Proc. Indian natn. Sci. Acad., **60**, A, No. 1, 1994, pp. **181-192**.

LONG RANGE WEATHER FORECASTING TECHNIQUES IN INDIA USING EOF APPROACH

H N SRIVASTAVA
India Meteorological Department, Pune 411 005, *India*
AND
S S SINGH
India Meteorological Department, New Delhi 110 003, *India*

(*Received* 21 *June* 1993; *Revised* 3 *September* 1993;
Accepted 6 *September* 1993)

Empirical Orthogonal Functions (EOF) have been used to study the interdependence of 16 parameters in power regression and parametric models used by the India Meteorological Department for predicting the summer monsoon rainfall. Alternate models, using an eigen index have been discussed for the country as a whole based on 16 parameters, and for the onset of the monsoon over Kerala using ten other parameters. Distinctive aspects during excess and drought years (± 10%) have been brought out. An attempt has also been made to evolve models for smaller areas using rainfall correlations with eigen indices enabling us to demarcate three broad regions in the country.

Key Words: Long Range Forecast; Empirical Orthogonal Functions; Principal Component Analysis; Southwest Monsoon

Introduction

Seasonal monsoon (June to September) rainfall has a direct bearing on the Indian economy. Attempts have, therefore, been made by the India Meteorological Department to predict monsoon rainfall since 1886, when the first operational Long Range Forecast (LRF) was issued by Blanford[1]. Subsequently, multiple regression techniques based on regional and global data have been used starting with the work of Walker[2], who evolved the concept of a southern oscillation—a seesaw fluctuation of atmospheric pressure from the Indian to the Pacific ocean. The parameters in multiple regression models were changed due to changing correlations between one or the other with the progress of time. Keeping in view these limitations, Thapliyal[3] introduced a dynamic stochastic model which was also used for operational LRF for peninsular India. However, limitations were noticed as the method was based on only one parameter, namely, the position of a 500hPa ridge over peninsular India first reported by Banerjee *et al.*[4], Gowarikar *et al.*[5], evolved a 'parametric model', wherein the favourable parameters out of a total of 16 are expressed as a percentage to, qualitatively, forecast monsoon rains over the country. Gowarikar *et al.*[6] also developed a power regression model based on the same global and regional parameters to provide a quantitative approach. The authors felt that some of the parameters are inter-related, but without presenting a scientific analysis. Keeping in view the considerable discussion among scientists about this aspect, Srivastava and Singh[7] employed Empirical Orthogonal Functions (EOF) to evaluate the interdependence of the parameters in these two models.

This technique was earlier used by Bedi and Bindra[8] to explain different rainfall patterns over India. Iyengar[9] also applied the method to understand the variability of rainfall in Karnataka.

The aim of this paper is to discuss the interdependence of the LRF parameters, and to present models using EOF to predict seasonal rainfall for the country as a whole and for smaller regions using additional data. A model for the monsoon's onset over Kerala has been also proposed using the same technique.

Methodology

Principal Component Analysis (PCA) has been extensively used in Meteorology after Lorenz[10] introduced the technique in 1956. It expresses a meteorological parameter, such as rainfall as a linear combination of orthogonal functions. Expressing different meteorological parameters *m* for a series of *n* years, the eigenvectors of the covariance matrix are computed. The Empirical Ortho-

Table I

Parameters used for long range forecast of southwest monsoon

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

gonal Function (EOF) is the set of coefficients appearing in the first principal component which is the linear function explaining maximum variance. To define the principal components uniquely, the eigenvalues are normalised (Preisendorfer[11]).

Long Range Forecasting Parameters Over India

Das[12] reported Long Range parameters for the onset date as well as the quantum of rainfall in multiple regression models used by the India Meteorological Department. These are constantly reviewed and changed with the passage of time. Table I gives a list of recent predictors for predicting the monsoon onset date and the southwest monsoon rainfall.

Results

Table II provides an analysis of first 4 EOFs, which explain more than 62% of rainfall over the country. Bedi and Bindra[8] however, reported only 47% of variance of 'monsoon rain' through the first 4 EOF's. However, the first EOF which explains 25.9% variance is in agreement with the earlier study. A detailed analysis was undertaken (Table III) for the excessive and deficit monsoon rainfall (> or < 10% of normal rainfall) to find out any distinctive aspects. Table IV presents an analysis for the onset date over Kerala. The correlation matrix of subdivisional monsoon rainfall and EOF for three regions, namely, SE India, NE India and India-(NE + SE) was also worked out and given in Table V.

Eigen Indices corresponding to the meteorological parameters for the onset date as well as southwest rainfall were computed separately from the linear combination of their respective values and the loading factors for each year are interpreted with respect to the actual onset date or the quantum of rainfall (Srivastava and Singh[7]).

Discussion

We note from Table II that the first EOF shows a high correlation of 0.83 with the summer monsoon rainfall for the country if we consider the data of 34 years (1958-91). It accounts for a variance of about 26%. The parameters which make a singificant contribution to explain the variability were found to be (*i*) 50hPa E-W ridge, (*ii*) Eurasian snow cover, (*iii*) 500hPa ridge, (*iv*) 10hPa zonal wind, (*v*) Central India temperature, (*vi*) east coast temperature, (*vii*) northern India temperature, (*viii*) 40° northern hemisphere pressure, and (*ix*) Argentina pressure. Four other parameters, namely, Northern Hemisphere temperature, Equatorial pressure, El-Nino of the current year and the Darwin pressure had relatively lesser influence. The southern oscillation index (Tahiti minus Darwin pressure) and the Himalayan snow cover were found to have least loading followed by the El-Nino of previous year.

Paranjpe and Gore[13] have used a parsimonious logistic regression model and found that only 5 variables namely the 500hPa ridge, 10hPa wind with temperature in northern hemisphere, the temperature in north India, the tem-

Table II

Coefficient vectors for first four EOFs for 16 parameters

Serial Number	E-WRIG	EURASIONSNOW	RIG. 500 MB	ZONAL 10 MB	VID IN MAY	CVMM TN MAR	PRESS 40 J	N-H TEMP J	ARGET PRS	TD PRS	ELNI(P)F	EUQAT PRS	NORT.I.TEM	ELNI(C)	DAR. PRS	HIMAL. SNOW	VARIANCE EXP. %	CORRELATION
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1.	0.38	0.38	0.31	0.31	0.26	0.29	0.30	0.18	0.31	0.07	0.10	0.16	0.25	0.14	0.15	0.08	25.98	0.83
2.	0.00	0.02	0.08	0.01	0.11	0.25	0.04	0.31	0.10	0.45	0.17	0.47	0.17	0.17	0.46	0.31	18.69	0.01
3.	0.42	0.11	0.19	0.02	0.18	0.12	0.19	0.29	0.09	0.25	0.30	0.01	0.43	0.25	0.06	0.44	9.14	0.27
4.	0.00	0.01	0.02	0.09	0.29	0.14	0.26	0.11	0.25	0.20	0.56	0.06	0.16	0.60	0.07	0.08	7.15	−0.15

Table III

EOF Analysis of excess and deficit monsoon years (± .10%)

(*a*) *Deficit Monsoon Years* (1965, 1966, 1972, 1974, 1979, 1982, 1986 &1987)

SERIAL NUMBER	ARGET. PRS.	CINDIA TEMP.	NORT.I.TEMP.	ZONAL 10 MB	T-D SPR.	E.W.RIG.	ELNI(P)	EURASION SNOW	EQUAT. PRES.	ELNI(C)	NH PRESS 40°N	RIG 500 MB	HIMAL. SNOW	C.I. TEMP. MAY	DAR. PRES.	NH TEMP.	VAR. EXPLAINED	CORRELATION
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1.	0.12	0.17	0.01	0.01	−0.26	−0.08	0.22	0.20	0.32	0.37	0.16	−0.40	−0.28	−0.23	0.27	0.40	33	.02
2.	−0.47	−0.36	0.01	−0.06	0.34	0.01	0.04	−0.18	0.31	0.22	0.30	0.12	0.25	−0.34	0.26	−0.07	23	.42
3.	−0.17	−0.19	−0.30	−0.48	0.10	0.47	0.21	0.28	0.13	0.07	−0.38	0.04	−0.24	0.19	0.03	−0.02	18	−.84
(*b*) *Excess Monsoon Years* (1959, 1961, 1970, 1975, 1983 &1988)																		
1.	0.11	−0.22	0.04	−0.25	0.24	0.75	−0.05	−0.11	−0.30	0.17	0.16	−0.28	0.39	0.33	0.35	0.37	30.11	−.47
2.	−0.24	0.10	0.35	0.17	−0.34	0.30	−0.34	−0.30	0.25	0.27	0.34	0.13	0.08	0.26	0.05	0.21	25.77	−.43
3.	0.42	−0.47	−0.04	−0.21	−0.06	0.21	−0.23	0.29	0.29	0.35	−0.18	0.02	−0.13	−0.06	0.31	−0.06	16.51	−.50

Table IV

Eigen vector for long range prediction of summer monsoon onset

	Eigen vectors of parameters											
EOF Parameters	COBAR WIND	THIRU.-MADRAS WIND	DARWIN WIND	DELHI JAN. WIND	FORREST (AUS) JAN. WIND	THIRU.NOV. KINETIC ENRG	FUKUOKA (JAPAN) FEB, KIN ENRG.	MEAN JAN.WIND DARWIN	MEAN MERID. DEC. WIND	KINE. ENRG. JAN. DARWIN	VARIANCE EXPLAINED	CORR. COEF. WITH THE ONSET DATES SW MONSOON OVER KERALA COAST
1	2	3	4	5	6	7	8	9	10	11	12	13
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	0.00	−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

Table V

EOF correlation matrix and broad sub-divisions in India

Broad sub-divisions	Correlation matrix of broad sub-divisional monsoon rainfall and EOF															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
India-(NE&SE)	.84	.02	.14	.18	.18	.01	.01	.15	.14	.21	0.40	.09	.05	.00	.13	0.7
SE India	.36	.18	.28	.39	.03	.04	.16	0.60	.01	.43	.10	.14	.20	.06	.19	.20
NE India	.04	.31	.18	.02	.27	.08	.13	0.70	.23	.31	0.06	.37	.17	.18	.00	.04

perature of the east coast and the interaction between temperature in north India and 10hPa wind may be used to assess the probability of a normal rainfall over the country. Keeping in view the advantage of EOF techniques, whereby the complicated variability of the original data set is reduced to a relatively few uncorrelated components, we found that an eigen index gives similar results as the parametric model on 25 occasions to bring out the correct rainfall. For three years, namely, 1967, 1978 and 1985, the actual rainfall could not be predicted by either of the two methods. During the years 1963 and 1966, a negative index supported deficient rainfall. This suggested an improvement over the parametric model.

It is interesting to note that the second EOF which explains about 19% of the variance has significant influence on the southern oscillation, equatorial pressure and the Darwin pressure, followed by the northern hemispherical temperature, Himalayan snow cover and the east coast temperature. However, the correlation with monsoon rainfall was found to be negligible. The influence of the third and the fourth EOF was about 9% each which gave correlations of 0.27 and −0.15 respectively, thus explaining the total variance of 62% through the first four EOF's. It may however, be noted that Shukla and Paolino[14] and Prasad and Singh[15] found the Darwin pressure tendency from winter to spring to be a good indicator of monsoon rainfall. Although IMD is using Darwin pressure (spring) instead of its pressure tendency, the use of one or two parameters to predict seasonal rainfall has not been successful in the past. Parthasarthy *et al.*[16], using 11 circulation parameters (5 parameters used by Gowariker *et al.*[6] and 6 new parameters) found that three EOF's explain 69% variance with a multiple correlation of 0.82.

By considering the excess and deficit monsoon years (rainfall more than or less than 10% of the normal) separately, it may be noted from Table III that the 3rd EOF shows a significant correlation of −0.84 during the deficit monsoon years. The variance explained was only 18%. But if the second EOF is included, which shows a correlation of −0.42, the variance explained increased to 41%. On the other hand, during excess monsoon years, first three EOF's explain about 72% of the variance and their correlations with rainfall were −0.47, −0.43 and −0.50 respectively. Considering the first EOF during excess monsoon years and in common with drought years based on the third EOF (Table III), it may be noted that northern India temperature, Eurasian snow cover, northern Hemispheric pressure and the El-Nino of the previous year influences the deficit rainfall years. On the other hand, central India and east coast temperatures, equatorial pressure, 500hPa ridge, northern hemispheric temperature, Tahiti-Darwin pressure and Darwin pressure influence excess monsoon years. However, the poor correlation with the first EOF during deficit monsoon years and its exclusion from the above discussion needs to be kept in view.

An attempt was made to identify smaller areas in the country after studying the correlation coefficients of monsoon rainfall with the eigenvectors in different meteorological sub-divisions (Singh *et al.*[17]). The meteorological sub-divisions were grouped together which showed a significant correlation (at 5% level of significance) with the EOF's. It may be noted that although EOF's were

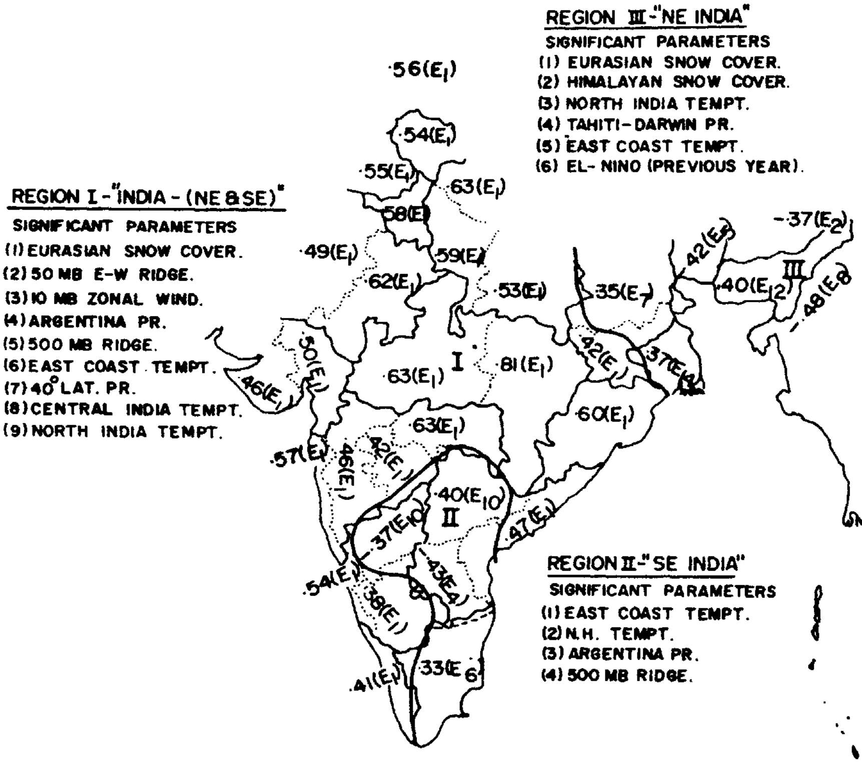

Fig 1 Broad divisions in India based on EOF analysis for Long Range Forecasting

computed for all the parameters, some of the parameters which do not make a significant contribution in the variability of the function were excluded. Thereafter, different regions of the country and the corresponding EOF's were also identified (Fig. 1) such that the interannual variability patterns of the monsoon rainfall and the EOF's were found similar to each other over the respective region. The results are shown in Fig. 1. After detailed analysis of the first EOF and their linear functions, it was found that the main parameters to explain interannual variability in total rainfall during the monsoon season over sub-division I namely, India minus (NE + SE) parts, are the 10hPa zonal wind, 500hPa ridge, 50hPa east west ridge, Vidarbha temperature, east coast temperature, North India temperature, Eurasian snow, Argentina pressure and pressure at 40°N latitudinal belt.

It was interesting to note that the first EOF was not significantly correlated with Tamil Nadu rainfall during this season which provides weightage to the methodology, keeping in view the climatological and synoptic pattern of rainfall. EOF 4 and EOF 10 for the broad sub-division II explained interannual variability through east-coast temperature, northern hemisphere temperature, Argentina pressure and the 500hPa ridge. In the northeastern part of India, EOF 7 (Bihar Plain), EOF 14 (Gangetic West Bengal), EOF 2 (Arunachal Pradesh), EOF 12 (Assam and Meghalaya), EOF 5 (Sub Himalayan Gangetic

West Bengal), and EOF 8 (Nagaland, Manipur and Tripura) gave the following predominant factors: as (*i*) Tahiti-Darwin pressure, (*ii*) North India temperature, (*iii*) East coast temperature, (*iv*) Eurasian snow, (*v*) Himalayan snow and (*vi*) El-Nino (previous year).

It may be mentioned that Bedi and Bindra[8] have shown an opposition in the rainfall pattern between northeast and western sectors of India. The present study brings out that maximum correlation has been found with the first EOF in northwest India while over the northeast India EOF's 2, 5, 7, 8, 12 and 14 show higher correlation. Expressed through regional and global parameters, and excluding the common parameters, it may be noticed from Fig. 1 that the northwestern sector of India has predominant loading due to (*i*) 50hPa E-W ridge, (*ii*) 10hPa zonal wind, (*iii*) Argentina pressure, (*iv*) 50hPa ridge, (*v*) 40°N latitude Pressure, and (*vi*) Central India temperature. On the other hand, northeast India has loading influence due to (*i*) Himalayan snow cover, and (*ii*) Tahiti-Darwin pressure and El-Nino (previous year). It is interesting to note that Gadgil *et al.*[18] found that the variation of the monsoon rainfall over the Indian region is characterised by two spatial scales corresponding to about 11 and 31 zones based on EOF analysis. However, the authors suggested further studies to understand the mechanism for two scales in relation to the dynamics of the monsoon. As the present paper is based on the rainfall predictive parameters, the present study to divide the country into three broader zones appears to be justified.

EOF analysis was also applied for the onset of monsoon over Kerala based on parameters given in Table I. It shows that the eigen index derived from the first EOF gives 92% accuracy as compared to 80% based on the parametric method.

Future Developments

Although parameteric and power regression models have provided more accurate forecasts during the last 5 years but they are statistical in nature and in some year may provide divergent forecast (Gowariker *et al.*[5,6]). Considering the atmosphere as a complex dynamical system which transforms different inputs into outputs, such as the monsoon rainfall, transfer dynamic relationship between the continuous input and output series has been expressed by a differential equation[3], (given by seasonal rainfall) with 500hPa as the lead forcing to the atmosphere (Banerjee *et al.*[4]). During recent years, this model gave lesser accuracy, if we compare it with the power regression model. For example, the official forecast based on this model during 1990 and 1991 when compared with actual rainfall gave a difference of 10% and 7% during the last two years, which is contrary to the results of Thapliyal and Kulshrestha[19]. Considerable efforts to include other forcing parameters may be needed to improve the model. Another dynamical approach based on deterministic chaos found lower accuracy in prediction, which was attributed to limited rainfall observations (Kulkarni and Verma[20]).

Among the new parameters considered for the monsoon rainfall prediction over India could be sea surface temperature over the Arabian Sea, Bay of

Bengal and Indian Ocean. On extension of EOF analysis for these regions in relation to two contrasting sets of monsoon years, Balakrishnan *et al.*[21] found that higher eigenvectors in May over northeast Arabian sea may signal good monsoon and vice versa. Also the temperature gradient over the Bay of Bengal shows a contrast between excess and deficient monsoons for which satellite observations through OLR (outgoing longwave radiation) could provide data. A question arises whether a regional monsoon pattern could be influenced by persistent volcanic dust in the stratosphere emitted through volcanic eruptions similar to those of 1991 in the Phillipines. This aspect needs investigation in relation to LRF.

Development of long range forecast models for smaller areas over the Indian region, based on EOF analysis suggests that new predictive parameters for the country, excluding the northeast and southeast parts (Fig. 1) can be identified by the first EOF. However, in northeast India, where the rainfall variability is only 20% or slightly less, different EOF's have been found to correlate with the rainfall implying complexity in choosing predictive parameters. Considerable research is, therefore, needed to evolve models for smaller areas, namely, the three broad divisions in the country (Fig. 1) through the identification of more physically linked regional parameters.

The Climate Analysis Centre, Washington employs a modified statistical technique called Canonical Correlation Analysis (CCA) to predict Sea Surface Temperature (SST) for the central Pacific area, which has been found to be the most useful in long range prediction with 1 season, 2 seasons and 3 seasons lead forecasts (Ropelewski and Barnston[22]). However, their accuracy at present is slightly better or almost similar to numerical models. Improvements in coupled ocean atmosphere models and their testing over longer periods are still needed before they can be used for routine operational LRF.

Acknowledgement

The authors are grateful to Dr N Sen Roy, Director-General of Meteorology for permission to publish the paper. The authors also express their gratitude to Dr P K Das, FNA for helpful suggestions.

References

1 H F Blanford *Proc R Soc London* **37** (1884) 322
2 G T Walker *Mem India Met Dept* (1910) 21
3 V Thapliyal *Mausam* **33** (1982) 399
4 A K Banerjee, P N Sen and C R V Raman *Indian J Met Hydrl Geophys* **29** (1978) 425
5 V Gowarikar, V Thapliyal, S M Kulshrestha, G S Mandal, N Sen Roy and D R Sikka *Mausam* **42** (1991) 25
6 V Gowarikar, V Thapliyal, R P Sarkar, G S Mandal and D R Sikka *Mausam* **40** (1989) 115
7 H N Srivastava and S S Singh *Mausam* **44** (1993) 29
8 H S Bedi and M M S Bindra *Tellus* **32** (1980) 296
9 R N Iyengar *Proc Indian Acad Sci (Earth Planet Sci)* **100** (1991) 105
10 E N Lorenz *Empirical Orthogonal Functions and Statistical Weather Prediction.* Sc. Report I. Statistical Forecasting Project, M.I.T., Deptt. of Meteorology, U.S.A. (1956) 49
11 R W Preisendorfer *Principal Component Analysis in Meteorology and Oceanography* Elsevier Publns, New York (1988)