

Southwest monsoon rainfall in Assam: An application of principal component analysis for understanding of variability

PIJUSH BASAK

Assistant Professor & HOD, Narula Institute of Technology, Kolkata, India

(Received 31 August 2015, Accepted 19 December 2016)

e mail : pijushbasak@rediffmail.com

सार – भारत में असम राज्य में मॉनसूनी वर्षा की स्थानिक और कालिक परिवर्तिता को समझने के लिए मूल अवयव विश्लेषण (पी सी ए) का अनुप्रयोग किया गया है। परिवर्तिता को समझने के लिए राज्य में दूर दूर तक फैले 12 केंद्रों के दक्षिण पश्चिम मॉनसून (एस डब्ल्यू एम) वर्षा के आँकड़ों का विश्लेषण 60 वर्षों की अवधि के लिए किया गया। कुछ केंद्रों और उनके तदनु रूप मूल अवयवों (पी सी) के लिए सांख्यिकीय रूप से महत्वपूर्ण प्रवृत्ति और अधिक/कम संक्रमण संकेत प्रेक्षित किए गए। संभावित महत्वपूर्ण संकेतों का पता लगाने के लिए मूल अवयव विश्लेषण के माध्यम से उत्तरी और दक्षिणी असम के संबद्ध क्षेत्रों की पहचान की गई। यह देखा गया कि राज्यवार और संबद्ध क्षेत्रों के लिए कुछ मूल अवयव (पी सी) की सकारात्मक अथवा नकारात्मक प्रवृत्ति तथा महत्वपूर्ण अधिक/कम संक्रमण है।

ABSTRACT. The principal component analysis (PCA) is applied to understand the spatial and temporal variability of monsoonal rainfall in the state Assam in India. The Southwest Monsoon (SWM) rainfall data over 12 widely spread stations located over the state has been analyzed for a period of 60 years for understanding variability. A statistically significant trend and a above/below transition signal has been observed for a few stations and the corresponding principal components (PCs). Coherent regions of Northern and Southern Assam have been identified through PCA to bring out the possible significant signals. It is observed that some of PCs for state-wise and coherent regions have positive or negative trend and significant above/below transition.

Key words – Southwest monsoon, Variability, Principal component analysis, Inter-annual association, Empirical orthogonal function, Eigen value, Rainfall anomaly, Spatial structure, Drought, Flood.

1. Introduction

Southwest Monsoon (SWM) rainfall is an important phenomenon connected with the weather and agriculture of the state of Assam. The amount of rainfall in different stations of the state over considerable range fluctuates. Variability may be understood on different time scales such as day, month and season and on diverse spatial scales namely, station, district, state and country. As SWM is an organized spatial phenomenon on a large scale and is existent over a months' time, it may be useful to study the data in some optimally chosen scale. The state of Assam and SWM is chosen for the study. A very few investigations in the state based on statistical analysis on daily, monthly and seasonal data has been undertaken by investigators. A few literatures are available for other states (Basu *et al.*, 2004; Chanda and Dhar, 1975; Basu, 2001). The paper discusses randomness of the SWM time series. However, as the inter-station data are spatially correlated, there may be some kind of trend that could be identified. The present work is connected with the

decomposition the series into spatial and temporal variations by the principal components in time and empirical orthogonal functions (EOFs) in space. A few works in this respect in the All India level are by Bedi and Bindra (1980); Hastenrath and Rosen (1983); Iyenger and Basak (1994), for West Bengal Basak (2014), for Karnataka Iyenger (1991) and for North East India Mahapatra *et al.* (2011). The emphasis in the paper is to choose the spatial structure in the field and temporal pattern detectable in the data.

2. Data

In the present study, the data analyzed are Southwest Monsoon (SWM) rainfall (June-September) over 12 stations of Assam for the period of 60 years (1900-1960). The state of Assam along with the stations considered is presented in Fig. 1. It is, however, understandable to include more stations along with longer period of time. But, there are some limitations due to data gaps and unequal length of data of station time series. However,

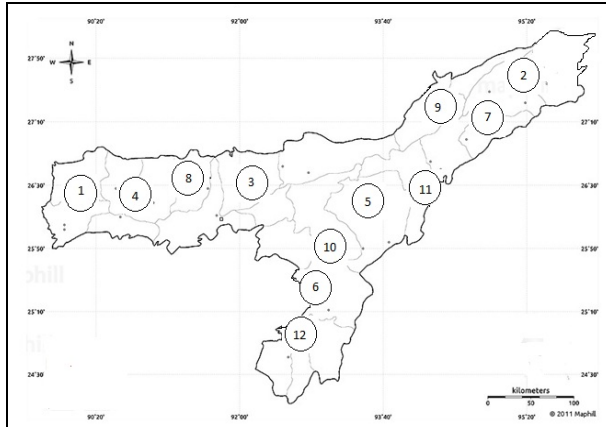


Fig. 1. Station data network of Assam

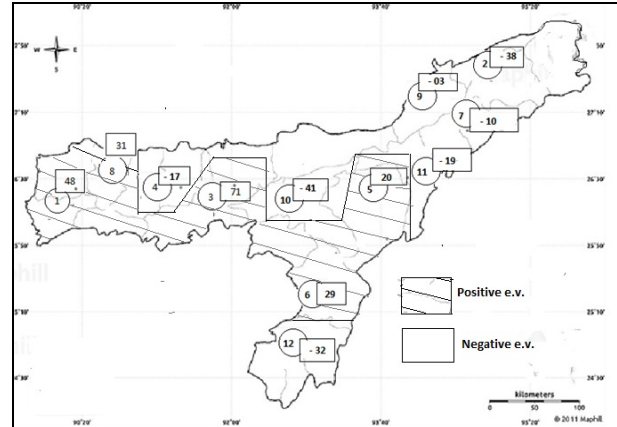


Fig. 2. Station network with first eigen-vector

inclusion of more stations may or may not enhance the desired signal. Meanwhile, an optimum number stations and length of time has been considered. The rainfall in Assam is of considerable interest. The state Assam situated in the rainfall subdivision of India Meteorological Department (IMD), namely, Assam and Meghalaya. Both the plane lands and hilly areas are located in different parts of the state. The SWM rainfall accounts for the bulk of the rainfall over the state overall and accounts for about 70% of total annual rainfall.

3. Principal component analysis

If R_{it} be the actual SWM rainfall at station i ($i = 1, 2, \dots, M$) in the year t ($t = 1, 2, \dots, N$), then, the anomaly data series are

$$r_{it} = (R_{it} - m_i); m_i = \left(\frac{1}{N}\right) \sum_{t=1}^N R_{it}$$

From the symmetric covariance matrix defined as

$$C_{ij} = (1/N) \sum_{t=1}^N r_{it} r_{jt}$$

the eigenvalues (λ_j) and the eigenvectors (ϕ_{ij}) are extracted such that the j -th vector (ϕ_{ij}) corresponds to the j -th largest eigenvalue λ_j of the covariance matrix.

The rainfall anomaly at station i in year t can be represented as orthogonal decomposition in terms of principal components in time and empirical orthogonal function (EOF) in space, namely,

$$r_{it} = \sum_{j=1}^M p_{jt} \phi_{ij}$$

and the rainfall series r_{it} are transformed to principal component series defined as

$$p_{jt} = \sum_{i=1}^M r_{it} \phi_{ij}$$

and reflects the spatial variation of the original series. The first few principal components series p_{jt} account for a large spatial variation contained in the data set. These p_{jt} series can be utilized to extract the temporal variability in the data while the eigenvectors (ϕ_{ij}) represent spatial patterns contained in the data.

Spatial organization of the rainfall field can be viewed by the structure of eigenvectors or EOF pattern for first principal component explaining 24.27% of the total variance is presented in Fig. 2. It may be thought of mainly as Northern (*i*) and Southern (*ii*) contrast having negative loadings for Northern stations, namely Goalpara, Nowgong, North Lakhimpur, Dibrugarh, Jorhat and Sibsagar and also positive loading for Southern stations, namely, Dhubri, Nalbari, Gwahati, Golaghat and Halflong. Also, an isolated station in the Plateau of Assam, namely Silchar (*iii*) is observed with negative loading.

As an interpretation, it may be thought of above/below normal rainfall with the largest weight in the Northern stations and Silchar would indicate a trend of below/above normal in the Southern stations. This EOF pattern may be considered as the dominant pattern of summer precipitation with distinct phase of monsoon. The pattern behavior perhaps results from the North/South oscillation of monsoon trough. In fact, the Northern and Southern stations of Assam do not experience similar SWM rainfall exactly at the same period; it may be an interpretation of that. No other specific quantitative

conclusions can be drawn from the EOF patterns as it is difficult to verify the significance of the sign and values of the station weights given by the eigenvectors.

EOF of second principal component indicate positive loading for all the stations. Consequently, it infers that all the stations of Assam possess similar pattern of rainfall. However, as EOF of second principal component accounts for only 13.66% of total variance, it is perhaps a local feature not related to emergent large scale SWM rainfall.

The EOF's of the third, fourth and fifth principal components explain less variances and the concerned importance are not presented. First five principal components of the SWM precipitation data contribute above 66% of the total variance. It is interesting to note that not more than three components are required to represent the SWM rainfall over the whole state.

4. Study of inter-annual variability of station rainfall

The statistical properties of all the rainfall time series r_{it} ($i = 1, 2, \dots, M; t = 1, 2, \dots, N$) have been examined for their statistical properties.

For the purpose of studying the Gaussianness of the series, the Kolmogorov-Smirnov (K-S) test (Hays & Winkler, 1970) is applied. All the 12 station rainfall series have been studied for the existence of Gaussianness (Table 1). Except for the station Halflong (No. 6) and North Lakhimpur (No. 9) which are marginally significant, all the SWM time series may be considered as Gaussian at 5% significant level.

For further analysis, each of the 12 rainfall series is assumed to be ergodic. Thus, presence of any trend is to be detected in the data series r_{it} from the data. The Mann-Kendal rank statistic test (WMO 1966b) is applied to test the existence of linear or non-linear trend, if any (Table 1). The rainfall series of Dhubri (No. 1), Dibrugarh (No. 2) and Silchar (No. 12) exhibit a negative trend at 1% level whereas Goalpara (No. 4), Jorhat (No. 7) and Sibsagar (No. 11) exhibit a negative trend at 5% level. The time series of North Lakhimpur, Dhubri and Dibrugarh are presented in Figs. 3-5.

As a further test of annual association, all the rainfall series r_{it} ($i = 1, 2, \dots, M; t = 1, 2, \dots, N$) have been studied to test the presence of auto-correlation for a maximum lag of 3 years (Table 2). Inspection of the table reveals that the auto-correlation at lag 1 and at lag 2 is significant at 5% level for North Lakhimpur (No. 9) and Nowgong (No. 10); it is significant at lag 1 at 5% level for Silchar

(No. 12) and finally, at lag 3 at 5% level for Gauhati (No. 3). However, the auto-correlations when significant are too small and marginal to be of any help in forecasting the rainfall. The time series of North Lakhimpur (No. 9) and Nowgong (No. 10) do not possess significant trend but possessing significant auto-correlation. Those stations may be attempted for possible prediction exercise.

Another aspect of the rainfall series is the presence of periodicity. The Power Spectral Density (PSD) function using Hamming Window (Blackman and Tukey, 1958) is utilized to test the existence of periodicity. The test indicates reveals that all the data series as white at 5% level. It may, however, be noted that no prominent temporal pattern emerges at the station level.

5. Further study of inter-annual variability of station rainfall

It has been studied that station rainfall mostly does not show any significant year-to-year variability as evidenced by auto-correlation and PSD function. However, the possibility of a year-to-year association existing in SWM rainfall series still persists. A particular pattern predominant in the rainfall anomaly (fluctuation about long term mean), namely the transition in sign of the anomaly is studied (Iyenger, 1991).

With respect to long term mean, the rainfall anomaly such as ++ (+ve and +ve anomaly in two consecutive years) etc. are presented in two way contingency table (Table 3). The significance of the association is tested against the expected number, as if the change in sign in anomaly is purely due to chance. The tabulated chi-square at 1 degree of freedom at 5% level (3.84) is utilized for acceptance or rejection of Null hypothesis: 'there is no dependence in the year-to-year sign changes'. The results of Table 3 clearly indicate that for the stations Nowgong (No. 10), the year-to-year change in rainfall anomaly clearly exhibit a pattern, which is not purely due to chance. The time series of the SWM rainfall of Nowgong (Fig. 6) exhibit a predominant period of 5 to 8 years before a change in anomaly pattern takes place.

It is very interestingly observed that the station Nowgong (No. 10) which does not exhibit significant trend (Table 1) and significant autocorrelation (Table 2) at 5% level indicates significant anomaly pattern.

6. Study of inter-annual variability of principal components

It may be noted that the principal components (PCs) in descending order of magnitude carries out the temporal

TABLE 1
Stations detail with tests of Gaussianness and trend

S. No.	Station Name	Lat./Long.	Zone	K-S Statistic	Mann-Kendall τ
1.	Dhubri	26.02° N, 89.98° E	S	-0.2073	- 0.3118***
2.	Dibrugarh	27.47° N, 94.92° E	N	0.4146	- 0.2576***
3.	Gauhati	26.18° N, 91.75° E	S	-1.2073	0.0078
4.	Goalpara	26.18° N, 90.63° E	N	1.0364	-0.2226**
5.	Golaghat	26.52° N, 93.98° E	S	-0.2073	- 0.1017
6.	Halflong	25.17° N, 93.02° E	S	-1.7619*	- 0.1751**
7.	Jorhat	26.90° N, 94.20° E	S	-1.1400	-0.2282**
8.	Nalbari	26.45° N, 91.43° E	S	-1.2020	-0.1012
9.	North Lakhimpur	27.23° N, 94.12° E	N	- 1.7619*	- 0.0452
10.	Nowgong	26.37° N, 92.70° E	N	0.2073	-0.1175
11.	Sibsagar	26.98° N, 94.63° E	S	1.3473	- 0.0926
12.	Silchar	24.75° N, 92.08° E	I	-0.5182	-0.2994***

* Significant at 10% level, ** Significant at 5% level, *** Significant at 1% level, N = Northern, S = Southern, I = Isolated

TABLE 2
Stations detail with auto-correlation up to lag 3

S. No.	Station Name	Zone	Auto-correlation		
			Lag 1	lag 2	lag 3
1.	Dhubri	S	0.2500	0.2349	0.1783
2.	Dibrugarh	N	0.1259	0.1752	0.1952
3.	Gauhati	S	0.1517	0.1283	0.3217**
4.	Goalpara	N	0.0807	0.1614	0.0127
5.	Golaghat	S	0.1671	0.1267	0.0022
6.	Halflong	S	0.0696	-0.2720	0.2696
7.	Jorhat	S	0.3321	-0.0855	-0.0331
8.	Nalbari	S	0.1985	0.1715	0.1294
9.	North Lakhimpur	N	0.4600**	0.4563**	0.2723
10.	Nowgong	N	0.4367**	0.3007**	0.1382
11.	Sibsagar	N	-0.0841	0.0143	0.0975
12.	Silchar	I	0.2609**	0.2123	0.2441

**Significant at 5% level, N = Northern, S = Southern, I = Isolated

variability of the rainfall at the station level. Each dominant p_{jt} ($j = 1, 2, \dots, 5$) is, however, a time series sampled annually and carry out the information on inter-annual variability and carry out the information on inter-annual variability.

An analysis similar to the SWM rainfall series has been carried out on the principal components series p_{jt} ($j = 1, 2, \dots, 5$) in question. All the other p_{jt} series are

found to be Gaussian as per the results of the K-S test (Table 4). Also, the Mann-Kendall rank statistics test indicates that PC1 possess significant trend at 1% level and so does PC3 at 5% level (Table 4). The PC series p_{jt} ($j = 1, 2, \dots, 5$) have been studied to test the presence of auto-correlation for a maximum lag of 3 years. PC2 is significant at lag 1, lag 2 and lag 3 at 5% level, whilst PC1 is significant at lag 1 and lag 3 at 5% level; also, PC4 and PC5 are significant at 5% level at lag 2; though PC4

TABLE 3
Frequency of sign sequences in SWM rainfall of 12 stations (N = 60 years)

Station name	Sign ++		Sign +-		Sign -+		Sign --		Chi-sq. obs.
	Obs.	Expt.	Obs.	Expt.	Obs.	Expt.	Obs.	Expt.	
Dhubri	16	13.76	12	14.23	13	15.23	18	15.76	1.3590
Dibrugar	17	15.25	13	14.74	13	14.74	16	14.25	0.8265
Gauhati	16	15.25	14	14.74	14	14.74	15	14.25	0.1517
Goalpara	16	15.25	14	14.74	14	14.74	15	14.25	0.1517
Golaghat	18	16.81	14	15.19	13	14.19	14	12.81	0.3878
Halflong	12	10.59	13	14.41	13	14.41	21	19.59	0.5651
Jorhat	18	15.76	12	14.24	13	15.24	16	13.76	1.3646
Nalbari	13	11.90	14	15.10	13	14.10	19	17.90	0.3352
North Lakhimpur	14	10.59	11	14.41	11	14.41	23	19.59	0.5651
Nowgong	15	11.90	11	14.10	12	15.10	21	17.90	2.6624**
Sibsagar	11	10.59	14	14.41	14	14.41	20	19.59	0.0478
Silchar	16	14.25	13	4.75	13	14.74	17	15.25	0.8287

**Significant at 5% level

TABLE 4
Principal components detail with tests of Gaussianness, trend and auto-correlation

S. No.	Station name	K-S statistics	Mann-Kendall τ	Auto-correlation		
				Lag 1	Lag 2	Lag 3
1.	PC1	-0.8291	- 0.4102***	0.4122**	0.2295	0.3716**
2.	PC2	1.3473	0.0531	0.4262**	0.4905**	0.4004**
3.	PC3	-0.2073	0.1684**	0.1351	0.0510	0.2250
4.	PC4	1.3473	0.0667	0.0915	0.3107**	0.0423
5.	PC5	1.0364	0.0486	-0.0378	0.2691**	-0.3432
6.	PC6	-0.2073	0.1333	0.0911	-0.1125	-0.1034

***Significant at 1% level, **Significant at 5% level

TABLE 5
Frequency of sign sequences in principal components of SWM rainfall of 12 stations (N = 60 years)

Station name	Sign ++		Sign +-		Sign -+		Sign --		Chi-sq. obs.
	Obs.	Expt.	Obs.	Expt.	Obs.	Expt.	Obs.	Expt.	
PC1	18	13.29	10	14.71	10	14.71	21	16.29	6.0520**
PC2	22	15.25	9	14.74	9	14.74	19	14.25	8.8934**
PC3	17	15.25	15	14.74	15	14.74	12	14.25	0.0349
PC4	20	15.25	14	14.74	15	14.74	10	14.25	0.0083
PC5	12	16.81	17	15.19	17	14.19	13	12.81	1.3788
PC6	15	10.59	17	14.41	11	14.41	16	19.59	0.5651

**Significant at 5% level

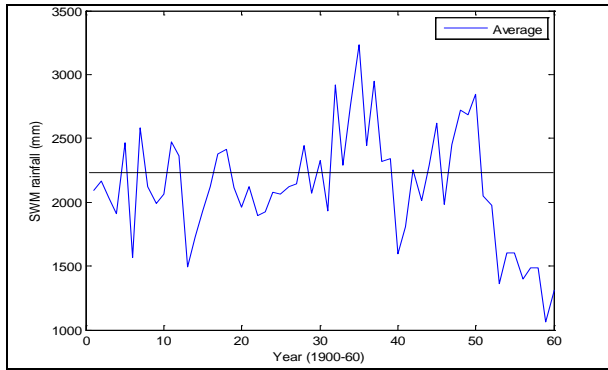


Fig. 3. SWM rainfall of north Lakhimpur

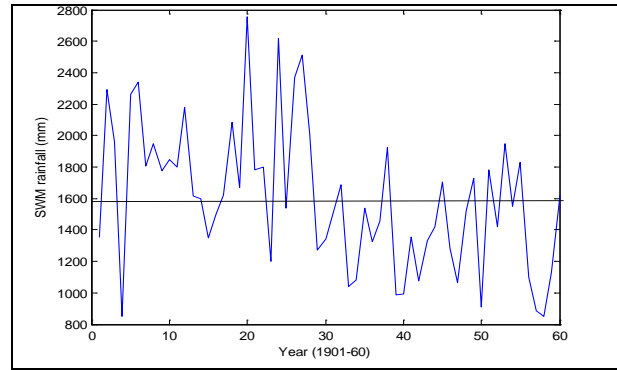


Fig. 4. SWM rainfall of Dhubri

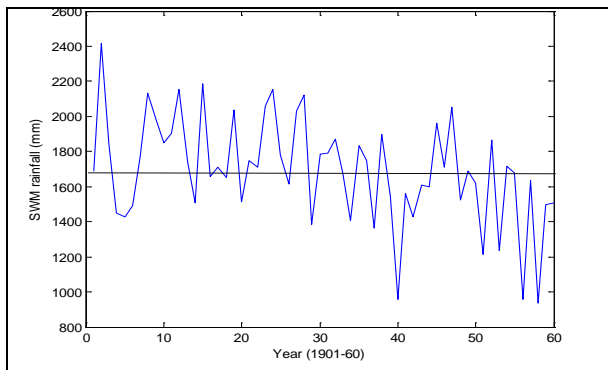


Fig. 5. SWM rainfall of Dibrugarh

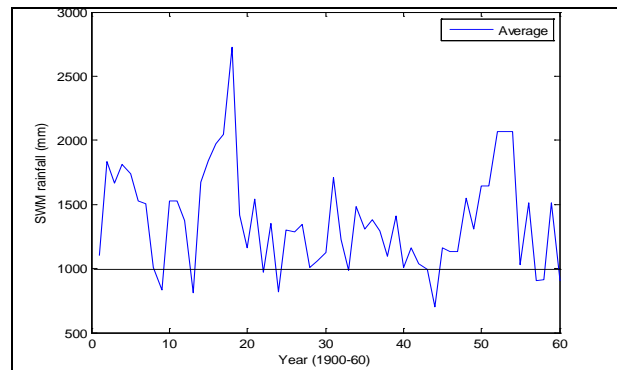


Fig. 6. SWM rainfall of Nowgong

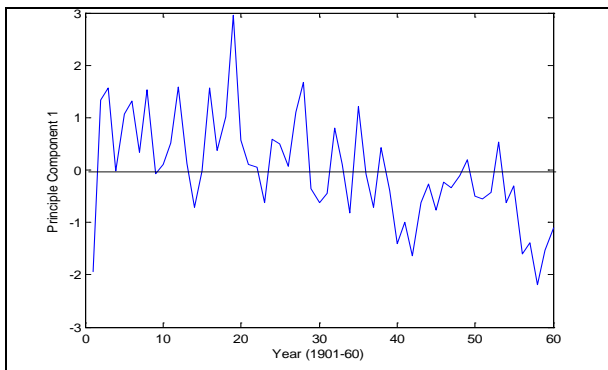


Fig. 7. Time series of first principal component of Assam

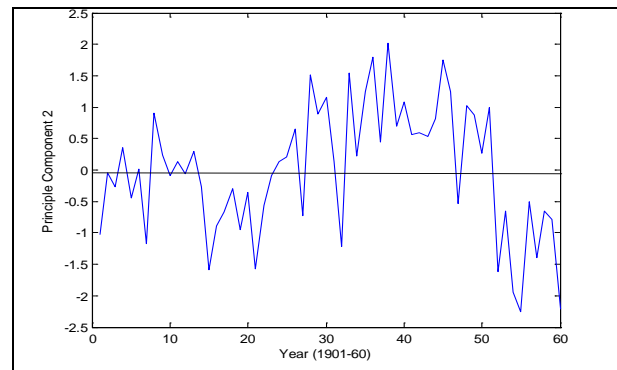


Fig. 8. Time series of second principal component of Assam

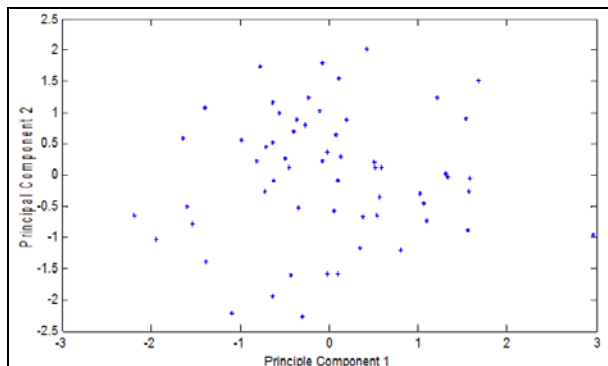


Fig. 9. Variability of the principal components of the SWM rainfall in Assam (1901-1960)

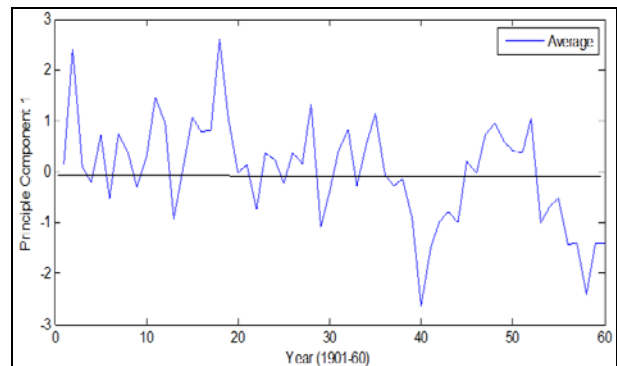


Fig. 10. Time series of First Principal Component of Northern Assam

and PC5 contribute merely 9.45% and 8.07% of total variance respectively. Time series of PC1 and PC2 which are possible candidates for a prediction exercise. None of the PSD functions are found significant at 5% level indicating absence of periodicity in the PC series.

7. Further study of inter-annual variability of principal component series

The PC time series p_{jt} ($j=1, 2, \dots, 5$) are optimally area-weighted SWM rainfall series, although the question remains whether the PCs representing the size of the state Assam would bring out the feature of year-to-year anomaly.

A similar analysis for the sign sequence changes as in Section 5 is performed for the PC series p_{jt} ($j = 1, 2, \dots, 5$). It is surprisingly observed that both main PCs, namely PC1 and PC2 possess significant transition at 1% level. All the other PC series are clearly identified as purely due to chance. Thus, the first and second components of PC analysis of the SWM rainfall of Assam contributing 24.27% and 13.66% respectively of total variance represents a pattern with characteristic term as a year or a multiple of it. The time series of PC1 shows a predominant period of nearly 4 to 10 years meaning that the same sign persist for 4 to 10 years before a change in sign takes place (Fig. 7). Similarly, PC2 shows a predominant period of nearly 5 to 8 years (Fig. 8). It is interesting to note that PC1 and PC2 having significant auto-correlations exhibit significant anomaly pattern.

8. Grouping the years

Whenever, we consider the area rainfall, it means the weighted average of the individual station rainfall in the optimal way. The first two dominant PCs, namely PC1 and PC2 on any time scale would indicate the main feature of rainfall are the most important characteristic rainfall for a particular year for the network of station. Thus, with PC1 and PC2 when presented in a diagram would bring the feature of a particular year. The normal years are near origin and the drought/flood years are away from origin. Years with excessive rainfall (flood), so called flood years such as 1910, 1911, 1915, 1916, 1931, 1954 have large positive PC1 and PC2 values and away from the origin (Fig. 9). Also, years with deficit rainfall years (droughts), namely 1901, 1905, 1911, 1918, 1920, 1941, 1951 possess negative PC1, PC2 values are also away from the origin in other direction. Those years are presented in Fig. 9. Moreover, one can mark years with prescribed percentage variations about the normal rainfall on this figure. Nearness of two or more years on this diagram indicates that for these

years atmospheric conditions are similar. Such information could help in foreshadowing droughts and floods.

9. Analysis in coherent zones

The study so far presented is considered all the SWM data of state-wise 12 rainfall stations spread over Assam. The stations are located in the regions of different climatic conditions and/or topography and are certainly not homogeneous. Consequently, the significant variability patterns or signals that emerged for the state as a whole state may not be valid for smaller regions. Moreover, the signals which are too weak to be detected in the state-wise network of stations may become stronger if principal components are extracted for a smaller coherent region. The EOF pattern of the first principal component of network has been utilized for the emergence of three coherent zones (Fig. 2). Keeping this in view, the stations situated in the northern region in the state indicating negative EOF loading is considered as Northern stations. These five stations are: Golagaht, Nowgong, North Lakhimpur, Dibrugarh and Sibsagar. The stations are at the foothills of Himalaya and perhaps, the orography plays the role for the Northern stations. Also, the set of other 6 stations situated southern part of the state is viewed as Southern stations. The stations are Dhubri, Gauhati, Goalpara, Halflong, Jorhat and Nalbari. The stations are all at the plane lands of Assam with apparently similar properties. An isolated station Silchar is situated at the Southern fringe of the state. The principal component analysis (PCA) for the set of Northern stations and Southern stations (excluding isolated station) have been carried out separately for the period 1901-1960 as explained in the earlier sections. The statistical properties are extracted for the concerned PCs for the Northern stations and Southern stations separately. For Northern stations, all the PCs are Gaussian with PC1 having negative trend at 1% level. PC1 and PC2 (explaining about 24% and 15% of total variance respectively) possess significant auto-correlations at lag 1, lag 2 and lag 3 (at 5% level) (Table 6). The time series of PC1 and PC2 of Northern stations are presented in Figs. 10 and 11 respectively. However, for Southern stations, all the PCs are Gaussian with PC1 having negative trend at 1% level. PC1 and PC2 possess significant auto-correlations at lag 1, lag 2 and lag 3 (at 5% level) whilst PC3 possess significant auto-correlations at lag1 and lag2 at 5% level. The characters of the time series of Southern stations are presented in Figs. 12 and 13 respectively.

The procedure for extracting sign sequence of year-to-year anomaly pattern is applied for the PC series for the set of Northern stations and Southern stations separately. Inspection of the results reveals that for Northern

TABLE 6

Principal components detail with tests of Gaussianness, trend and auto-correlation for Northern stations

S. No.	Station Name	K-S statistics	Mann-Kendall τ	Auto-correlation		
				Lag 1	Lag 2	Lag 3
1.	PC1	0.2073	-0.3458***	0.5145**	0.3450**	0.2964**
2.	PC2	-0.8291	- 0.0859	0.4100**	0.4453**	0.3762**
3.	PC3	0.4145	- 0.0903	0.3997**	0.2821**	0.1194
4.	PC4	0.1703	- 0.1244	0.1244	0.1107	-0.0356

***Significant at 1% level, **Significant at 5% level

TABLE 7

Frequency of sign sequences in principal components of Northern stations ($N = 60$ years)

Station Name	Sign ++		Sign +-		Sign -+		Sign --		Chi-sq. obs.
	Obs	Expt	Obs	Expt	Obs	Expt	Obs	Expt	
PC1	26	20.19	9	14.83	8	13.83	16	10.83	7.9273***
PC2	16	10.59	9	14.41	9	13.83	25	19.51	8.0265***
PC3	12	9.76	12	14.24	12	14.24	23	20.76	1.4605
PC4	11	10.59	14	14.41	14	14.41	20	19.59	0.0478
PC5	15	12.81	13	15.19	12	14.19	19	16.81	1.3134

***Significant at 1% level

TABLE 8

Principal Components detail with tests of Gaussianness, trend and auto-correlation for Southern stations

S. No.	Station Name	K-S statistics	Mann-Kendall τ	Auto-correlation		
				Lag 1	Lag 2	Lag 3
1.	PC1	-0.2073	-0.2813***	0.4919**	0.3112**	0.3211**
2.	PC2	1.0364	- 0.0282	0.4540**	0.4604**	0.3919**
3.	PC3	1.0364	- 0.0723	0.2501**	0.2004**	- 0.0708
4.	PC4	- 0.2073	0.0011	0.1248	0.1621	-0.1180
5.	PC5	1.0364	- 0.1288	0.1362	0.2691**	0.0260

***Significant at 1% level, **Significant at 5% level

TABLE 9

Frequency of sign sequences in principal components of Southern stations ($N = 60$ years)

Station Name	Sign ++		Sign +-		Sign -+		Sign --		Chi-sq. obs.
	Obs.	Expt.	Obs.	Expt.	Obs.	Expt.	Obs.	Expt.	
PC1	20	16.81	12	15.19	11	14.19	16	12.81	2.7868*
PC2	21	13.76	7	14.24	8	15.24	23	15.76	14.2559***
PC3	15	11.46	11	14.54	11	15.54	22	18.46	3.4961**
PC4	12	12.35	15	14.64	15	14.64	17	17.35	0.0347
PC5	17	14.45	12	14.25	13	15.25	17	14.74	1.3769

***Significant at 1% level, ** Significant at 5% level, * Significant at 10% level

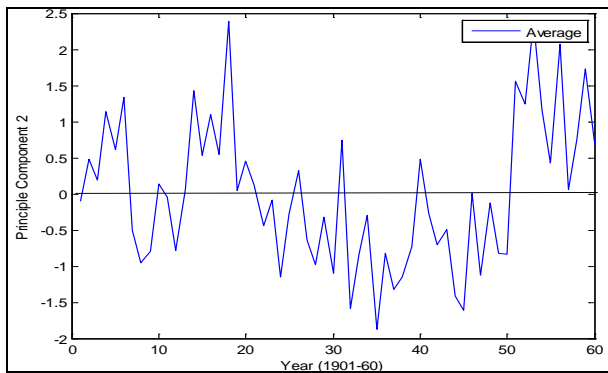


Fig. 11. Time series of second principal component of Northern Assam

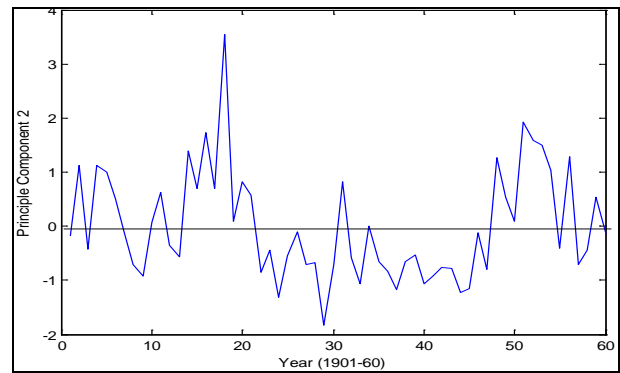


Fig. 13. Time series of second principle component of Southern Assam

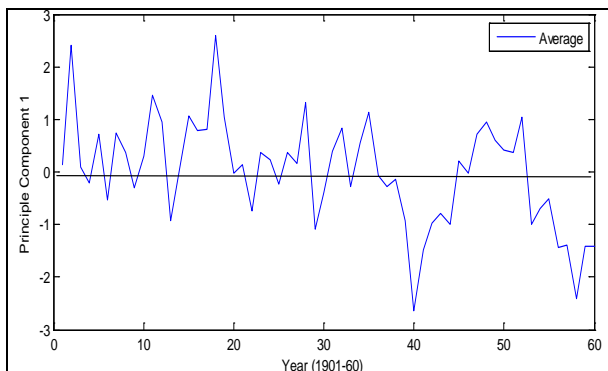


Fig. 12. Time series of first principal component of Southern Assam

stations PC1 and PC2 exhibits significant year-to-year transition pattern at 1% level, clearly showing that the year-to-year patterns cannot be dismissed as simply due to chance in those cases (Tables 7).

Inspection of the results also reveal that for Southern stations, PC2 exhibits significant year-to-year transition pattern at 1% level whilst PC1 and PC3 exhibits that year-to-year transition pattern at 10% level and 5% level respectively, Consequently the year-to-year patterns of PCs cannot be dismissed as due to chance (Tables 8 and 9).

10. Emergence or diffusion of sign sequences of SWM rainfall and inter-annual variability

The review of results indicates that all the 12 SWM rainfall series, except a few (such as Nowgong and North Lakhimpur) are white noise processes. The series is mostly normal and having no trend except for a few cases (such as Dhubri, Dibrugarh, Goalpara and Halflong). No

oscillations are observed in any of the series as indicated by their respective PSD functions. But, the examination of test of significant sign sequence shows year-to-year rainfall anomaly pattern for Nowgong. It is to be noted that the station Nowgong is located at hilly areas of 5 Northern stations of Himalayan Assam.

Both the dominant PCs, namely PC1 and PC2 explaining 24.27% and 13.66% of total variance respectively show significant year-to-year transition in sign sequence of anomaly pattern.

11. Discussion

A conventional approach for time series analysis (SWM rainfall or PCA series) is extraction of auto-correlation and power spectral density. However, difficulty arises to analyze the time series through classical spectral analysis as most of auto-correlations and power spectral densities are insignificant at the station level. However, as the rainfall stations are widely spread and may be correlated among themselves, the straightforward time series considering all the stations are difficult and complicated in the vast topographic area like a state. The analysis of the time series of the stations would obviously neglect the inherent spatial structure that is emergent in the state that play a vital role in emerging or diffusing any signal. The SWM rainfall is large scale organized phenomenon spread on a long length of time (SWM) and so the area rainfall instead of station rainfall are more preferable. This requires demarcation of the area connected to station, as indicated by Iyenger and Basak (1991) and Basak (2014). PCA is a solution to the difficulties. In lieu of handling with a large number of station data, only a few significant PCs are to be considered. PCA can be thought of as a generalized Fourier decomposition of a random field. Although, the identification of the periodicity is not directly possible, the energy contained in different components is extracted as

the eigenvalue and eigenvector of the covariance matrix. In the present study, for the Assam rainfall stations PCA has been demonstrated in understanding the SWM rainfall variability. The station data which are neither perfectly spatially correlated, nor exactly uncorrelated, gets transformed into PCA but contain the entire temporal characteristic for complete station network. The advantage of this is very apparent when it is observed that for Assam SWM rainfall, the first PC explain less than 30% of spatial variance; but a representative of all the stations and possess all the characteristics. It is because all the stations are highly correlated with PC1. Another advantage of the analysis is the delineation of different zones of the state as per its topography and geographical location such as Northern stations and Southern station (also North-South oscillation).

12. Summery and conclusion

In the current study, the SWM rainfall spread over 12 stations of Assam for 60 years has been considered. It may be mentioned that study of temporal variability of rainfall (SWM or otherwise) did not receive much attention in the past except a few, particularly for the state of Assam. In the studies, the statistical tools namely, PCA, regression analysis and others are utilized. The SWM rainfall is a large scale spatially organized phenomenon and leads to existence of variation at spatial and temporal scales along with considerable spatial correlations among the stations. A common approach for time series analysis (SWM or PCA) is the study of auto-correlation and power spectral analysis. In the state of Assam, at the station level, the autocorrelations at different lags are mostly insignificant or of very small magnitude when significant; also, as the rainfall stations are widely spread and may be correlated among themselves, the straight forward time series analysis is complicated. However, efforts are made to identify some kind of temporal association for the SWM series and PC time series. The SWM station wise data are, except a few are white noise; but the 2×2 contingency table yield significant relationship for a station namely, Nowgong (No. 10), indicating that anomaly pattern of year-to-year rainfall of the station are not purely due to chance. In the same way, when applied to PC series, the above/below average contingency table also yield similar trends such as PC1, PC2 of All Assam stations; PC1, PC2 of Northern stations and PC1, PC2 and PC3 of Southern stations exhibit similar relationships.

The process is an interesting one and may be useful in identifying impending signal in SWM rainfall in Assam. However, intense analysis is required for establishing the signals of the rainfall fluctuation.

Acknowledgement

The author sincerely thanks National Data Centre ADGM (R) Pune for providing SWM rainfall data of 12 stations of Assam.

References

- Basak, P., 2014, "Variability of south west monsoon rainfall in West Bengal: An application of principal component analysis", *Mausam*, **65**, 4, 559-568.
- Basu, G. C., 2004, "A feature of monsoon rainfall and its variability with comparative study in meteorological sub-divisions of West Bengal", *Mausam*, **52**, 4, 736-746.
- Basu, G. C., Bhattacharjee, U. and Ghosh, R., 2004, "Statistical analysis of rainfall distribution and trend of rainfall anomalies district wise during monsoon period over West Bengal", *Mausam*, **55**, 409-418.
- Bedi, H. S. and Bindra, M. M. S., 1980, "Principal components of monsoon rainfall", *Tellus*, **32**, 296-298.
- Blackman, R. B. and Tukey, J. W., 1958, "The measurement of power spectra", Dover Publication, New York, p190.
- Chanda, B. N. and Dhar, O. N., 1975, "A study of incidence of droughts in the Gangetic West Bengal", *Proc. Indian National Academy of Sciences*, **42**, 22-33.
- Hastenrath, S. and Rosen, A., 1983, "Patterns of Indian monsoon rainfall anomaly", *Tellus*, **A35**, 324-331.
- Hays, W. H. and Winkler, R. L., 1970, "Statistics: Probability, Inference and Decision", Rinehart and Winston, inc., New York, 937.
- Iyenger, R. N., 1991, "Application of principal component analysis to understand variability of rainfall", *Proc. Ind. Acad. Sc. (Earth Planet Sc.)*, **100**, 2, 105-126.
- Iyenger, R. N. and Basak, P., 1994, "Regionalization of Indian monsoon rainfall and long term variability signals", *Int. J. Climatol.*, **14**, 1095-1114.
- Mohapatra, N., Biswas, H. R. and Sawaisarje, G. K., 2011, "Spatial variability of daily rainfall over Northeast India during summer monsoon season", *Mausam*, **62**, 2, 215-228.
- World Meteorological Organization, 1966b, "Some methods in climatological analysis", *WMO Technical Note*, No. **81**, WMO No. 199, TP.103,