skill, since even with phase difference between the data and the model the correlation coefficient may be high which would be spurious. However PP_m is a measure of the variance explained by the model over the complete range of the sample and goes to unity in the ideal model. Hence for a model to be accepted as useful in forecasting PP_m should be high both in the modeling stage and in an independent verification stage. From Table 3 it is observed that the preliminary ANN selected above explains about 70% of the data variance, which is almost same as the variance of the short period component x_i in the empirical mode decomposition approach of Iyengar and Raghukanth (2004). Hence one can expect further improvement by enhancing the freedom of the ANN architecture to have six more parameters.

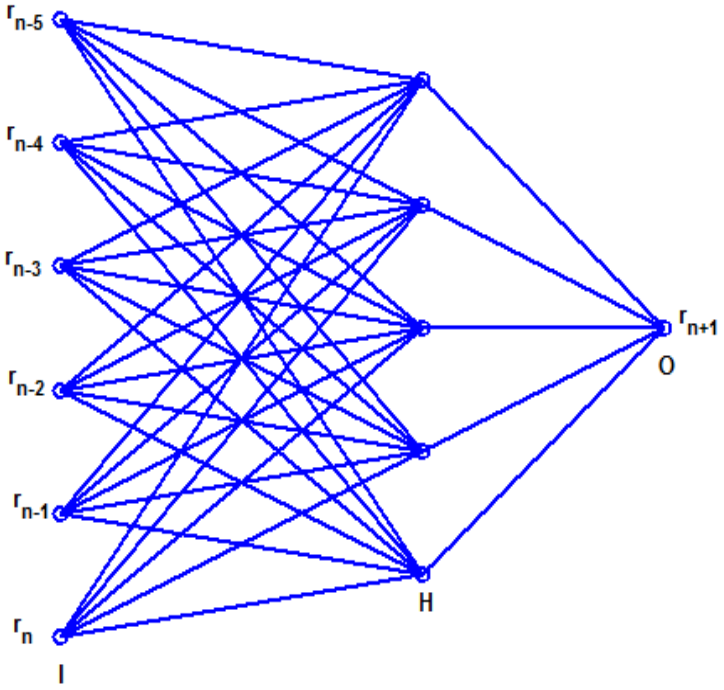


Fig 4: ANN with six input nodes, five neurons in the hidden layer and one output model

The simplest way, keeping in view the long period component y_i of the above authors, is to add one extra input node to the previous ANN architecture. This is symbolically shown in

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

Figure.4. This increases the number of parameters to be estimated from 36 to 41, but this is still less than the half the length of the sample.

These parameters are found using the back propagation algorithm in MATLAB toolbox. The training process of ANN employed here is explained in Appendix 1. In the training process 100 samples of 41 random initial weights as parameters are selected. These initial weights are iterated till the mean square error (MSE) between the actual data r_i and the simulated $\hat{r}_i$ converges. It is found that 250 to 400 iterations are required for the MSE to converge to a small value of the order 1×10^{-3}. One such convergence result requiring 300 iterations for the data series of AIRF is shown in Figure 5. Out of such 100 samples the one which produces the least MSE is taken as the best ANN model for r_i. ***(See Figure:-5)***

This leads to the representation for the original rainfall series as

$R_i = m_R e^{(\hat{r}_i + \varepsilon)}$.....(3)

where ε is the error between r_i and $\hat{r}_i$ characterized as a random variable. It is found that in all the six cases of SWM rainfall data ε_i is a white noise and hence can be treated as a random error, uncorrelated with $\hat{r}_i$.

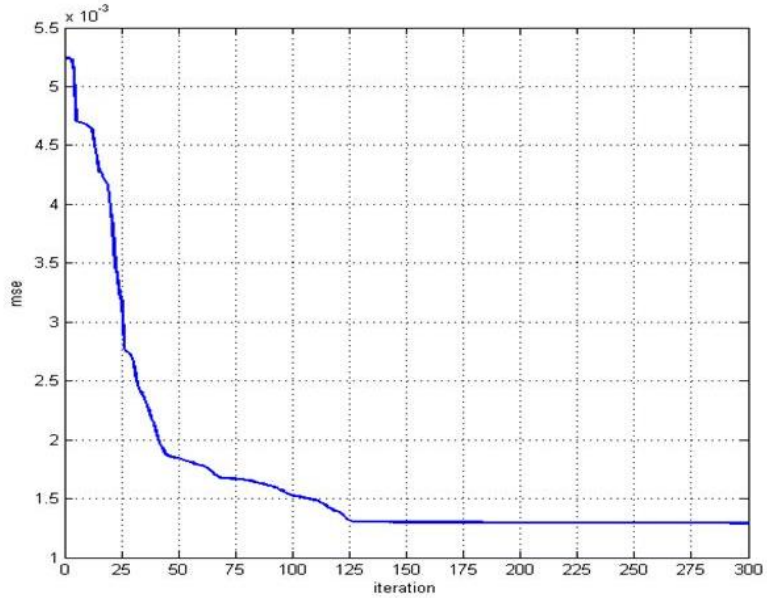


Fig 5: Convergence of the MSE with the iterations using back propagation algorithm for AIRF data r_i

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

Equation (3) is a strictly positive quantity, but in each year it is a random variable due to ε. The distribution of this random variable is not known and it would also be different for each data series. To overcome this difficulty, the model for R_i is proposed in terms of the most probable value or mode of e^{ε} as,

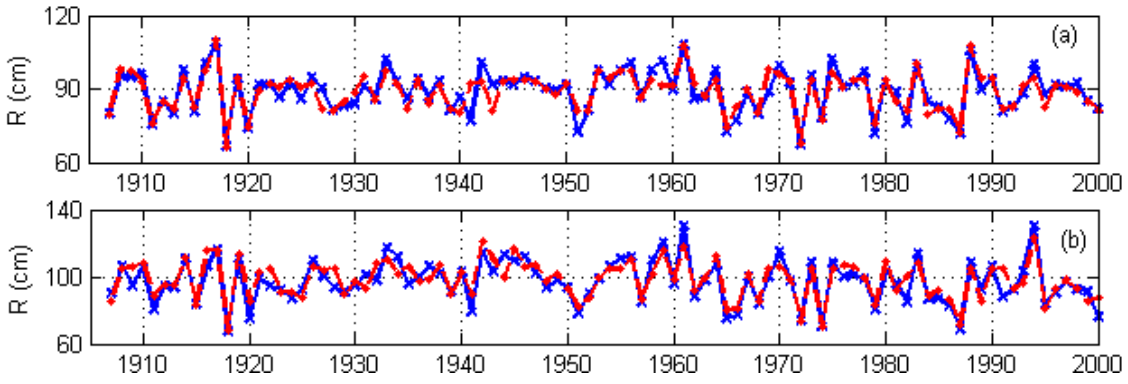
$\hat{R}_i = m_R e^{\hat{r}_i} mode(e^{\varepsilon})$…….. (4).

In the modeling period for a given sample length, $mode(e^{\varepsilon})$ remains a constant very nearly equal to unity. The estimated mean (m_ε) and standard deviation σ_ε of ε and $mode(e^{\varepsilon})$ for the present data series is ***shown in Table 4***.

The skill of the new ANN model is presented in Table 5. It is seen that between the previous five input node model and the present one, there is a distinct 10% increase in the performance parameter. This new model is capable of explaining 80% of the variation in all cases. Comparison between the data and sample simulation is shown in Figure 6 (a-f) for a visual appreciation of the statistical skill of the ANN model***.(See Figure:-6)***

Table 4: Statistics of the error variable ε

Region	m_ε	σ_ε	$mode(e^{\varepsilon})$
AIRF	0	0.0436	1.0001
CEIND	0	0.0608	1.0145
NWIND	0	0.0894	1.0006
NEIND	0	0.0497	0.9923
PEIND	0	0.0624	1.0098
COKNT	0	0.0748	0.9838



Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

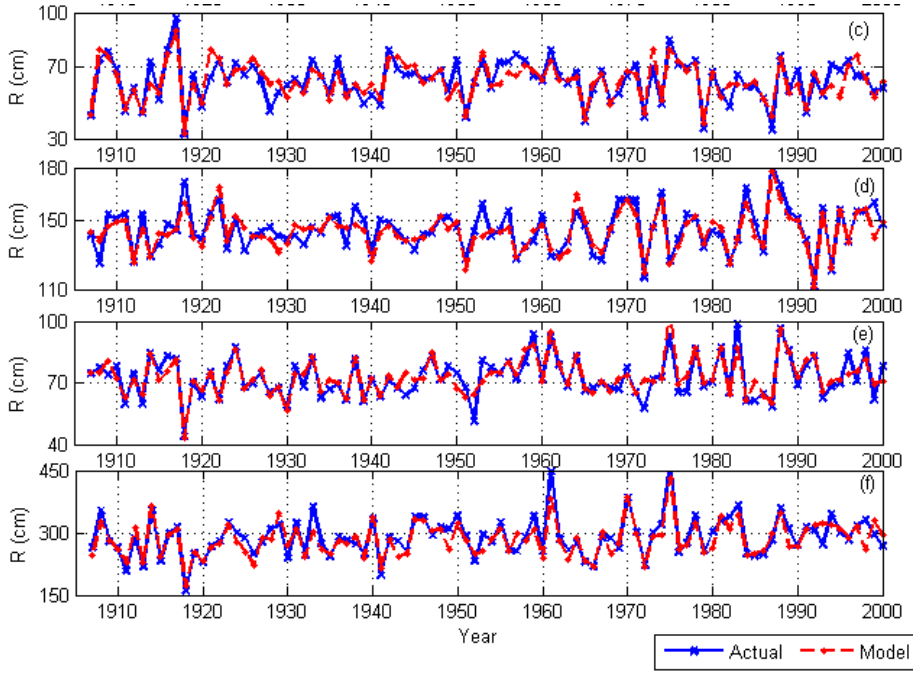


Fig 6: Comparison between the actual data and the model (a) AIRF, (b) CEIND, (c) NWIND, (d) NEIND,(e) PEIND and (f) COKNT

Table 5: The performance skill of the new ANN model

Region	Modeling Period (1907:2000)		
	σ_m (cm)	CC_m	PP_m
AIRF	3.85	0.90	0.81
CEIND	5.98	0.89	0.78
NWIND	5.54	0.88	0.77
NEIND	5.76	0.89	0.79
PEIND	4.51	0.89	0.79
COKNT	21.81	0.89	0.79

FORECASTING:

The ANN modeling demonstrated above can be thought of as a curve fit with high skill for the rainfall time series. Since the monsoon phenomenon recurs annually it can be claimed that the ANN model has been able to capture inter annual variability pattern robustly. Hence as with any well defined function, extrapolation to estimate the next year value would be meaningful. The only limitation of such an exercise is that the parameters found in the training period will not remain valid for a long time. But this can be overcome by updating

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

the model parameters each year. To verify the skill of the present ANN model in forecasting SWM rainfall, a set of observed values for the period (2001-2017) not included in the previous training exercise are chosen. The forecast for any year $(N_y + 1)$ is based only on the parameters computed for the interval $(1900,\ N_y)$ retaining the same ANN architecture as in Figure 4. This way the end point approximation inherent in the EMD method (Iyengar and Raghukanth 2004) can be overcome and also the non-stationarity of data (Figure 2) can be addressed effectively. The forecast has the same form as Equation (4), where the error ε is a random variable for which only the sample values for the period $(1907,\ N_y)$ will be known. Hence the forecast is also a random variable, but for operational purposes one can give the most probable value in terms of $mode(e^{\varepsilon})$ which also changes each year. Since the data is non-stationary, the above exercise has to be repeated for the next year with the length of the data series increasing by one step. In Table 6 one year ahead forecast for years (2001 to 2017) as given by the ANN model is shown along with the observed value, except for 2018 in five cases.

Table 6: Comparison of forecast with independent data in cm

Region	PEIND		NEIND		COKNT	
Year	Actual	Forecast	Actual	Forecast	Actual	Forecast
2001	64.39	64.17± 4.51	128.44	137.23± 5.87	267.07	261.21± 21.80
2002	48.30	52.50± 4.50	134.19	140.88± 5.67	207.20	198.68± 22.10
2003	63.32	68.12± 4.54	138.08	148.21± 5.68	306.29	283.50± 23.42
2004	60.82	58.49± 4.32	127.72	135.20± 5.88	228.53	227.69± 23.27
2005	79.85	84.15± 4.20	110.03	121.49± 5.71	270.33	295.31± 23.13
2006	67.76	77.49± 4.31	122.69	133.56± 6.88	264.68	276.17± 20.96
2007	89.29	82.02± 4.48	158.79	162.78± 6.39	336.39	357.13± 20.92
2008	68.54	70.49± 4.88	140.23	149.59± 6.11	237.38	272.63± 20.96
2009	67.54	76.06± 4.92	110.61	119.57± 6.85	281.89	262.58± 23.40
2010	83.93	85.58± 4.89	117.51	125.74± 6.34	290.01	289.14± 24.37
2011	71.30	76.57± 4.89	125.54	137.11± 6.97	360.96	333.74± 24.73
2012	64.38	71.05± 4.94	130.36	138.45± 6.87	282.02	263.92± 22.36
2013	82.28	86.93± 4.94	104.99	116.06± 6.85	338.35	308.09± 21.55
2014	66.54	65.19± 5.00	126.77	129.50± 6.27	291.98	276.46± 23.86
2015	60.57	66.25± 4.96	131.75	139.86± 6.48	214.50	250.74± 22.66
2016	66.77	83.51± 4.94	128.28	149.79± 6.55	NA	296.86± 23.33
2017	71.76	58.46±5.08	138.64	141.24±7.25	NA	NA
2018	NA	74.86±4.92	NA	121.56±7.06	NA	NA

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

For each year $(N_y + 1)$ the standard error of the point forecast will be equal to the RMS error between the actual data and the model fit for the period $(1907, N_y)$. This is also shown in the above table. The statistical forecast skill of the present approach in terms of σ_f, CC_f the correlation between R and $\hat{R}$ and the performance parameter PP_f is shown in Table 7.

Table 7: Performance of the model in the testing period

Region	Testing Period (2001:2017)		
	σ_f	CC_f	PP_f
AIRF	5.56	0.88	0.77
CEIND	6.92	0.83	0.68
NWIND	6.84	0.82	0.67
NEIND	8.85	0.85	0.70
PEIND	5.29	0.87	0.75
COKNT	21.55	0.87	0.75

It is observed that even with a small sample length of 15 years the forecast skill of the ANN technique is very promising. A measure of forecast widely used in India is the rainfall percentage above or below the long-term average. Such a comparison of the actual and forecast percentage deviations above and below normal value (LTA) in the independent verification stage is presented in Fig. 7(a-f).

DISCUSSION

The available time series data of SWM rainfall does not exhibit any significant year to year linear correlation. However, since the data is strongly non-Gaussian, one cannot conclude that there are no short- and long-term patterns in the oscillations of the data series. In fact, the EMD analysis of Iyengar and Raghukanth (2004) highlights the existence of a dominant period of 2-3 years accounting for nearly 60-70% of inter annual variability in the rainfall. This situation indicates that a complex non-linear relationship is required to model and further forecast Indian SWM rainfall. Since ANN technique has the flexibility of handling unstructured nonlinear relations, this approach has been selected here. Before finalizing the

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

six-input node model, simple architectures with 3, 4, 5 and 6 input nodes keeping a fixed hidden layer with 5 nodes was investigated as a model for the (1901-2000) AIRF data series. The efficiency of the ANN model in terms of PP_f was found to be 0.36, 0.55, 0.74 and 0.81 respectively. Further addition in the number of input nodes leads to saturation of the statistical skill and also increases the number of parameters beyond fifty, which is not desirable.

Even though the modeling skill of the five-input node ANN is as high as 0.74, in the forecasting stage with independent data it was found that the PP_f is too poor being around 0.2. After such preliminary exercises the present ANN model of Figure 4 has been arrived at as the optimal for use in forecasting of season rainfall. The correlation coefficient (CC_f) between the data and the forecast has to be at least 0.52 to be taken as significant in the independent test period of 15 years. As seen from Table 7 the forecasting skill is well beyond this threshold level. However, a point forecast is always a probable value and hence comes with a standard error, like the standard deviation from the mean. This, forecast error varies slightly in time depending on the length of data up to the antecedent year. These are also computable in the present approach as shown in Table 6.

The average forecast errors for the period 2001-2017 for the five data and 2001-2015 for COKNT considered here are (3.95, 6.65, 5.81, 6.34, 4.70, and 22.68) respectively. A simple comparison with the LTD value shown in Table 1 brings out the reduction achieved in the error between climatic mean (LTA) as a forecast versus the present ANN forecast. A pictorial performance of the present forecast model is shown in figure 7. It is observed that the forecast is able to capture the same sign of departure in 74 of the 90 cases. In the remaining 16 cases, the absolute error to obtain the same sign of departure is within the error bands. Since empirical forecast $\widehat{R}_N$ for any year N is statistical in nature, any forecast in Table 6 has to be treated as a random variable with its forecast error as standard deviation. The probability density function for this variable can be found in terms of the sample distribution of the model error.

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

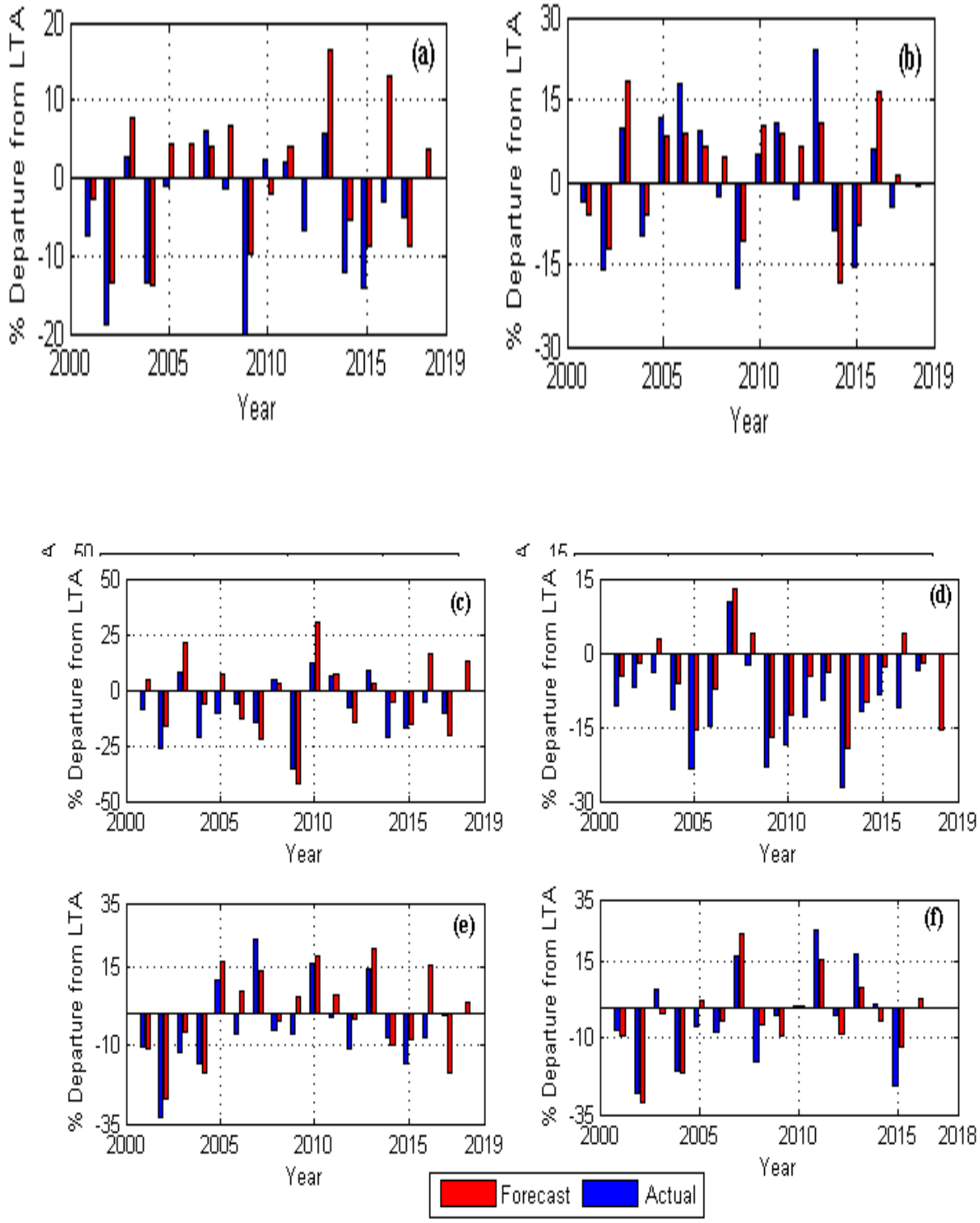


Fig 7: Percentage Departure from the normal (LTA) for the period of 17 years (2001-2017) (a) AIRF, (b) CEIND, (c) NWIND, (d) NEIND,(e) PEIND and (f) COKNT (2001-2015)

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

SUMMARY AND CONCLUSION

A new approach for modeling and forecasting the total rainfall of the monsoon season in India is presented in this paper. After a brief literature review it is demonstrated that a simple ANN architecture with six nodes at the input layer, a hidden layer with five neurons and an output, is capable of explaining about 80% of the observed inter-annual variability of observed SWM rainfall data. This has been demonstrated on six sets of data for the period (1901-2000). The model is capable of updating itself as the length of the sample increases. This property can be effectively used in forecasting, by extrapolating the results of the model by one step. The statistical skill of the forecasts have been verified on six sets of observed data not used in the ANN training exercise and shown to be highly significant. The data series considered here is the sum of the June, July, August and September month precipitation on annual basis. Hence the present work has focused on inter-annual variability only. However, the model developed is capable of improvement to include intra-annual variability in terms of pre-monsoon and post-monsoon seasonal data. Such generalizations for three seasons per year and monthly rainfall time series will be presented in a future publication.

APPENDIX 1

The training process of the neural networks is presented as follows:

Step 1: Input node receives information from the input data form r_i and will be transmitted to the hidden
layer nodes h_j

Step 2: The output information of the hidden nodes h_j is computed by using the activation function F as:

$$h_j = F\left(b_{oj} + \sum_i r_i b_{ij}\right)$$

where b_{oj} is the bias on j^{th} hidden node and b_{ij} is the bias on connecting the i^{th} input layer node to the j^{th} node of the hidden layer. Hence the same is sent to the output layer node. The activation function used here is a tan sigmoid function.

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

Step 3: The output information is computed by applying the activation function F as:

$$O_k = F\left(c_{oj} + \sum_j h_j c_{jk}\right)$$

where c_{oj} is the bias on jth output node, and c_{jk} is the bias on connecting the j^{th} hidden layer node to the k^{th} node of the output layer. The activation function used here is a linear function.

Step 4: The weight c_{jk} for error correction is adjusted which is occurred due to output node O_k. This information is back-propagated to the hidden layer nodes h_j , which provides information again on output layer node O_k.

Step 5: The weight b_{ij} is adjusted due to back-propagation of error to the hidden layer nodes h_j for error correction and it is repeated until the error is minimized.

Step 6: When the error is minimized between calculated output and the desired output to a specified level, stop the whole steps of the training process.

REFERENCES

I. Ashok Kumar, D. S. Pai, J. V. Singh, Ranjeet Singh, D. R. Sikka (2012): Statistical Models for Long-range Forecasting of Southwest Monsoon Rainfall over India Using Step Wise Regression and Neural Network, Atmospheric and Climate Sciences, 2, 322-336

II. Chattopadhyay, S (2007): Feed forward Artificial Neural Network model to predict the average summer monsoon rainfall in India, Acta Geophysica, 55(3), 369-382.

III. Delsole T and Shukla J (2002): Linear prediction of the Indian monsoon rainfall, Centre for Ocean–Land–Atmosphere Studies (COLA); Tech. Rep. 46 52.

IV. Eisner, J. B., Tsonis AA (1992): Nonlinear prediction, chaos, and noise, Bull Amer Meteor Soc 73, 49–60

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

V. Goswami, B. N., (1998): Interannual variations of Indian summer monsoon in a GCM: External conditions versus internal feedbacks. Journal of Climate, 11, 501-522.

VI. Gadgil, S. and S. Sajani, (1998): Monsoon precipitation in the AMIP runs, Climate Dynamics, 14, 659-689.

VII. Gadgil S, Rajeevan M, Nanjundiah R (2005): Monsoon prediction – Why yet another failure?, Current Science, 88, 1389-1400.

VIII. V. Gowariker, V. Thapliyal, R. P. Sarker, G. S. Mandal and D. R. Sikka, (1989): Parametric and Power Regression Models: New Approach to Long Range Forecasting of Monsoon Rainfall in India, Mausam, 40, 115- 122.

IX. Gowariker, V., V. Thapliyal, S. M. Kulshrestha, G. S. Mandal, N. Sen Roy, and D. R. Sikka, (1991): A power regression model for long range forecast of southwest monsoon rainfall over India. Mausam, 42, 125-130.

X. Guhathakurta P., (2008): Long lead monsoon rainfall prediction for meteorological sub-divisions of India using deterministic artificial neural network model, Meteorol Atmos Phys, 101, 93-108.

XI. Hsieh, W. W., Tang, B (1998): Applying neural network models to prediction and data analysis in meteorology and oceanography, Bulletin of American Meteorological Society, 79(9), 1855–1870

XII. Iyengar R.N and Raghukanth, S.T.G., (2004): Intrinsic mode functions and a strategy for forecasting Indian monsoon rainfall, Meteorology and Atmospheric Physics, 90, 17-36.

XIII. Iyengar R.N and Raghukanth, S.T.G., (2006): Forecasting of seasonal monsoon Rainfall at subdivision level, Current Science, 91, 350-356.

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

XIV. Jarque, C.M., and A.K. Bera, (1987): A test for normality of observations and regression residuals, International Statistical Review, 55(2), 163-172.

XV. Krishnamurthy V. and J. Shukla, (2000): Intraseasonal and interannual variability of rainfall over India, Journal of Climate, 13, 4366-4377.

XVI. Kang, I.-S., K. Jin, B. Wang, K.-M. Lau, J. Shukla, V. Krishnamurthy, and coauthors, (2002b): Intercomparison of the climatological variations of Asian summer monsoon precipitation simulated by 10 GCMs, Climate Dynamics, 19, 383-395.

XVII. Krishna Kumar, K Hoerling, M; Rajagopalan, B, (2005): Advancing dynamical prediction of Indian monsoon rainfall, Geophysical Research Letters, 32, 1-4.

XVIII. K. Rajendran, A. Kitoh, R. Mizuta, S. Sajani, and T. Nakazawa, (2008): High Resolution Simulation of Mean Convection and its Intraseasonal Variability over the Tropics in MRI/JMA 20-km Mesh AGCM, Journal of Climate, 21(15), 3722-3739.

XIX. Latif, M., A. Sterl, M. Assenbaum, M. M. Junge, and E. Maier-Reimer, (1994): Climate variability in a coupled GCM. Part II: The Indian Ocean and monsoon, Journal of Climate, 7, 1449-1462.

XX. Pritpal Singh and Bhogeswar Borah, (2013): Indian summer monsoon rainfall prediction using artificial neural network, Stochastic Environmental Research and Risk Assessment, 1436-3240.

XXI. Rajeevan M, Guhathakurta P, Thapliyal V (2000): New models for long range forecasts of summer monsoon rainfall over Northwest and Peninsular India, Meteorol Atmos Phy, 73 211–225

XXII. Rajeevan M, Pai. D S, Anil Kumar R, (2005): New statistical models for long range forecasting of southwest monsoon rainfall over India, NCC Research Report, 1, Indian Meteorological Department.

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved

XXIII. Sahai AK, Soman MK, Satyan V (2000): All India summer monsoon rainfall prediction using an artificial neural network, Climate Dynamics 16, 291–302.

XXIV. Sahai AK, Grimm AM, Satyan V, Pant GB (2003): Long lead prediction of Indian summer monsoon rainfall from global SST evolution, Climate Dynamics, 20, 855–863

XXV. Sajani S., T. Nakazawa, A. Kitoh and K. Rajendran, (2007): Ensemble simulation of Indian summer monsoon rainfall by an atmospheric general circulation model, Journal of Meteorological Society, Japan , 85 (3), 213-231.

XXVI. Thapliyal, V., (1990): Long range prediction of summer monsoon rainfall over India: Evolution and development of new models, Mausam, 41, 339-346.

XXVII. Wang, B., Q. Ding, X. Fu, I.-S. Kang, K. Jin, J. Shukla, and F. Doblas-Reyes, (2005): Fundamental challenges in simulation and prediction of summer monsoon rainfall, Geophysical Research Letters, 32, L15711.

XXVIII. Walker GT (1923): Correlations in seasonal variations of weather, VIII, Mem, Indian Meteorological Department, 24, 75–131

XXIX. Walker, G. T., (1924): Correlation in seasonal variations of weather, IX: A further study of world weather, Mem. India Meteorological Department, 24, 275-332.

Page 1-21 © MANTECH PUBLICATIONS 2019. All Rights Reserved