MAUSAM, 49, 1 (January 1998), 71-78

551.553.21:551.577(540)

Southwest monsoon rainfall in India: Part I-Spatial variability

A.B. MAZUMDAR

Meteorological Office, Pune- 411005, India

(Received 9 June 1994, Modified 20 September 1996)

सार— साप्ताहिक वर्षा की विसंगतियों के 10 वर्षों (1977-86) के आँकड़ों का उपयोग करके स्थानीय वर्षा के (एस-मोड) मुख्य घटकों के विश्लेषण (पी. सी. ए.) के परिणामों को इसमें प्रस्तुत किया गया है। अध्ययन की अवधि में वर्षा औसत से कम रही तथा वर्षा के मानक वितरण में मुख्यतः कम वर्षा वाले क्षेत्रों में विभिन्नता अधिक पाई गई है। उपखंडों के आंतरिक मौसमी संबंधों से पता चलता है कि देश के अधिकांश भागों में बड़े पैमाने पर मौसम प्रणाली पहले से प्रभावी थी। पहला मुख्य घटक (पी. सी.) माध्य पैटर्न से मिलता जुलता है। दूसरा मुख्य घटक (पी. सी.) सक्रिय मानसून की दशाओं से संबंधित है। तीसरा और चौथा मुख्य घटक क्रमशः मैडॅन-जुलिअन दोलन के उत्तर की ओर आरोह-अवरोह से और क्षीण मानसून परिस्थिति से संबंधित है।

विशिष्ट प्रबल और क्षीण मानसून गतिविधियों के साथ बारी-बारी से मुख्य घटकों की स्थानीय वर्षा के पैटर्न के साम्य से समग्र मानसून वर्षा के लिए देश के कुछ भागों (उत्तरपूर्वी, दक्षिणपूर्वी) से मानसून वर्षा के कम होने की स्थिति में महत्वपूर्ण योगदान मिलने का पता चलता है।

ABSTRACT. Results of Principal Component Analysis (PCA) in spatial mode (S- Mode) applied on a 10 years' (1977-86) data set of weekly rainfall anomalies are presented in this study. The rainfall activity has been below average in the period under study with high standard deviations mostly over low rainfall areas. Inter- subdivisional correlation values suggest predominance of broad- scale weather systems over most parts of the country. The first principal component (PC) has resemblance with the mean pattern. The second PC has been associated with active monsoon condition. The third and fourth PCs have been related to the northward progression of the Madden-Julian oscillation and the weak monsoon condition respectively.

Resemblance of spatial patterns of alternative PCs with typical strong and weak monsoon activities suggest significant contributions from certain parts of the country (northeast, southeast) towards overall rainfall activity during weak monsoon situation.

Key words - Spatial variability of southwest monsoon rainfall, Principal component analysis, Broad-scale systems of southwest monsoon.

1. Introduction

During the last two decades, eigen-techniques like Empirical Orthogonal Functions (EOF), Principal Component Analysis (PCA) and Factor Analysis (FA) have been widely used in meteorology with a view to detecting elementary features of physical phenomena embedded in huge sets of meteorological data over various regions of the globe. The PCA has been found to be more useful and elegant by many workers than other eigen techniques, as a result majority of multivariate analysis requiring use of eigen techniques has employed this type of analysis. In view of the significance of precipitation in a climatological/meteorological context, majority of studies involving application of the PCA have utilised rainfall data to study the spatio-temporal variability of the precipitation regime over various regions of the globe (e.g., Drosdowsky 1993, Sumner *et al.* 1993, etc.).

The large scale features of the rainfall variability over the Indian region have been studied previously by several workers. They employed EOF's to bring out the features of variability in annual, seasonal and monthly rainfall data. Studies have been made by various authors, *viz.*, Rakhecha and Mandal (1981) on monthly total rainfall of stations, 60 years, Bedi and Bindra (1980) on monthly total of stations, 60 years, Hastenrath and Rosen (1983) on annual total of subdivisions, 80 years, Rasmusson and Carpenter (1983) on seasonal total of subdivisions, 105 years, Shukla (1987) on annual total of subdivisions, 80 years, Prasad and Singh

TABLE 1

Distribution of number of weeks in each year after the establishment of the SW monsoon over the entire country

Year	First week	Last week	Total number of weeks
1977	30 June-6 July	22-28 Sept	13
1978	29 June-5 July	21-27 Sept	13
1979	12 July-18 July	13-19 Sept	10
1980	26 June-2 July	11-17 Sept	12
1981	25 June-1 July	10-16 Sept	12
1982	22 July-28 July	2-8 Sept	7
1983	30 June-6 July	8-14 Sept	11
1984	5 July-11 July	13-19 Sept	11
1985	11 July-17 July	5-11 Sept	9
1986	26 June-2 July	11-17 Sept	12

Total number during 1977-86 = 110 Weeks

(1988) on monthly total of sub-divisions, 83 years, Gregory (1989) on annual total of stations, 115 years, Iyengar (1991) on monthly total of stations in Karnataka, 80 years, etc., have studied the spatial variability of Indian rainfall data pertaining to various temporal and spatial dimensions, by employing EOFs. The applications of EOFs in medium range scale has been made only by Kripalani *et al.* (1991) who used pentad rainfall data of 53 blocks of $2.5^{\circ} \times 2.5^{\circ}$ latitude-longitude grids corresponding to 80 years of total southwest (SW) monsoon seasonal rainfall while Gadgil and Iyengar (1980) have used 50 years' mean pented rainfall data (total 73) pertaining to 53 stations in peninsular India. The PCA, which is slightly different from EOF, has not been applied on Indian rainfall data, though the difference between EOF and PCA is small and the structure of variability expected to be similar. Such analyses on weekly scale have also not been undertaken, so far. Considering the fact that the rainfall account is maintained on weekly intervals on routine operational basis in medium range scale, a PCA, as undertaken in this work on weekly rainfall data would be important.

In the present study S-mode PCA has been applied on weekly rainfall anomaly data corresponding to various meteorological sub-divisions during the SW monsoon period for the years 1977 to 1986, to bring out broad-scale features of spatial variability of rainfall during the monsoon period. In a subsequent study the PCA has been used in the temporal mode (T-mode) to explore association of rainfall distribution with the broad-scale synoptic systems of southwest (SW) monsoon. This study enables us to discern the spatial variability of precipitation climate in the medium range and its possible association with the broad- scale synoptic features during SW monsoon over India. Such concepts could be useful for designing statistical forecast models in the medium range for the monsoon rainfall.

The short rainfall series of 10 years used to study spatial variability of SW monsoon could be considered as a limitation of the work and the results of analysis have a bias towards deficient monsoon conditions. Nevertheless, the study brings out some broad-scale features which would contribute towards better understanding of the highly complex monsoon system over India.

2. Principal component model

Principal Component (PC) analysis is multivariate statistical technique that linearly transforms a collinear data array into a set of orthogonal areas or components. In addition to this orthogonality property, the components are ordered such that the first component explains the greatest proportion of the variance in the data and the second accounts for as much of the residual variance as possible while remaining orthogonal to the first. The remaining components explain less variance than preceding components. Thus a sizeable portion of the variance in the data can be explained by a relatively small number of PCs. It provides a powerful method of investigation of spatial and/or temporal behaviour of the meteorological variables. A number of text books and papers present the theory of PCA (Mather 1976, Richman 1986, etc.). Nevertheless, a short outline of the important features of PCA is provided in this paper.

If the data matrix is defined as :

$$\mathbf{Z} = \{Zij: i=1,........n, j = 1, ...p\} \qquad (1)$$

where 'i' and 'j' represent individuals or cases and variables, respectively, the 'p' PCs of $\mathbf{Z}$ are the columns of an $(n \times p)$ matrix $\mathbf{F}$ containing PC scores which are related to $\mathbf{Z}$ by:

$$\mathbf{F} = \mathbf{ZA} \qquad (2)$$

The matrix $\mathbf{A}$ contains unit length eigenvectors and provides PC loadings. The columns of $\mathbf{F}$ are obtained by linear combination of the original variables with the conditions of mutual orthogonality and maximum variance.

The correlation matrix $\mathbf{R}$ which is a symmetric matrix is defined by:

$$\mathbf{R} = \mathbf{Z}^{\mathrm{T}}\,\mathbf{Z} \qquad (3)$$

(where superscript T represents the transpose of the matrix)

The basic eigenstructure of the correlation matrix can be represented as $\mathbf{A}\ \mathbf{D}\ \mathbf{A}^{\mathrm{T}}$, where $\mathbf{D}$ is the diagonal matrix of eigenvalues of $\mathbf{R}$ arranged in descending order of magnitude ($d_1 > d_2 > ......> d_p$).

The variance of $\mathbf{F}$ can be found from :

$$\mathbf{F}^T\mathbf{F} = \mathbf{A}^T\,\mathbf{Z}^T\,\mathbf{Z}\,\mathbf{A}$$

$$=\mathbf{A}^T (\mathbf{A}\ \mathbf{D}\ \mathbf{A}^T)\ \mathbf{A} \quad (4)$$

$$=\mathbf{I}\ \mathbf{D}\ \mathbf{I} = \mathbf{D}$$

(Since **A** is orthonormal, $\mathbf{A}\ \mathbf{A}^T = \mathbf{A}^T\ \mathbf{A} = \mathbf{I}$ identity matrix)

which shows that variances of column vectors of **F** are numerically equal to the elements of matrix **D** of eigenvalues of **R**, the sum of variances is thus trace (**D**) or trace (**R**) and proportion of the total variance due to each PC is $\{d_i$ / trace $(\mathbf{D}) \times 100\}$ per cent.

The matrices **A** and **F** are rescaled according to:

$$\mathbf{L} = \mathbf{A}\mathbf{D}^{1/2} \quad (5)$$

$$\mathbf{F}^* = \mathbf{F}\ \mathbf{D}^{-1/2} \quad (6)$$

so that, elements of matrix **L** provide correlation between PCs and the original variables and matrix $\mathbf{F}^*$ contains standardised PC scores having zero means and unit variances. Such rescaling allows for easier interpretation of spatial and temporal coefficients and provides a slightly more elegant mathematical model compared to EOF.

The eigentechniques are commonly referred to as EOF or PCA sometimes in a confusing manner. The EOFs are the unit length eigenvectors and corresponding eigenvalues obtained by the decomposition of a dispersion matrix **R** as **RA=AD**, where **A** is the orthogonal matrix of eigenvectors and **D** is the diagonal matrix of eigenvalues. The corresponding time series **F** is given by fundamental EOF equation $\mathbf{Z}=\mathbf{F}\mathbf{A}^T$, where **Z** is the standardised input data matrix and **F** is called empirical orthogonal functions of time due to their temporal orthogonality.

A PCA can be specified in atleast six basic operational modes, depending on which parameters are chosen as variables, individuals and fixed entries, like, S-mode, T-mode, R-mode, etc. (Richman 1986). In the S-mode, stations are selected as variables, time as individuals and parameters as fixed entries, whereas, in T-mode, times are selected as variables, stations as individuals and parameters as fixed entries and in R-mode, parameters are taken as variables, stations as individuals and times as fixed entries. In the present S-mode PCA, the meteorological subdivisions have been selected as variables and the weekly rainfall departures as individuals.

3. Data utilised

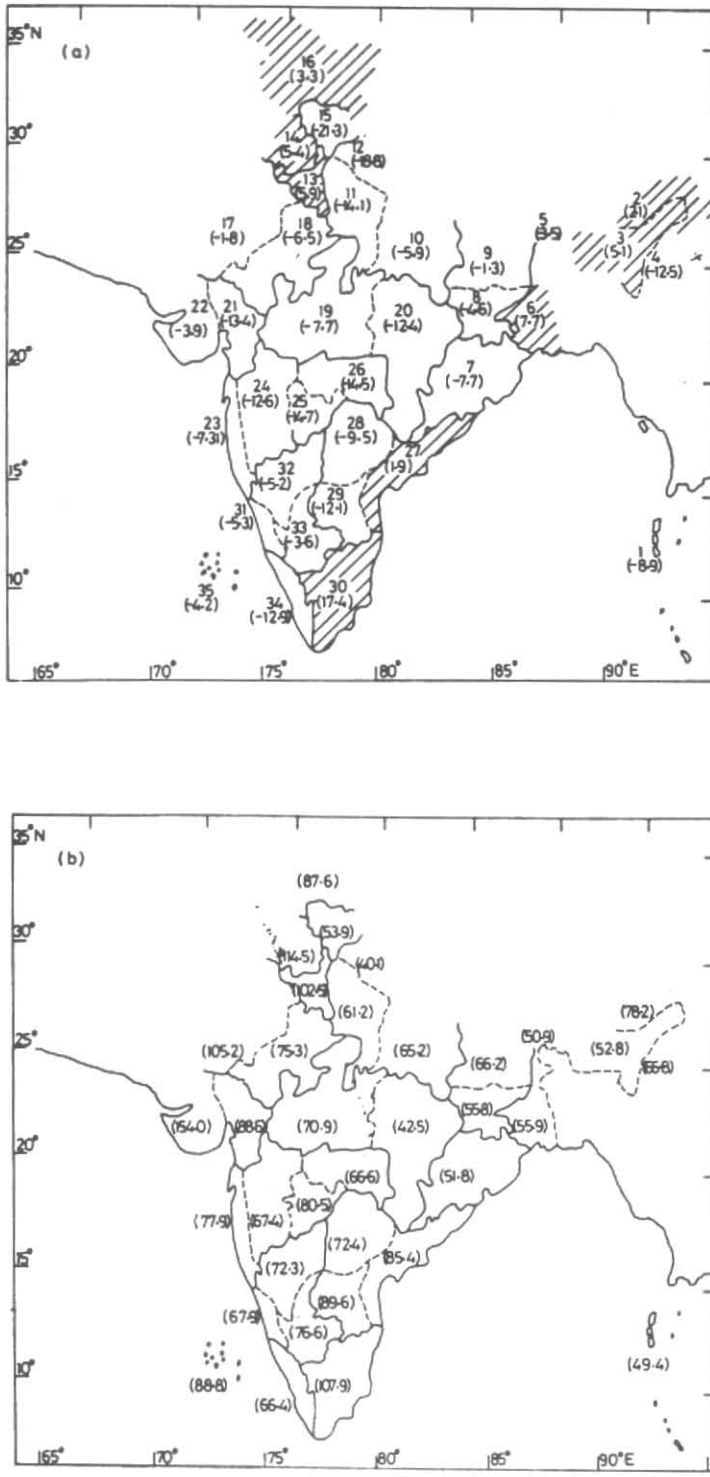
Weekly rainfall departures from normal (WRDN) for all the 35 meteorological sub-divisions from 1977 to 1986 during the southwest monsoon period, have been utilised for the study. These data have been collected from the Weekly Weather Reports published by the India Meteorological Department. The WRDN series for each subdivision have been generated considering only the period during which the entire country was under the grip of the monsoon. The first WRDN value in any year corresponds to the week after the establishment of SW monsoon over the entire country (the week when onset of SW monsoon has taken place over West Rajasthan) and the last WRDN value in the same year corresponds to the week of the first withdrawal from any part of the country (mostly from West Rajasthan). The total number of such weeks occurring during the ten years under study has been shown in Table 1.

In this way 110 WRDN data have been selected for each subdivision to form a 110(WRDN) × 35 (meteorological subdivision) data matrix on which S-mode PCA has been applied.

4. Result and discussion

The means and standard deviations of WRDN for all the 35 subdivisions have been shown in Figs. 1(a & b). In Fig. 1(a) regions having positive values of mean WRDN have been shaded. It can be seen that during the 10-year period under study the rainfall has been less than normal in most parts of the country. Positive values of mean WRDN have been distributed over the northern, north-eastern and south-eastern parts of India. Such distributions have always been associated with the weak monsoon activity over the Indian region. The standard deviations during the 10-year period have been maximum in the north western and south-eastern parts of the country (Fig. 1b). These regions get low rainfall during the monsoon period.

The correlation matrix reveals some high values of correlation between adjoining sub-divisions. Inter-subdivisional correlation coefficient values of more than 0.5 have been indicated with arrows in Fig. 2. Earlier studies by De and Joshi (1993) have shown similar pattern though the authors have used a simpler statistical approach. Large negative correlation coefficients (< -0.5) have not been noticed in the correlation matrix. High correlation coefficients have been found mostly in the adjoining sub-divisions over peninsula, northwestern India and over the groups of sub-division lying close to the normal position of the monsoon trough. Such distribution could be because of simultaneous occurrence of precipitation in association with large scale synoptic systems, like, monsoon trough, mid-tropospheric cyclones, off-shore troughs and easterly systems. Such high correlations have been absent in the sub-divisions of northeast India, Jammu & Kashmir and Punjab, perhaps, because rainfall in these sub- divisions are controlled by orography, mid-latitude systems and preponderance of comparatively small scale weather systems.



Figs. 1(a&b). (a) Average values of WRDN. Areas with positive values of WRDN are shaded and mean values are shown in parenthesis and (b) standard deviation of WRDN

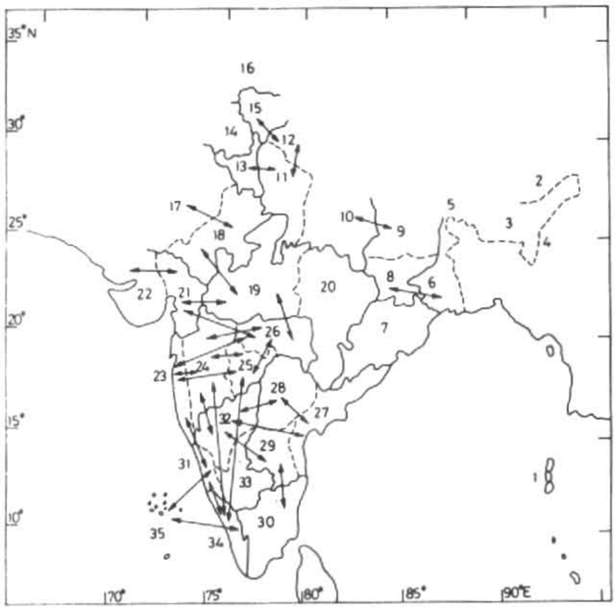


Fig.2. Values of correlation coefficients more than 0.5 indicated by arrows

TABLE 2

Eigenvalues and cumulative proportion of total variance accounted by 15 PCs

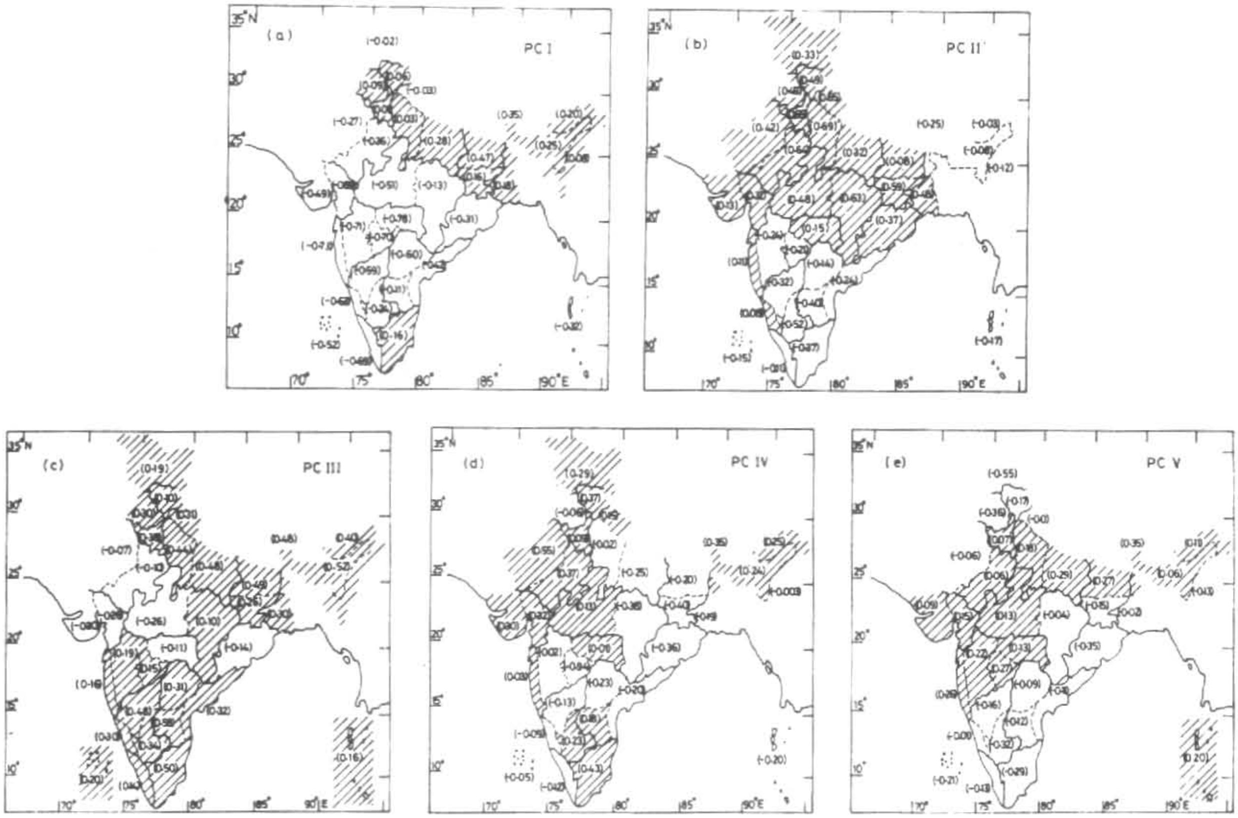
PC	Eigen value	Cumulative% of total variance
1	6.25	17.9
2	5.09	32.4
3	3.54	42.5
4	2.15	48.7
5	1.72	53.6
6	1.66	58.3
7	1.36	62.2
8	1.29	65.9
9	1.08	69.0
10	1.00	71.8
11	0.96	74.6
12	0.84	77.0
13	0.76	79.1
14	0.74	81.3
15	0.69	83.2

Eigenvalues and cumulative proportion of total variance accounted by first 15 PCs have been tabulated in Table 2.

Cumulative variance explained (VE) by the first fifteen PCs have been about 83% of the WRDN data for the ten-year period under study. The first five PCs (cumulative VE about 54%) have been selected for detailed study. The loadings of these PCs represent correlations between the PCs and rainfall activities in each meteorological sub-division. These have been plotted in Figs. 3(a -e).

The PC I (VE about 18%) has positive loadings over the northern, northeastern and southeastern parts of the country. High negative loading have occurred over the western parts of peninsular India (Fig. 3a). This pattern shows that there exists an out-of-phase variation in rainfall activities between these regions. As the overall activity during the period under study has been less than normal, the PC I has brought out a typical distribution reflecting below average performance of monsoon and, as such, the distribution of PC I has resemblance with the mean pattern (Fig. 1a).

The VE by the PC II (14%) and PC I are comparable which indicate that the second component has also been equally dominant. The PC II has positive loadings over the central, northern, northwestern parts of the country and over the northern parts of west coast (Fig. 3b). Active monsoon with the trough in a relatively southern location, Mid-Tropospheric Cyclones (MTC) and low level trough off the west coast could be the broad-scale features responsible for such pattern. In addition, cyclonic systems originating from the Bay of Bengal could have had westerly or west-northwesterly courses. The pattern corresponds to the active monsoon rainfall distribution. Negative loadings to the east of the Western Ghats and over the northeastern parts of the country



Figs. 3 (a-e). Values of PC. I, II, III, IV & V loadings

are well known features normally associated with the active monsoon conditions.

The PC III (VE about 10%) has positive loadings over the northern and the southern parts and negative values over the central parts (Fig. 3c). Such north-south patterns could be due to northward advancements of broad-scale cloud patterns termed as the Maximum Cloud Zone (MCZ) during the monsoon as described by Sikka and Gadgil (1980), arising due to the Madden-Julian oscillation, the effect of which have been brought out in a number of studies in the Indian region during the monsoon period (e.g., Ramasastry *et al.*, 1986, De *et al.* 1988, etc).

The VE by PC IV has been 6% which has positive loadings over the northern, northwestern, northeastern and southeastern parts of peninsula (Fig. 3d). Such distribution could be due to the movement of westerly systems in the north and easterly system in the south, the pattern has some sort of resemblances with weak monsoon condition. The PC V (VE about 5%) is having positive loadings over the northern, northeastern and western parts (Fig. 3e). Such distribution is difficult to interpret. However, similar patterns are sometimes encountered when cyclonic systems take a rather northerly course. It may be pointed out that the sampling errors become large enough to result in mixing of higher eigen components, making interpretation of higher modes of variation very difficult (North *et al.* 1982).

The spatial patterns obtained in the earlier studies of the southwest monsoon pentad rainfall using EOFs by Kripalani *et al.* (1991) can only be compared with the present study. Such comparison shows some similarities with those in the present study, even though the data sets are different. The major differences in the patterns have been as follows:

(*i*) The VE by PC I (18%) has been slightly smaller than the VE by first EOF (20%) in the above study.

(*ii*) The pattern of distribution of PC I and EOF I are almost opposite. This could be attributed to the below average performance of monsoon during the period of study and probably because of the utilisation of shorter rainfall series in the present study. The signatures of below average performance of monsoon have been absent in the spatial variability in the above mentioned study.

(*iii*) The PC II of this study resembles EOF I of the above study. The variance explained by PC II (14%) is higher than the variance explained by EOF II(10%) in the above study.

(*iv*) The PC IV of this study resembles EOF III.

(*v*) North-south pattern of loading appears in EOF II, whereas, the same is seen in PC III.

The above comparison shows out of phase spatial patterns between the two analysis, which could be attributed to the difference in the data sets and length of the rainfall series used in these two studies. In both the studies alternate occurrence of patterns corresponding to strong and weak monsoon conditions have been quite evident, suggesting that even the weak monsoon conditions make significant contribution towards overall rainfall variance during the monsoon by increased rainfall activities in some parts of the country (e.g., northeastern and southeastern parts). Such alternate patterns have also been noticed in the EOFs of other studies employing various longer scale (monthly, seasonal, yearly) rainfall data (*e.g.*, Bedi & Bindra 1980).

The ten years period under study consisted on one excess, 3 deficient/drought and 6 normal monsoon years and as such, has a bias towards below normal performance of the monsoon. Therefore, the inference drawn based on the small sample of data may not be true representative of actual spatial variability of monsoon rainfall. However, the similarities in dominant patterns of PCs between the present study and other earlier studies employing longer periods of data suggest the presence of remarkable stability and consistency of southwest monsoon rainfall patterns.

5. Conclusion

(*i*) A special data set of WRDN has been prepared considering rainfall activity pertaining exclusively to the period when monsoon was fully established over the entire country in each of the years during 1977-86 to study spatial variability over India during the SW monsoon. S-mode PCA has been applied to bring out the structure of rainfall variability. Possible association of the distribution of such a spatial pattern with broad-scale synoptic weather patterns has been suggested.

(*ii*) The mean pattern of WRDN has shown that the overall rainfall activity during the peak monsoon time has been below average. The standard deviation has been high over the regions having low rainfall. Most of the adjoining meteorological sub-divisions, except the northeastern subdivisions Jammu & Kashmir and Punjab have been found to have a correlation coefficient of more than 0.5 which could be attributed to the effects of broad scale systems in controlling rainfall activity in these regions and absence of such systems in the north-eastern parts and perhaps, the effect of mid-latitude systems in the northern parts of the country.

(*iii*) The distribution of loading of the PC I has resemblance with the mean pattern and which in turn corresponds to below average performance of monsoon. The PC II has been associated with the active monsoon condition with active trough system, MTC, off-shore vortices and trough and westerly or west-northwesterly tracks of cyclonic systems originating from the Bay of Bengal.

(*iv*) Spatial patterns of PC III have been suggested to be due to the northward progression of Madden-Julian oscillation and the spatial variability associated with PC IV has resemblance with the weak monsoon condition. The comparison of the spatial patterns of PCs with the patterns of EOFs in a similar study by Kripalani *et al.* (1991) reveals out of phase spatial distribution between the two, perhaps, due to the difference in the length of rainfall series. But, the occurrence of alternate patterns corresponding to strong and weak monsoons in both the studies and in other studies shows the importance of increased rainfall in some parts of the country (northeast, southeast, etc.) during the weak phase making significant contribution towards overall rainfall activity.

References

Bedi, H.S. and Bindra, M.M.S., 1980, "Principal components of monsoon rainfall", *Tellus*, **32**, 296-298.

De, U.S., Chatterjee, S.N. and Sinha Ray, K.C., 1988, "Low frequency modes in summer monsoon circulation over India", *Mausam*, **39**, 167-178.

De, U.S. and Joshi, K.S., 1993, "A climatological-cum-statistical analysis of monthly rainfall in meteorological subdivisions of India", Presented at the National Symp, on Meteorol. and Natl. Development, Indian Met. Soc., New Delhi, 17-19 March 1993.

Drosdowsky, W., 1993, "An analysis of Australian seasonal rainfall anomalies: 1950-1987 I : spatial patterns", *Int. J. Climatol.*, **13**, 1-30.

Gadgil, S. and Iyengar, R.N., 1980 "Cluster analysis of rainfall stations of the Indian Peninsula", *Quart. J. Roy. Meteor. Soc.*, **106**, 873-886.

Gregory, S., 1989, "Macro-regional definition and characteristics of Indian summer monsoon rainfall", 1971-1985', *Int. J. Climatol.*, **9**, 465-484.

Hastenrath, S. and Rosen, A., 1983, "Patterns of Indian monsoon rainfall anomalies", *Tellus A*, **35**, 324-331.

Iyengar, R.N., 1991, "Application of principal component analysis to understand variability of rainfall", *Proc. of Indian Acad. of Sci. (Earth and Planetary Sci.)*, 100-2, 105-126.

Joshi, K.S., Vaidya, D.V. and De, U.S., 1981, "An objective aid to forecasting weekly rainfall over India", *Vayu Mandal*, **11**, 19-22.

Kripalani, R.H., Singh, S.V. and Arkin, P.A., 1991, "Large-scale features of rainfall and outgoing longwave radiation over Indian and adjoining regions", *Contr. to Atoms. Phys.*, **64**, 159-168.

Mather, P.M., 1976, "Computational methods of multivariate analysis in Physical Geography", John Wiley & Sons, N.Y., pp. 524.

North, G.E., Bell, T.A., Cahalan, R.F. and Moeng, F.J., 1982, "Sampling errors in the estimation of empirical orthogonal functions", *Mon. Wea. Rev.*, **10**, 699-706.

Prasad, K.D. and Singh, S.V., 1988, "Large-scale features of the Indian summer monsoon rainfall and their association with some oceanic and atmospheric variables", *Adv. Atmos. Sci.*, **5**, 499-513.

Rakhecha, P.R. and Mandal, B.N., 1981, "The use of empirical orthogonal function for rainfall estimates", *Monsoon Dynamics*, J. Lighthill and R. Pierce (eds.), Cambridge Press, 627-638.

Ramasastry, A.A., De, U.S., Vaidya, D.V. and Sundari, G., 1986, "Forty-day mode and medium range forecasting", *Mausam*, **37**, 305- 312.

Rasmusson, E.M. and Carpenter, T.H., 1983, "The relationship between eastern Equatorial Pacific sea surface temperatures and rainfall over India and Sri Lanka", *Mon. Wea. Rev.*, **111**, 517-528.

Richman, M.B., 1986, "Rotation of principal components", *J. Climatol.*, **6**, 293-335.

Shukla, J., 1987, "Interannual variability of monsoon", *Monsoons*, eds., J.S. Fein and P.L. Stephens, Wiley, N.Y., 399-464.

Sikka, D.R. and Gadgil, S., 1980, "On the maximum cloud zone and the ITCZ over Indian longitudes during the southwest monsoon", *Mon.Wea.Rev.*, **108**, 1840-1853.

Sumner, G., Ramis, C. and Guijarro, J.A., 1993, "The spatial organisation of daily rainfall over Mallorca, Spain", *Int. J. Climatol.*, **13**, 89-109.