Mausam, (1993), 44, 1, 29-34

551. 509. 33: 551. 509. 314

Empirical orthogonal functions associated with parameters used in long range forecasting of Indian summer monsoon

H. N. SRIVASTAVA and S. S. SINGH

Meteorological Office, Pune

(*Received 1 April 1991, Modified 13 March 1992*)

सार – भारतीय ग्रीष्म मानसून के आरंभ होने तथा मौसमी वर्षा के दीर्घ अवधि पूर्वानुमान के प्राचलों से संबद्ध अनुभूत मूलक स्वतंत्र फलनों पर विचार किया गया है। यह पाया गया कि स्पष्ट किये गये प्रसरण की प्रतिशतता पहले चार आनुभविक लांबिक फलनों (ई. ओ. एफ.) के माध्यम से क्रमशः 77 और 67 थी। आरंभ होने की तिथि के साथ उच्चतम सहसंबंध गुणांक पहले फलन के लिये पाया गया जिससे आरंभ होने की तारीख को कॉबर (आस्ट्रेलिया) तथा डारविन (आस्ट्रेलिया) क्षेत्रीय पवनों के अधिकतम प्रभाव का पता चला है। वर्षा प्रागुक्ति के लिये पहले ई. ओ. एफ. पर प्रबल प्रभाव उत्तरी गोलार्ध के ऊपर 50 एच. पी. ए., यूरेसियन हिम आच्छादन, अर्जनटाइना दाब (ऋणात्मक रूप से सहसंबंधित) तथा 500 एच. पी. ए., 10 एच. पी. ए. बालबोआ पवन, उत्तर, मध्य भारत तथा पूर्वी तट न्यूनतम तापमान और उत्तरी गोलार्ध तापमान का था हालांकि एलनीनो, भूमध्यरेखीय दाब तथा डारविन दाब (ताहिती ऋण डारविन सहित) और हिमालय हिम आच्छादन का प्रभाव लगभग नगण्य था। ईगन सूचकांक ग्रीष्म मानसून आरंभ होने की तिथि के दीर्घ अवधि प्रागुक्ति में इसके अनुप्रयोग के लिये पूरक विधि का सुझाव देता है।

ABSTRACT. Empirical Orthogonal Functions (EOF), associated with the parameters for long range forecasting of Indian summer monsoon onset and seasonal rainfall have been discussed. It was found that the percentage of variance explained was 77 and 67 respectively through the first four EOF. The highest correlation coefficient with the onset date was found for the first function which showed the maximum influence of Cobar (Australia) and Darwin (Australia) zonal winds on the onset date. It was interesting to note that for rainfall prediction predominant effect on the first EOF was noticed of 50 hPa ridge over northern hemisphere, Eurasian snow cover, Argentina pressure (negatively correlated) and 500 hPa ridge, 10 hPa Balboa wind, north, central India and east coast minimum temperatures, and northern hemisphere temperature. However, the influence of *El Nino*, equatorial pressure and Darwin pressure (including Tahiti *minus* Darwin) and Himalayan snow cover was almost negligible. The eigen index for the onset date suggests a complementary method for its application in long range prediction of summer monsoon onset date.

Key words — Long range forecast, Empirical Orthogonal functions (EOF), Principal component analysis, Southwest monsoon, Monsoon onset.

1. Introduction

The southwest monsoon which is the principal rain bearing season in India from June to September has attracted the attention of the meteorologists to devise methods for long range forecasting during the past 100 years. Two types of predictions are made namely; forecast of the date of onset of rains and the total quantum of rainfall during the entire season. These predictions are needed for the agriculturists and decision makers. Statistical techniques have been devised for forecasting the rainfall and its onset over Kerala coast starting with Blanford (1884), Walker (1910) and more recently by Bhalme and Mooley (1980), Verma (1980) and others. Kung and Sharif (1980, 1982) attempted to forecast the onset date and the total quantum of rainfall based on regression equations. Shukla and Paolino (1983) found Darwin pressure tendency from winter to spring as a good indicator of monsoon rainfall. It was found that the relationship between the monsoon rainfall and some of the predictors either ceased to exist or showed considerable decline with the passage of time. The need has, therefore, always been felt for developing a suitable model which could be validated over a long period of time. Thapliyal (1982) developed a stochastic transfer model which showed 5% more accuracy compared with the multiple regression model. More recently a parametric and power regression model has been developed by Gowariker *et al.* (1989, 1991) which has shown promise to provide forecast of monsoon rains well in advance and has been validated during the last four years.

India Meteorological Department is regularly issuing long range forecast for the onset of monsoon over the Kerala coast based on a few local and global parameters making use of the regression equation. The object of this paper is to discuss empirical orthogonal functions (EOF) of long range forecasting parameters predicting the onset of summer Indian monsoon and rainfall (June to September). Results have been compared by making use of the eigen index to arrive at the long range prediction of onset of monsoon over Kerala and the seasonal rainfall over the country during this season.

TABLE 1

Parameters used for long range forecast of southwest monsoon

(a) Onset	(b) Rainfall
(1) Cobar (Australia), November zonal wind (500 hPa)	(1) 10 hPa westerly winds over Balboa (January)
(2) Thiruvananthapuram–Madras, February wind (200 hPa)	(2) Darwin pressure (April)
(3) Darwin (Australia), January zonal wind (200 hPa)	(3) Eurasian snow cover (previous December)
(4) Delhi, January wind (300 hPa)	(4) *El-Nino* in current year (October of previous year to May of current year)
(5) Forrest (Australia), January wind (200 hPa)	(5) Equatorial pressure (January to May)
(6) Thiruvananthapuram, November kinetic energy (200 hPa)	(6) Himalayan snow cover (January to March)
(7) Fukuoka (Japan), February kinetic energy	(7) Location of 500 hPa ridge over India (April)
(8) Mean January zonal wind over Darwin (Australia) (300 hPa)	(8) Central India minimum temperature (May)
(9) Mean meridional December wind (previous year) over Calcutta (200 hPa)	(9) 50 hPa ridge-trough extension over northern hemisphere (January and February)
(10) Kinetic energy of January over Darwin (Australia) (200 hPa)	(10) North India minimum temperature (March)
	(11) East coast minimum temperature (March)
	(12) Northern hemispheric temperature (January and February)
	(13) *El-Nino* in previous year
	(14) Argentina pressure (April)
	(15) Tahiti-Darwin pressure (March to May)
	(16) Northern hemispheric pressure (January to April)

TABLE 2

Eigen vectors for long range prediction of summer monsoon onset

EOF Para-meters*	Eigen vector of parameter										Variance explained	Cor. coef. with the onset dates of SW monsoon over Kerala coast
	1	2	3	4	5	6	7	8	9	10		
1	0.46	0.17	−0.43	−0.17	0.33	0.20	0.13	−0.37	0.35	0.35	31.13	0.79
2	0.15	0.45	0.33	−0.53	0.34	−0.16	−0.06	−0.15	−0.26	−0.38	20.45	0.31
3	0.15	−0.29	0.28	0.C0	−0.04	−0.12	0.70	−0.09	0.40	−0.37	13.75	0.11
4	0.04	−0.13	−0.09	−0.27	0.48	−0.79	0.00	0.08	−0.05	0.18	11.70	−0.21

*Same sequence as in (a) of Table 1.

TABLE 3

Eigen vectors for LRF of summer monsoon rainfall

EOF/ Para-meters*	Eigen vector of parameter																Vari-ance ex-plain-ed	Cor. coef. with the rainfall of SW monsoon over country as a whole
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		
1	0.32	−0.14	−0.37	−0.12	−0.14	0.07	0.30	0.27	−0.37	0.26	0.30	0.19	0.12	−0.32	0.05	0.29	30.47	0.83
2	0.07	0.48	0.03	0.17	0.48	−0.35	−0.07	0.04	−0.04	0.09	0.22	0.29	0.20	−0.04	−0.44	0.02	16.68	−0.05
3	−0.16	0.04	0.05	−0.46	−0.05	−0.42	0.25	−0.23	−0.36	−0.30	−0.13	0.36	0.04	0.00	0.21	−0.23	10.67	−0.22
4	−0.18	−0.09	0.00	0.44	0.00	−0.10	0.19	0.34	−0.05	−0.16	−0.22	0.06	0.58	0.29	0.28	0.15	9.29	0.17

*Same sequence as in (b) of Table 1.

It may be mentioned that Paranjpe and Gore (1991) examined the probability of a normal rainfall using the parasiomonious logistic regression model to examine which of the variables out of 15 can best describe the rainfall over the Indian region. Compared to this study, the method adopted in this paper employs a better approach of using empirical orthogonal functions which is the most standard technique presently available to work out the inter-dependence cf the parameters. The results have also been extended to include 16 parameters (Gowariker *et al.* 1991) in the study presented in this paper.

2. Methods of analysis

The principal component analysis provides a simple means due to its restrictions to linear functions of the original variables expressing a linear functions of the p variable of the form given by :

$$z = a_1x_1 + a_2x_2 + \ldots\ldots + a_px_p$$

where, a_1, a_2, ... , a_p are constants.

As we change $a_1, a_2 \ldots\ldots\ldots a_p$ we get different linear functions and we can calculate the variance of any such linear function. The first principal component (PC) is that linear function which has the maximum possible variance; the second PC is the linear function with maximum possible variance subject to being uncorrelated with the first PC; the third PC is the linear function which maximises variance subject to being uncorrelated with the first and second PCs, and so on. Thus, it is easy to construct p principal components providing optimal m-dimensional representation of the data for each $m = 1, 2, \ldots\ldots, p$ for various different definitions of optimality. In particular, at each stage the sum of the variance of the PCs is as large as possible. In other words, with PCA we get for each $m = 1, 2, \ldots, p$ the m linear functions $_1, x_2, \ldots\ldots, x_p$ which account for the maximum possible proportion of the original variations. Expressing in more general terms, the kth principal component is given by :

$$z_k = a_{k1}x_1 + a_{k2}x_2 + \ldots\ldots + a_{kp}x_p$$

for $k=1, 2, \ldots, p$. Here a_k are vectors consisting of the weights to different variables. We compute eigen vectors of the $(p \times p)$ covariance matrix.

Thus the empirical orthogonal function (EOF) is the set of coefficients a_{11} , a_{1p} appearing in the first PC. Similarly subsequent EOF consist of coefficients of x_1 $x_2, \ldots, x_p$ in each successive PC. The first eigen value is the variance of the first PC and so on. To define the principal components uniquely the normalisation constant is imposed. Method of normalisation used in this paper is given by :

$$\sum_{i=1} a^2_{ki} = \frac{1}{\lambda_i}$$

which gives Var $(z_k) = 1$ for all $k = 1, 2, \ldots\ldots, p$.

The primary advantage of the principal component solution is its ability to compress the complicated variability of the original data set into a relatively few temporally uncorrelated components. However, the spatial orthogonality of the eigen vectors is a strong and often undesirable constraint imposed on the principal component solution. While the first principal component and its eigen vector are not influenced by this constraint, the remaining eigen vectors often bear predictable geometric relationships to the first eigen vector.

3. Predictive parameters and data

A number of predictors have been used from time to time to predict the date of onset of the monsoon over the southern tip of India and also the total quantum of rainfall during the period between the first of June and September end. A few of these have been reported by Das (1986) in WMO publication on *Monsoon* which are constantly under revision as and when high correlation with new parameters is established. Gowariker *et al.* (1989, 1991) have adopted 16 parameters to develop parametric and power regression model for prediction of seasonal rainfall. These parameters are broadly divided into four main groups related to temperature, pressure, wind and snow cover keeping in view their possible physical linkages with the monsoon rainfall (Table 1).

Empirical orthogonal functions have been computed to the two sets of observations, namely, the onset date over Kerala and the monsoon rainfall based on the data available in the India Meteorological Department.

4. Results and discussion

The results based on the computation of empirical orthogonal functions are given in Tables 2 and 3 along-with the period for which the data has been analysed.

It may be seen from Table 2 that the first four empirical orthogonal functions explain more than 76% variance. The first EOF gives the highest correlation coefficient of 0.79 with the onset date and explain 31% variance. The influence of Cobar (Australia) and Darwin zonal winds is maximum on this EOF though oppositely correlated while the effect of Delhi and Thiruvananthapuram-Madras winds and Fukuoka (Japan) kinetic energy is least. The second EOF explains about 20% variance with a correlation coefficient of 0.31 with the maximum influence of Delhi wind followed by Thiruvananthapuram-Madras wind. The third EOF explains about 14% variation with correlation coefficient of 0.11 showing least influence of Darwin, Delhi and Forrest zonal wind directions. The influence of Thiruvananthapuram kinetic energy is maximum in the fourth EOF which explains about 12% of the variance.

Table 3 shows that the first four empirical orthogonal functions explain 67% variance with reference to rainfall over country. The first EOF which explains 31% of the variance has the highest correlation coefficient of 0.83 with the monsoon rainfall. Among the 16 parameters used in the model, the 50 hPa ridge over northern hemisphere, Eurasian snow cover and Argentina pressure were negatively correlated. The predominant influence was noted for Balboa 10 hPa

TABLE 4

Comparison of eigen index with the parametric model for predicting the onset dates of southwest monsoon over Kerala

Year	Onset departure from the normal (1 June – in days)	Predicted dates in terms of departure from the normal (1 June – in days)	
		P. C. analysis	Parametric method
1966	+0	+(—0.23)	+(60%)
1967	+8	+(—1.44)	+(70%)
1968	+7	+(—0.64)	+(70%)
1969	—7	—(2.99)	—(30%)
1970	—6	—(0.33)	—(30%)
1971	—5	—(0.52)	+(50%)
1972	+17	+(—5.12)	+(100%)
1973	+3	+(—1.06)	+(70%)
1974	—6	—(3.78)	—(30%)
1975	—1	+(—0.2)	+(80%)
1976	—1	—(0.34)	(70%)
1977	—2	—(1.50)	—(30%)
1978	—4	—(1.28)	—(40%)
1979	+10	+(—0.63)	+(50%)
1980	+0	+(—0.11)	+(50%)
1981	—2	—(0.36)	(50%)
1982	—2	—(1.18)	—(30%)
1983	+12	+(—2.51)	(100%)
1984	—1	—(1.51)	—(20%)
1985	—4	—(1.34)	—(10%)
1986	—3	+(—0.1)	(60%)
1987	—1	+(0.61)	(60%)
1988	—6	—(0.99)	—(30%)
1989	+2	+(—1.56)	(50%)
1990	—4	+(—1.99)	(70%)
	Probability of success	92%	80%

wind, 500 hPa ridge, Eurasian snow cover, north and central India and east coast minimum temperatures, and northern hemispheric temperature. Of these, the influence of the Eurasian snow cover and 50 hPa ridge was found to be the largest. The influence of *El-Nino* in the current year, *El-Nino* in previous year, Darwin pressure (April), equatorial pressure (January to May) is relatively small. The influence of the Himalayan snow cover and Tahiti *minus* Darwin pressure is the least on the first EOF. This is an interesting observation because monsoon circulations are largely determined by the intensity of Walker and Hadley cells, affecting the pressure and wind parameters. The result is nevertheless supported by the fact that during the period 1951-1987 when moderate and severe

TABLE 5

Tropical cyclones in Bay of Bengal *vis-a-vis* onset date (1901 to 1990)

Cyclone onset	May (4th week)		May	
	Storm	No storm	Storm	No storm
Late	10	16	1	25
	(15.5)	(10.5)	(6.1)	(20.9)
Early	21	5	11	15
	(15.5)	(10.5)	(5.9)	(20.1)
χ^2cal	=11.2*		χ^2cal	=9.66*

Note : Figures within brackets are the expected frequencies.
*Significant at 1% level of significance.

El-Nino were observed on 7 occasions, the monsoon was normal during the subsequent years except during 1966.

The influence of the second EOF on monsoon rainfall is relatively small and accounts for about 17% variance with the Darwin pressure, equatorial pressure and Tahiti-Darwin pressure (negatively correlated) followed by Himalayan snow cover, northern hemispheric temperature and east coast minimum temperature besides *El-Nino* in previous and current year. The third EOF gives correlation coefficient of 0.22 and brings out greater influence of *El-Nino* in current year and Himalayan snow cover followed by 50 hPa ridge and northern hemispheric temperature. Fourth EOF, however, shows the greatest influence of *El-Nino* of previous year followed by the *El-Nino* of the current year, central India minimum temperatures, east coast minimum temperature and Argentina pressure and Tahiti *minus* Darwin pressure. The analysis enables to get an insight about the behaviour of the parameters and their inter-dependence.

5. Eigen index as predictant

A comparison of the onset of monsoon namely, early, normal (first June) or late over Kerala for the year 1966 to 1990 is given in Table 4. The composite value derived from the first EOF is defined as eigen index for the onset date. Early onset is expressed as negative while late onset as positive for the eigen index based on principal component analysis. The parametric method is expressed as percentage taking 50% for the normal onset date. In general, there is a good agreement between the two with 92% and 80% accuracy for eigen index and parametric method respectively. During the year 1980, the onset of monsoon was 1 June and the index was nearly zero (0.1) according to the principal component analysis while it was 50% as per parametric model. When the percentage was more than 50 and the eigen index positive, the onset was late over India. Both the eigen indices as well as parametric method showed agreement in 17 cases with

TABLE 6

Comparison of PC model with the parametric model for predicting the southwest monsoon rainfall

Year	Rainfall % departure from normal	Predicted rainfall in terms of departure from normal	
		P. C. model	Parametric model
1958	+10	+(1.50)	+(64%)
1959	+14	+(3.11)	+(78%)
1960	+ 1	+(0.30)	+(57%)
1961	+22	+(3.02)	+(64%)
1962	— 3	—(—1.29)	—(43%)
1963	— 2	—(—0.12)	+(64%)
1964	+10	+(2.27)	+(71%)
1965	—18	—(—4.76)	—(0%)
1966	—13	—(—2.22)	+(57%)
1967	+ 0	—(—0.51)	—(47%)
1968	—10	—(—2.79)	—(20%)
1969	+ 0	+(0.60)	+(60%)
1970	+12	+(1.32)	+(67%)
1971	+ 4	+(0.62)	+(60%)
1972	—24	—(—3.03)	—(20%)
1973	+ 8	+(1.23)	+(60%)
1974	—12	—(—2.06)	—(40%)
1975	+15	+(2.43)	+(80%)
1976	+ 2	+(0.00)	+(60%)
1977	+ 4	+(0.68)	+(60%)
1978	+ 9	—(—1.18)	—(40%)
1979	—19	—(—3.08)	—(13%)
1980	+ 4	+(2.21)	+(60%)
1981	+ 0	+(0.49)	+(53%)
1982	—15	—(—3.02)	—(20%)
1983	+13	+(0.56)	+(60%)
1984	— 4	—(—0.48)	—(40%)
1985	— 7	+(2.68)	+(67%)
1986	—13	+(0.24)	—(40%)
1987	—19	—(—2.39)	—(33%)
1988	+19	+(2.19)	+(87%)
1989	+ 1	+(0.42)	+(73%)
1990	+ 6	+(1.06)	+(53%)

Note : (1) Figures given in brackets under column 3 are the eigen indices based upon the parameters.

(2) Figures given in brackets under column 4 are the percentage of parameters favourable for the good monsoon.

the onset date. During 1990, eigen index as well as parametric method suggested late onset but the monsoon arrived four days earlier. This was attributed to the formation of a tropical cyclone in May which affects the onset date in medium range (Table 4). During 1979, when the monsoon was late, the parametric model suggested almost normal onset while the eigen index gave better results. Thus the eigen index provides a complementary method for long range prediction of onset date of southwest monsoon over India using the same set of parameters.

In order to study the influence of the formation of cyclonic storms during May over the onset date of monsoon over Kerala, the data for the period 1901 to 1990 was examined. It was found that during the year, when a cyclonic storm formed over the Bay of Bengal, there was, in general, an early onset of monsoon. The results were found to be significant using chi-square test at 1% level (Table 5). The inclusion of this parameter for long range prediction alongwith the other parameters was, therefore, examined. It was found that the probability of success decreased to 84% as compared to 92% for the eigen index suggesting the exclusion of this parameter to compute the onset date. However, the probability of success remarkably increased in the medium range when the cyclone formed in the Bay during fourth week of May (Table 5).

It may be noted that out of 33 years, the eigen index as well as parametric method (taking 50% parameters for normal rainfall) give identical results on 25 occasions bringing out normal, below normal, or above normal rainfall correctly (Table 6). During the years 1967, 1978 and 1985 both the methods could not predict the actual rainfall. However, during the years 1963 and 1966 the parametric models showed 64% and 57% of the parameters being favourable but the monsoon was on the negative side. The eigen index, however, showed negative index validating deficit rainfall over the country. The eigen index, however, did not correlate with the rainfall during 1986 when the parametric model showed 40% parameters being favourable implying deficit rainfall. The percentage accuracy works out as 88% and 85% respectively for both the methods and suggests that the eigen index could also be used to provide additional weightage to the conclusions based on the parametric model. During the years, when the parametric model and eigen index show opposite results, greater caution is needed to interpret the results. It may be mentioned that Prasad and Singh (1988) have tried to associate eigen vectors with northern hemispheric surface temperatures (January, February and average for January and February), storm and hurricane frequencies in west Atlantic, *El-Nino* events off the west coast of south America, Darwin sea surface pressure tendencies (May-June), Arctic region (65 to 85°N) surface and temperature (January) and southern hemisphere (2.5°S to 62.5°S) sea surface temperature (January). They found that the strongest association is that with Darwin pressure tendency which is the best predictor of the Indian monsoon fluctuations among the variables examined. The present study has shown negligible influence of the

Darwin pressure on the first EOF, if we consider only 15 parameters. It, however, increases slightly to 0.05 by considering all the 16 parameters (Gowariker *et al.* 1991) suggesting the need to include all of them for operational long range forecasting.

6. Conclusions

The study has brought out the following results :

(*i*) The first four empirical orthogonal functions derived from the parameters for the onset date of monsoon explains more than 77% variance. The influence of Cobar and Darwin (Australia) zonal winds is maximum on the first eigen vector which shows the highest correlation coefficient. The eigen index suggests a complementary method for long range prediction of onset date.

(*ii*) The first four empirical orthogonal functions derived from the 16 parameters for monsoon rainfall explain 67% of variance. Of interest is the small influence of *El-Nino* and Darwin pressure on the first eigen vector which shows the highest correlation with the rainfall.

References

Blanford, H.F., 1884, *Proc. Roy. Soc. London*, **37**, 322.

Bhalme, H.N. and Mooley, D.A., 1980, *Mon. Weath. Rev.*, **108**, 1197.

Das, P. K., 1986, *Monsoon*, WMO No. 613.

Gowariker, V., Thapliyal, V., Sarker, R.P., Mandal, G.S. and Sikka, D.R., 1989, *Mausam*, **40**, 2, 115.

Gowariker, V., Thapliyal, V., Kulshrestha, S. M., Mandal, G.S., Sen Roy, N. and Sikka, D.R., 1991, *Mausam*, **42**, 125.

Kung, E.C. and Sharif, T.A., 1980, *J. Appl. Met.*, **19**, 370.

Kung, E.C. and Sharif, T.A., 1982, *J. Met. Soc. Japan*, Ser. II, **60**, 672.

Paranjpe, S.A. and Gore, A.P., 1991, *Curr. Sci.*, **60**, 446.

Prasad, K.D. and Singh, S.V., 1988, *Advances in Atmos. Sci.*, **5**, 499.

Shukla, J. and Paolino, D., 1983, *Mon. Weath. Rev.*, **111**, 1830.

Thapliyal, V., 1982, *Mausam*, **33**, 399.

Verma, R.K., 1980, *Mon. Weath. Rev.*, **108**, 1072.

Walker, G. T., 1910, Mem. India Met. Dep., **21**, **22**.