Reprinted from MONTHLY WEATHER REVIEW, Vol. 114, No. 8, August 1986
American Meteorological Society

Application of Extended Empirical Orthogonal Function Analysis to Interrelationships and Sequential Evolution of Monsoon Fields

S. V. SINGH AND R. H. KRIPALANI

Indian Institute of Tropical Meteorology, Pune: 411 008, India

(Manuscript received 27 July 1985, in final form 6 March 1986)

ABSTRACT

Extended empirical orthogonal function (EEOF) analysis has been employed to study linear relationships among the mean sea level pressure, 700 mb height and rainfall over India, and their low-frequency sequential evolution during the peak summer monsoon months. The interrelationships between these fields are strongest over central India and, while the rainfall activity is colocated with the corresponding changes in the 700 mb heights, it is displaced southward with respect to the pressure changes. The first two EEOF's of all the three fields (averaged over 5 or 7 days) show that the dominant low-frequency sequential evolution is associated with north and northeastward movement of the anomaly centers with a recurrence period of about 40 days. In addition, the presence of a westward moving wave in sea level pressure anomalies located roughly near 15°N latitude is revealed by the third EEOF.

1. Introduction

A large variety of techniques is employed to uncover the significant climatic structures from the large data sets of atmospheric–ocean fields. The techniques followed for the study of only temporal or spatial features are rather well developed. The spatio–temporal variation of the fields is usually studied by fitting some mathematical (or empirical) functions (e.g., Fourier series, polynomials, empirical orthogonal functions) to the individual spatial fields and analyzing their coefficients for temporal behavior. However, the techniques for studying the sequential evolution of the fields are less well developed. The procedure of standard EOF analysis has been employed by several workers, e.g., Kutzbach (1967) and Prohaska (1976) to study the linear relationships between two meteorological fields. Weare and Nasstrom (1982) extended the usual EOF procedure for studying the sequential evolution of equatorial Pacific SST fields. They named this procedure involving the concurrent and lag correlations of the same meteorological field as extended empirical orthogonal function (EEOF) analysis (see also Lau and Chan, 1985).

In this paper, we shall present some results of the application of the EEOF analysis in studying the interrelationships between the daily circulation and the rainfall fields and their low-frequency sequential evolution during the peak monsoon months (July–August) over India. In section 2 we describe the data used and in section 3 a brief outline of the principles of EEOF analysis, including its advantages and disadvantages, is given. The results are presented in sections 4 and 5, followed by a brief discussion in section 6.

2. Data

The daily 700 mb height data for a 98-point 2.5° lat/, 5° long diamond grid (Fig. 1a) (for details of data see Singh et al., 1978) and that for mean sea level pressure anomaly (MSLPA) for 72 2.5° lat, 2.5° long blocks (Fig. 1b) were available for 21 (1958–78) and 30 (1946–75) peak monsoon periods, respectively. The daily rainfall data of 32 meteorological subdivisions (Fig. 1c), needed in section 4, are prepared from the data of 220 stations for eight monsoon seasons (1966–73). Weekly percentage departures of rainfall for the same 32 subdivisions are obtained for 20 peak monsoon periods (1963–82) from the weekly weather reports published by the India Meteorological Department. Each peak monsoon period consists of 62 calendar days, 1 July to 31 August.

3. Principles of extended empirical orthogonal function (EEOF) analysis

As mentioned previously, the EOF analysis can be extended to explore the linear relationships between two or more sets of variables (fields) or lag relationships between the same set of variables. The data matrix of the two sets of variables can be joined in the manner shown in Fig. 2(i). The rows represent the number of observations and the columns the total number of variables. When we are interested in analyzing sequential evolution over one lag only and if the number of variables are p, then the total columns (variables) in the data matrix become $2p$, i.e. $(p + p)$. When we are interested in studying sequential evolution over many lags then the total columns are equal to the number of

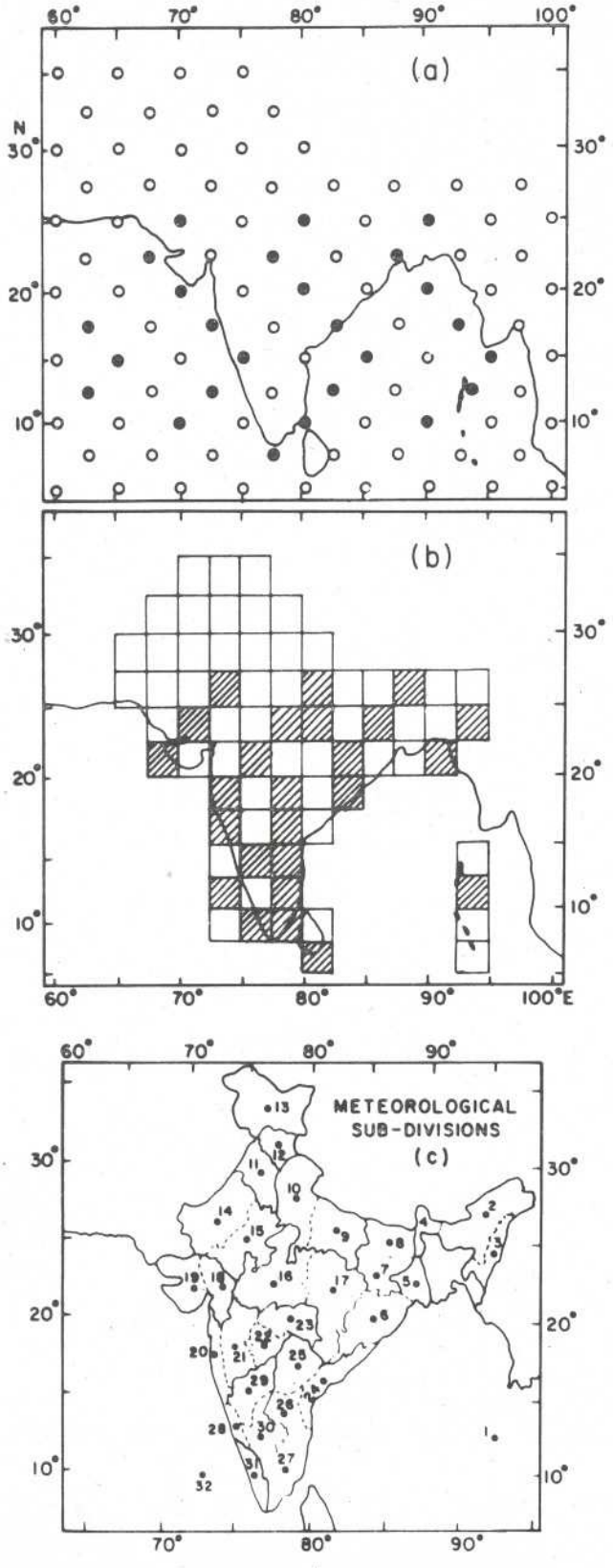

FIG. 1. (a) The 98 grid points of the 700 mb height fields. Dots represent the 25 points considered for the EEOF analysis. (b) The 72 blocks for the MSLPA fields. Shaded blocks are considered for the EEOF analysis. (c) The 32-meteorological subdivisions considered for representing the rainfall fields.

variables (p) times the number of the lags considered. The correlation matrix then is derived in the usual manner.

The correlation matrix for the case of only one lag is shown in Fig. 2(ii). The dimension of the matrix is $2p \times 2p$. The top left quadrant contains the intercorrelation matrix computed by using only part (a), i.e., present observations of the data matrix. [See Fig. 2(i).] Likewise the bottom right quadrant contains the intercorrelation matrix computed by using only part (b), i.e., lagged observations of the data matrix. For a continuous time sequence of length N, the first $N-1$ observations [first to $(N-1)$] comprise the present observations and the last $N-1$ observations (second to Nth) the lagged observations, in the case of lag 1 being

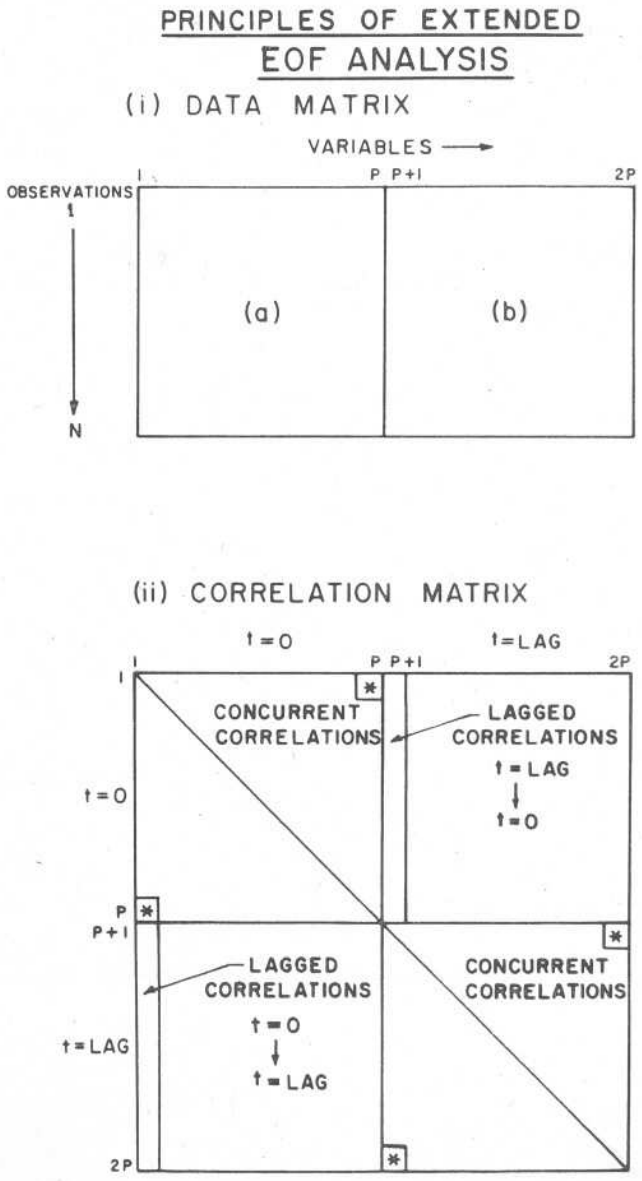

FIG. 2. The principles of EEOF analysis: (i) data matrix; (ii) the requisite correlation matrix for one lag. (See text for details.)

discussed here. The remaining two quadrants contain lag correlations and hence the information about the sequential evolution of the fields. The matrix is asymptotically symmetric and the EOFs can be determined by any standard method—the Jacobi method in the present case. The first half of the elements of the resultant eigenvector corresponds to the observations belonging to the present period (lag 0) and the second half correspond to the lagged period.

Two shortcomings of the EEOF analysis may be noted. First, it is seen that the concurrent intercorrelations and the lag correlations are repeated four times compared to two times in a symmetric correlation matrix used in the ordinary EOF analysis. This fact is indicated in Fig. 2(ii) by 4 asterisks, all corresponding to the concurrent correlation between the first and the pth variable. This redundancy of correlation will tend to dilute the evolutionary signal which we are seeking. The second problem is that the dimension of the correlation matrix becomes large. Third, it is obvious that the elements of the normalized eigenvector will be of smaller absolute magnitude for the EEOF analysis as compared to the standard EOF analysis because the sums of squares of these weights has to be equal to one in both the cases. For more details see Weare and Nasstrom (1982).

4. Interrelationships among the circulation and rainfall fields

We note from Figs. 1a, b that the grids for the 700 mb heights and MSLPA fields are different. In EOF analysis it is desirable to give equal weighting to all fields by considering similar grids in extent and resolution. Besides, a straightforward EOF analysis of these fields could not be possible due to the limited memory of the computer. Hence we determined daily averages of these fields over the 32 subdivisions. Simple averages of the value of the fields for the grid points/blocks overlying a particular subdivision are taken to represent the subdivisional averages of these fields. This averaging has also enabled us to study the interrelationship between the circulation and rainfall fields, the rainfall fields being originally available only for the subdivisions. For studying the interrelationships, the data of only eight peak monsoon periods (1966–73; $8 \times 62 = 496$ days), for which the rainfall data were available, are considered.

For studying the relationship between the MSLPA and 700 mb height fields we joined the matrices of MSLPA and 700 mb height, each of dimension (496×32) to form a joint data matrix of dimension (496×64) and obtained a 64×64 correlation matrix. The first 32 elements of each of the eigenvectors of this matrix correspond to the 32 points of MSLPA and the remaining 32 correspond to the 700 mb height. The ratio of observations to the variables is about 8 and the eigenvectors are expected to be stable. The variance explained by the first eigenvector (see Fig. 3) of this matrix is 48%. From the figure it is seen that the two fields show maximum intercorrelation over central India and this relationship decreases both in the south and in the north direction. Similar results were obtained by Singh et al. (1979) for 5-day fields by computing simple intercorrelations. This represents higher barotropicity of the atmosphere over central India. The normal position of the semipermanent monsoon trough is about 20°N at the 700 mb level and generally more intense and coherent fluctuations in the circulation fields take place around this latitude. Better quality of data over this region may be another contributing factor for these higher loadings over the central region.

In Fig. 4 are presented the first eigenvectors of the EOF analysis conducted by combining MSLPA and rainfall fields (variance explained = 30%), and 700 mb height and rainfall (variance explained = 28%) fields. Again, better interrelationships are obtained for the central parts of the country. The area of maximum loadings for rainfall fields is displaced relatively southward as compared to the area of maximum loadings for the MSLPA fields. Such displacement is much smaller for the 700 mb height fields. This result is consistent with the climatological fact that the area of large-scale rainfall coincides with the areas of low 700 mb heights and are comparatively located southwards of the areas of low pressure at the surface. The reason for the better spatial coincidence of the changes in the 700 mb height fields and rainfall is that the monsoon trough slopes southward with height and that the 700 mb level is the nondivergent level in this monsoon region at which maximum vertical velocities are obtained.

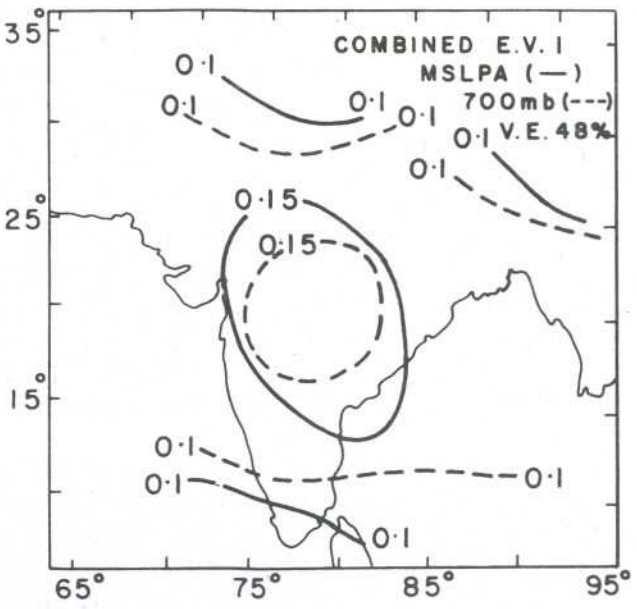

FIG. 3. The first combined eigenvector of the daily MSLPA and the 700 mb height fields.

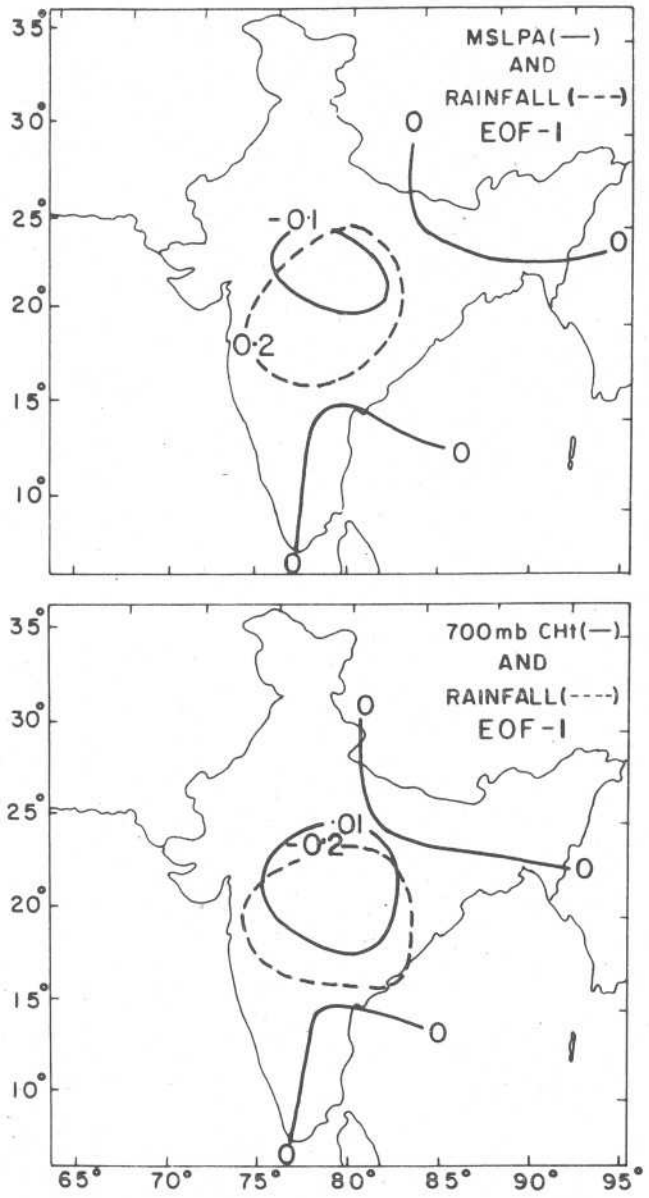

FIG. 4. (i) The first combined eigenvector of MSLPA and rainfall fields. (Variance explained is 30%.) (ii) The first combined eigenvector of the 700 mb height and rainfall fields. (Variance explained is 28%.)

5. Low frequency sequential evolution of circulation and rainfall fields

For studying the low-frequency evolution of the MSLPA and 700 mb height fields we consider 5-day (pentad) averages of these fields. The rainfall data were already available in the form of weekly percentage departures from normal. Sequences of four consecutive pentads or weeks are considered for studying the sequential evolution of the circulation and rainfall fields.

a. 700 mb heights

A total of 252 nonoverlapping, pentad, 700 mb height charts were prepared from the data of daily heights available for 21 peak monsoon periods. From these the pentad anomaly charts were obtained by subtracting the long-term (based on 21 yr) average for the corresponding pentad. Initially the EEOF analysis was conducted by using all the 98 grid points and by joining the data matrices corresponding to the current pentads and only one more data matrix corresponding to the pentad lagging the current pentad by one, two or three pentads, since it was not possible to accommodate in the computer memory the large correlation matrices resulting from joining of the data matrices corresponding to all the three lags together. This gave us a reasonable idea about the geographical locations of the grid points showing high loadings in the EEOF analysis. Later, for conducting the final EEOF analysis for understanding the principal modes in the sequential evolution of the fields over a continuous sequence of four pentads, 25 grid points were selected. Four data matrices corresponding to the lags of zero (present pentad), one, two, and three pentads were joined together and the EEOF analysis conducted.

The first two functions, which explained respectively, 31.3 and 18.5% of the variance and showed some systematic evolution are presented in Fig. 5. In the first EEOF the centers of loadings are located near 11°, 15°, 19° and 22.5° lat on the charts corresponding to the sequence of four consecutive pentads, showing northward movement of the principal anomaly fields at the rate of 0.8° lat per day. The second EEOF also shows northward movement of the strongest center of loadings located near 17.5°N, 92.5°E on the chart corresponding to the current pentad. In addition the centers of loadings exhibit a westward (eastward) component in their movement when lying south (north) of 20°N lat.

b. MSLPA fields

Pentad anomaly charts were obtained after simple averaging of the daily data. A total of 360 (12 × 30) pentad charts and 270 (9 × 30) sequences of four consecutive pentads were obtained from the 30 yr of data. Due to the limitations of the computer memory, only 25 blocks were selected for EEOF analysis. The first three EEOF's which explained 24.3, 21.1 and 12.3% of the variance and showed some systematic evolution of pressure fields are shown in Fig. 6. The first EEOF first shows northward and then northeastward movement of the centers of the loadings; the center of loadings is located near 13°, 17°, 23° and 25°N latitude on the charts corresponding to the sequences of four successive pentads. The northward progression of pressure anomalies thus occurs at the rate of about 0.8° lat day^{-1}. It may be noted that on all charts of the first EEOF, only one (negative in this case) sign of anomaly prevails. The second EEOF also shows north and northeastward movement of the center of positive

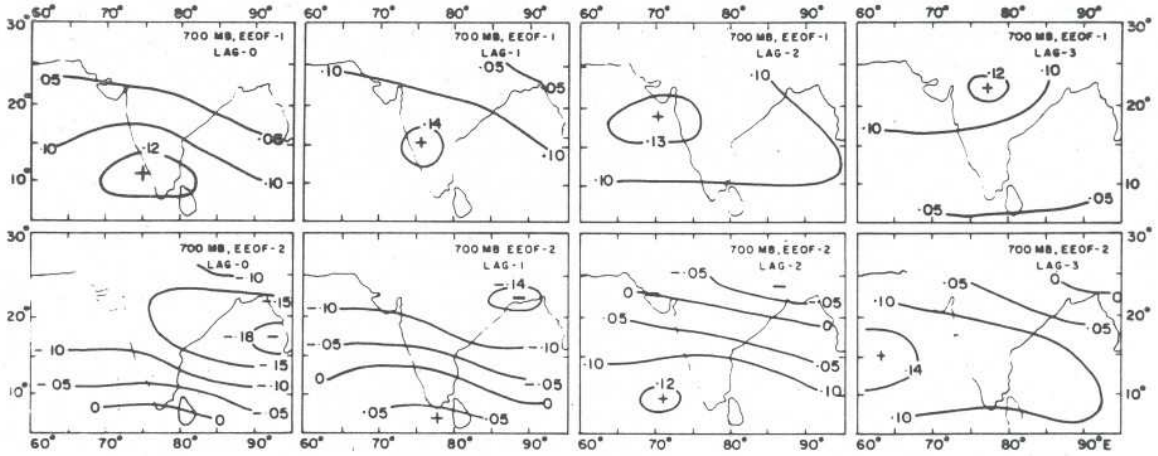

FIG. 5. The first two functions of the EEOF analysis of the 700 mb height fields. The variances explained by these two functions are 31.3 and 18.5%, respectively.

loadings located near 22.5°N on the chart corresponding to the present pentad. As this center of positive loadings moves northward, a center of negative loadings appears on the southern peninsula which also moves northward. The third EEOF shows a westward propagating wave centered near 15°N lat and nearly completing its cycle in 15 days. The role of creating active and weak monsoon conditions over the Indian region by this westward propagating wave of 15-day period has been discussed by Krishnamurti et al. (1985).

c. *Rainfall fields*

The EEOF analysis of weekly rainfall fields is conducted by considering three lags together. Thus each observational sequence consists of 4 weeks; the first one corresponds to the present week and the second, third and fourth correspond to lags of 1, 2 and 3 weeks, respectively. A total of 126 such observational sequences are obtained from the 20 peak monsoon periods. The first two EEOF's which showed some systematic evolution are shown in Fig. 7. In the first EEOF

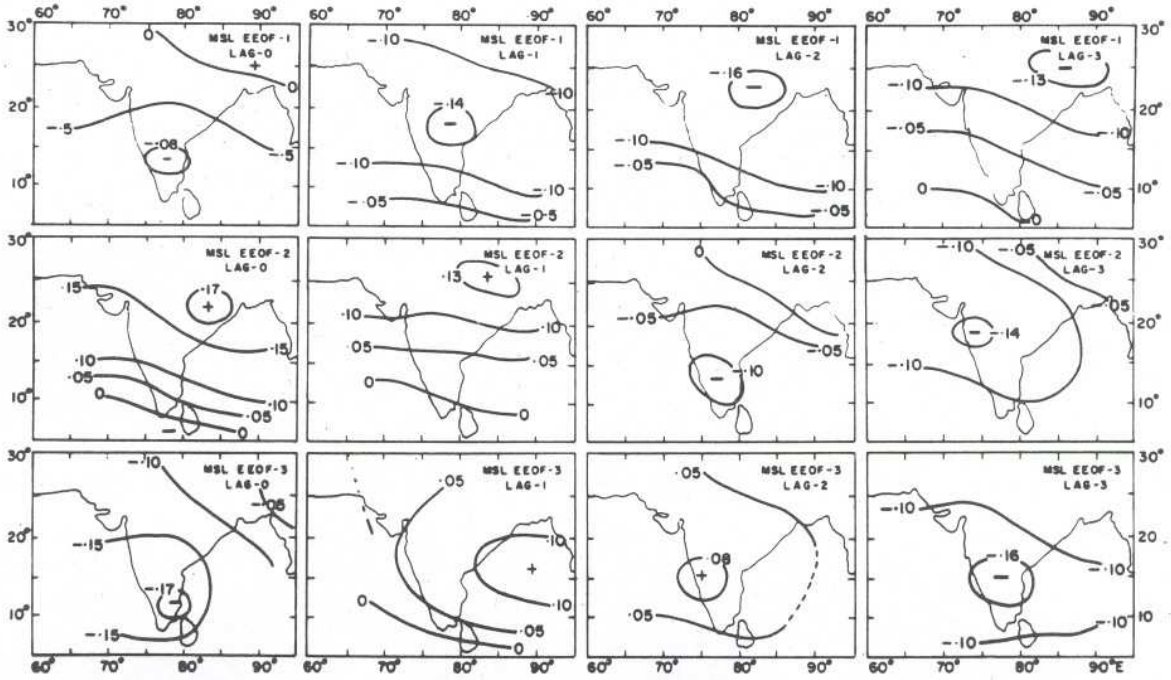

FIG. 6. The first three functions of the EEOF analysis of the MSLPA fields explaining 24.3, 21.1 and 12.3% of the variance respectively.

(variance explained = 11.4%), the dominant centers of positive loadings are centered near 18°, 22°, 23° and 27°N latitude during the 4 successive weeks. The centers of loadings move northward until about 22°N latitude and northeastward thereafter, at the rate of about 0.5° lat. day^{-1}. The chart corresponding to the present week is nearly in an opposite phase to that obtained after 3 weeks showing a quasi-periodicity of about 40 days. We notice that as a center of loading moves northward from the central parts of the country, another center of opposite anomaly appears over the southeast peninsula, which also moves northward in subsequent weeks. On all the charts, two or three centers of loadings with alternating signs in the north–south direction are present. The centers of loadings lying near the tip of the peninsula or in the extreme northeast region are not identified properly on many occasions when the centers lay further away from the country. Generally the distance between the centers of loadings of opposite sign is 6° (9°) over the areas of country lying towards the north (south) of 20°N latitude. The second EEOF (variance explained = 10.3%) shows different phases of the same evolutionary feature as was revealed by the first EEOF, i.e., the north and northeastward movement of centers of loadings. Again, the chart corresponding to the present week is nearly in opposite phase to the chart for 3 weeks later, suggesting a quasi-periodicity of about 40 days in the rainfall fields.

It may be noted that the principal components corresponding to the identified phases (Fig. 7) of the 40-day oscillation were used by us for developing multiple regression equations for predicting weekly monsoon rainfall over India. On verification of an independent sample of about 50 cases, these equations showed higher skill, ranging from 0.1 to 0.2 in terms of Heidke skill score, for the subdivisions lying over the parts of the country (central parts) over which the EEOF analysis showed higher loadings (Kripalani et al., 1986).

6. Discussion

In section 4 we presented some examples of the application of the EOF analysis for understanding the interrelationships among the circulation and the rainfall fields. The first EOF brought out the regions of high intercorrelations and the relative disposition of the major circulation and rainfall anomalies. Some examples of the application of the EEOF analysis related to the low-frequency sequential evolution of the circulation and rainfall fields were presented in section 5. The EEOF analysis not only enabled us to identify the principal modes of the sequential evolution of the fields, but also quantified the variances contributed by the identified modes. The spatial display of the loadings also enabled us to know the specific geographic locations over which the more intense changes related to the identified evolutionary process (40-day oscillation in the present case) take place.

Generally, the first two EEOF's were found to represent the systematic evolution of these fields and both functions represented the south to north propagation of the rainfall anomalies. The two EEOF's are out of phase by 90° to satisfy the constraints of orthogonality. Lorenc (1984) also found two orthogonal functions of the global velocity potential fields to represent the same (only one) feature. Both these functions explain nearly equal variances for MSLPA and rainfall fields and

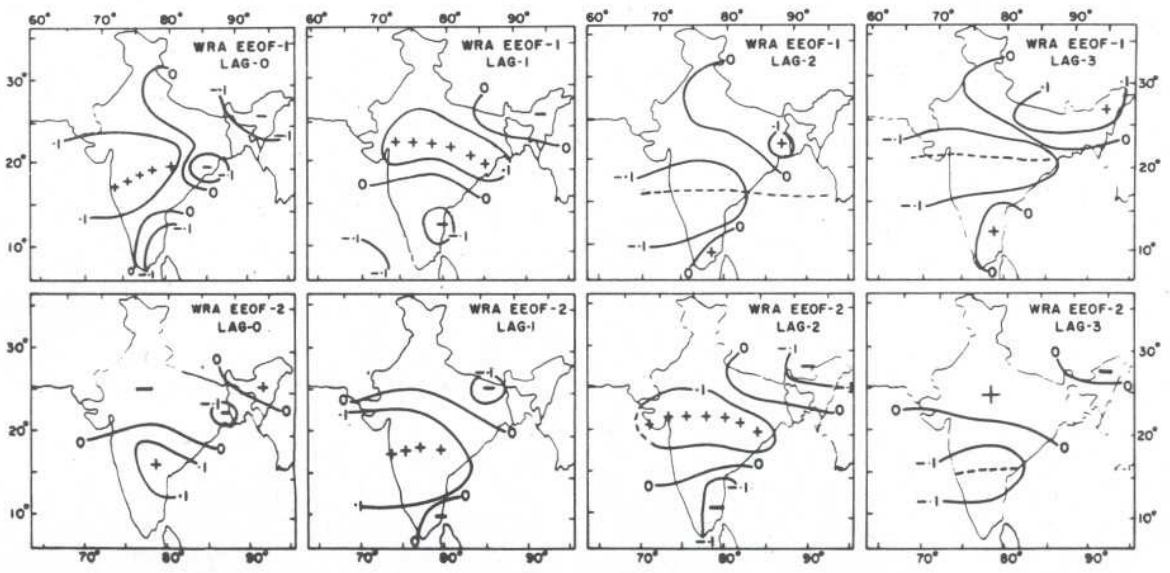

FIG. 7. The first two functions of the EEOF analysis of the weekly rainfall anomaly (WRA) fields explaining 11.4 and 10.3% of the variance, respectively.

comparable ones for the 700 mb height fields. This again confirms that the first two functions represent the two phases in quadrature of the same phenomena and the small difference in the variance explained of the two functions is due to the mathematical constraint of analysis. For both the circulation fields the two EEOF's are distinctly separate and for the rainfall fields nearly so, as tested by the thumb rule proposed by North et al. (1982). It may, however, be noted that in the present case the two EEOF's need not be distinct as they represent the quadrature phases of the same process.

The rate of south to north progression of the major anomalies is about 0.75° lat. day^{-1} for the circulation fields and about 0.5° lat. day^{-1} for the rainfall fields. This disparity in the movement of the anomalies in the rainfall and the circulation fields does not appear to be merely due to the irregular shape and size of the subdivisions. When we analyzed the rainfall of regular sized blocks lying along a longitudinal transect by the EEOF method, similar rate of movement was indicated (Singh and Kripalani, 1986). We may note that the rate of movement could be compared strictly only if some prefiltering was done.

However, for all the fields the progression appears to recur in the same phase after a period of about 40 days. This feature is part of the global 40–50 day oscillation first identified in the tropical atmosphere by Madden and Jullian (1971) showing definite eastward phase propagation over the tropics (Lorenc, 1984; Weickmann, 1983) and northward propagation over the Indian ocean region (Krishnamurti and Subramanyam, 1982; Murakami et al., 1984). The northward propagation of the 700 mb height anomalies over the Indian region with a recurrence period of about 40 days was first noted by the authors in 1976 (Singh et al., 1976). An indirect evidence of the progression was further noted by Singh et al. (1978) who found significant correlations between the 700 mb height anomalies and the rainfall during the next 10 days over the areas located northward of the 700 mb heights. In recent years the recurring northward propagation has been observed in many circulation and weather fields by several workers (Alexander et al., 1976; Sikka and Gadgil, 1980; Krishnamurti and Subramanyam, 1982; Yasunari, 1980). This subject is presently of great topical interest and a prolific number of papers continue to appear either explaining the mechanism (Chang, 1977; Webster, 1983; Yamagata and Hayashi, 1984; Simmons et al., 1983; Goswami and Shukla, 1984) or structure (Anderson et al., 1984; Murakami et al., 1984; Lau and Chan, 1985) of this 40-day oscillation.

7. Conclusions

It is clear from these results that the EEOF analysis not only can identify the hidden structures in the atmospheric/ocean fields but can also quantify the contribution of these structures towards the total variance of the sequentially evolving fields. Spatial regions or channels over which a particular mode evolves more intensity are also identified. The method requires only standard subroutines used to determine the eigenvectors of symmetric matrices and is much simpler than the more versatile complex EOF analysis (Trenberth and Shin, 1984; Barnett, 1983). The application of this technique for different regions, weather parameters, levels and time periods may provide useful evolutionary information in quantitative terms.

Acknowledgments. The authors are grateful to Dr. Bh. V. Ramanamurty, Director, and Dr. S. S. Singh, Assistant Director and Head, Forecasting Research Division of the Indian Institute of Tropical Meteorology for encouragement and also for the facilities provided. Furthermore, the authors wish to express their sincere thanks to Shri K. D. Barne for typing the manuscript.

REFERENCES

Alexander, G., R. N. Keshavamurthy, U. S. De, R. Chellapa, S. K. Das and P. V. Pillai, 1978: Fluctuations of monsoon activity. *Indian J. Meteor. Hydrol. Geophys.*, **29,** 76–87.

Anderson, J. R., D. E. Stevens and P. R. Julian, 1984: Temporal variations of the tropical 40–50 day oscillation. *Mon. Wea. Rev.*, **112,** 2432–2438.

Barnett, J. P., 1983: Interaction of the monsoon and Pacific trade wind systems at interannual time scales. Part I: The equatorial zone. *Mon. Wea. Rev.*, **111,** 750–773.

Chang, C. P., 1977: Viscous internal gravity waves and low frequency oscillations in the tropics. J. Atmos. Sci., **24,** 901–910.

Goswami, B. N., and J. Shukla, 1984: Quasi-periodic oscillations in a symmetric general circulation model. *J. Atmos. Sci.*, **41,** 20–37.

Kripalani, R. H., S. D. Bansod and S. V. Singh, 1986: The spatio-temporal evolutionary features of weekly monsoon rainfall over India. Submitted to *Mausam.*

Krishnamurti, T. N., and D. Subramanyam, 1982: The 30–50 day mode at 850 mb during Monex. *J. Atmos. Sci.*, **39,** 2088–2095.

——, P. K., Jayakumar, J. Sheng, N. Surgi and A. Kumar, 1985: Divergent circulations on the 30 to 50 day time scale. *J. Atmos. Sci.*, **42,** 364–375.

Kutzbach, J., 1967: Empirical eigenvectors of sea level pressure, surface temperature, and precipitation complexes over North America. *J. Appl. Meteor.*, **6,** 791–802.

Lau, K. M., and P. H. Chan, 1985: Aspects of the 40–50 day oscillation during the northern winter as inferred from outgoing longwave radiation. *Mon. Wea. Rev.*, **113,** 1889–1909.

Lorenc, A. C., 1984: The evolution of planetary scale 200 mb divergent flow during the FGGE year. *Quart. J. Roy. Meteor. Soc.*, **110,** 427–442.

Madden, R., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28,** 702–708.

Murakami, T., T. Nakazawa and J. He, 1984: On the 40–50 day oscillation during the 1979 northern hemisphere summer. Part I: Phase propagation. *J. Meteor. Soc. Japan,* **62,** 250–271.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110,** 699–706.

Prohaska, J. T., 1976: A technique for analyzing the linear relation-

ships between two meteorological fields. *Mon. Wea. Rev.*, **104,** 1345–1353.

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

Simmons, A. J., J. M. Wallace and G. W. Branstator, 1983: Barotropic wave propagation and instability and atmospheric teleconnection patterns. *J. Atmos. Sci.*, **40,** 1363–1392.

Singh, S. V., and R. H. Kripalani, 1986: The south to north progression of rainfall anomalies across India during the summer monsoon. *Pure Appl. Geophys.*, **123**(4), (in press).

——, K. D. Prasad and P. D. Ubale, 1976: Evolution of 5-day 700 mb flow patterns and its relationship with rainfall, Part I. *Proc. of the Symposium on Tropical Monsoon,* Pune: 411005, India, Indian Institute of Tropical Meteorology, 515–520.

——, D. A. Mooley and R. H. Kripalani, 1978: Synoptic climatology of the daily 700 mb summer monsoon flow patterns over India. *Mon. Wea. Rev.*, **106,** 510–523.

——, —— and K. D. Prasad, 1979: Prediction of 10-day summer monsoon rainfall over India. *Arch. Meteor. Geophys. Bioklim., Ser. A.*, **27,** 317–331.

Trenberth, K., and W. T. K. Shin, 1984: Quasi-biennial fluctuations in sea level pressure over the southern hemisphere. *Mon. Wea. Rev.*, **112,** 761–777.

Weare, B. C., and J. S. Nasstrom, 1982: Examples of extended empirical orthogonal function analysis. *Mon. Wea. Rev.*, **110,** 481–485.

Webster, P. J., 1983: Mechanisms of low-frequency variability: Surface hydrological effects. *J. Atmos. Sci.*, **40,** 2110–2124.

Weickmann, K. M., 1983: Intraseasonal circulation and outgoing longwave radiation modes during northern hemisphere winter, *Mon. Wea. Rev.*, **111,** 1838–1858.

Yamagata, T., and Y. Hayashi, 1984: A simple diagnostic model for the 30–50 day oscillation in the Tropics. *J. Meteor. Soc. Japan.*, **62,** 709–717.

Yasunari, T., 1980: A quasi-stationary appearance of 30 to 40 day period in the cloudiness fluctuations during the summer monsoon over India. *J. Meteor. Soc., Japan,* **58,** 225–229.