MAUSAM, **69**, 4 (October 2018), 615-634

L E T T E R S

551.577 : 551.553.21 (540)

EMPIRICAL MODELLING AND FORECASTING OF MONSOONAL RAINFALL OF EAST AND NORTH EAST SUBDIVISIONS OF INDIA

1. During the South West Monsoon (SWM) or the summer monsoon rainfall, good amount of rainfall is received by the meteorological subdivisions of Gangetic West Bengal (GWB: No.6), Sub Himalayan West Bengal (SHWB: No.5), Assam and Meghalaya (AM: No.3) and also Nagaland, Manipur, Mizoram and Tripura (NMMT: No.4). The variability of rainfall amount has significant impact on the agriculture, economy and society. Over the years, efforts are made to quantify the variability and forecast of monsoonal phenomenon at various temporal and spatial scales. A number of literatures are available on analysis of variability of SWM rainfall data; namely, the architectural works of Mooley and Parthasarathy (1984); Thapliyal (1990); Iyenger and Basak (1994); Iyenger and Raghukant (2003) and Basak (2014, 2017). A detailed review of literature is available in the works of Hastenrath and Greischar (1993) and Gadgil *et al.* (2002).

The basic characteristics of rainfall data is non-Gaussianness on several temporal and spatial scales, namely, SWM or a sub division even though the data can be treated as sum of large number of random variables. The linear time scale models based on past rainfall reflect the behaviour near the mean value reasonably well, but the extreme events of flood and drought like situation are not identifiable (Iyenger and Raghukant, 2003).

Kedem and Chiu (1987), however, recognizing the property of the rainfall, proposed that rain rate follows lognormal random distribution for small time scale following law of proportion, namely,

$$R_{j+1} - R_j = \varepsilon_j R_j \qquad (1)$$

where, ε_j's are independent identically distributed random variables and are independent of R_j's.

The authors claim the fitness of equation (1) to the hourly rainfall. Since, the rainfall series of different time scales such as monthly, SWM exhibit strong non-Gaussianness, the application of equation (1) to model SWM rainfall for subdivisions of GWB, SHWB, AM and NMMT is justified. The locations of the subdivisions are presented in the All India map of Fig. 1. It is explained that the equation (1) is systematically utilized for

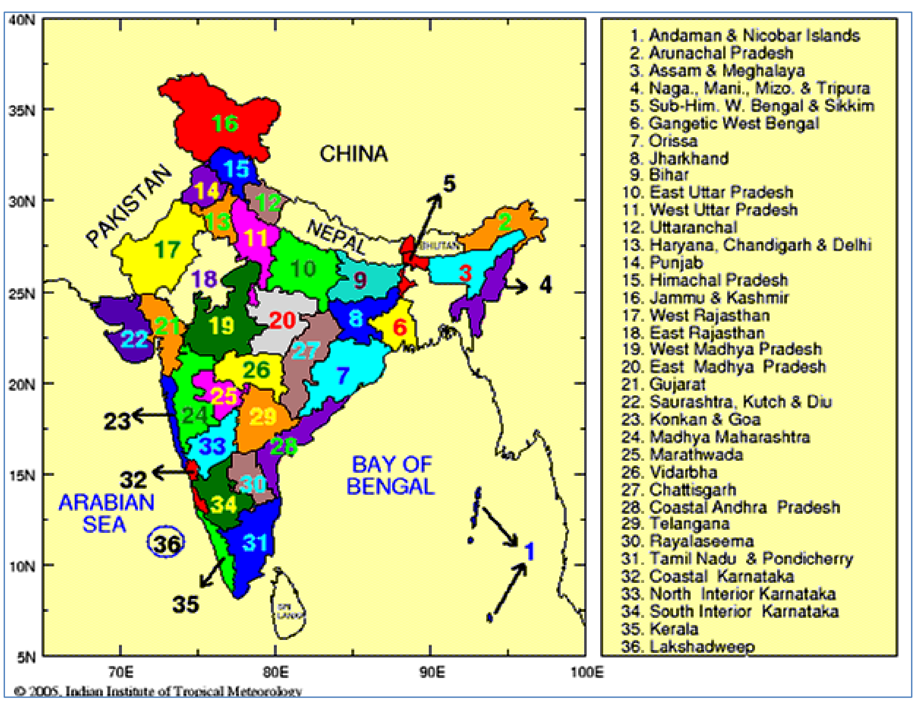

Fig. 1. All India Meteorological subdivisions

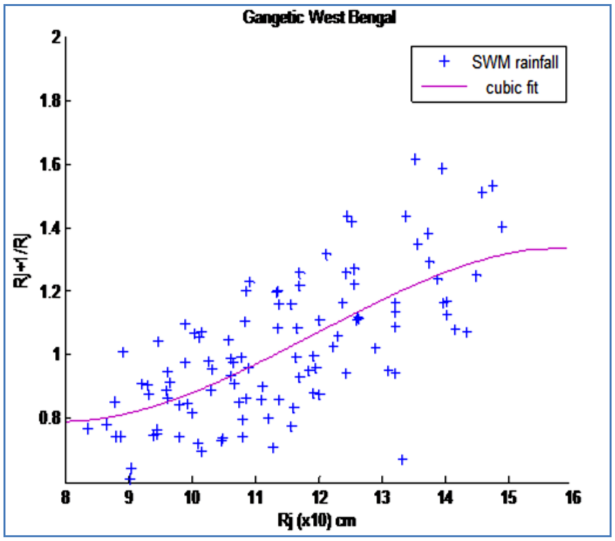

Fig. 2. (R_{j+1}/R_j) *vs.* R_{j+1} plot of Gangetic West Bengal (GWB) SWM rainfall

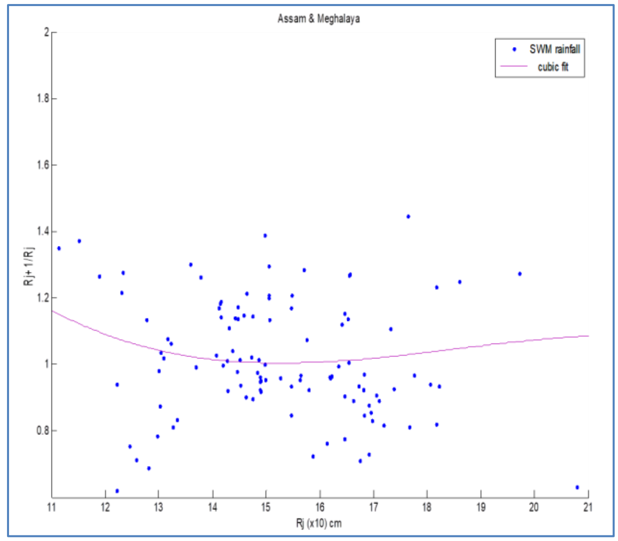

Fig. 4. (R_{j+1}/R_j) *vs.* R_{j+1} plot of Assam and Meghalaya (AM) SWM rainfall

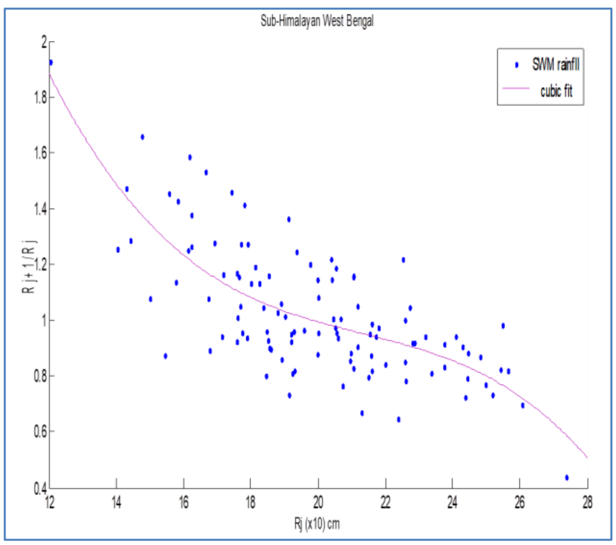

Fig. 3. (R_{j+1}/R_j) *vs.* R_{j+1} plot of Sub-Himalayan West Bengal (SHWB) SWM rainfall

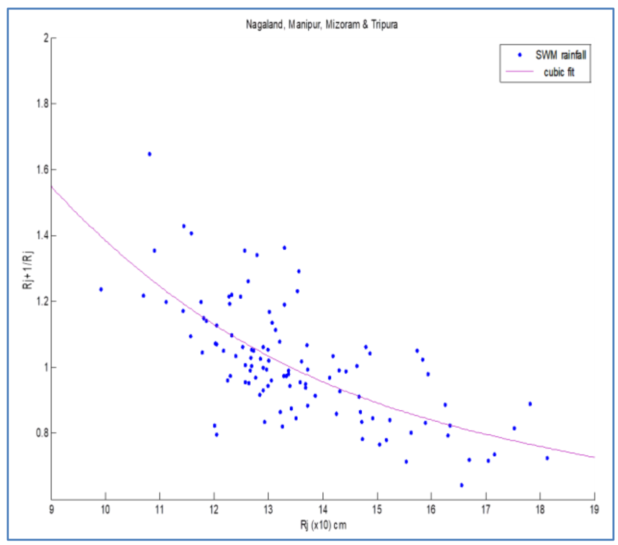

Fig. 5. (R_{j+1}/R_j) *vs.* R_{j+1} plot of Nagaland, Manipur, Mizoram and Tripura (NMMT) SWM rainfall

year-to-year and long term relationship known to persist in monsoonal rainfall. The proposed model is understood to behave as a 'useful statistical tool' with detailed approach is presented subsequently.

2. The SWM rainfall data of the subdivisions such as GWB, SHWB, AM and NMMT are extracted from IMD data base available at IITM-IMD site (http://www.tropmet.res.in/). The details of data assembly are also available in the concerned website.

3. *Modeling*: The previews of statistical properties of the sub divisional SWM rainfall are presented in Table 1. The mean μ*R* is the long term time average indicating climatic normal and the standard deviation δR reflects year-to-year variability of the SWM rainfall.

The autocorrelations and other salient features helpful for modeling are extracted for all the series. As the data is non-Gaussian, except for long-term average as useful reference quantities, other characteristics for Gaussian series are inapplicable. Thus, data series is treated as generalized lognormal data in view of above discussions.

The equation (1) can be adjusted to

TABLE 1

Statistical details of SWM rainfall data (1871-2014)

Region	Area (square km)	μR	σR	Skewness	Kurtosis
GWB	44300	106.93	10.2734	-0.4425	2.9845
SHWB	21625	199.1141	20.0709	0.1255	3.1025
AM	109096	152.7796	1.4568	-0.2355	2.9615
NMMT	70495	199.1141	20.0709934	0.0588	3.0978

GWB = Gangetic West Bengal; SHWB = Sub Himalayan West Bengal; AM = Assam and Meghalaya; NMMT = Nagaland, Manipur, Mizoram and Tripura

TABLE 2

Parameters of f (*R*) and R_c

Region	a	b	c	d	σ_ε	R_c
GWB	-0.0022	0.8499	-1.1597	6.4694	0.1913	114.2501
SHWB	-0.0007	0.0504	-1.1283	9.5568	0.1925	199.1141
AM	0.0003	-0.0156	-0.3554	3.7484	0.1921	145.6818
NMMT	0.0017	-0.0682	-0.8492	-2.1548	0.2861	200.7099

GWB = Gangetic West Bengal; SHWB = Sub Himalayan West Bengal; AM = Assam and Meghalaya; NMMT = Nagaland, Manipur, Mizoram and Tripura

$$R_{j+1}/R_j = f(R_j) + \varepsilon_j \tag{2}$$

where $f(R_j)$ = an appropriate function of R_j.

The equation (2) is reduced to equation (1) when $f(R_j) = 1$. It indicates that given j^{th} year rainfall R_j, the annual change in R_j is an unknown function of R_j itself.

In Figs. (2-5), the relation between the (R_{j+1}/R_j) and R_j, for sub divisional SWM rainfalls are presented for the period 1871-1990; the figures indicate that there is a clear cut cubical trend of the form

$$f(R) = aR^3 + bR^2 + cR + d \tag{3}$$

The cubic equation is selected on the basis of least variance error. The equation (2) is now a non-linear difference equation for the data in the form

$$R_{j+1} = R_j . f(R_j) + \varepsilon_j . R_j = R_j (aR_j^3 + bR_j^2 + cR_j + d) + \varepsilon_j . R_j \tag{4}$$

If ε_j is taken to be independent of R_j, it follows that the conditional expectation of R_{j+1} given R_j would be

$$< R_{j+1} > = R_j (aR_j^3 + bR_j^2 + cR_j + d) \tag{5}$$

TABLE 3

Percentage reduction in climatic variance

Region	Equation (5)	Equation (7)
GWB	1.6562	1.7711
SHWB	3.1163	3.2049
AM	1.7791	1.9848
NMMT	1.6085	1.6598

GWB = Gangetic West Bengal; SHWB = Sub Himalayan West Bengal; AM = Assam & Meghalaya; NMMT = Nagaland, Manipur, Mizoram & Tripura

which defines the natural point predictor for the rainfall in year (j+1) if only the previous year value is known.

The error in the predictor is perhaps the conditional standard deviation of R_{j+1}, namely,

$$\sigma R_{j+1} = \sigma_\varepsilon . R_j \tag{6}$$

σ_ε being the standard deviation of errors in model (5).

The parameters of the proposed cubic polynomial $f(R)$ & the fixed point (R_c) of equation (5) are presented for the SWM series of the sub divisions in Table 2.

TABLE 4

Root mean square error (RMSE) between observed SWM rainfall and model rainfall for the model building period (1891-1990)

Area / RMSE	GWB		SHWB		AM		NMMT	
	Eqn. (5)	Eqn. (7)	Eqn. (5)	Eqn. (7)	Eqn. (5)	Eqn. (7)	Eqn. (5)	Eqn. (7)
RMSE	1.6633	1.5921	2.9790	2.9161	2.7946	2.7498	2.6737	2.5014

GWB = Gangetic West Bengal; SHWB = Sub Himalayan West Bengal; AM = Assam and Meghalaya; NMMT = Nagaland, Manipur, Mizoram and Tripura

TABLE 5

Correlation Coefficient between (R_{j+1}/ R_j) and SWM rainfall at different lags $R_{j-1}, R_{j-2},...,R_{j-20}$

Lag-RF	GWB	SHWB	AM	NMMT
R_j	**-0.7068***	**0.7096***	**-0.7315***	**-0.7314***
R_{j-1}	-0.0757	-0.0336	-0.0458	-0.0458
R_{j-2}	**0.1372***	0.0823	**0.1854***	0.0854
R_{j-3}	0.0032	0.0978	-0.0883	-0.0883
R_{j-4}	-0.0097	**-0.1556***	**0.1580***	-0.0479
R_{j-5}	-0.0332	-0.0379	0.0272	0.0271
R_{j-6}	0.0144	0.0559	-0.0161	- 0.0161
R_{j-7}	-0.0627	-0.1002	0.1553*	0.1452
R_{j-8}	**0.1356***	-0.0984	-0.0646	-0.0646
R_{j-9}	0.0115	**0.0244***	-0.0388	-0.0388
R_{j-10}	-0.0469	0.0922	0.0501	0.2258*
R_{j-11}	-0.1341	-0.0326	-0.0381	-0.0381
R_{j-12}	**0.2433***	-0.0680	**0.1731***	0.1731
R_{j-13}	-0.0803	0.0211	-0.1141	-0.0704
R_{j-14}	0.0120	0.0668	-0.1088	-0.1197
R_{j-15}	**0.2019***	0.0171	0.2025	-0.0576
R_{j-16}	0.0761	-0.0564	-0.0661	0.0005
R_{j-17}	-0.0942	-0.0684	-0.0667	0.0668
R_{j-18}	-0.0173	0.0652	0.0179	0.1034
R_{j-19}	0.0893	0.0361	-0.0309	-0.0571
R_{j-20}	**-0.2329***	**-0.1573***	**-0.1573***	-0.0431

*Significant GWB = Gangetic West Bengal; SHWB = Sub Himalayan West Bengal; AM = Assam and Meghalaya; NMMT = Nagaland, Manipur, Mizoram and Tripura

The fixed point R_c of Table 2 is compared with the mean value μ_R of Table 1. The variability band in any year (j+1) depends on the previous rainfall values R_j given by equation (6) which may be large enough although σ_ε is too small. It implies that equation (5) in its current form would not yield good one-year ahead forecast due to plausible attraction towards long-term mean value (fixed point R_c).

4. *Improvement in forecast skill*: An improvement in forecast has to be acquired by reduction of variance with respect to σ_R^2. The error between the observed rainfall and the value predicted by equation (5) is compared for the years 1872-1990. The percentage reduction in variance with respect to climatic normal μ_R is computed (Table 3) indicating that a small reduction in variance has been taken place and to be improved by incorporating the connection of longer inter-annual

relations by finding correlation between (R_{j+1}/R_j) and lagged yearly values R_j. The investigators (Moolay and Parthasarathy, 1984; Iyenger and Basak, 1994) reported that SWM rainfall possess significant correlations at several lags at different regions. On the ground that correlations are small numbers and series concerned are non-negative and non-Gaussian, the statistical significance level for the SWM rainfall series are different and less than that of Gaussian series (Johnson and Kotz, 1972). The small correlations which are rejected for Gaussian series at 5% level (±0.20 for a data series of length around 120 at 5% level of significance) may still be valid for non-Gaussian series relevant to our study; consequently, the level is accepted as ±0.13 at the same level (Iyenger and Raghukant, 2003).

It evolves an improvement of the model (5) by incorporating up to twentieth lagged terms, namely, $R_{j-1}, R_{j-2,\ldots,} R_{j-20}$ (1/6 of data length).

In Table 5, significant lagged correlations between (R_{j+1}/R_j) and rainfall at different lags are presented. For GWB rainfall series, the significant lag correlation persists for 2nd, 8th, 12th, 15th and 20th lags. Consequently, the analysis indicates that for GWB, an improvement of the model (3) and (5) leads to

$$R_{j+1} = R_j(aR_j^3 + bR_j^2 + cR_j + d + g_2R_{j-2} + g_8R_{j-8} + g_{12}R_{j-12} + g_{15}R_{j-15} + g_{20}R_{j-20}) + \delta_j.R_j \quad (7a)$$

With similar consideration, significant correlations for 4th, 9th and 20th lags for SHWB (Table 5) suggests a model of the form

$$R_{j+1} = R_j(aR_j^3 + bR_j^2 + cR_j + d + g_4R_{j-4} + g_9R_{j-9} + g_{20}R_{j-20}) + \delta_j.R_j \quad (7b)$$

In the same way, the model for the subdivision of AM and NMMT (Table 5) are respectively,

$$R_{j+1} = R_j(aR_j^3 + bR_j^2 + cR_j + d + g_2R_{j-2} + g_4R_{j-4} + g_7R_{j-7} + g_{12}R_{j-12} + g_{20}R_{j-20}) + \delta_j.R_j \quad (7c)$$

and

$$R_{j+1} = R_j(aR_j^3 + bR_j^2 + cR_j + d + g_{10}R_{j-10}) + \delta_j.R_j \quad (7d)$$

The coefficients of the equations [7(a-d)] are computed with minimization of error δ_j and are presented in Table 6.

TABLE 6

Coefficients of equation (7)

Coefficients	GWB	SHWB	AM	NMMT
a	0.0018	0.0005	4.2177e-4	1.3487e-3
b	-0.0582	-0.0256	-0.0204	-0.0484
c	0.5152	0.3606	0.2538	0.4739
d	0.0166	1.3039e-3	8.5856e-5	2.4320e-4
g_1	-	-	-	-
g_2	0.1255	-	5.8754e-3	-
g_3	-	-	-	-
g_4	-	2.5144e-3	-1.0235e-2	-
g_5	-	-	-	-
g_6	-	-	-	-
g_7	-	-	-2.4633e-3	-
g_8	0. 0463	-	-	-
g_9	-	0.0357	-	-
g_{10}	-	-	-	6.1722e-3
g_{11}	-	-	-	-
g_{12}	0.0053	-	2.7847e-3	-
g_{13}	-	-	-	-
g_{14}	-	-	-	-
g_{15}	-0.0012	-		-
g_{16}	-	-	-	-
g_{17}	-	-	-	-
g_{18}	-	-	-	-
g_{19}	-	-	-	-
g_{20}	2.6210e-2	2.6201e-3	0.0125	-
σ_δ	0.1828	0.1665	0.1852	0.2568

GWB = Gangetic West Bengal; SHWB = Sub Himalayan West Bengal; AM = Assam and Meghalaya; NMMT = Nagaland, Manipur, Mizoram and Tripura

It is evident may be interestingly observed that the standard deviations σ_δ's are less than the previous error value σ_ε's for all the corresponding cases (Tables 2 and 6) that indicates increment of the skill of the forecast from model (5) to model (7).

Equations 7(a-d) provide a non-linear map of higher dimension. Initial conditions of 20 past values are required for solving the equations 7(a-d) to achieve steady state of the model. For that purpose, we start from the data of year 1891 (with 1871 starting year) as 20 years of initial conditions for evaluation of coefficients of the equations 7(a-d).

TABLE 7

Test forecasting with independent data for the test period (1991-2014)

Year	GWB		SHWB		AM		NMMT	
	Actual	Prediction	Actual	Prediction	Actual	Prediction	Actual	Prediction
1991	113.57	119.21 ± 21.94	237.671	195.00 ± 31.5	146.25	151.55 ± 27.66	122.67	131.52 ± 34.73
1992	93.42	123.75 ± 22.64	152.97	199.8 ± 39.57	128.05	149.37 ± 27.02	105.48	135.14 ± 34.50
1993	160.10	109.94 ± 20.09	188.85	201.38 ± 25.46	215.24	159.85 ± 23.71	151.87	134.36 ± 31.00
1994	131.10	107.24 ± 19.60	135.58	203.96 ± 31.39	132.83	145.17 ± 39.86	104.13	125.07 ± 34.16
1995	154.62	121.57 ± 22.22	269.91	202.75 ± 22.57	163.38	139.17 ± 24.60	144.94	133.03 ± 31.80
1996	119.72	112.06 ± 20.48	200.01	194.02 ± 44.94	119.26	148.72 ± 30.25	120.62	123.88 ± 34.44
1997	145.27	124.26 ± 22.72	195.55	253.29 ± 33.30	146.62	159.49 ± 22.09	146.99	134.13 ± 34.35
1998	97.03	131.26 ± 23.99	269.89	197.78 ± 32.56	163.56	146.08 ± 27.15	110.61	133.80 ± 34.35
1999	142.39	128.05 ± 23.41	224.23	200.50 ± 44.93	152.90	153.59 ± 30.29	131.51	133.79 ± 33.12
2000	119.91	111.69 ± 20.42	212.66	231.51 ± 37.33	160.94	139.64 ± 28.31	108.65	128.97 ± 34.78
2001	109.50	117.18 ± 21.42	177.33	197.66 ± 35.41	114.98	159.98 ± 29.80	131.96	135.44 ± 28.60
2002	138.55	122.77 ± 22.44	198.99	194.30 ± 29.52	139.33	164.34 ± 21.29	147.19	127.60 ± 32.76
2003	112.33	120.83 ± 22.09	180.97	203.14 ± 33.13	131.10	149.35 ± 25.80	142.22	135.44 ± 34.78
2004	114.46	119.40 ± 21.83	185.93	199.91 ± 30.13	135.22	149.61 ± 24.28	171.94	133.76 ± 36.52
2005	103.32	118.43 ± 21.65	172.69	201.10 ± 30.96	129.42	149.27 ± 25.04	118.96	134.56 ± 44.15
2006	138.85	108.24 ± 19.79	169.34	202.38 ± 28.75	107.31	147.35 ± 23.97	113.52	135.19 ± 30.54
2007	169.52	126.51 ± 23.13	192.38	202.88 ± 28.19	147.33	164.96 ± 19.87	146.56	133.24 ± 34.21
2008	115.03	121.20 ± 22.16	225.55	203.71 ± 32.03	140.94	135.09 ± 27.28	102.94	130.75 ± 37.63
2009	96.94	161.85 ± 29.58	156.07	197.47 ± 37.55	106.41	155.55 ± 26.10	107.40	133.86 ± 26.43
2010	78.92	107.16 ± 19.67	226.95	193.47 ± 25.98	119.61	143.78 ± 19.71	117.87	122.76 ± 27.58
2011	137.80	131.26 ± 23.99	186.49	201.69 ± 37.78	106.64	137.51 ± 22.15	97.65	126.63 ± 30.26
2012	93.19	96.56 ± 17.65	209.47	194.01 ± 31.91	148.86	149.96 ± 19.75	100.18	132.82 ± 25.07
2013	115.20	126.88 ± 16.24	171.10	198.93 ± 34.87	99.27	136.41 ± 18.38	88.71	117.15 ± 25.73
2014	118.20	107.22 ± 19.60	201.26	196.46 ± 32.71	151.30	136.41 ± 18.38	111.23	119.97 ± 25.19

GWB = Gangetic West Bengal; SHWB = Sub Himalayan West Bengal; AM = Assam and Meghalaya; NMMT = Nagaland, Manipur, Mizoram and Tripura

Further, fitness of the models is noticed by calculation of Root Mean Square Error (RMSE). For all the subdivisions under consideration, the RMSE of the equations [7(a-d)] are lower than that of equation (5) with the actual data for the model building period (1891-1990) as presented in Table 4 indicating that the equations 7(a-d) are better than equation (5) attesting the inclusion of the lagged correlation term.

5. *Test forecasting* : The models 7(a-d) proposed for SWM rainfall of the subdivisions exhibit a clear deterministic part and a random part. The deterministic part $<R_{j+1}>$ is a point predictor for the rainfall in the year (j+1) whereas $\pm\ \sigma_{\delta}.R_j$ is the statistical random part. As a forecast we understand that the rainfall in year (j+1) will be 67% probability in the interval $[<R_{j+1}> - \sigma_{\delta}.R_j,\ <R_{j+1}> + \sigma_{\delta}.R_j]$. Obviously, it is expected that $\sigma_{\delta}.R_j$ is less than σR for a year. The rainfall data is available over the period 1871-2014. First 120 years data was utilized for the modeling purpose and the last 24 years data (1991-2014) was utilized for verification of the forecasts. In the prediction exercise, for the first prediction year, 1991, the previous model has been directly used but for subsequent years, the actual rainfall of the previous years is utilized. In Table 7, the predictions of the SWM rainfall are presented. Totally 96 predictions have been made and the observed numbers of correct predictions are 72, which is about 75% of total prediction. In Chi-square

test, the hypothesis that the correct predictions are purely due to chance gets rejected at 5% level of significance.

6. It is expected that the model has to be updated every year with current data before the next year forecast is produced. The achieved models have been presented in equations 7(a-d).

The model developed captures only the inter-annual variability of SWM rainfall of the subdivisions. However, the variability due to inter-seasonal, monthly, weekly and smaller scale fluctuation are ignored. Due to non-linear oscillating nature of SWM phenomenon, bifurcation and sensitivity to parameter errors may also be present but avoided. The predicted models [7(a-d)] appear as 'good statistical tool' for one year ahead prediction. The values of the coefficients are available in the Table 6.

7. The model developed in the paper carries importance in the sense that SWM rainfalls of the subdivisions are modeled as generalized lognormal random variables following basic law of proportion. It is also shown that long-term climatic mean is essentially controlled by one-step annual connection which is insufficient to simulate observed variability. These connections are established with lagged years are incorporated to shape and modify the model. Figs. 3&4, the rates of change of year wise SWM rainfall have inflection points; clearly indicating SHWB and AM sub divisional rainfalls may be sensitive to parameter values. For these cases, the mean part represented by equation (2) may contain two stable fixed points for oscillation of the system.

It would be relevant to mention the works of the pioneering researchers. Sahai *et al.* (2000) suggested an ANN model for the All India SWM rainfall data on the basis of past rainfall data whereas Iyenger and Raghukant (2003) suggested a regression model for All India and different zonal SWM rainfall series namely, West Central India, Southern India and Northern India. However, little attention is given to SWM rainfall of particular regional or meteorological subdivision series of Eastern and North Eastern subdivisions. In our study, regression models with the effect of back correlation are undertaken for the meteorological sub divisions of GWB, SHWB, AM and NMMT.

The author sincerely thanks National Data Centre ADGM (R), Pune for providing South West Monsoon rainfall data of subdivisions of Gangetic West Bengal (No. 6), Sub-Himalayan West Bengal (No. 5), Assam and Meghalaya (No. 3) and also Nagaland, Manipur, Mizoram and Tripura (No. 4).

The contents and views expressed in this letter are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Basak, P., 2014, "Variability of south west monsoon in West Bengal: An application of principal component analysis", *Mausam*, **65**, 559-568.

Basak, P., 2017, "Southwest monsoon rainfall in Assam: An application of principal component analysis for understanding variability", *Mausam*, **68**, 357-366.

Gadgil, S., Srinivasan, J., Nanjundiah, R. S., Krishna Kumar, K., Munot, A. A. and Rupa Kumar, K., 2002, "On forecasting the Indian summer monsoon: the intriguing season of 2002", *Current Science*, **83**, 394-403.

Hastenrath, S. and Greischar, L., 1993, "Changing predictability of Indian monsoon rainfall anomalies?", *Proc. Ind. Acad. Sci. (Earth Planet Sci.)*, **102**, 35-47.

Iyenger, R. N. and Basak, P., 1994, "Regionalization of Indian monsoon rainfall and long-term variability signals", *Int. J. Climatol.*, **14**, 1095-1114.

Iyenger, R. N. and Raghukant, S. T. G., 2003, "Empirical modelling and forecasting of Indian monsoon rainfall", *Current Science*, **8**, 1189-1201.

Johnson, N. L. and Kotz, S., 1972, "Distribution in Statistics: Continuous multivariate Distribution", New York, Wiley.

Kedem, B. and Chiu, L. S., 1987, "On the log normality of rain rate", *Proc. Natl. Acad. Sci., USA*, **84**, 901-905.

Mooley, D. A. and Parthasarathy, B., 1984, "Fluctuations in all-India summer monsoon rainfall during 1871-1978", *Climatic Change*, **6**, 287-301.

Sahai, A. K., Soman, M. K. and Satyan, V., 2000, "All India summer monsoon rainfall prediction using an artificial neural network", *Climate Dynamics*, **16**, 291-302.

Thapliyal, V., 1990, "Large scale prediction of over India: Evaluation and development of new model", *Mausam*, **41**, 339-346.

PIJUSH BASAK

Narula Institute of Technology, Kolkata,
West Bengal – 700 109, India
(*Received 10 July, 2017, Accepted 24 September, 2018*)

e mail: pijushbasak@rediffmail.com