Meteorol Atmos Phys 92, 33–43 (2006)
DOI 10.1007/s00703-005-0136-9

Meteorology and Atmospheric Physics
Printed in Austria

India Meteorological Department, Pune, India

Empirical prediction of Indian summer monsoon rainfall with different lead periods based on global SST anomalies

D. S. Pai and **M. Rajeevan**

With 3 Figures

Received July 5, 2004; revised September 28, 2004; accepted January 20, 2005
Published online: September 15, 2005

Summary

The main objective of this study was to develop empirical models with different seasonal lead time periods for the long range prediction of seasonal (June to September) Indian summer monsoon rainfall (ISMR). For this purpose, 13 predictors having significant and stable relationships with ISMR were derived by the correlation analysis of global grid point seasonal Sea-Surface Temperature (SST) anomalies and the tendency in the SST anomalies. The time lags of the seasonal SST anomalies were varied from 1 season to 4 years behind the reference monsoon season. The basic SST data set used was the monthly NOAA Extended Reconstructed Global SST (ERSST) data at $2^\circ \times 2^\circ$ spatial grid for the period 1951–2003. The time lags of the 13 predictors derived from various areas of all three tropical ocean basins (Indian, Pacific and Atlantic Oceans) varied from 1 season to 3 years. Based on these inter-correlated predictors, 3 predictor sub sets A, B and C were formed with prediction lead time periods of 0, 1 and 2 seasons, respectively, from the beginning of the monsoon season. The selected principal components (PCs) of these predictor sets were used as the input parameters for the models A, B and C, respectively. The model development period was 1955–1984. The correct model size was derived using all-possible regressions procedure and Mallow's "Cp" statistics.

Various model statistics computed for the independent period (1985–2003) showed that model B had the best prediction skill among the three models. The root mean square error (RMSE) of model B during the independent test period (6.03% of Long Period Average (LPA)) was much less than that during the development period (7.49% of LPA). The performance of model B was reasonably good during both ENSO and non–ENSO years particularly when the magnitudes of actual ISMR were large. In general, the predicted ISMR during years following the El Niño (La Niña) years were above (below) LPA as were the actual ISMR. By including an NAO related predictor (WEPR) derived from the surface pressure anomalies over West Europe as an additional input parameter into model B, the skill of the predictions were found to be substantially improved (RMSE of 4.86% of LPA).

1. Introduction

The long range prediction of the seasonal Indian Summer Monsoon Rainfall (ISMR) has been one of the first targets of the tropical climate prediction research. However, for more than one century, the prediction for the ISMR has been mainly based on empirical models (Walker, 1914; 1923; Thapliyal, 1982; Gowariker et al, 1989; 1991; Hastenrath, 1995; Krishna Kumar et al, 1995; Navone and Cecatto, 1995; Singh and Pai, 1996; Rajeevan et al, 2000; 2004; Delsole and Shukla, 2002; Sahai et al, 2002; 2003 etc.). In spite of the inherent problems in these empirical models such as epochal variation in the predictand-predictor relationship, inter-correlation between the predictors, changing predictability etc., the main approach towards the prediction of ISMR is still based on empirical models. This is because the alternate approach of prediction based on dynamical models has not yet shown the level of skill required to accurately simulate salient features of

the mean monsoon and its interannual variability (Latif et al, 1994; Gadgil and Sajini, 1998; Goddard et al, 2001; Kang et al, 2002).

The year to year variation of ISMR is attributed to its natural variability and to its association with slowly varying climate boundary conditions (Charney and Shukla, 1981). Among the slowly varying boundary conditions, sea surface temperatures (SSTs) have the most significant association with the ISMR as SSTs are strongly coupled to the deep convection and thereby to the large scale dynamics of the atmosphere. The association between SSTs in the Pacific Ocean and ISMR has been known for a long time (Flohn and Fleer, 1975; Sikka, 1980; Angell, 1981; Rasmusson and Carpenter, 1983). In general, the warm (cold) SSTs over central and east Pacific Ocean are associated with a deficient (excess) ISMR. The ISMR is also associated with the SSTs over other ocean basins. Joseph and Pillai (1984), and Rao and Goswami (1988) observed warm (cold) SSTs over parts of the Arabian Sea during pre-monsoon months prior to an excess (deficient) ISMR. The existence of warm SST anomalies over the equatorial Indian Ocean was attributed as one of the reasons for the deficient ISMR during 1987 (Krishnamurti et al, 1989). Rajeevan et al (2002) observed a positive relation between ISMR and the pre-monsoon SST anomalies over the Arabian Sea and the southeast Indian Ocean. Nicholls (1995) found positive correlations between April SSTs over the Indonesia-northern Australia area and ISMR. Soman and Slingo (1997) using GCM studies observed a strong association between SST anomalies over western equatorial Pacific with ISMR. Ose et al (1997) found time lagged C.C. between SSTs in the South China Sea and ISMR.

However, the association between ISMR and SSTs over the Pacific and other ocean basins is not one sided. Yasunari (1990) showed that a weaker (stronger) than normal Asian summer monsoon is favorable for triggering the El Niño (La Niña) state of the equatorial Pacific through weaker (stronger) than normal eastwest circulation in the tropics. He also suggested that the interannual variability over the Indian and Pacific sector is predominantly controlled by quasi biennial periodicity. Joseph and Pillai (1984) observed warming (cooling) of the Arabian Sea after a deficient (excess) monsoon and suggested a triennial oscillation in the SSTs over the Arabian Sea and ISMR. Webster et al (1998) observed significant covariance in the 2–8 year period band in the cross wave-let analysis of ISMR and Niño-3 SST index. These results therefore, not only indicate the strong mutual association between SSTs over the various ocean basins and ISMR but also indicate the possibility of such a mutual association with large lag periods (say 2–4 years).

Recently Pai (2003), using global grid point monthly surface temperature anomaly data for the period 1901–98, observed significant teleconnections between ISMR and SST anomalies over various geographical areas of the Pacific, Indian and Atlantic Oceans. For the last 15 years, the India Meteorological Department has been using the SST conditions over Niño regions during the previous monsoon season as one of the parameters for its operational models for the long range prediction of ISMR (Gowariker et al, 1989; Rajeevan et al, 2004). Delsole and Shukla (2002) used monthly Niño-3 and NAO indices during six months preceding a monsoon season as the predictors for the long range prediction of ISMR during that season. Sahai et al (2003) observed predictive signals for ISMR in SSTs over different ocean basins with lags up to 4 years. They have demonstrated the prospects for the long range prediction of ISMR using only global SST anomalies. They have developed a multiple linear regression model based on objectively selected 14 predictors with time lags of 4 seasons to 4 years. The model was able to provide reasonably correct model predictions during the independent test period (1980–2001).

In the present study, the main objective was to develop empirical models for the long range prediction of ISMR using predictors derived from global seasonal SST anomalies pertaining to the recent half century (1951–2003). We have explored the prediction of ISMR with three different lead time periods (0, 1 and 2 seasons) and found that the model with a lead time of 1 season (3 months) had the best predictive skill among the 3 models. In this context, it should be mentioned that Sahai et al (2003) considered only one prediction lead time period, i.e., 3 seasons (9 months). Further we have also explored the prospect for the improvement in the predictions based on only global SSTs by including a

predictor derived from NAO linked pressure anomalies over West Europe during the fall season (September to November) of the previous year.

In the remainder of this paper Sect. 2 describes the various data sets used in this study and Sect. 3 discusses the methodology used for the derivation and selection of predictors from the global SST anomalies, and the development of prediction models with different lead time periods. Sections 4–6 discuss the various results and finally Sect. 7 gives the conclusions of the study.

2. Data

The main data set used was the monthly NOAA Extended Reconstructed Global Sea Surface Temperature (ERSST) data at $2^\circ \times 2^\circ$ latitude $\times$ longitude grid (Smith and Reynolds, 2003). These data were obtained from the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA (http://www.cdc.noaa.gov/). This data set was produced based on the latest version of the Comprehensive Ocean Atmosphere Data Set (COADS) release 2 observations (Woodruff et al, 1998). The monthly ERSST data are available from 1854 onwards. In this study we have used the ERSST data for the period from January 1951 to May 2003.

The ISMR series used in this study was based on the seasonal (June–September) monsoon rainfall data of the 36 meteorological subdivisions in India. The seasonal summer monsoon rainfall over the country as a whole was calculated as the area weighted average of seasonal rainfall of all 36 subdivisions. The ISMR was expressed as the percentage departure from the long period average (LPA) of seasonal summer monsoon rainfall over the entire country, which is equal to 88 cm. When the ISMR during a year is $>10\%$ ($<10\%$) of LPA, the year is termed an excess (deficient) monsoon year. All other years are termed normal monsoon years.

Another data set used was the monthly surface sea-level pressure at a $2.5^\circ \times 2.5^\circ$ spatial grid for the period 1951–2003. These data were taken from National Center for Environmental Predication (NCEP) reanalysis (Kalnay et al, 1996). The monthly mean North Atlantic Oscillation (NAO) index (normalized sea-level pressure difference between Gibraltar, Azores and Stykkisholmur, Iceland) data (Jones et al, 1997) and Niño-3.4 index data (Climate Prediction Centre, NOAA) for the period 1955–2003 were also used.

3. Methodology

3.1 Derivation of predictors and selection of best predictor sets

At first, the grid point anomalies of the SSTs were computed using the 1951–2000 climatology. To identify the predictive signals for ISMR, maps of correlation coefficient (C.C.) between ISMR and the seasonal SST anomalies, and between ISMR and tendency in the seasonal SST anomalies were prepared. Time lags of the SST anomalies were varied from 1 season to 4 years behind the reference monsoon season. In this study, the time lag of a season in years with respect to the reference monsoon year is indicated by 0 or a negative number in a bracket along with the season name. For example, the DJF (December to February) season belonging to the year of reference monsoon year is indicated as DJF (0), that belonging to one year previous to the reference year as DJF (-1), and so on. Seasonal tendency in the grid point SST anomalies were computed as the difference in the SST anomalies during one season (say MAM (-1)) from that during the previous season (say DJF (-1)).

In the C.C. maps, areas of significant C.C. were identified and over each geographical region, the SST anomalies within rectangular boxes drawn to enclose maximum areas of most significant C.C. were averaged to form time series of seasonal SST anomaly and SST anomaly tendency indices. However, the areas of the rectangular boxes were restricted to a minimum of $10^\circ \times 10^\circ$ (latitude $\times$ longitude) and a maximum of $20^\circ \times 20^\circ$.

When the number of available predictors is large it is essential to apply some criteria for the selection of the best predictor set out of them. We have used the criteria that the 21 years C.C. between the selected predictor and ISMR remained stable (significant at 5% and above) during the development period.

3.2 Formulation of the prediction model

We have used Multiple Regression Analysis (MRA) for the development of the prediction model. However, models based on MRA have

an inherent problem of multi-colinearity caused by inter-correlation among predictors. To alleviate the problem of multi-colinearity, Principal Component Analysis (PCA) of the predictors prior to MRA is a useful technique. PCA also reduces the dimensionality and noise level in the data set as the first few significant components account for most of the variation of the original data (Rao, 1964). Singh and Pai (1996) have used PCA followed by MRA to develop a Principal Component-Regression (PCR) model for the prediction of the seasonal (June–September) ISMR for the country as a whole based on predictors from the Indian Ocean only. Rajeevan et al (2000) on the other hand used a PCR model for the prediction of seasonal summer monsoon rainfall over the homogeneous rainfall regions of India.

If PCA is applied to a set of time series of "m" inter-correlated standardized predictors for "n" years and select any "p" Principal Component (PC) modes ($p < m$), we can write the PCR model as $\mathbf{Y} = \mathbf{BF'} + \boldsymbol{\varepsilon}$, where $\mathbf{Y}$ is the $(n \times 1)$ predictand matrix, $\mathbf{B}$ is the $(1 \times p)$ matrix of regression coefficients, $\mathbf{F'}$ is the $(n \times p)$ matrix of selected PC scores and $\boldsymbol{\varepsilon}$ is the $(n \times 1)$ error matrix.

In selecting suitable PC modes, first we retained "k" PCs which together explained at least 80% of the total variability of the predictor set. However, the best model is one that would use the lowest possible number of PCs and explain the most variance during the development period. The selection of the best model was achieved by the use of all-possible regressions procedure along with Mallow's "Cp" statistics (Delsole and Shukla, 2002). In this procedure, all possible simple linear regressions between each of the "k" PCs and the ISMR were first constructed. Then, the model with least root mean square error (RMSE) during the development period was selected as the best model among all the possible 1-parameter models.

The PC corresponding to the best 1-parameter model was then used in combination with each one of the remaining "$k-1$" PCs to construct "$k-1$" two parameter multiple regression models. The best model among all possible 2-parameter models was selected using the above criteria. In the same way best 3-parameter model was selected and so on. The procedure was stopped when there was no further reduction in the RMSE. However, the correct model size (number of PCs used as input) was determined by plotting Mallow's "Cp" against $p+1$. Here p is the number of PCs used as input to the regression model and Mallow's "Cp" is given as

$$\mathrm{Cp} = \left(\frac{e_X^2}{\frac{e_Y^2}{(N-p-2)}} \right) - (N - 2(p+1)),$$

where X and Y are the models with "p" and "$p+1$" PCs, respectively, as the input, and e_X^2

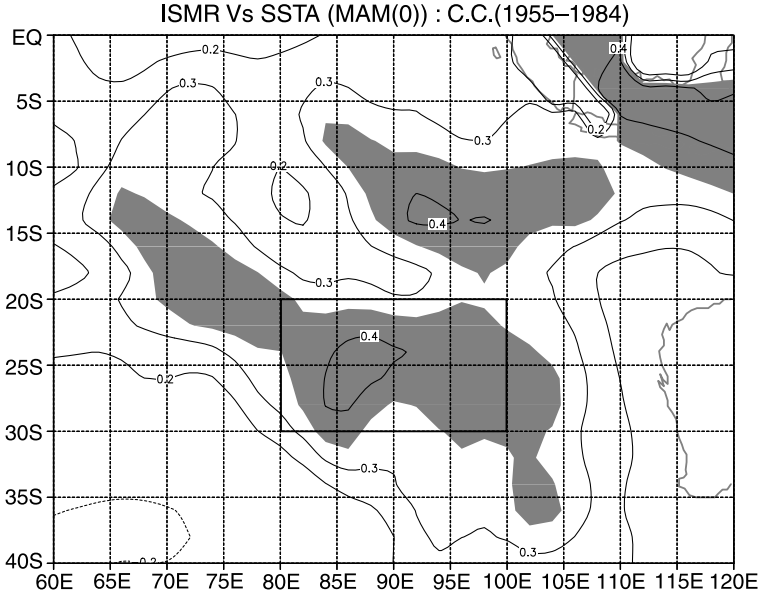

Fig. 1. Spatial pattern of correlation coefficient (C.C.) between ISMR and SST anomalies during the MAM (0) season over the South Indian Ocean. The C.C. was computed using the data for the period 1955–1984. The areas of C.C. significant at 5% level and above are shaded

and e_Y^2 are the mean squared errors of models X and Y, respectively, during the model development period consisting of N cases. When the correct model size is near $p+1$, Cp will roughly be equal to $p+1$. Therefore, when the value of Cp corresponding to a $p+1$ value was below but close to the 45° line, the correct model size was taken as p.

The skill of the models was tested as an independent period (1985–2003) and was measured by calculating model statistics such as the C.C. between actual and predicted ISMR, RMSE and bias (BIAS). The RMSE and BIAS were calculated using the following equations;

$$\mathrm{RMSE} = \left[\sum_{i=1}^{i=M}(R_{ip} - R_i)^2/M\right]^{1/2},$$

$$\mathrm{BIAS} = \sum_{i=1}^{i=M}(R_{ip} - R_i)/M,$$

where R_{ip} and R_i are the predicted and actual ISMR values for the ith year and M is the number of years.

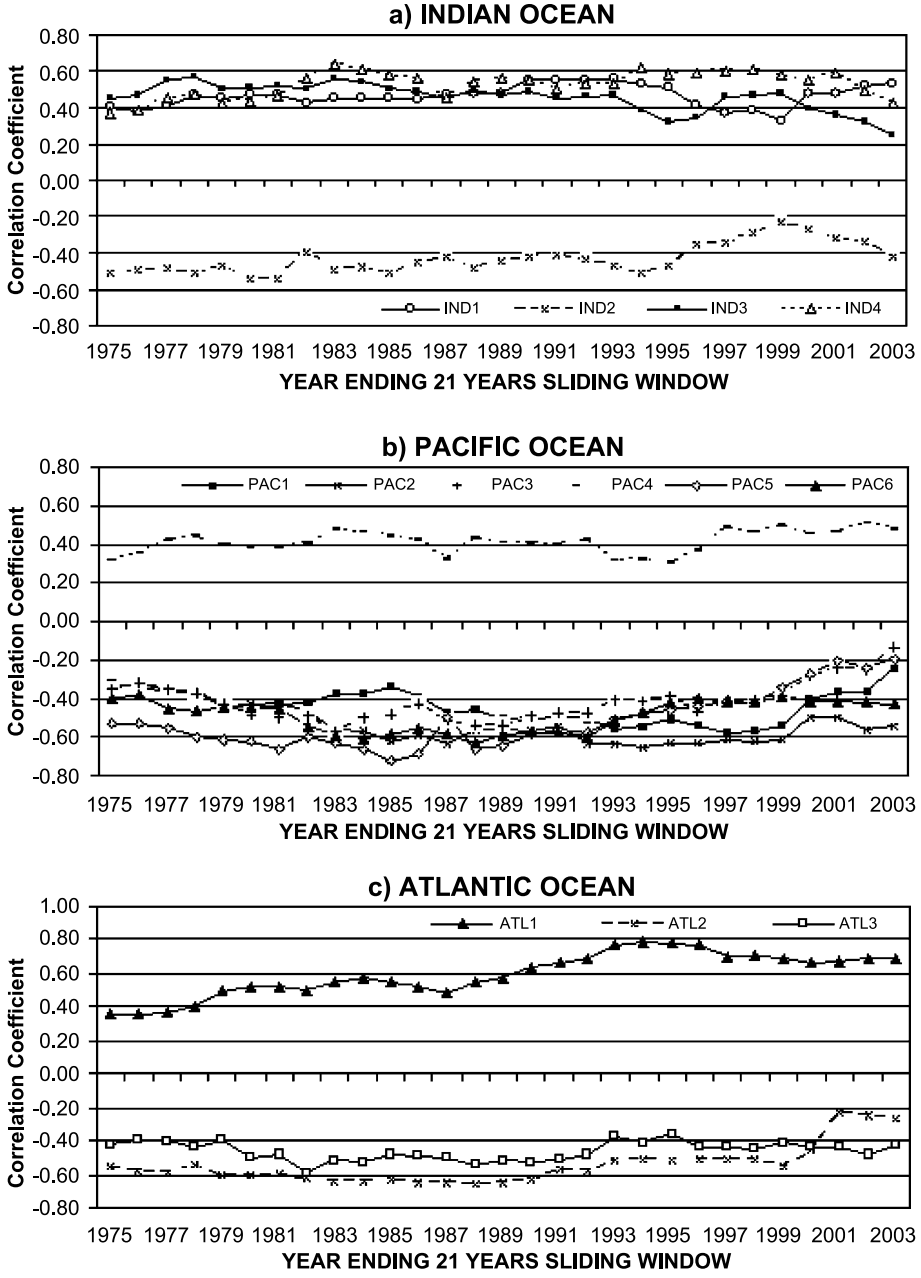

Fig. 2. 21 years moving C.C. between ISMR and selected 13 predictors from (**a**) Indian, (**b**) Pacific, and (**c**) Atlantic Oceans for the period 1955–2003. The horizontal dotted lines represent the C.C. significant at 5% level

4. Predictive signals from SST anomalies

4.1 Correlation Coefficient (C.C.) maps

Seasonal SST anomalies and seasonal SST anomaly tendencies with lags from one season to 4 years were correlated with ISMR using data for the period 1955–1984. The beginning year of the C.C. period was taken as 1955, accounting for the maximum lag of 4 years whilst retaining a constant period of 30 years (1955–1984) for the preparation of all the C.C. maps. Altogether 35 C.C. maps (18 for seasonal SST anomalies and 17 for seasonal tendency in the SST anomalies) were prepared. Figure 1 shows the C.C. map over the South Indian Ocean for the MAM (0) season. The areas where C.C. is significant at 5% or above are shaded. Rajeevan et al (2002) have discussed the physical linkage between the SST anomalies over the Southeast Indian Ocean during pre-monsoon and ISMR in detail. In Fig. 1, the rectangular box indicates the area selected for averaging. In a similar way, time series of 28 potential predictors from various oceanic regions with different seasonal lags were obtained from the 18 correlation maps of seasonal SST anomalies. From the 17 correlation maps of tendency in seasonal SST anomalies, another 23 predictors were also obtained.

4.2 Predictor sets with different prediction lead times

The 21 year moving C.C.s between the above derived 51 predictors and ISMR showed that there were 13 predictors which have stable relationships with the ISMR. Figure 2 shows the moving 21 year C.C. of ISMR with these 13 selected predictors. As can be seen in the Fig. 2, for all predictors, the relationship with the ISMR is stable not only during the development period but also during most of the test period. Of these 13 predictors, 7 were derived from the seasonal SST anomalies and 6 were derived from the seasonal tendency in the SST anomalies. The geographical domains of these 13 predictors and their C.C.s with ISMR are given in Table 1. As seen in Table, 4 predictors were derived from the Indian Ocean, 6 from the Pacific Ocean and 3 from the Atlantic Ocean. However, the longest lag time of any of the predictors was of 3 years behind the reference monsoon year.

From Table 1, it is clear that by the collective use of all 13 predictors we can predict ISMR at the end of MAM (0). This means that the prediction lead time is nil or 0 season with respect to the beginning of the reference monsoon season. Hereafter, this set of 13 predictors is termed set A. By removing 3 predictors (IND4, PAC5 and PAC6) pertaining to the MAM (0) from set A, a second predictor set (set B) containing 10 predictors with prediction

Table 1. Details of the selected 13 predictors. The predictors from Indian, Pacific and Atlantic Oceans are indicated by the short names IND, PAC and ATL, respectively. The time lag of the predictor is indicated by the season name and a number in bracket. The season name corresponds to the season over which the SST anomalies were averaged to derive the predictors and the number in the bracket is the lag in years behind the reference monsoon year

No.	Predictors	Time lag	Geographical domain	C.C. with ISMR 1955–1984		C.C. with Niño-3.4 1955–1984	
				Previous year	Reference year	JJA (−1)	JJA (0)
1	IND1	SON (−3) − JJA (−3)	40° S–30° S, 50° E–70° E	0.23	0.36*	−0.20	−0.28
2	IND2	DJF (−1)	28° S–18° S, 104° E−114° E	0.44*	−0.46**	−0.27	−0.02
3	IND3	DJF (0)	20° S–10° N, 80° E–100° E	−0.46**	0.43*	0.68**	−0.18
4	IND4	MAM (0)	30° S–20° S, 80° E–100° E	−0.02	0.40*	0.36*	−0.24
5	PAC1	DJF (−2)	30° S–20° S, 100° W–80° W	0.20	−0.42*	−0.35	0.19
6	PAC2	JJA (−1) − MAM (−1)	30° N–40° N, 170° E–170° W	0.42*	−0.48**	−0.36*	0.11
7	PAC3	JJA (−1) − MAM (−1)	10° S–0° , 120° E–140° E	0.00	−0.46**	−0.31	0.38*
8	PAC4	SON (−1)	20° S–10° S, 130° W–110° W	−0.57**	0.36*	0.78**	−0.12
9	PAC5	MAM (0)	40° N–50° N, 170° W–150° W	0.05	−0.50	−0.22	0.44*
10	PAC6	MAM (0) − DJF (0)	6° S–6° N, 150° W–130° W	0.66**	−0.43*	−0.87**	0.37*
11	ATL1	MAM (−3) − DJF (−3)	38° S–28° S, 40° W–20° W	−0.08	0.47**	−0.02	−0.16
12	ATL2	SON (−1) − JJA (−1)	50° S–40° S, 60° W–40° W	0.40*	−0.57**	−0.39*	0.35
13	ATL3	DJF (0)	20° N–30° N, 100° W–80° W	0.13	−0.36*	−0.27	0.28

* Significant at ≥5% level, ** Significant at ≥1% level

lead time of 1 season (3 months) was created. By removing 2 more predictors pertaining to DJF (0) from set B (IND3, ATL3), a third predictor set (set C) was formed containing 8 predictors with a prediction lead time of 2 seasons (6 months). Thus, there are 3 predictor sets A, B and C with prediction lead time periods of 0, 1 and 2 seasons with 13, 10 and 8 predictors, respectively.

5. Prediction models of ISMR with different lead time periods

In this section, we describe the details of the prediction models for ISMR developed using the above 3 predictors sets (A, B and C). We used data for the first 30 years (1955–1984) for the development of the models and data for the next 19 years (1985–2003) for the independent verification.

5.1 Significant PC modes of the predictors sets

Table 2 shows the lower triangle of the correlation matrix of the 13 predictors used for forming the three predictor sets. There were significant intercorrelations among many of the predictors. PCA was carried out on each of the 3 predictor sets for the development period and the first few PCs (6, 5 and 5 from sets A, B and C, respectively) that together explained at least 80% of the total variability of the respective data set were retained. The retained PCs from each of the predictor sets A, B and C were then used to develop the best multiple regression prediction models (models A, B and C). Following the all-possible regressions procedure and Mallow's "Cp" statistics described in Sect. 3.2, the correct model size of model A was 3 (PC1, PC2 and PC4), that of model B was 2 (PC1, PC2) and that of model C was 1 (PC1). It should also be mentioned that the eigen values of all the PCs selected to construct the best models were ≥ 1. PC scores of all the selected PCs were then computed for the test period (1985–2003).

As discussed above, the common parameter in all the 3 models is PC1. On examining the principal loading vector corresponding to PC1, it was found that in all the 3 cases, magnitude of loadings by the predictors pertaining to period from JJA (-1) season to MAM (0) season (seasonal lags ≤ 1 year) was large compared to that by other predictors. The signs of the loadings were same as that of the C.C.s between the corresponding predictor and the ISMR (Table 1). Thus, PC1 can be assumed as the composite predictive signal evolving in the global (particularly tropical) SST anomalies from the beginning of the previous monsoon season. As seen in Table 1, most of the predictors with seasonal lags ≤ 1 year were significantly correlated with Niño-3.4 of JJA (-1) season. However, these predictors were poorly correlated with the Niño-3.4 of JJA (0) season. The correlations of PC1 with Niño-3.4 index during previous JJA season (JJA (-1)) and that during the concurrent JJA season (JJA (0)) were also calculated for the period 1955–1984. Corresponding to sets A, B and C, the C.C.s of PC1 with Niño-3.4 of JJA (-1) were 0.70 (0.1%), 0.59 (1%), and 0.47 (1%), respectively, and that with Niño-3.4 of JJA (0) were

Table 2. Lower triangle of the correlation matrix of the selected 13 predictors calculated using data for the period 1955–1984. The C.C.s significant at 5% and above are shown by bold numbers

Predictors → ↓	IND1	IND2	IND3	IND4	PAC1	PAC2	PAC3	PAC4	PAC5	PAC6	ATL1	ATL2	ATL3
IND1	1.00												
IND2	−0.08	1.00											
IND3	0.20	**−0.36**	1.00										
IND4	0.22	−0.27	**0.63**	1.00									
PAC1	−0.05	0.23	−0.24	−0.05	1.00								
PAC2	−0.11	**0.59**	−0.22	−0.21	**0.47**	1.00							
PAC3	0.00	0.04	**−0.37**	**−0.36**	0.26	0.27	1.00						
PAC4	−0.10	−0.30	**0.70**	**0.40**	−0.18	**−0.37**	−0.28	1.00					
PAC5	−0.10	0.19	−0.28	−0.32	**0.38**	0.32	**0.48**	**−0.47**	1.00				
PAC6	0.06	**0.41**	**−0.68**	**−0.36**	**0.39**	**0.39**	0.30	**−0.79**	0.31	1.00			
ATL1	**0.61**	−0.33	0.17	0.18	−0.12	**−0.36**	−0.17	−0.06	−0.01	−0.07	1.00		
ATL2	**−0.37**	0.23	−0.31	−0.11	**0.36**	**0.47**	0.22	−0.27	0.18	**0.47**	−0.28	1.00	
ATL3	−0.03	**0.36**	−0.18	−0.26	**0.37**	0.35	0.31	−0.30	**0.43**	**0.37**	−0.26	0.22	1.00

−0.39 (5%), −0.33, −0.32, respectively. Thus, one factor that can be significantly associated with PC1 is the ENSO of the previous year. However, the C.C. between the Niño-3.4 index during JJA (−1) was not significantly correlated with ISMR. The C.C. between these two quantities was 0.30. This result, therefore, indicates the possibility of factors other than ENSO which can be associated with the PC1.

From Table 1, it is evident that most of the predictors with seasonal lags ≤ 1 year were significantly related to the ISMR of previous year. Similarly, the C.C.s between ISMR of the previous year and PC1 corresponding to models A, B and C during the period 1955–1984 were −0.49 (1%), −0.48 (1%), −0.44 (5%). Thus, both the monsoon and ENSO of the previous year can act as driving forces behind the composite predictive signal (for the subsequent monsoon) evolving in the global SST anomalies between previous and reference monsoon seasons. Furthermore, the predictive relationship of ISMR of reference year with PC1 was very much stronger than that with ENSO (Niño-3.4) of previous year, although there is a strong concurrent relationship between ISMR and ENSO (the C.C. between ISMR and Niño 3.4 of JJA (0) was 0.61 (1%)). The C.C.s between ISMR of the reference year and PC1 corresponding to models A, B, and C were 0.71, 0.73, and 0.74 all at 1% significant level. These results also indicate significant association between ISMR and the global SST anomalies in the quasi-biennial scale.

5.2 The performances of the best prediction models

In the previous section, the construction of 3 models, A, B and C, with prediction lead time periods of 0, 1, and 2 seasons, respectively, for the prediction of ISMR were described. The model regression equations computed for the development period (1955–1984) are given below.

$$R_A = 0.962 + 7.953 * PC1 + 3.394 * PC2 + 1.539 * PC4$$

$$R_B = 0.955 + 8.206 * PC1 + 1.915 * PC2$$

$$R_C = 0.957 + 8.307 * PC1,$$

where R_A, R_B, and R_C are the predicted values of ISMR corresponding to the models A, B and C. Various details and statistics of these models are given in the Table 3. The skill of a model is mainly measured by its performance during the independent test period. As seen from the various model statistics during the test period given in the Table 3, model B has the better skill in comparison to other two models. It is also important to note that RMSE of model B during the test period (6.03% LPA) is smaller than that during the development period (7.49% of LPA).

Figure 3 shows the performance of model B. It can be seen that the model was able to correctly indicate the sign of the actual ISMR during most of the years (particularly during the extreme years) of both development and independent test periods. In Fig. 3, the El Niño and La Niña years are marked by letters E and L, respectively. As expected a general inverse relationship is observed between the phases of ENSO and ISMR. The predicted ISMR during years succeeding El Niño (La Niña) years were mostly above (below) LPA. During the development period, there were 7 El Niño years (1957, 1963, 1965, 1969, 1972, 1977 and 1982) and 8 La Niña years (1956, 1964, 1967, 1970, 1971, 1973, 1975 and 1984). During the years (except in 1966) following

Table 3. Details and statistics of the models A, B, C and B′

Model →	A	B	C	B′
Input parameters to the model	PC1, PC2 and PC4 of set A	PC1, PC2 of set B	PC1 of set C	PC1, PC2 of set B and WEPR
RMSE (1955–1984) as % of Long Period Normal (LPA)	7.07	7.49	7.62	6.71
RMSE (1985–2003) as % of LPA	7.15	6.03	6.62	4.86
C.C. between observed and model ISMR (1985–2003)	0.74	0.83	0.73	0.86
BIAS (1985–2003) as % of LPA	3.24	2.45	1.34	0.71

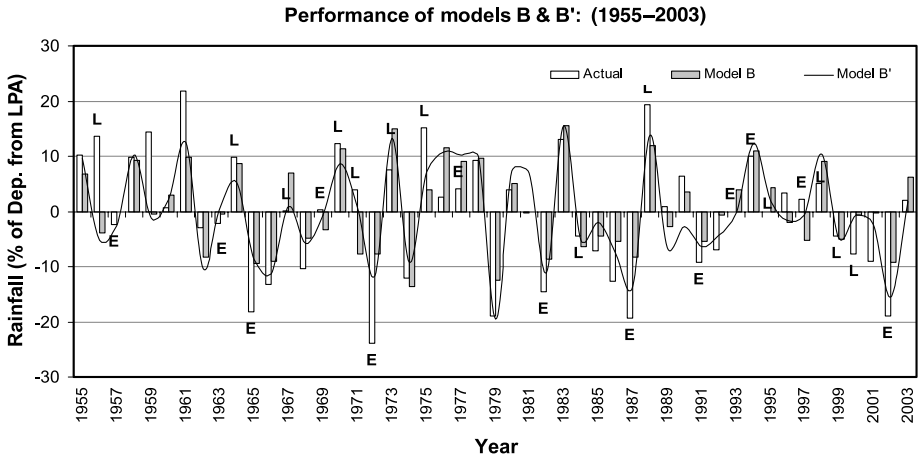

Fig. 3. Actual and predicted ISMR by the Model B with lead time of one season during the period 1955–2003. The El Niño and La Niña years are indicated by letters E and L, respectively. The predicted ISMR by model B′ is shown by solid line

El Niño years, the predicted ISMR values were above LPA. But in all these years (including 1966), the signs of both the actual and the predicted ISMR were the same. On the other hand, during years (except in 1971 and 1976) following the La Niña years, both the actual and the predicted ISMR values were below LPA. The association between ENSO events during previous year and predicted ISMR was not surprising as the most significant PC mode (PC1) of the predictor set (input parameter to the model) was strongly related to the Niño-3.4 index of the previous monsoon season (see earlier section). But the important point was that during years following El Niño (La Niña) years, even when the actual ISMR was below (above) LPA, the signs of the actual and predicted ISMR were the same.

During the independent test period the signs of the predicted ISMR during the years following the El Niño (La Niña) years were mostly positive (negative) and the same as that of the actual ISMR. The exceptions were during 1989 and 1996, which followed La Niña years, and 1992 which followed an El Niño year. During 1989 and 1996, actual (predicted) ISMR was above (below) LPA. But both actual and predicted ISMR were very close to the LPA. During 1992, both actual and predicted ISMR were below LPA.

6. Inclusion of NAO linked surface pressure anomalies over West Europe in the model based on SST anomalies

Rajeevan (2002) showed an association between NAO related winter mean sea-level pressure anomalies over northwest Europe and ISMR on an inter-annual time scale. Pai (2004) showed that the weakening of ENSO-ISMR during the recent decade (1990's) was due to the persistant positive phase of the winter NAO and below normal winter and spring Eurasian snow cover. However, the NAO is quite independent of ENSO. In Sect. 5.1, we have shown that the most significant PC mode (PC1) of the predictor data used for developing prediction models was strongly associated with ENSO. Now we examine the performance of model B, which was the best among the 3 models based on only SST anomalies, by including an NAO related predictor as one of the inputs to model B. The NAO related predictor (WEPR) was derived by averaging the NCEP grid point surface pressure anomalies (base period 1951–2000) over West Europe (40° N–50° N, 10° W–10° E) during the SON (−1) season. As WEPR has a seasonal lag of one season, the prediction lead period of the modified model B (model B′) was the same as that of model B. The C.C. between WEPR and ISMR for the development period (1955–84) was −0.46 (1%) and the 21 years moving C.C.s between these two parameters were above the 5% significant level during both the development and training periods. The WEPR also showed a strong and stable association with concurrent NAO index with a C.C. of 0.44 (5%) for the period 1955–84.

The WEPR was standardized before using it as a predictor in the model. It should be mentioned that WEPR showed no correlation with the 2 PCs used in model B. The model equation of the new model B′ developed using the data for the period 1955–84 is given below.

$$R_{B'} = 0.954 + 7.923 * PC1 + 1.059 * PC2 - 3.508 * WEPR,$$

where $R_{B'}$ is the predicted ISMR. The model statistics of model B′ are given in Table 3 along with that of the models A, B and C. The performance of model B′, particularly in the independent test period, was better than the models based on SST predictors only. In Fig. 3, the predicted ISMR for model B′ is shown by the solid line. In Fig. 3, it can be seen that during most of the years, the error in the prediction (predicted – actual) of ISMR by model B′ was smaller than that of model B. In 16 of the 19 years of the independent test period (except 1989, 1990 and 1996), the sign of the ISMR predicted by model B′ was the same as that of the actual ISMR. Moreover, during 9 of these 16 years which included year 2002, a highly deficient monsoon year (ISMR $= -19\%$), the absolute prediction errors were $\leq 3\%$ of LPA. These results, therefore, showed significant improvement in the accuracy of ISMR predictions by including NAO related WEPR in the models based on SST anomalies only.

7. Conclusions

A search for the potential predictors in the global seasonal SST anomalies and tendency in the seasonal SST anomalies for the long range prediction of ISMR showed that there were 13 predictors with stable and significant relationships with ISMR. From these predictors, derived from various areas of three tropical oceans (Indian, Pacific and Atlantic) with time lags varying from 1 season to 3 years, 3 predictor sets A, B and C were formed with prediction lead time periods of 0, 1 and 2 seasons, respectively, with respect to the beginning of the monsoon season. The selected principal components (PCs) of these predictor sets were used as the input parameters for the models A, B and C, respectively. The correct model size was derived using all-possible regressions procedure and Mallow's "Cp" statistics.

From various model statistics computed for the independent period, it was found that model B with a lead time of 1 season had the best predictive skill among the three models. The RMSE of model B during the independent test period (6.03% of LPA) was much less than that during the development period (7.49% of LPA). Model B performed well during both ENSO and non–ENSO years particularly during years when the magnitudes of actual ISMR were large.

In general, the predicted ISMR during the years following El Niño (La Niña) years was above (below) LPA as was the actual ISMR. This was because, the most significant principal component mode of the predictor data (PC1) was found to be significantly associated to the state of ENSO during the previous year. PC1 represented a composite predictive signal for ISMR evolving in the global (particularly tropical) SST anomalies during the period from the previous monsoon season to the pre-monsoon season of the reference year. However, the predictive relationship between PC1 and ISMR was very much stronger than between ENSO itself and ISMR. This was because PC1 was also associated with the ISMR of previous year. In short, a quasi-biennial relationship was evident between the composite state of global SST anomalies represented by PC1 and ISMR.

Model B′ developed by using an NAO related predictor (WEPR) derived from surface pressure anomalies over West Europe as an additional input parameter to model B performed much better than model B. The RMSE of model B′ during the independent test period (4.86% of LPA) was substantially lower than that of model B (6.03% of LPA). This result provides the encouraging option of including predictors derived from climatic fields other than SST as additional input parameters into models based on SSTs only for better prediction of ISMR.

Acknowledgements

The authors are grateful to Dr. S. K. Srivastav, DGM, India Meteorological Department (IMD), for granting permission to submit this research paper in this journal and his encouragement. The authors also express their thanks to Dr. S. K. Dikshit, retired ADGM (R), IMD, Pune for his encouragement. We also sincerely thank Prof. Stefan Hastenrath for his valuable comments which helped in the overall improvement of the quality of this paper.

References

Angell JK (1981) Comparison of variations in atmospheric quantities with SST variations in equatorial eastern Pacific. Mon Wea Rev 109: 230–243

Charney JG, Shukla J (1981) Predictability of monsoons. In: Monsoon dynamics (Lighthill J, ed). Cambridge University Press, pp 99–110

Delsole T, Shukla J (2002) Linear prediction of the Indian monsoon rainfall. J Climate 15: 3645–3658

Flohn H, Fleer H (1975) Climate teleconnections with the equatorial Pacific and the role of ocean/atmosphere coupling. Atmosphere 13: 96–109

Gadgil S, Sajani S (1998) Monsoon precipitation in the AMIP runs. Clim Dyn 14: 659–689

Goddard L, Mason J, Zebiak SE, Ropelewski CF, Basher R, Cane MA (2001) Current approaches to seasonal and interannual climate predictions. Int J Climatol 21: 1111–1152

Gowariker V, Thapliyal V, Sarker RP, Mandal GS, Sikka DR (1989) Parametric and power regression models: new approach to long range forecasting of monsoon rainfall in India. Mausam 40: 115–122

Gowariker V, Thapliyal V, Kulshrestha SM, Mandal GS, Sen Roy N, Sikka DR (1991) A power regression model for long range forecast of southwest monsoon rainfall over India. Mausam 42: 125–130

Hastenrath S (1995) Recent advances in tropical climate prediction. J Climate 8: 1519–1532

Jones PD, Jonsson T, Wheeler D (1997) Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and Southwest Iceland. Int J Climatol 17: 1433–1450

Joseph PV, Pillai PV (1984) Air sea interaction on a seasonal scale over north Indian Ocean, part I: Interannual variation of sea-surface temperature and Indian monsoon rainfall. Mausam 35: 323–330

Kalnay E, co-authors (1996) The NCEP/NCAR Reanalysis Project. Bull Amer Meteor Soc 77: 437–471

Kang IS, co-authors (2002) Inter comparison of the climatological variations of Asian summer monsoon precipitation simulated by 10 GCMs. Clim Dyn 19: 383–395

Krishna Kumar K, Soman MK, Rupa Kumar K (1995) Seasonal forecasting of Indian summer monsoon rainfall. Weather 50: 449–467

Krishnamurti TN, Bedi HS, Subramaniam M (1989) The summer monsoon of 1987. J Climate 2: 321–340

Latif M, Sterl A, Assenbaum M, Junge MM, Maier-Reimer E (1994) Climate variability in a coupled GCM – part II: the Indian Ocean and monsoon. J Climate 7: 1449–1462

Navone HP, Ceccato HA (1995) Predicting Indian monsoon rainfall: a neural network approach. Clim Dyn 10: 305–312

Nicholls N (1995) All India summer monsoon rainfall and sea surface temperatures around northern Australia and Indonesia. J Climate 8: 1463–1467

Ose T, Song Y, Kitoh A (1997) Sea surface temperature in the South China sea and index for the Asian monsoon and ENSO system. J Meteorol Soc Japan 75: 1091–1107

Pai DS (2003) Teleconnections of Indian summer monsoon with global surface air temperature anomalies. Mausam 54: 407–418

Pai DS (2004) A possible mechanism for the weakening of the El Niño-monsoon relationship during recent decade. Meteorol Atmos Phys 86: 143–157

Rajeevan M (2002) Winter surface pressure anomalies over Eurasia and Indian summer monsoon. Geophys Res Lett 29: 941–944

Rajeevan M, Guhathakurta P, Thapliyal V (2000) New models for long range forecasting of monsoon rainfall over northwest and peninsular India. Meteorol Atmos Phys 73: 211–225

Rajeevan M, Pai DS, Thapliyal V (2002) Predictive relationships between Indian Ocean sea-surface temperatures and Indian summer monsoon rainfall. Mausam 53: 337–348

Rajeevan M, Pai DS, Dikshit SK, Kelkar RR (2004) IMD's new operational models for long range forecast of southwest monsoon rainfall over India and their verification for 2003. Curr Sci 86: 422–431

Rao CR (1964) The use and interpretation of principal component analysis in applied research. Sankhya (Indian J Statistics) A26: 329–358

Rao Kusuma G, Goswami BN (1988) Interannual variation of sea surface temperature over the Arabian Sea and the Indian monsoon. A new perspective. Mon Wea Rev 116: 558–568

Rasmusson EM, Carpenter TH (1983) The relationship between eastern equatorial Pacific sea-surface temperature and rainfall over India and Sri Lanka. Mon Wea Rev 111: 517–528

Sahai AK, Grimm AM, Satyan V, Pant GB (2002) Prospects of prediction of Indian summer monsoon rainfall using global SST anomalies. IITM Res Rep India RR-093: 1–44

Sahai AK, Grimm AM, Satyan V, Pant GB (2003) Long-lead prediction of Indian summer monsoon rainfall from global SST evolution. Clim Dyn 20: 855–863

Sikka DR (1980) Some aspects of the large scale fluctuations of summer monsoon rainfall over India in relation to fluctuations in the planetary and regional scale circulation parameters. Proc Ind Acad Sci (Earth and Planetary Sci) 89: 179–195

Singh OP, Pai DS (1996) An oceanic model for the prediction of SW monsoon rainfall over India. Mausam 47: 91–98

Smith TM, Reynolds RW (2003) Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). J Climate 16: 1495–1510

Soman MK, Slingo JM (1997) Sensitivity of the Asian summer monsoon to aspects of sea-surface temperature anomalies in the tropical Pacific Ocean. Q J R Meteorol Soc 123: 309–336

Thapliyal V (1982) Stochastic dynamic model for long range forecasting of summer monsoon rainfall in peninsular India. Mausam 33: 399–404

Walker GT (1914) A further study of relationships with Indian monsoon rainfall, II. Mem India Meteorol Dept 23: 123–129

Walker GT (1923) Correlation in seasonal variations of weather, VIII, A preliminary study of world weather. Mem India Meteorol Dept 24: 75–131

Woodruff SD, Diaz HF, Elms JD, Worley SJ (1998) COADS release 2 data and metadata enhancements for improvements of marine surface flux fields. Phys Chem Earth 23: 517–527

Yasunari T (1990) Impact of Indian monsoon on the coupled atmosphere/ocean system in the tropical Pacific. Meteorol Atmos Phys 44: 29–41

Corresponding author's address: D. Sivananda Pai, India Meteorological Department, LRF Division, National Climate Centre Office of ADGM (R), 411 005 Pune, India (E-mail: sivapai@hotmail.com)