# Long range prediction of Indian summer monsoon rainfall

A A Munot and K Krishna Kumar

*Indian Institute of Tropical Meteorology, Pune 411 008, India.*

The search for new parameters for predicting the all India summer monsoon rainfall (AISMR) has been an important aspect of long range prediction of AISMR. In recent years NCEP/NCAR reanalysis has improved the geographical coverage and availability of the data and this can be easily updated. In this study using NCEP/NCAR reanalysis data on temperature, zonal and meridional wind at different pressure levels, few predictors are identified and a prediction scheme is developed for predicting AISMR. The regression coefficients are computed by stepwise multiple regression procedure. The final equation explained 87% of the variance with multiple correlation coefficient (MCC), 0.934. The estimated rainfall in the El-Niño year of 1997 was $-1.7\%$ as against actual of 4.4%. The estimated rainfall deficiency in both the recent deficient years of 2002 and 2004 were $-19.5\%$ and $-8.5\%$ as against observed $-20.4\%$ and $-11.5\%$ respectively.

## 1. Introduction

The agricultural economy of India is closely linked to the performance of summer monsoon rainfall over India. Variations in the total amount of rainfall have strong socio-economic consequences. Parthasarathy *et al* (1988a) for the first time noted the positive relationship between crop yields and Indian summer monsoon rainfall. In view of the dependence of crop yield on monsoon rainfall, forecasting of monsoon rainfall becomes an issue of immense importance. An accurate long-lead prediction of summer monsoon rainfall will help to improve the planning to mitigate the adverse impact in case of monsoon failure. The first official seasonal monsoon forecast was issued by Sir Henry Blanford in 1886 which was based entirely on Himalayan snowfall. John Eliot used extra-Indian factors, viz., pressure of Mauritus, Zanzibar and Seychelles in the monsoon forecast of 1896. Sir Gilbert Walker who laid the basis for a forecast on a statistical association, was the first meteorologist who systematically examined the relationship between Indian monsoon rainfall and global circulation parameters and selected 28 predictors to issue forecast based on regression equation during the year 1906 (Jagannathan 1960; Rao and Rama Murthy 1960; Rao 1965). Savur (1931) showed that 7 out of 28 parameters have lost their significance in course of time. Since then, extensive research work has been done on empirical seasonal forecasting of Indian summer monsoon rainfall. Some of the noteworthy studies are those of Banerjee *et al* (1978); Kung and Sharif (1982); Bhalme *et al* (1986); Gowariker *et al* (1989, 1991); Parthasarathy *et al* (1988b, 1991); Krishna Kumar *et al* (1995, 1997); Rajeevan *et al* (2004), etc. In spite of many efforts in the long range prediction of all-India summer monsoon rainfall (AISMR) the success we have achieved is not adequate and there is much scope to investigate new predictors. The present study is an attempt in this direction. In this study using data on temperature, zonal and meridional wind at different pressure levels, few predictors are identified and a prediction scheme is developed for the prediction of AISMR.

## 2. Data and methodology

In this study NCEP/NCAR reanalysis data on mean sea level pressure, temperature, zonal and

Table 1. *Predictors for AISMR.*

| Predictors | Month/ season | Location | CC with AISMR (1961–1995) |
|---|---|---|---|
| $p1$: 1000 hPa temperature (K) | PJJA | 20N–25N, 62.5E–67.5E | 0.51 |
| $p2$: 850 hPa temperature (K) | PJJA | 30N–35N, 112.5E–117.5E | −0.63 |
| $p3$: 700 hPa temperature (K) | POCT | 15N–20N, 15E–25E | 0.65 |
| $p4$: 500 hPa zonal wind (m/s) | MAY | 5N–10N, 105W–95W | −0.55 |
| $p5$: 500 hPa meridional wind (m/s) | PJJA | 15N–25N, 180–170W | −0.60 |
| $p6$: 200 hPa meridional wind (m/s) | PNOV | 10N–17.5N, 112.5E–117.5E | −0.64 |

meridional wind for 1000, 850, 700, 500 and 200 hPa level for the period 1949–2005 have been used. The NCEP/NCAR data is available on $2.5 \times 2.5$ grid size. The monthly all-India area weighted rainfall series for the period 1949–2005, as prepared by Parthasarathy *et al* (1995) and updated subsequently in IITM is used in this study.

### 2.1 *Identification of the predictors*

For identification of predictors and to prepare spatial correlation maps, we have used NCEP/NCAR reanalysis and AISMR data for the period 1949–2003. In addition to monthly data for the period 1949–2003, we have also prepared seasonal series for all the parameters as mentioned above for the seasons June–July–August (JJA), September–October–November (SON), December–January–February (DJF) and March–April–May (MAM) seasons. The DJF to MAM tendency (MAM–DJF) series are also prepared. As we are interested in identifying the predictors for AISMR with sufficient lead time, we have selected May as the cut-off month and monthly/seasonal series for each of the months June to May prior to monsoon season as well as JJA, SON, DJF, MAM and MAM–DJF series prior to monsoon are correlated with AISMR. Spatial correlation maps are prepared for all the parameters and the regions having significant correlation with AISMR are identified. The statistical significance of the correlation coefficients (CCs) was evaluated after taking the serial correlations (or persistence) of the series involved for adjusted number of degrees of freedom (Quenouille 1952). The parameters which are strongly related to AISMR at 1% level of significance are identified as the predictors for AISMR. In this way about 73 predictors are identified.

### 2.2 *Prediction of AISMR*

Linear regression analysis is the most commonly used statistical technique for long range forecasting (LRF) of AISMR. A large number of regression models are proposed so far which include simple as well as multiple regression methods. In the present study, a stepwise multiple regression procedure based on the algorithms of Jennrich (1977) which is an objective way to select the predictors has been used with 73 predictors as identified above. An optimum number of parameters in a statistical regression model has been a matter of debate. More number of parameters in the final equation will improve the variance but it will decrease the degrees of freedom which in turn will decrease the accuracy in the prediction during the independent period. In this scheme the first variable entered the regression equation with multiple correlation coefficients, 0.65 and variance explained 43%. On subsequent steps up to six steps, increase in variance explained is 22%, 11%, 6%, 3%, 2% respectively. Afterwards increase in variance explained becomes 1% or less. In view of this, in this scheme only six predictors are allowed to enter the regression scheme. The regression equation is based on the training period 1961–1995. The mean and standard deviation of rainfall for training period 1961–1995, are 838.4 mm and 89.3 mm respectively. The period 1949–1960 and 1996–2005 is used for independent verification of the regression equation. The final multiple regression equation is:

$$\text{AISMR} = -7417.7 + 48.87934 \times p1$$
$$- 45.56182 \times p2 + 23.98553 \times p3$$
$$- 10.45115 \times p4 - 28.85484 \times p5$$
$$- 21.00083 \times p6$$

where AISMR is the all-India summer monsoon rainfall (mm/season) and the predictors $p1, p2, \ldots, p6$ are indicated table 1 (in column 2, PJJA, POCT, PNOV, ..., P denotes JJA prior to monsoon, October prior to monsoon, ..., etc.).

This equation explained 87% of the variance with MCC, 0.934, F-ratio, i.e., ratio of the variance explained by regression to the residual variance, 31.8 significant at 1% level and standard error of estimate 36 mm. Spatial correlation maps of the
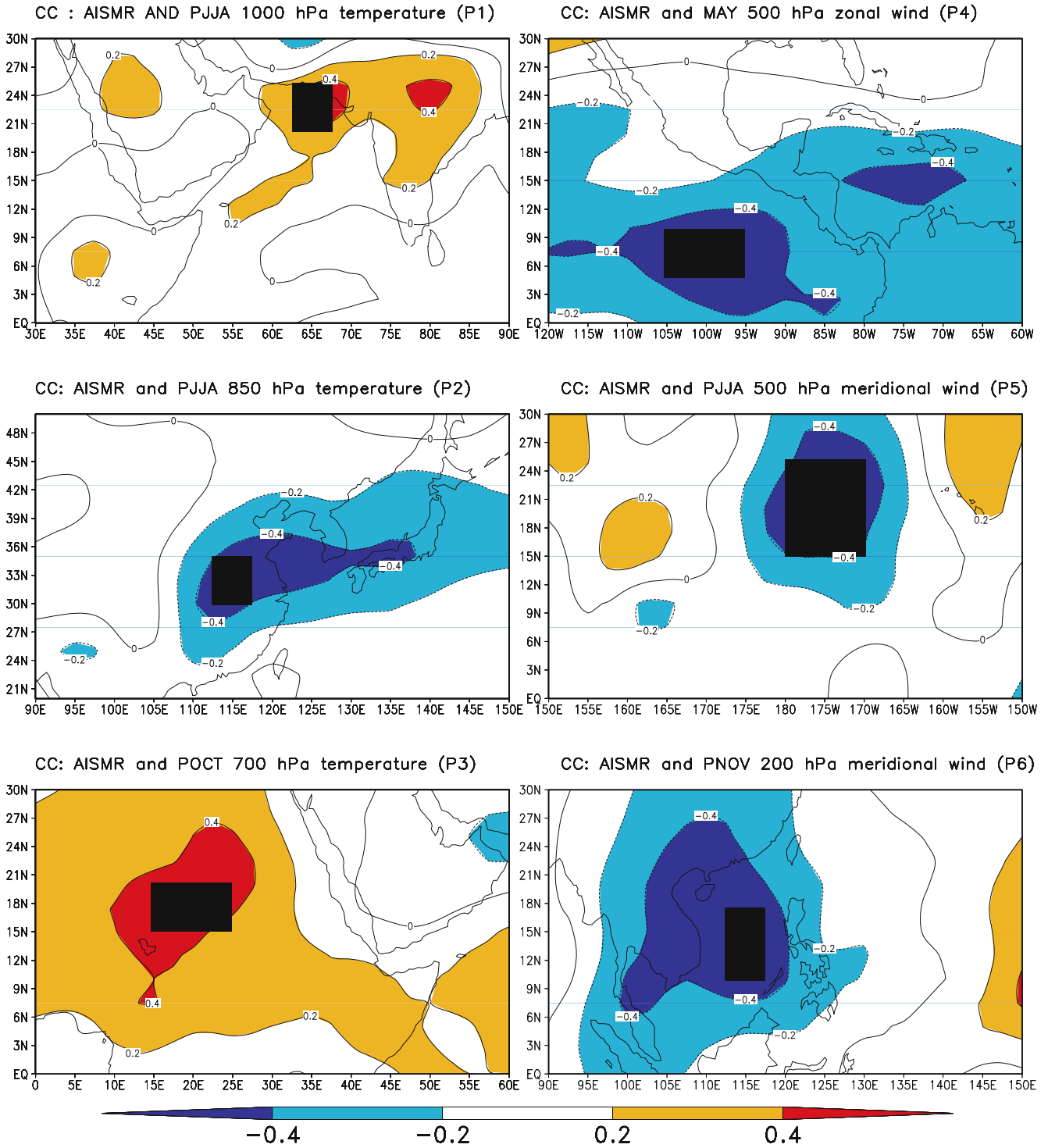
Figure 1.   CC between AISMR and predictors.

six parameters, and the location (dark squared region) from which the predictors entered in to regression equation are extracted, are shown in figure 1.

To check the multi-collinearity between the predictors we have computed correlation matrix between the predictors and presented it in table 2.

We are aware that there is scope for multi-collinearity which can be seen from the cross-correlation matrix shown. However, inclusion or exclusion of any additional variable in to regression model using Jennrich approach is based on the partial-correlation of the new variable with respect to the already included variables. This process, in

Table 2.  *Correlation coefficients among the predictors.*

| Predictors | $p1$ | $p2$ | $p3$ | $p4$ | $p5$ | $p6$ |
|---|---|---|---|---|---|---|
| $p1$ | 1.000 | | | | | |
| $p2$ | −0.194 | 1.000 | | | | |
| $p3$ | 0.382* | −0.462** | 1.000 | | | |
| $p4$ | −0.122 | 0.317 | −0.319 | 1.000 | | |
| $p5$ | −0.489** | 0.314 | −0.452** | 0.168 | 1.000 | |
| $p6$ | −0.160 | 0.216 | −0.305 | 0.361* | 0.276 | 1.000 |

*Significant at 5%; **significant at 1%.

Table 3.  *Examination of the relationship between predictors and AISMR: 1949–2005.*

| | | Relationship between AISMR and | | | | | |
|---|---|---|---|---|---|---|---|
| | | $p1$ | $p2$ | $p3$ | $p4$ | $p5$ | $p6$ |
| 1 | Percentage of total number of years under group $F$ | 63.2 | 71.9 | 66.7 | 71.9 | 73.7 | 68.4 |
| 2 | Sum of the products of the variables for years under group $F$ | 33.2 | −36.1 | 32.9 | −34.8 | −38.1 | −38.4 |
| 3 | Sum of the products of the variables for years under group $UN$ | −8.7 | 5.4 | −6.5 | 4.3 | 4.8 | 4.4 |
| 4 | Numerical value of the ratio of the sum (2) to sum at (3) | 3.8 | 6.9 | 5.1 | 8.1 | 7.9 | 8.7 |
| 5 | Correlation coefficient | 0.430 | −0.537 | 0.463 | −0.535 | −0.583 | −0.596 |

our opinion, eliminates multi-collinearity problem to a large extent though not fully.

## 2.3 *Critical examination of relationship between each of the predictors and AISMR*

A critical examination of the correlation coefficients of each of the six predictors which are entered into regression equation and AISMR would be useful in getting an insight into the relationships between predictors and AISMR. In this section we have critically examined the relationship between AISMR and each of the six predictors. We have considered data of the standardized variables. When we are dealing with the two standardized variables then,

$$\text{Covariance} = (1/n) \sum \text{ product of standardized}$$

$$\text{variables}$$

$$= \text{Correlation coefficient}$$

$n$ is the number of years of the data of the variables.

We divide the total number $n$ of the years into two groups, $UN$ and $F$, $UN$ is the group of years for which data are observed to be unfavourable for the relationship and $F$ is the group of years for which data are observed to be favourable for the relationship. The number of years under group

Table 4.  *Number of parameters favourable (F)/ unfavourable (UN) to deficient and excess rainfall during deficient and excess rainfall years respectively.*

| Deficient rainfall | | | Excess rainfall | | |
|---|---|---|---|---|---|
| | No. of parameters | | | No. of parameters | |
| Year | $F$ | $UN$ | Year | $F$ | $UN$ |
| 1951 | 5 | 1 | 1956 | 5 | 1 |
| 1965 | 5 | 1 | 1959 | 5 | 1 |
| 1966 | 5 | 1 | 1961 | 6 | 0 |
| 1968 | 6 | 0 | 1970 | 6 | 0 |
| 1972 | 6 | 0 | 1975 | 6 | 0 |
| 1974 | 5 | 1 | 1983 | 6 | 0 |
| 1979 | 6 | 0 | 1988 | 6 | 0 |
| 1982 | 5 | 1 | 1994 | 5 | 1 |
| 1985 | 4 | 2 | | | |
| 1986 | 5 | 1 | | | |
| 1987 | 5 | 1 | | | |
| 2002 | 6 | 0 | | | |
| 2004 | 4 | 2 | | | |

$F$ expressed as a percentage of the total number of years is computed for each of the relationships. Sum of the products of the two standardized variables is obtained for the years under each of the groups $UN$ and $F$. Finally we computed the numerical value of the ratio of the sums of the
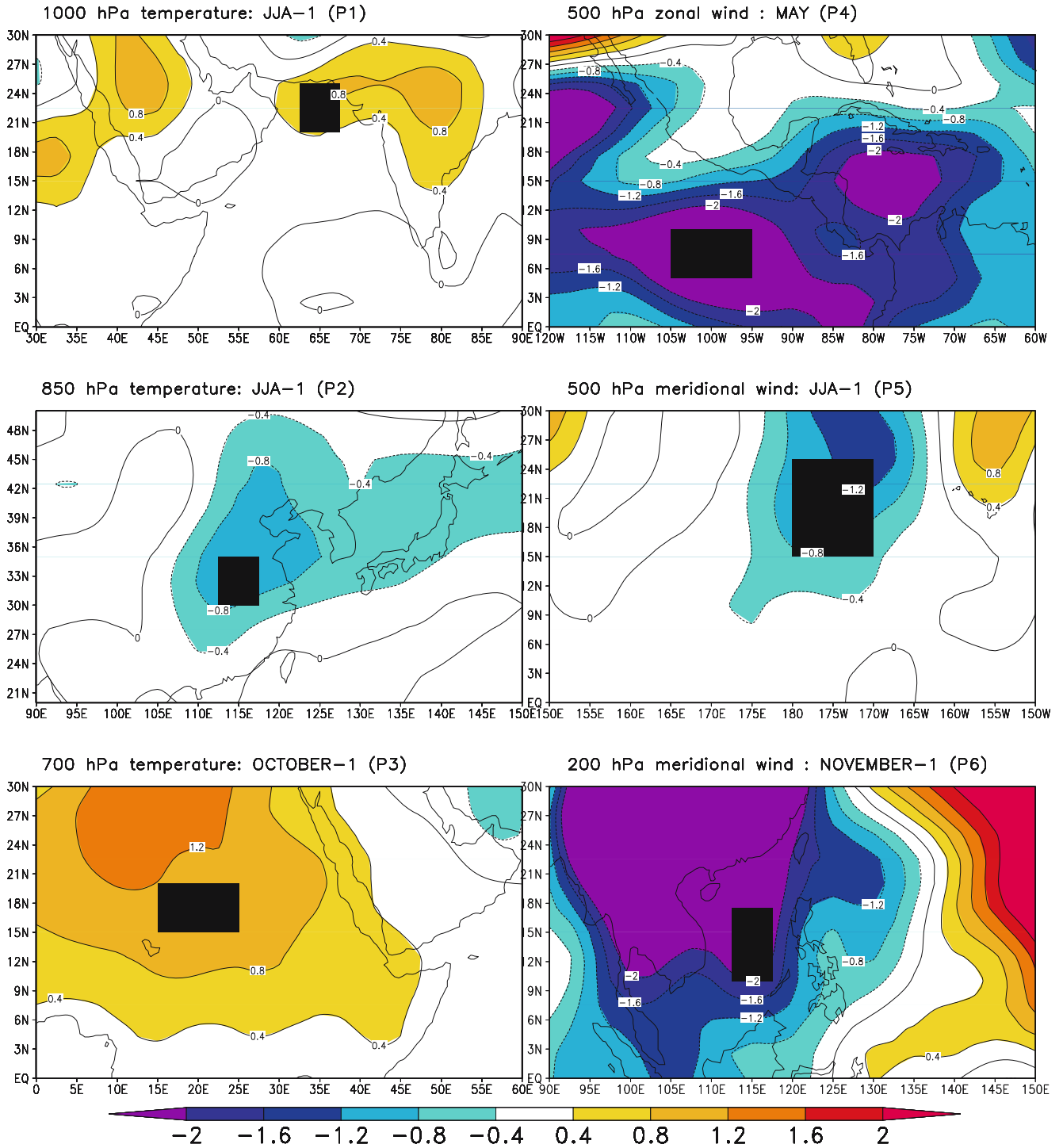
Figure 2. Composite excess–deficient year anomaly.

products of the standardized variables for years in group $F$ to that in group $UN$. Table 3 gives for each of the six relationships, the percentage of years under group $F$, sum of the products of the variables for years under group $F$, sum of the products of the variables for years under group $UN$, numerical value of the ratio of the sum of the products of the variables for years under group $F$ to

that under group $UN$ and the correlation coefficient between AISMR and the variable. This table shows that for quite a large percentage of years the data are favourable to the relationships, and that for the years for which data are favourable to the relationship the sum of the products of the two variables, i.e., contribution to the relationship is very large in comparison to the

sum of the products of the two variables for the years unfavourable to the relationship. This clearly brings highly significant relationships between AISMR and each of the six predictors and the potential utility of the predictors for prediction of AISMR.

Total number of parameters making $F/UN$ contribution to the relationship, only for deficient and excess rainfall years are tabulated in table 4. It is observed that for all the deficient years out of six predictors four or more (more than 60 per cent) predictors were favourable for deficient rainfall whereas for all the excess rainfall years more than 80% of the predictors were favourable for excess rainfall.

To get more insight into the relationship between each of the predictors and the AISMR, the composite anomalies of the predictors during excessive monsoon years (rainfall more than mean plus one standard deviation) and deficient monsoon years (rainfall less than mean minus one standard deviation) are computed. The composite anomalies for excess minus deficient rainfall years are also computed for each of the predictors and plotted in figure 2. It is clearly seen from figure 2 that for the predictors which are positively (negatively) related with AISMR, composite excessive minus deficient years anomalies are strongly positive (negative). (In figure 2, JJA-1 means JJA of previous year, i.e., PJJA and so on). The significance of the difference between excess (8) and deficient (12) years anomalies are tested for significance by t-test for all the six predictors. The numerical t-values for predictors $p1, p2, p3, p4, p5$ and $p6$ are 4.2, 4.8, 4.6, 5.9, 4.1 and 3.6 respectively, which are significant (above 2.88, for 18 degrees of freedom) at 1% level which also strongly support the highly significant relationship between each of the predictors and AISMR.

## 3. Validation of the regression equation

To see the validity of the regression equation, its performance is tested for the independent period 1949–60 and 1996–2005. Four statistics are computed to measure the skill of the forecast for the test period. These are mean absolute error (ABSE), root mean square error (RMSE), Bias and Heidke score. These are computed as follows:

$$\text{RMSE} = \left[ \sum (\text{EST} - \text{OBS})^2/n \right]^{1/2}$$

$$\text{ABSE} = \sum |\text{EST} - \text{OBS}|/n$$

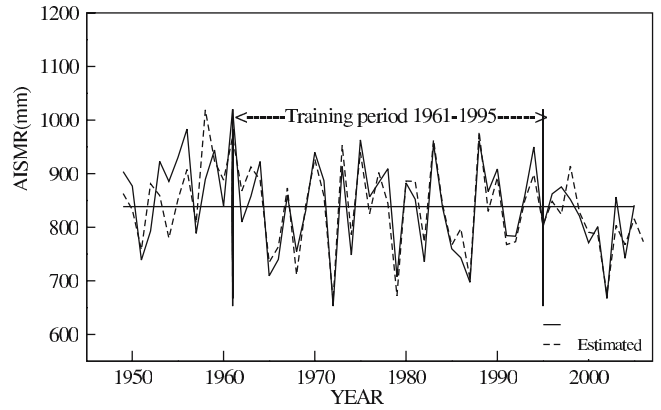$$\text{BIAS} = \sum (\text{EST} - \text{OBS})/n$$



Figure 3. Observed and estimated AISMR: 1949–2005. Continuous line corresponds to actual and dotted line corresponds to estimated rainfall.

Table 5. *Independent verification of regression equation.*

| Year | Observed rainfall (% departure) | Estimated rainfall (% departure) |
|---|---|---|
| 1949 | 7.8 | 2.9 |
| 1950 | 4.6 | −0.7 |
| 1951 | −11.9 | −9.6 |
| 1952 | −5.4 | 5.1 |
| 1953 | 10.1 | 2.5 |
| 1954 | 5.6 | −7.0 |
| 1955 | 10.9 | 1.6 |
| 1956 | 17.3 | 8.3 |
| 1957 | −6.0 | −3.7 |
| 1958 | 6.0 | 21.6 |
| 1959 | 12.6 | 10.2 |
| 1960 | 0.2 | 5.9 |
| 1996 | 2.9 | 1.2 |
| 1997 | 4.4 | −1.7 |
| 1998 | 1.6 | 9.1 |
| 1999 | −2.1 | −1.2 |
| 2000 | −8.0 | −5.7 |
| 2001 | −4.4 | −6.1 |
| 2002 | −20.4 | −19.5 |
| 2003 | 2.1 | −4.2 |
| 2004 | −11.5 | −8.5 |
| 2005 | 0.3 | −2.7 |

$$\text{Heidke score (\%)} = \frac{\text{Number of correct forecasts}}{\text{Total number of forecasts}}$$
$$\times 100.$$

where EST is estimated rainfall and OBS is observed rainfall.

It is found that for independent period 1949–1960 and 1996–2005, ABSE = 46 mm, RMSE = 57 mm and Bias = −7 mm, both RMSE and ABSE are less than the standard deviation. The ABSE, RMSE and Bias for training period 1961–1995 are

27 mm, 36 mm and 0.0 mm respectively. To calculate Heidke score we consider forecast to be correct if the error between actual and estimated rainfall is less than 0.6 times the standard deviation of the rainfall. The Heidke score for independent period 1949–1960 and 1996–2005 is 68%, for total period 1949–2005, it is 81% and for the recent 10 independent years 1996–2005 it is 90%. Thus, the performance of the model is good and even better for recent years. Using this prediction equation the forecast of AISMR for the year 2006 is found to be 92% of the normal. The actual and estimated rainfall is plotted in figure 3 (prediction for 2006 is also shown in the figure). The CC between actual and estimated rainfall is 0.86. From figure 3, it is seen that there is close agreement between actual and estimated rainfall. In table 5 actual and estimated rainfall during independent period 1949–1969 and 1996–2005 is given.

The estimated rainfall in the El-Niño year of 1997 was $-1.7\%$ as against actual of 4.4%. The estimated rainfall in both the recent deficient years of 2002 and 2004 were $-19.5\%$ and $-8.5\%$ as against observed $-20.4\%$ and $-11.5\%$ respectively.

## 4. Summary and conclusions

Using NCEP/NCAR data, an attempt is made to develop a prediction scheme for Indian Summer Monsoon Rainfall in hind cast mode. With only six predictors the prediction equation could explain 87% of the variance with MCC, 0.934. From figure 3, it is observed that there is close agreement between actual and estimated rainfall. The estimated rainfall in the El-Niño year of 1997 was $-1.7\%$ as against actual of 4.4%. The estimated rainfall in both the recent deficient years of 2002 and 2004 were $-19.5\%$ and $-8.5\%$ as against observed $-20.4\%$ and $-11.5\%$ respectively.

## References

Banerjee A K, Sen P N and Raman C R V 1978 On foreshadowing southwest monsoon rainfall over India with mid-tropospheric circulation anomaly of April; *Indian J. Meteorol. Hydrol. Geophys.* **29** 425–431.

Bhalme H N, Jadhav S K, Mooley D A and Ramana Murty Bh V 1986 Forecasting of monsoon performance over India; *J. Climatol.* **6** 347–354.

Gowariker V, Thapliyal V, Sarkar R P, Mandal G S and Sikka D R 1989 Parametric and power regression models – New approach to long range forecasting; *Masaum* **40** 115–122.

Gowariker V, Thapliyal V, Kulshrestha S M, Mandal G S, Sen Roy N and Sikka D R 1991 A power regression models for long range forecast of southwest monsoon rainfall over India; *Masaum* **42** 125–130.

Jagannathan P 1960 Seasonal forecasting in India: A review; Published by India Meteorological Department Pune, 67 pp.

Jennrich R 1977 Stepwise regression; In: Statistical models for digital computers; (eds) Euslering K, Relston A and Wilf H S (New York: John Wiley) 58–75.

Krishna Kumar K, Soman M K and Rupa Kumar K 1995 Seasonal forecasting of Indian summer monsoon: A review; *Weather* **12** 449–467.

Krishna Kumar K, Rupa Kumar K and Pant G B 1997 Pre-monsoon maximum and minimum temperature over India in relation to the summer monsoon rainfall; *Int. J. Climatol.* **17** 1115–1127.

Kung E C and Sharif T A 1982 Long-range forecasting of the Indian summer monsoon onset and rainfall with upper air conditions; *J. Meteorol. Soc. Japan* **60** 672–681.

Parthasarathy B, Munot A A and Kothawale D R 1988a Regression model for eastimation of Indian food grain production from Indian summer monsoon rainall; *Agric. For. Meteorol.* **42** 167–182.

Parthasarathy B, Diaz H Z and Eischeid J K 1988b Prediction of All-India summer monsoon rainfall with regional and large scale parameters; *J. Geophys. Res.* **93(D5)** 5341–5350.

Parthasarathy B, Rupa Kumar K and Deshpande V R 1991 Indian summer monsoon rainfall and 200 mb meridional wind index: Application for long range prediction; *Int. J. Climatol.* **11** 165–176.

Parthasarathy B, Munot A A and Kothawale D R 1995 Monthly and seasonal rainfall series for All-India homogeneous regions and meteorological subdivisions: 1871–1994; IITM Research Report No. RR-065.

Quenouille M H 1952 Associated measurement; (London: Butterworths) pp. 242.

Rao K N 1965 Seasonal forecasting-India; Proc. of symp. on 'Research and development aspects of long-range forecasting' WMO-IUGG Tech. Note No. 66 WMO-No. 162-TP-79 World Meteorological Organisation Geneva, pp. 17–30.

Rao K N and Rama Moorthy K S 1960 Seasonal (Monsoon) rainfall forecasting in India; Proc. of symp. on 'Monsoon of the World', held at New Delhi Feb. 1958, Published by India Meteorological Department, New Delhi, pp. 237–250.

Rajeevan M, Pai D S, Dikshit S K and Kelkar R R 2004 IMD's new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003; *Curr. Sci.* **86(3)** 422–431.

Savur S R 1931 The seasonal forecasting formulae used in the India Meteorological Department. Scientific Notes; Vol. 4 No. 37 Published by India Meteorological Department, New Delhi, pp. 57–68.