Kenya Journal of Science and Technology Series A (1984) 5: 63 – 83

Principal component analysis of sea-level pressure over the southwest Indian Ocean during the northern summer of 1979

Raphael E.A. Okoola,
Institute for Meteorological Training and Research, Nairobi, Kenya

SUMMARY

Sea-level pressure over the southwest Indian Ocean has been studied for the northern summer season during the first Garp global experiment (FGGE) period. Grid point pressure values, resulting from objective analysis, at 1200 GMT for the period 15 June to 13 August 1979 were selected for this study. This period is referred to as the "selected monsoon period" and it lay between the Indian subcontinent monsoon onset and withdrawal periods.

The space and time characteristics of sea level pressure in the region were estimated from principal component analysis and from spectral analysis respectively. These analyses revealed the following points:

1. Twelve eigenvectors were significant and these explained more than 80% of the total variance.
2. The first eigenvector, accounting for 31.8% of the total variance, comprised a zonal pattern. The time-dependent coefficients of this pattern displayed a statistically significant oscillation (at 95% C.L.) with period of 4.5 days.
3. The second eigenvector, accounting for 12.8% of the total variance, consisted of a meridional pattern with two cells. The most intense cell had its centre to the south of the Mozambique Channel near the point 35°S, 40°E. The time-dependent coefficients of this meridional pattern had a statistically significant oscillation with period 8.0 days.
4. The third eigenvector, explaining 9.8% of the total variance, consisted of a zonal pattern with centre of maximum positive departures located near 35°S, 80°E. These values decreased westwards. This pattern displayed oscillations with periods 13.0, 4.0 and 2.5 days. The 13.0 day period is suggested as due to the changes in pressure of the eastern and western centres (zonal overturnings), the 4.0 day period is part of the effect of middle-latitude disturbances moving eastwards, while the 2.5 day period is due to short-term persistence in the pressure patterns.
5. The fourth eigenvector, explaining 5.4% of the total variance, consisted of zonal pattern. The pattern displayed oscillations with a period of 4.5 days.

The trough later moves towards the East African coast causing marked rainy spells there.

Ramsay (1971) also reported a significant relationship between surface pressure gradient (between Mombasa and the region to the south near latitude 15°S) and the surface wind speed at Mombasa. Pressure rises to the south (near latitude 15°S) lead to increases in surface wind speed at Mombasa within 12 hours.

Findlater (1974) reported that during the period of the East African low-level jet excessive rainfall occurred at the East African coast (near Mombasa) close to the position where the jet axis approached the coast. Findlater (1972) and Okoola (1982) reported that the low-level jet core, found near the 850 mb level during the northern summer is associated with horizontal velocity convergence.

In view of the above relationships, forecasters in the East African meteorological services have routinely used surface pressure changes over Arabia, the Mascarene Islands region and the southeast of the Atlantic Ocean (to the east of St. Helena Island) to give both short-range and medium-range forecasts of rainfall in East Africa.

In this study, the space and time characteristics of sea-level pressure in the southwest Indian Ocean during the northern summer of 1979 were estimated from principal component analysis (PCA). The patterns of eigenvectors and time coefficients were related to the space and time characteristics of the sea-level pressure field. Since the coefficients are mutually uncorrelated, they can be used as predictors to estimate any related predictands over East Africa, in the knowledge that the variance predicted by any one coefficient is independent of and additive to that predictable by any other. Any or all the coefficients that are significantly correlated to the predictands can be used in a forecast model knowing that no better linear prediction technique is possible. PCA is used to identify those horizontal patterns of sea-level pressure which contribute to the total variance most efficiently. The periodicity of the patterns was estimated from spectral analysis.

DATA USED IN THIS STUDY

The study area is restricted to the southwest Indian Ocean (figure 1). The domain extends in latitude from 15°N to 35°S and in longitude from 30°E to 85°E.

The data consist of once-daily (12 GMT) sea-level pressures at grid points (2° latitude/longitude) derived from objective analysis. The period 15 June to 13 August 1979 was used. The above period is hereafter referred to as the "selected monsoon period". It lay between the onset and the withdrawal of the northern summer monsoon for that year.

FGGE – A unique international experiment

The monsoon experiment (MONEX – 79), performed in 1979 as part of the first Garp global experiment (FGGE), attempted to provide a comprehensive set of data over the Indian Ocean. These data are more numerous than those that have been available to previous workers (WMO/ICSU 1981), and the data is likely to provide a better understanding of the monsoon (Fein and Kuettner 1980). The surface-based observing system during the northern summer of 1979 consisted of the established station network as used in the world weather watch programme. In addition to this there were two other important data sources:

1. research and commercial ships; and
2. drifting buoys.

A network of research and other vessels, called tropical wind observing ships (TWOS), were deployed in tropical latitudes,

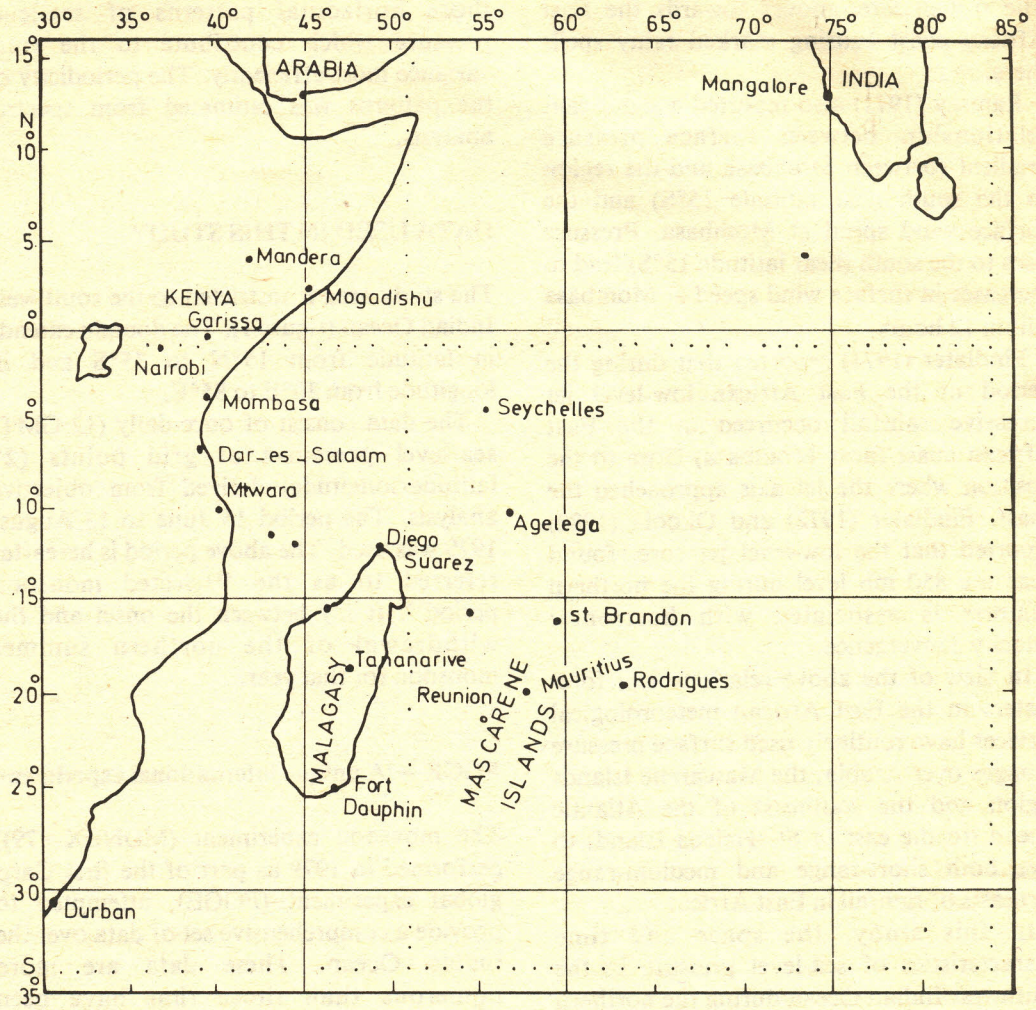

Figure 1. Area of study and stations used in this study are shown (underlined where only surface observations were available).

especially between 10°N and 10°S. Altogether, 22 ships provided sea-level pressure and other data during the months May to August 1979.

The southern hemisphere drifting buoys were designed to sample the vast data-void areas of the southern hemisphere oceanic regions between 20°S and 65°S. These measured surface pressure and sea-surface temperature. Their data were collected and their positions determined by means of the Argos system carried on board the polar-orbiting satellite TIROS N. Some buoys were also deployed in the tropical region (equatorward of 20°S). There were 49 buoys on 14 June, 47 buoys on 28 June and 48 buoys on 16 July 1979.

Considering the entire region from 20°S to 65°S, it has been deduced that between 70% and 80% of the area was within 500 km of a buoy transmitting pressure data. The summer MONEX data from drifting buoys

in the Indian Ocean was of particular importance for good definition of surface pressure field, particularly the strength and location of the Mascarene high.

Thus, during the FGGE year there was a unique coverage of observations over the MONEX area. This coverage coupled with the fact that data received even 30 hours later were still used in the analyses, has enabled preparation of analyses which are far superior to those that had been available before MONEX – 79.

For the principal component analysis the data at 2° latitude/longitude grid points were interpolated into a coarser grid of 5° latitude/longitude. The coarser grid was necessary in order to keep the data matrix within reasonable limits for the computer to handle.

Principal component analysis (PCA)

Principal component analysis enables fields of highly correlated data to be represented adequately by a small number of orthogonal patterns (eigenvectors) and corresponding orthogonal time coefficients. The first principal component or eigenvector is that pattern which explains the greatest fraction of the total variance. Subsequent eigenvectors account for the largest parts of the remaining variance. The variance accounted for by an eigenvector is known as the eigenvalue, and the time co-efficient gives its weight in any given time interval (a day in this case). In most applications of the technique it is found that a large portion of the variance can be accounted for by retaining only the first few eigenvectors. PCA has been widely applied in meteorology by among others, Lorenz (1956), Kutzbach (1967), and Paegle and Haslam (1982).

Lorenz (1956) was the first to realise the meteorological possibilities of this method. He suggested the use of eigenvector analysis to efficiently reduce the number of variables necessary to describe atmospheric patterns.

Kutzbach (1967) produced a principal component analysis of January temperature, pressure and rainfall for the period 1941 – 1965 over North America using 23 stations.

Recently, Paegle and Haslam (1982) applied PCA to 500 mb geopotential heights of seven years of winter data over a portion of the western United States of America.

A good account of principal component analysis is given by Lorenz (1956), Kutzbach (1967) and Kendall (1975). Only an outline of the method is given here for the sake of completeness. The mean sea-level pressure at grid point m ($\bar{P}_m$) is expressed as:

$$\bar{P}_m = L^{-1} \sum_{t=1}^{L} P_m \quad (t) \qquad (1)$$

and the covariance, σ, between grid points m, n as:

$$\sigma_{m,n} = L^{-1} \sum_{t=1}^{L} P_m(t)\, P_n(t) - \bar{P}_m \bar{P}_n \qquad (2)$$

where the number of daily observations in the sample $L = 60$, and there are 110 grid points for the 5° × 5° latitude/longitude grid over the study area.

In the case where $m = n$, equation (2) gives the variance of the sea-level pressure field P at grid point m. The fields of mean sea-level pressure and its standard deviation (square root of variance) are given in figures 2 and 3.

The mean sea-level pressure field shows two highs near 30°S, one of the African continent and the other over the southern Indian Ocean with centre at 31°S, 59°E. There is a ridge over the eastern African coast. Pressure is generally low near the equator, especially to the east of 60°E.

The standard deviation (figure 3) has two dominating features: an area of large values

(greater than 3 mb) which lies to the south of about latitude 20°S; and an area of small values (less than 1.5 mb) lying in a northwest/southeast orientation and including the East African coast.

The covariance, σ, of a given field can be expressed in matrix form as:

$$\sigma = L^{-1} (\underline{P}^{*T} \underline{P}) \quad (3)$$

where $\underline{P}^*$ is a matrix of 110 columns (one for each grid point) and 60 rows (one for each day). $\underline{P}^*$ at grid point m is defined as:

$$P^*_m = P_m - \bar{P}_m \quad (4)$$

The matrix $\underline{P}^{*T}$ is the transpose of the matrix $\underline{P}^*$

The field P is now expressed in matrix form as the sum of products of a small

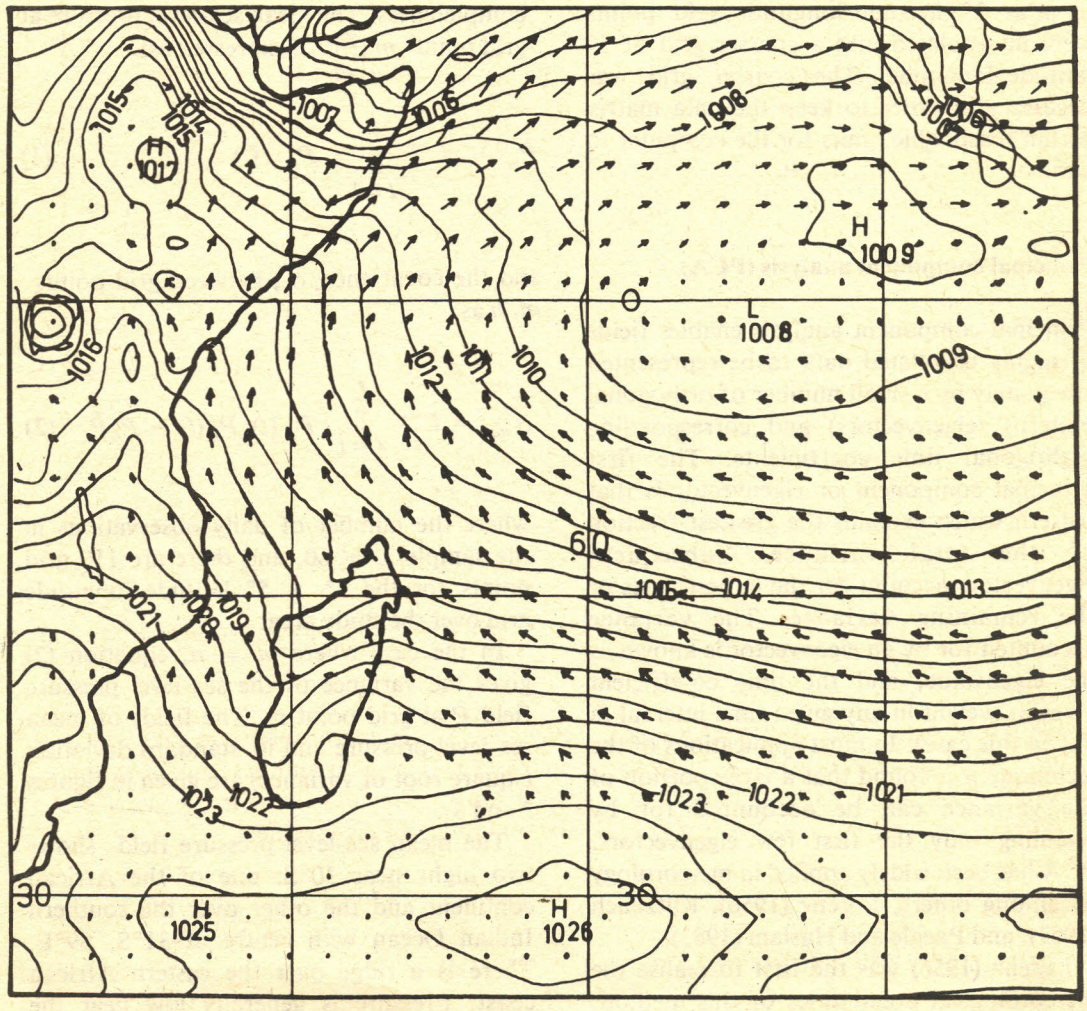

Figure 2. The time-averaged mean sea-level pressure at 1200 GMT for the "selected monsoon period" (15 June to 13 August 1979). Isobars at intervals of 1 mb. Also shown are mean surface wind vectors for the above period.

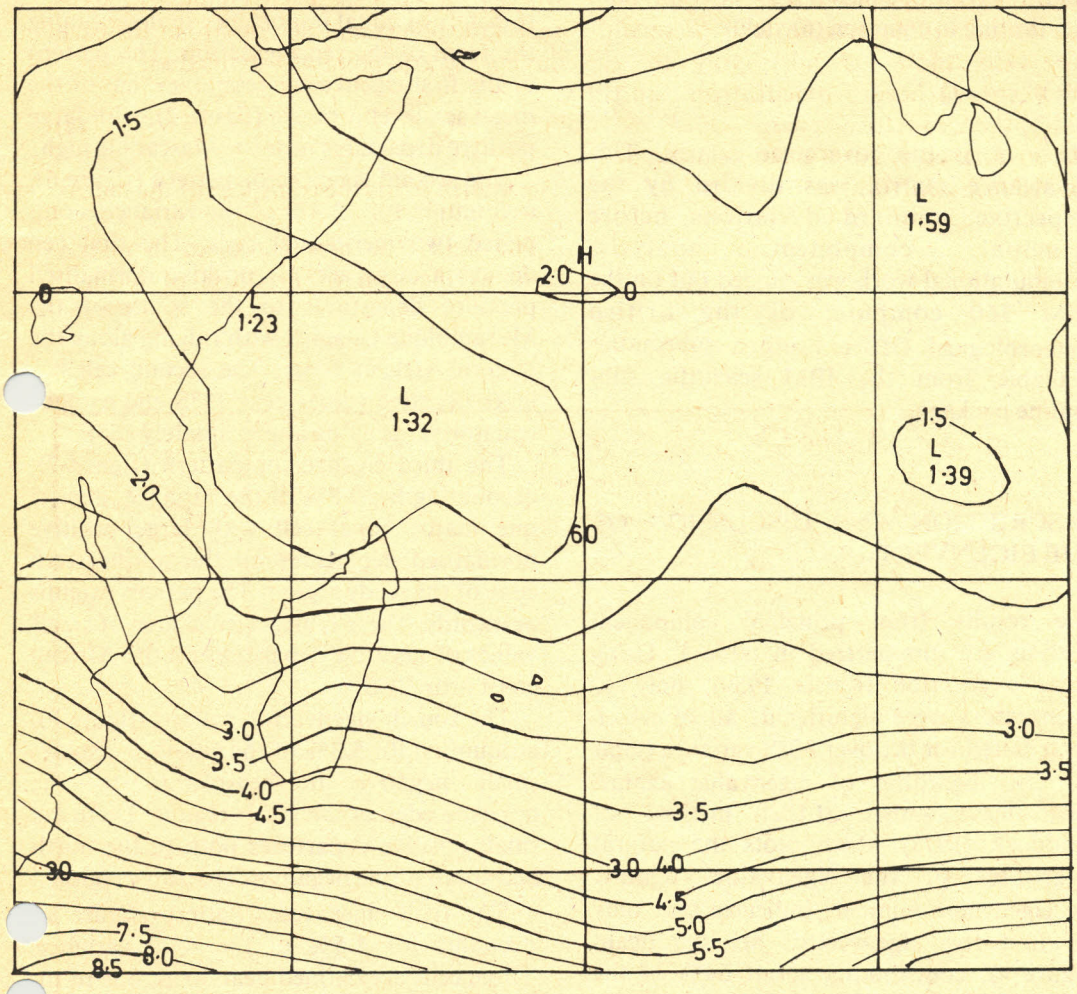

Figure 3. The standard deviation field of sea-level pressure for the selected monsoon period (15 June to 13 August 1979) at 1200 GMT. Isobars at intervals of 0.5 mb.

number of time-dependent coefficients, Q, and corresponding eigenvectors, Y, which are constant in time and orthogonal to each other; that is

$$\underline{P}^*(t) = Q^*(t)\,\underline{Y} \tag{5}$$

The eigenvectors are orthonormalized and

$$Q^{*T}(t)\,Q^*(t) = \underline{D} \tag{6}$$

where $\underline{D}$ is a diagonal matrix whose diagonal elements λ_j increase monotonically as j increases ($j = 1,110$) and are known as the eigenvalues. The total variance is given by the sum of all the eigenvalues.

$$\sum_{j=1}^{110} \lambda_j \quad \sum_{m=1}^{110} \overline{P^*_m 2} \tag{7}$$

Each eigenvalue gives the contribution of each eigenvector to the total variance of the

field. Solving equation (5) for Q^* and substituting in equation (6) yields:

$$Y \sigma Y^T = \underline{D} \quad (8)$$

where σ is the covariance matrix. This covariance matrix was divided by the respective standard deviations before principal component analysis. Computational work was carried out on the IBM 360 computer of the British Meteorological Office using a subroutine available from the IBM scientific subroutine package.

DESCRIPTION AND DISCUSSION OF THE EIGENVECTORS

The results from principal component analysis are summarized in table 1. Using Kaiser's criterion (Kaiser 1959), only 12 eigenvectors were significant, all of which contributed for the over 80% variance (table 1). The logarithm of eigenvalue against eigenvalue number (LEV) method of Craddock (1973) which plots the natural logarithms of the eigenvalues against the eigenvalue number, indicates that only the first five eigenvectors are significant (figure 4), contributing about 64% of the total pressure variance.

The first five eigenvectors are shown in figures 5, 6, 7, 8 and 9. The isolines are non-dimensional since the eigenvectors were derived from normalised departure fields (correlation matrix). Each eigenvector mathematically defines two departure patterns because the coefficients associated with each eigenvector may be positive or negative.

The first eigenvector (figure 5), accounting for 31.8% of the variance, comprised a quasi-zonal sea-level pressure departures pattern with the centre of maximum positive departures located near the point 35°S, 50°E. The use of normalised variables assures that each variable at each point in the region is of equal importance in determining the form of the representation. The pattern of the first eigenvector resembles that of the mean sea-level pressure (figure 2), with large positive departures near the Mascarene high.

The second eigenvector pattern (figure 6), accounting for 12.8% of the variance, comprised two meridional cells. The first cell depicts a large area of positive normalised pressure departures to the south of the Mozambique Channel with ridging along the eastern African coast. The second cell has large negative values near 35°S and extend equatorwards along approximately 75°E.

The third eigenvector pattern (figure 7), accounting for 9.8% of the variance, depicts one major zonal cell with large positive normalised departures on the southeastern edge of the study region. The pattern extends westwards. There is also a small area of small values of negative departures on the African coast near 30°S.

The fourth eigenvector pattern (figure 8), accounting for 5.3% of the variance displays small negative northeastwards to the opposite edge of the study region. There are small positive departures both to the north and south of the negative departures region.

The fifth eigenvector pattern (figure 9) accounts for 4.5% of the total variance. Three cells are defined near 30°S, one to the east and the other to the west, have negative departures, while the third cell in the centre has positive departures. The intensity of these cells decreases eastwards along 30°S. It is noteworthy that the distance between the zero isopleths along 30°S is approximately 25 degrees of longitude. This suggests a wavelength of 50 degrees of longitude for this pattern near 30°S.

Spectral analysis of the time-dependent coefficients

The time behaviour of the coefficients associated with the first five eigenvectors are

Table 1. Eigenvalues of the first 25 eigenvectors computed from the correlation matrix

Eigenvector number	Eigenvalue	Natural logarithm of eigenvalue	Total variance explained explained (%)	Cummulative total variance explained (%)
1	19.10	2.95	31.84	31.84
2	7.65	2.04	12.76	44.60
3	5.90	1.78	9.84	54.44
4	3.21	1.17	5.35	59.79
5	2.71	1.00	4.51	64.30
6	1.95	0.67	3.25	67.55
7	1.70	0.53	2.83	70.38
8	1.53	0.43	2.55	72.93
9	1.27	0.24	2.11	75.04
10	1.21	0.19	2.02	77.06
11	1.09	0.08	1.81	78.87
12	1.03	0.03	1.72	80.59
13	0.97	−0.03	1.61	82.20
14	0.85	−0.16	1.42	83.62
15	0.77	−0.26	1.28	84.90
16	0.73	−0.31	1.21	86.11
17	0.67	−0.40	1.12	87.23
18	0.62	−0.48	1.03	88.26
19	0.55	−0.60	0.91	89.17
20	0.51	−0.67	0.85	90.02
21	0.45	−0.80	0.76	90.78
22	0.42	−0.87	0.70	91.48
23	0.41	−0.89	0.68	92.16
24	0.36	−1.02	0.60	92.76
25	0.33	−1.11	0.55	93.31

(Kaiser's criterion indicates a cut-off value at eigenvalue number 12)

shown in figure 10. The magnitudes of these coefficients grow and decay.

The period 2 to 10 July was associated with an intense high near the Mascarene Islands, which achieved maximum intensity on 5 July. During this period (vertical lines in figure 10) the time-dependent coefficients of the first eigenvector (C_1) grew from the lowest value on 2 July to a maximum on 5 July and remained above 6 units up to 10 July. For the same period, the coefficients of the second eigenvector (C_2) decreased from a maximum on 3 July to a minimum on the 5 July. The coefficients of the third eigenvector (C_3) remained positive between 4 July and 8 July. The coefficients of the fourth eigenvector (C_4) showed an upward trend from a minimum on 3 July to a maximum on 9 July. While the coefficients associated with the fifth eigenvector (C_5) displayed a marked downward trend after 4 July.

Spectral analysis following Jenkins and Watts (1968) was applied to the time coefficients in order to identify the dominant periods associated with them. Time coefficients of the first eigenvector showed a marked peak with a periodicity of 4.5 days and a minor one of 2.7 days (figure 11).

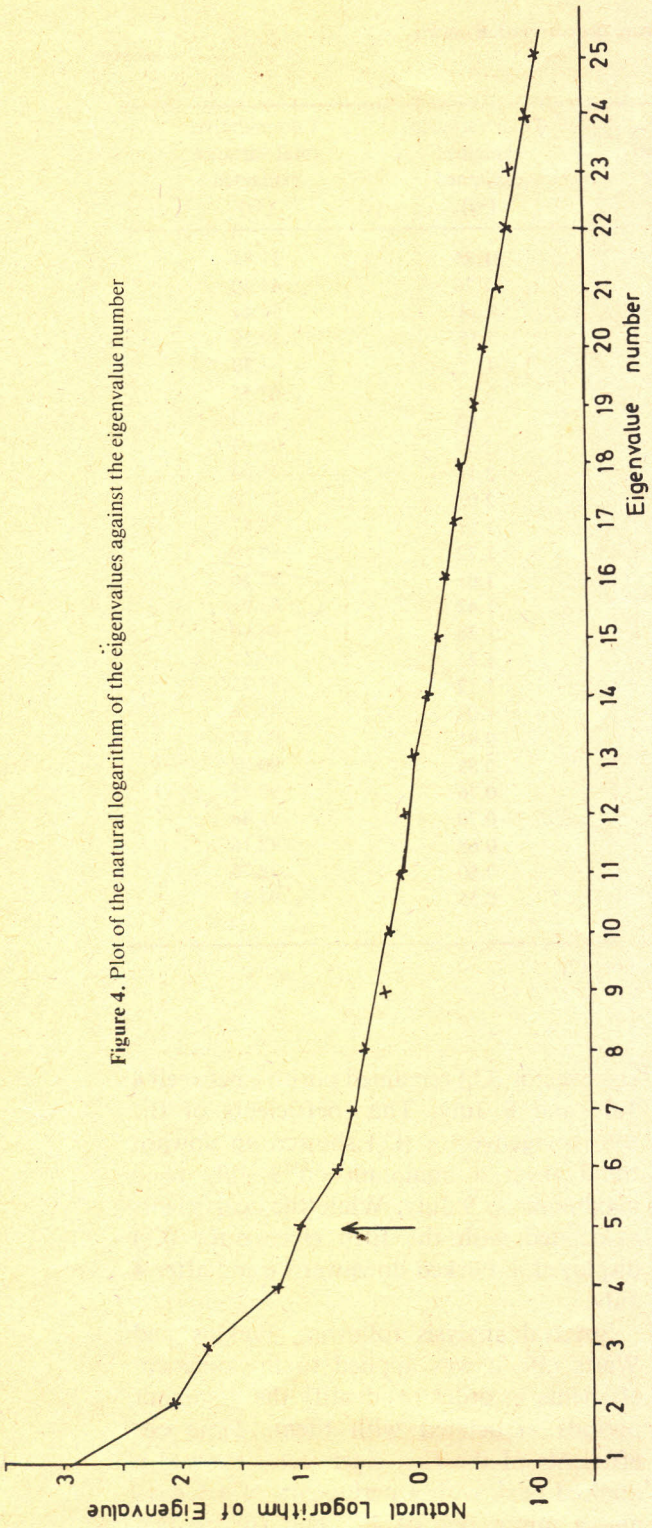

Figure 4. Plot of the natural logarithm of the eigenvalues against the eigenvalue number

Those of the second eigenvector showed a major peak with a periodicity of 8.0 days and a minor peak of 4.0 days (figure 12). The peaks associated with the coefficients of the third eigenvector had periods of 13.0 days, 4.0 days and 2.5 days (figure 13). The fourth eigenvector, like the first one, showed a major peak with period 4.5 days and a minor one with period 2.7 days (figure 14). The coefficients associated with the fifth eigenvector had a statistically significant peak (95% confidence limit) with period 5.0 days and a minor peak of 2.5 days (figure 15).

It may be observed that:

1. periods of 4 to 5 days and 2.5 to 2.7 days are common in all the five time-dependent coefficients;
2. periods of 8 days appear only in the coefficients of the second eigenvector;
3. periods of 13 days appear only in the time coefficients of the third eigenvector.

The first five eigenvectors and their time-dependent coefficients have been shown in figures 5, 6, 7, 8, 9 and 10. The spectra of the time coefficients have been shown in figure 11, 12, 13, 14 and 15.

These time coefficients resemble closely the raw surface-level pressure series at Mauritius and St. Brandon (figure 16). Thus the first eigenvector reflects standing oscillations with an approximate period of 4.5 days. Considering the porportion of the total variance that it explained, the coefficients of this pattern should be assigned the largest weight in any regression forecast equation derived from pressure changes near latitude 30°S.

The second eigenvector pattern (figure 6) suggests that whenever there are large positive (negative) normalised pressure departures to the south of Mozambique Channel, at the same time, large negative (positive) departures occur to the east near the point 35°S, 80°E. These centres will be

sources of meridional wind components responding to the observed pressure departure gradients. Thus, when we have southerlies up the Mozambique Channel there will be, at the same time, northerlies in the region near 35°S, 80°E. This result is consistent with synoptic experience in this region.

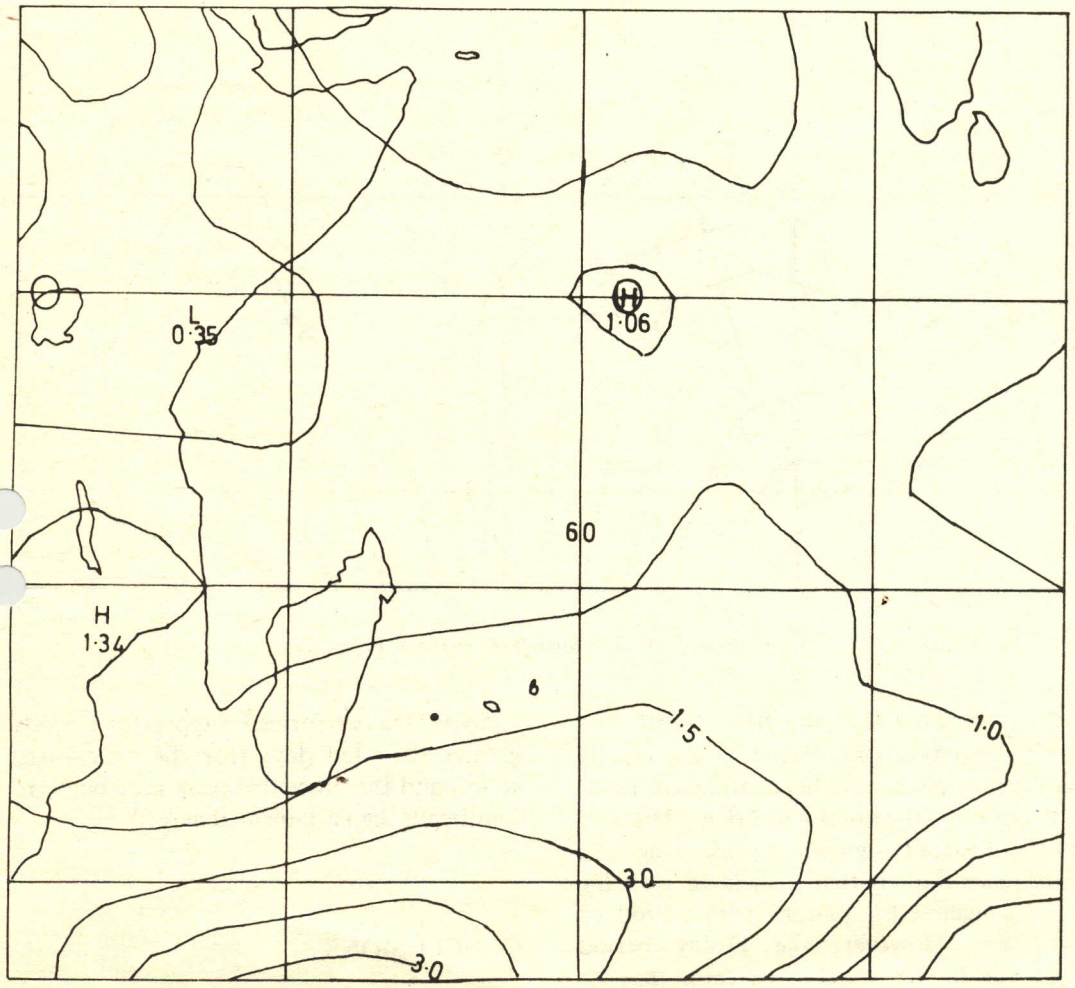

Figure 5. First eigenvector of sea-level pressure. Isopleth interval of 0.5×10^2

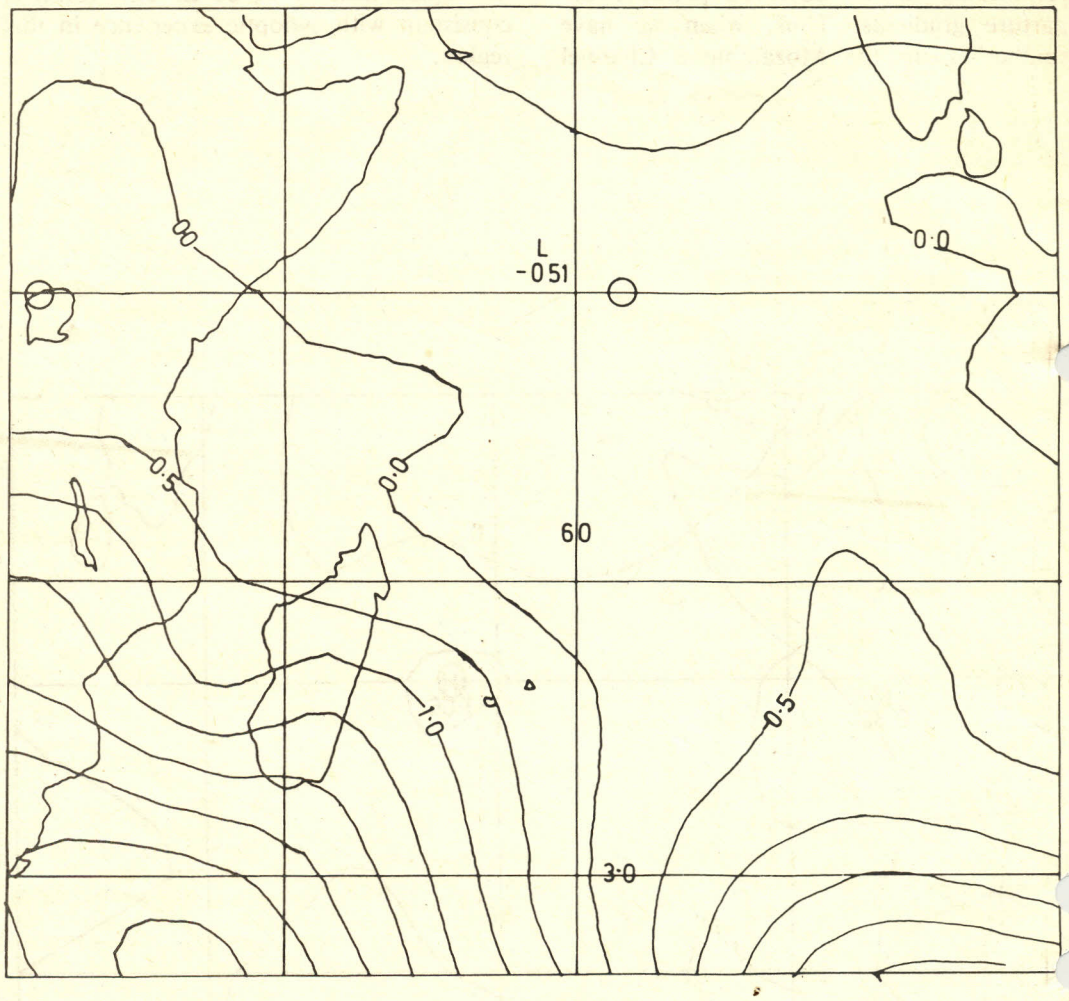

Figure 6. Second eigenvector of sea-level pressure. Isopleth interval of 0.5×10^2

It is observed that the first, fourth and fifth eigenvectors have statistically significant peaks (95% confidence limit) with period in the range 4 to 5 days. The only other statistically significant peak at the 95% confidence limit is that associated with the second eigenvector pattern with period of 8.0 days. However, the 13-day period displayed by the third eigenvector may be physically significant. Also, it should be noted that the difference filter that was used to detrend the raw series before the spectral analysis was performed suppresses periods greater than ten days (for the once-a-day series) and therefore this peak may be more significant than is indicated here.

CONCLUSIONS

Pressure data over the southwest Indian Ocean have been studied for the summer MONEX – 79 season during the FGGE

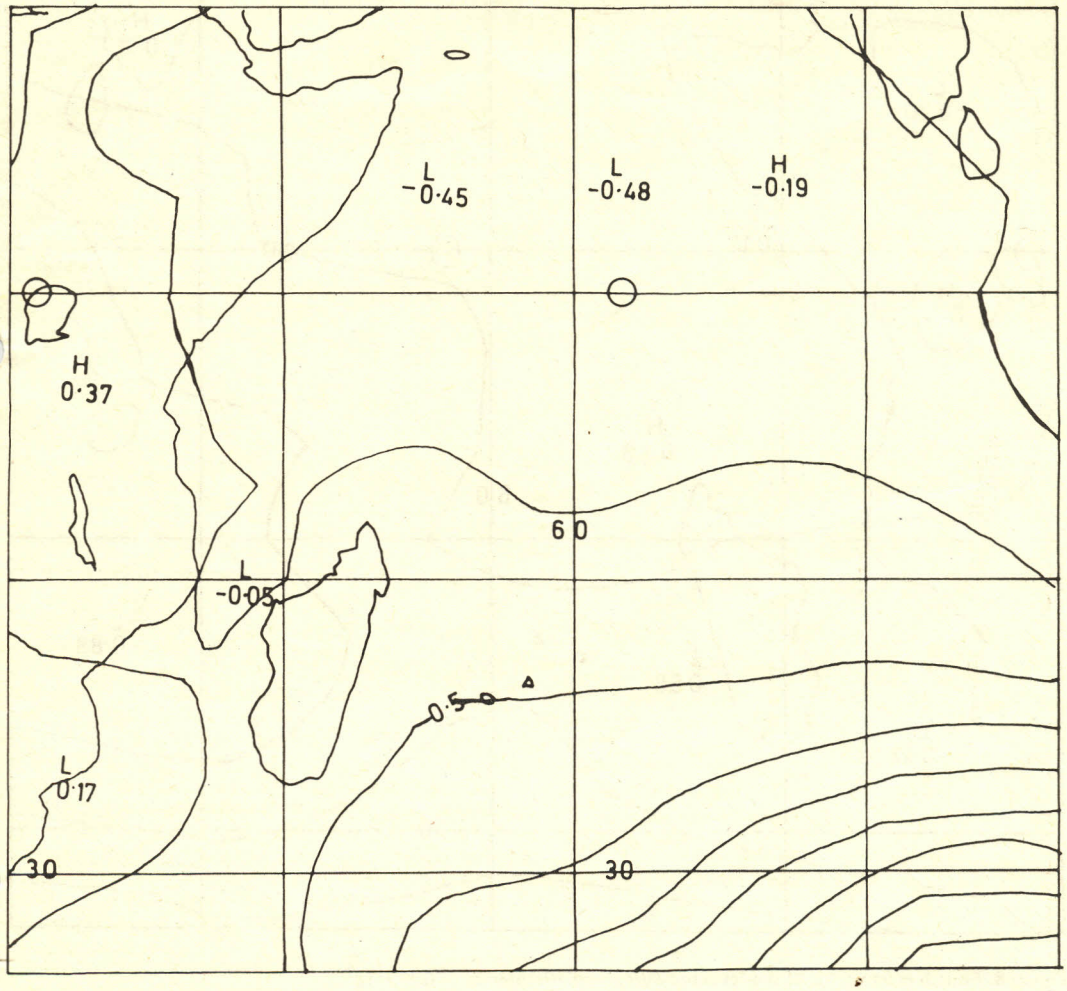

Figure 7. Third eigenvector of sea-level pressure. Isopleth interval of 0.5×10^2

period. From this study the following conclusions can be drawn:

1. Twelve eigenvectors were significant and these explained over 80% of the total variance.
2. The first eigenvector, accounting for 31.8% of total variance, comprised a zonal pattern. The time-dependent coefficients of this pattern revealed that the only statistically significant oscillation was that with a period of 4.5 days.
3. The second eigenvector, accounting for 12.8% of the total variance, consisted of a meridional pattern with two cells. The most intense cell had its centre to the south of the Mozambique Channel near the point 35°S, 40°E. The time-coefficients of this meridional pattern had a statistically significant oscillation with period 8.0 days.
4. The third eigenvector, accounting for 9.8% of the total variance, consisted of a zonal pattern with centre of maximum

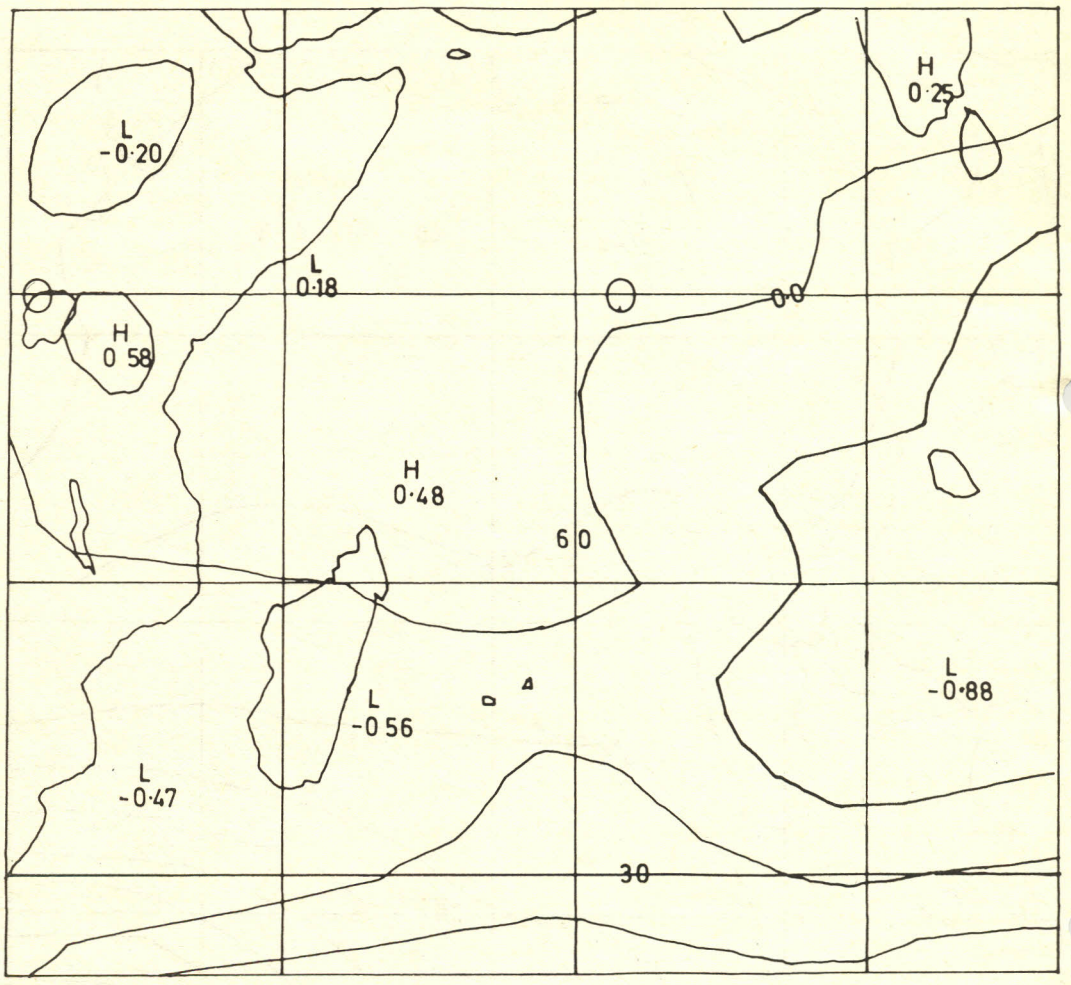

Figure 8. Fourth eigenvector of sea-level pressure. Isopleth interval of 0.5×10^2

positive departures located near 35°S, 80°E. These values decreased westwards. This pattern displayed oscillations with periods 13.0, 4.0 and 2.5 days.

5. The fifth eigenvector pattern displayed three cells. The time coefficients showed a significant oscillation with period of 5 days. This pattern may be due to the passage of middle latitude systems, which have an approximate wavelength of 50 degrees of longitude.
6. Spectral peaks with periods of 2.5 to 2.7 and 4 to 5 days appeared in all the time coefficients of the first five eigenvectors, but the spectral peaks with periods of 8.0 days and 13.0 days appeared only in the time coefficients of the second and third eigenvectors. The 13.0-day period is suggested as due to the changes in pressure of the eastern and western centres (zonal over-turnings), the 4.0 day period is part of the effect of middle-latitude disturbances moving eastwards, while the 2.5-day period is due to short-term persistence in the pressure patterns.

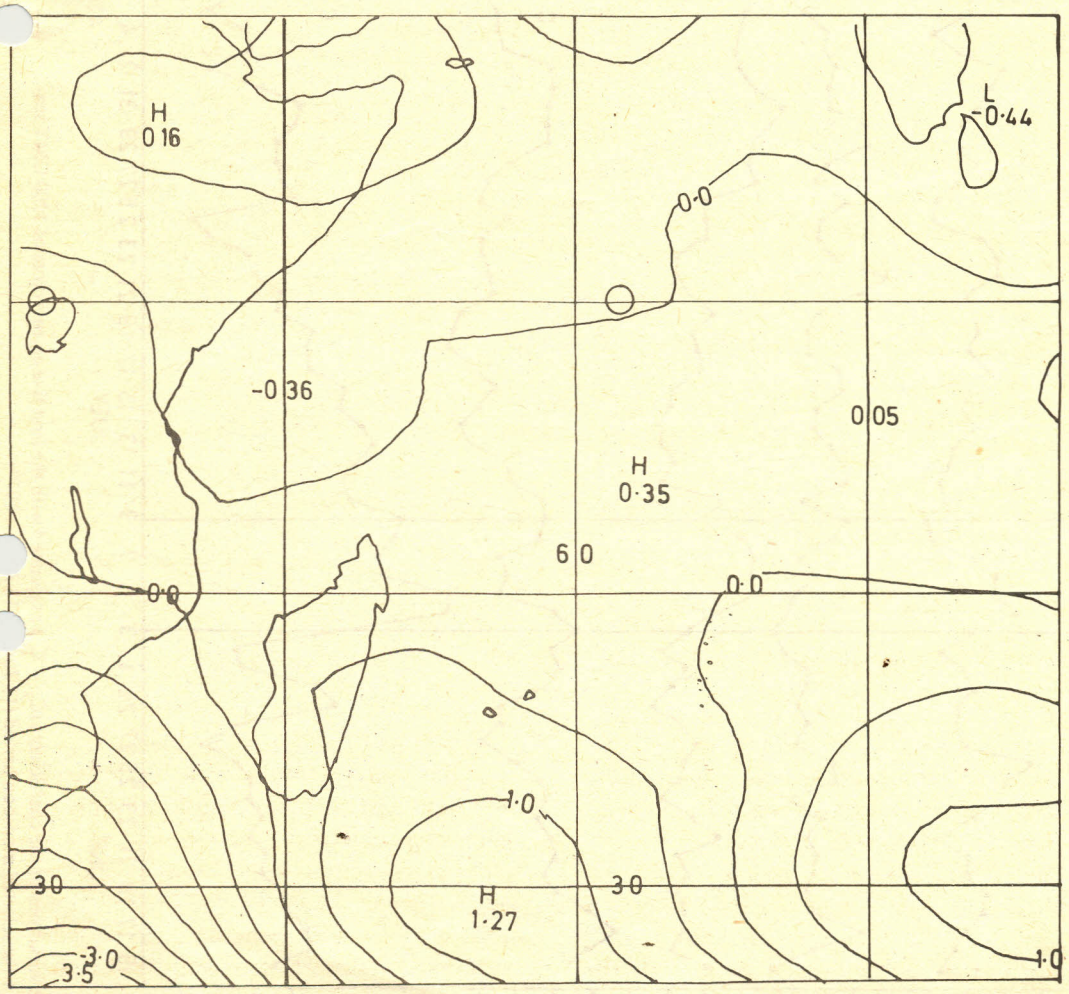

Figure 9. Fifth eigenvector of sea level pressure. Isopleth interval of 0.5×10^2

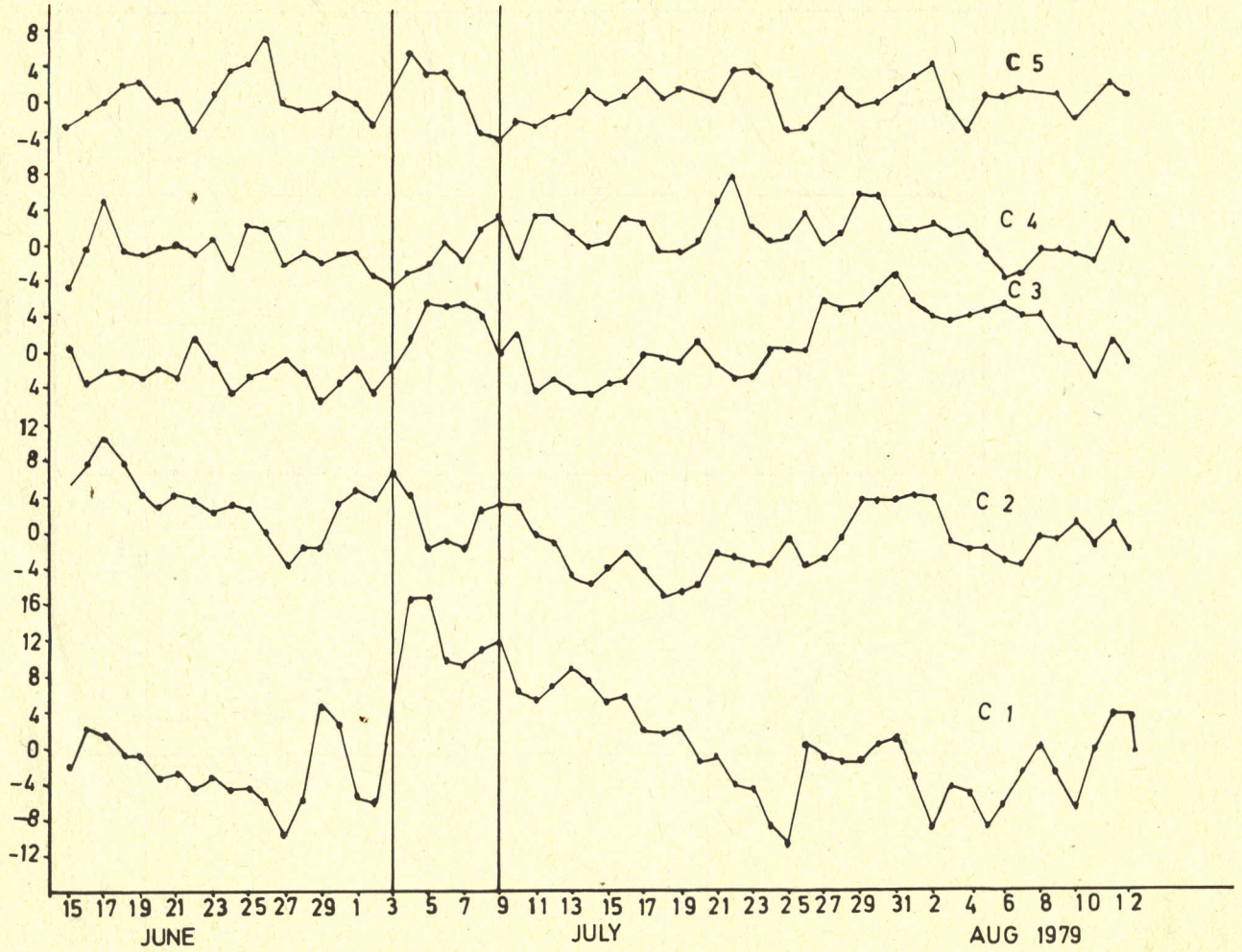

Figure 10. Time variations of the coefficients ($C_1 - C_5$) associated with the first five eigenvectors of 1200 GMT sea-level pressure for the selected monsoon period (15 Jur 13 A st 1979)

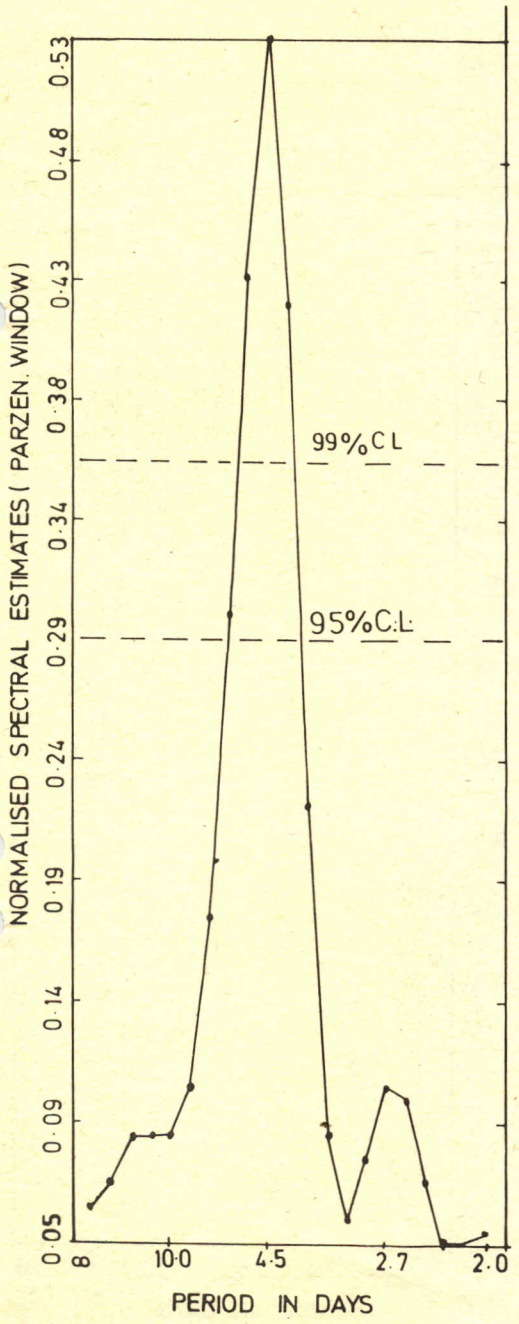

Figure 11. Spectral estimates of the time-coefficients associated with the first eigenvector pattern of sea-level pressure.

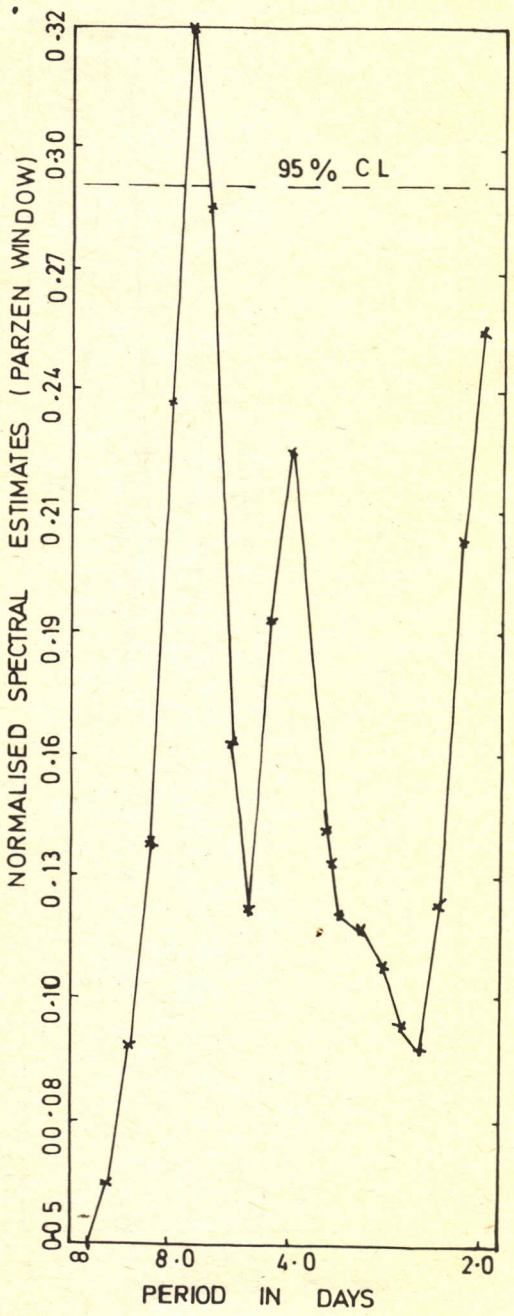

Figure 12. Spectral estimates of the time-coefficients associated with the second eigenvector pattern of sea-level pressure.

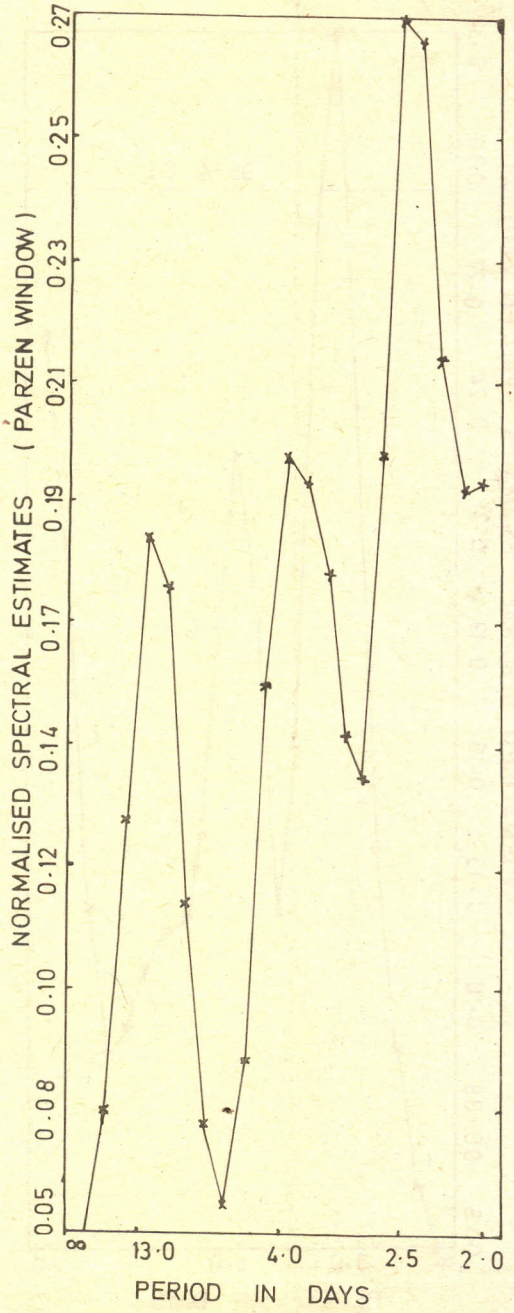

Figure 13. Spectral estimates of the time-coefficients associated with the third eigenvector pattern of sea-level pressure

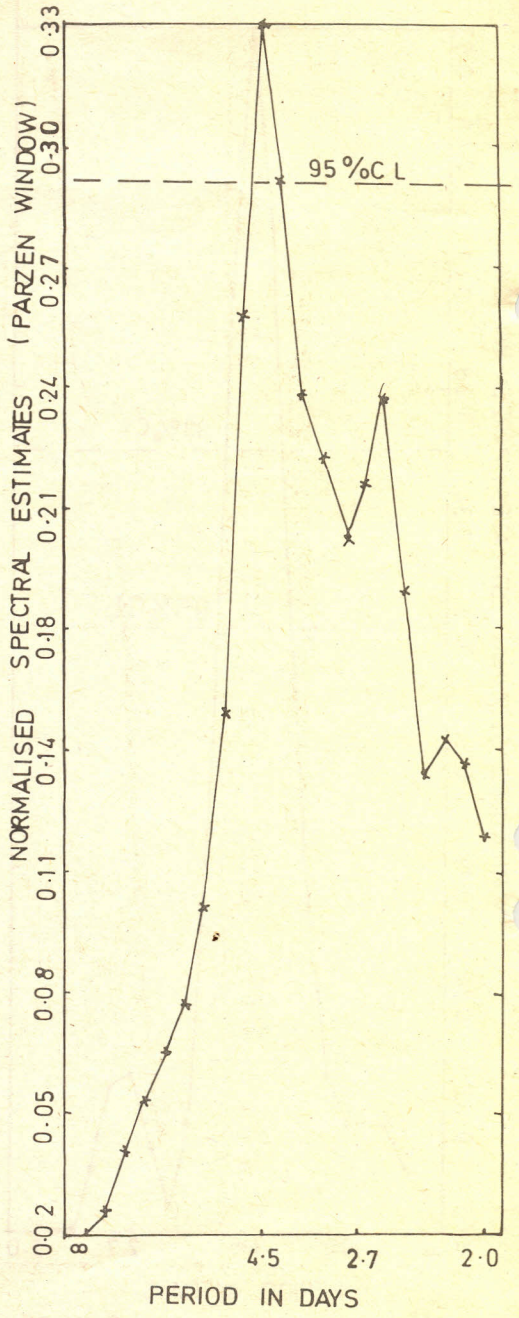

Figure 14. Spectral estimates of the time-coefficients associated with the fourth eigenvector pattern of sea-level pressure

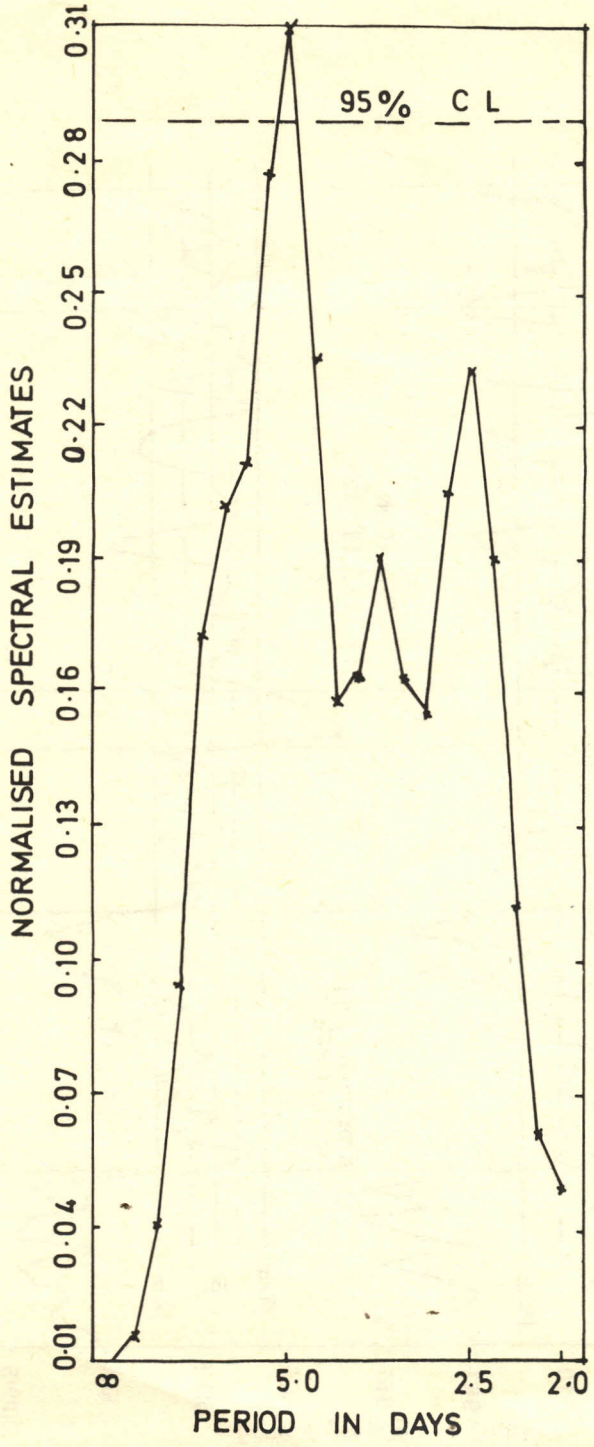

Figure 15. Spectral estimates of the time-coefficients associated with the fifth eigenvector pattern of sea-level pressure

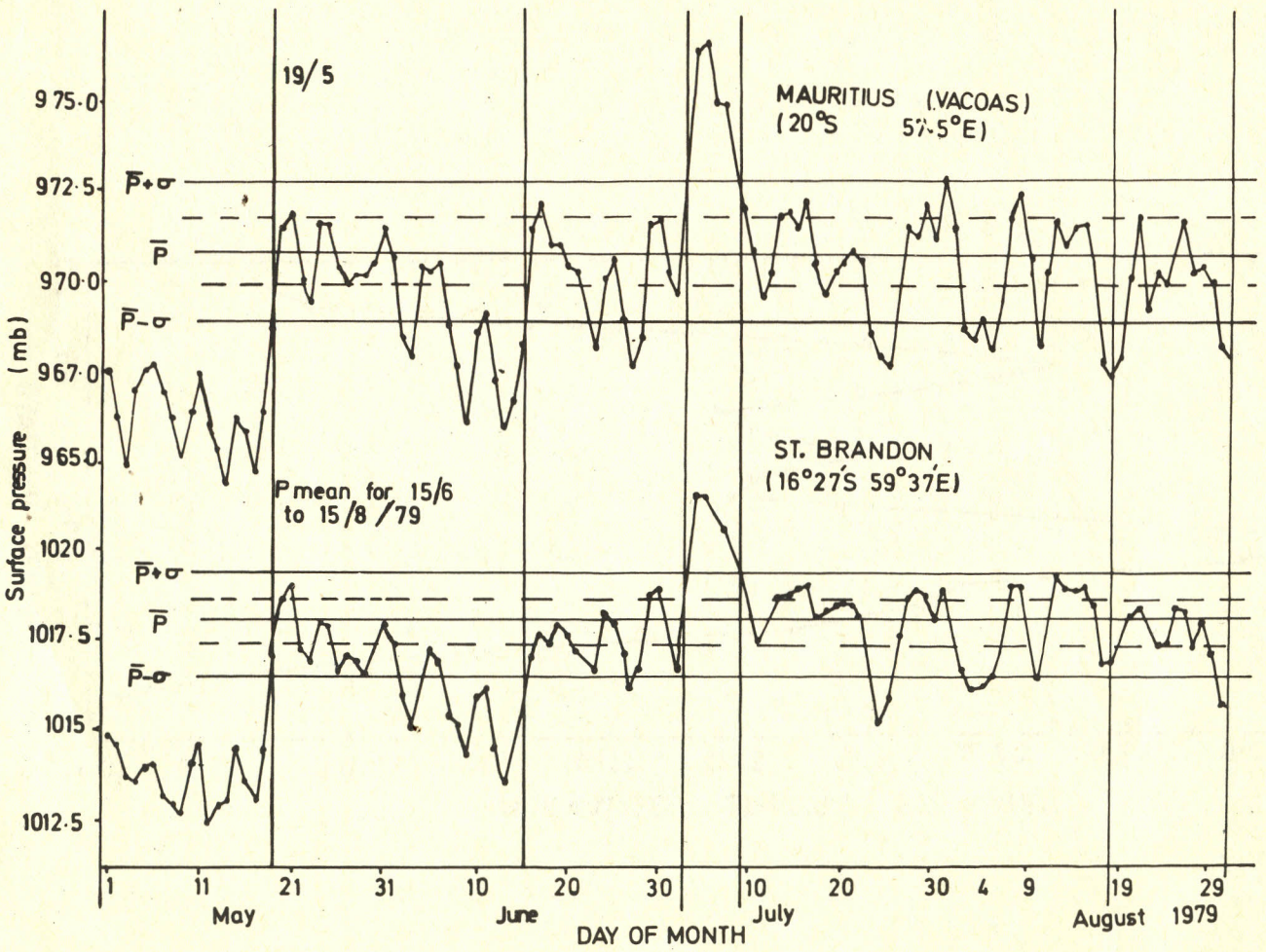

Figure 16. Time series pressure at 1200 GMT over M ritius op) and St. Brandon (bottom) from 1 May to 31 Aug 1979 (see figure 1 for location of stations)

ACKNOWLEDGEMENTS

Most of the analyses were carried out during a study leave period at the British Meteorological Office, Bracknell. The guidance and encouragement of Dr. A. Gadd and Mr. A. Lorenc are sincerely appreciated.

I am highly indebted to the World Meteorological Organization for financial support given under the visiting scientists programme. I am also grateful to the Director of the Kenya Meteorological Department for granting me permission to take up the visiting scientists fellowship in the United Kingdom.

REFERENCES

(1) Craddock, J.M. (1973). Problems and prospects for eigenvector analysis in meteorology. *Statis.* 22: 133 – 145.

(2) Fein, J.S. and J.P. Kuettner (1980). Report on the summer MONEX field phase. *Bull. Am. Meteorol. Soc.* 61: 461 – 474.

(3) Findlater, J. (1972). Aerial exploration of the low-level cross equatorial current over eastern Africa. *Q.J.R. Meteorol. Soc.* 98: 284 – 289.

(4) ———. (1974). An extreme wind speed in the low-level jetstream system of the western Indian Ocean. *Meteorol Mag.* 103: 201 – 205.

(5) Jenkins, G.M. and D.G. Watts (1968). *Spectral Analysis and Its Applications.* San Francisco: Holden-Day, 525 pp.

(6) Kaiser, H.F. (1959). Computer program for Varimax rotation in factor analysis. *Educ. Psychol. Meas.* 19: 413 – 420.

(7) Kendall, M.G. (1975). *Multivariate Analysis.* London: Charles Griffin and Co.

(8) Kutzbach, J.E. (1967) Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. *J. Appl. Meteorol.* 6: 791 – 802.

(9) Lorenz, E.N. (1956). Empirical orthogonal functions and statistical weather prediction. Science Report no. 1, Statistical Forecasting Project, Department of Meteorology, Massachussets Institute of Technology, 49 pp.

(10) Lumb, F.E. (1966). Synoptic disturbances causing rainy periods along the East African coast. *Meteorol Mag.* 95: 150 – 159.

(11) Okoola, R.E. (1982). Monsoon systems over southwest Indian Ocean during northern summer of 1979. M.Sc. thesis, Department of Meteorology, University of Nairobi, Nairobi, Kenya

(12) Paegle, J.N. and R.B. Haslam (1982). Empirical orthogonal function estimates of local predictability. *J. Appl. Meteorol.* 21 (2): 117 – 126.

(13) Ramsay, B. (1971). Surface winds at Mombasa, Kenya. *Meteorol Mag.* 100: 293 – 301.

(14) Sansom, H.W. (1955). Prediction of the seasonal rains of Kenya by means of correlation. *Weather* 10: 81 – 86.

(15) WMO/ICSU (1981) *Summer MONEX Field Phase Report. FGGE Operations Report* no. 8, April 1981.

Raphael E.A. Okoola
Institute for Meteorological Training and Research, P.O. Box 30259, Nairobi, Kenya

Acceptance date: 4 August 1984