Summer monsoon onset over Kerala: New definition and prediction

D S Pai* and Rajeevan M Nair

National Climate Centre, Pune 411 005, Maharashtra, India.
**e-mail: sivapai@hotmail.com*

The summer monsoon onset over Kerala (MOK) marks the beginning of the rainy season for the country. Associated with the MOK, significant transitions of large scale atmospheric and oceanic circulation patterns are observed over the Asia–Pacific region. In this study, a new method for the objective identification of MOK, based on large scale circulation features and rainfall over Kerala, is discussed. Further, a set of empirical models based on the principal component regression (PCR) technique was developed for the prediction of the date of MOK by keeping in mind the IMD's operational forecasting service requirements. Predictors for the models were derived using correlation analysis from the thermal, convective and circulation patterns. Only five predictors pertaining to the second half of April were used in the first model (Model-1) so that the prediction of MOK can be prepared by the end of April itself. The second model (Model-2) used four additional predictors pertaining up to the first half of May along with two predictors used in the Model-1 for update prediction at the end of the first half of May. To develop each of the PCR models, Principal Components Analysis (PCA) of the respective predictor data was carried out followed by regression analysis of first two principal components (PCs) with the date of MOK. Both these models showed good skill in predicting the date of MOK during the independent test period of 1997–2007. The root mean square error (RMSE) of the predictions from both the models during the independent test period was about four days which was nearly half the RMSE of the predictions based on climatology.

1. Introduction

The onset of Indian summer monsoon over Kerala is the most anxiously awaited weather singularity in the Indian subcontinent as it heralds the rainy season and marks the end of the hot summer. The onset of Indian summer monsoon represents significant transitions in the large scale atmospheric and oceanic circulation in the Indo-Pacific region. There is no widely accepted definition of this monsoon transition. However, at the surface, monsoon onset is recognized as a rapid substantial and sustained increase in rainfall. The first rains of monsoon occur over Burma and Thailand in mid-May and subsequently extend to the northwest. The monsoon sets over Kerala around 1 June with a standard deviation of about 8 days. In association with the monsoon onset, heavy rains lash south peninsula after the cross-equatorial low-level jet is established across the Somali coast into the near-equatorial Arabian Sea. This phenomenon is usually accompanied by the formation of a mid-troposphere shear zone across the Bay of Bengal to the south-east Arabian Sea in which a cyclonic vortex may be embedded. The northward progression of the monsoon is symptomatic of a large scale transition of a deep convection from the equatorial to continental regions (Rao 1976; Sikka and Gadgil 1980; Webster *et al* 1998). By middle of July, monsoon covers the whole country.

Large scale changes occur in the circulation features in association with the onset phase of

Keywords. Statistical prediction; Indian summer monsoon; monsoon onset over Kerala; principal component analysis; multiple linear regression.

J. Earth Syst. Sci. **118**, No. 2, April 2009, pp. 123–135

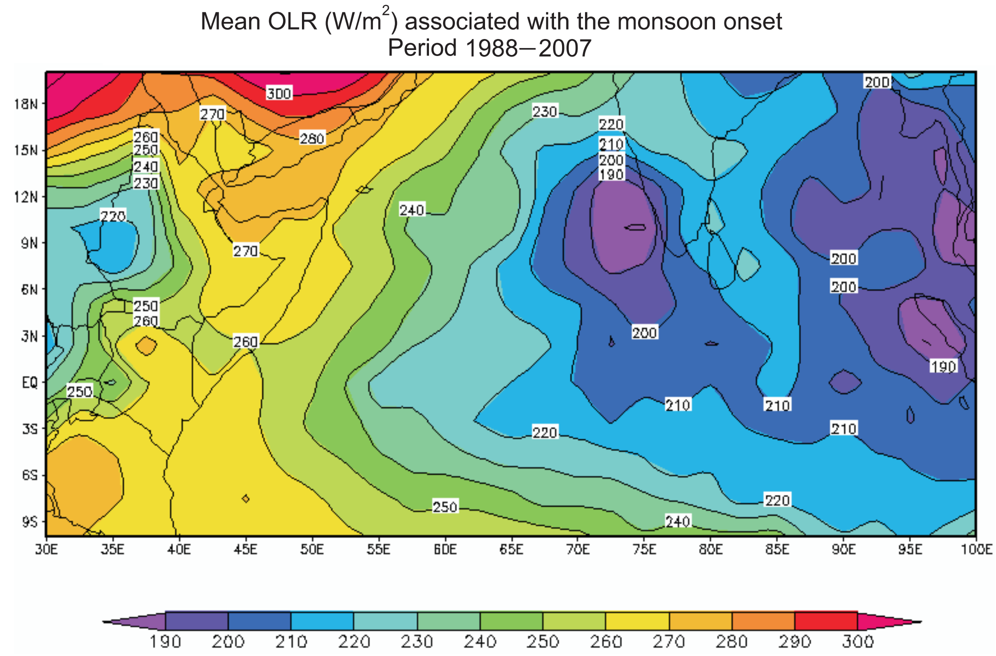

Figure 1. Composite mean OLR during the day of MOK. Period: 1988–2007. Unit: W m^{-2}.

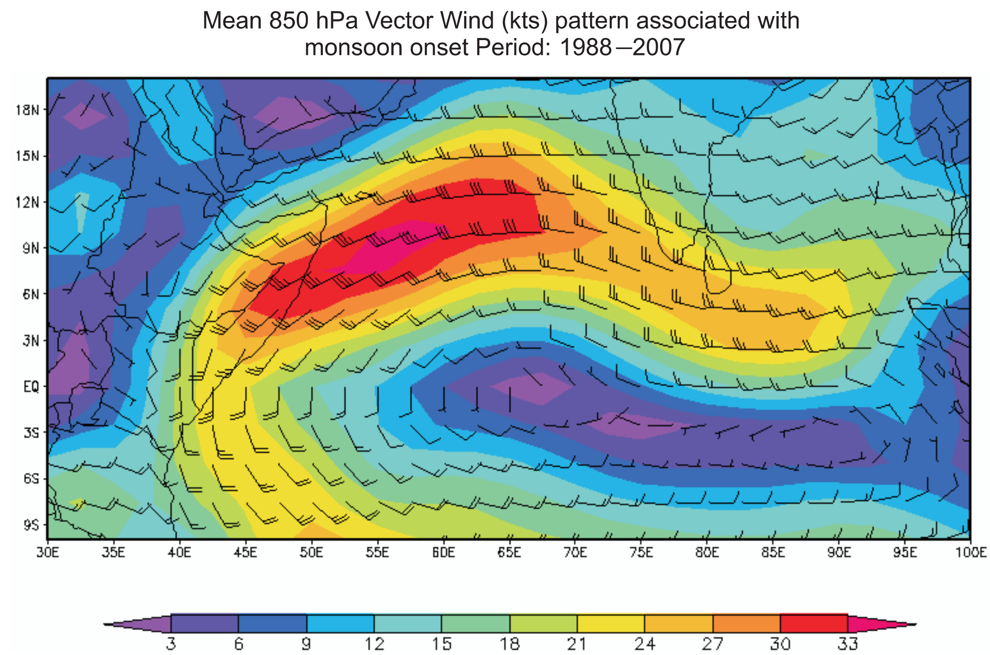

Figure 2. Composite mean 850 hpa zonal winds over Indian monsoon region during the day of MOK. Wind speed in knots. Period of data: 1988–2007.

Indian monsoon (Ananthakrishnan *et al* 1983; Pearce and Mohanty 1984; Ananthakrishnan and Soman 1988; Soman and Krishna Kumar 1993; Joseph *et al* 1994, 2006). At the date of monsoon onset, there is a band of deep convection (low OLR) in the east-west direction passing through southern tip of India, a maximum cloud zone as identified by Sikka and Gadgil (1980). Figure 1 shows the spatial structure of composite mean OLR during the onset phase. During the summer monsoon onset phase, major changes are also observed in the atmospheric wind flow at all levels. There is an appreciable acceleration of cross equatorial flow across the Somali coast and westerly zonal flow over the equatorial Indian Ocean (figure 2). The westerly zonal flow extends up to 600 hpa. The relative humidity of the air also increases at least up to 500 hpa (Rao 1976). At the upper tropospheric levels, the onset is generally associated with a northward shift in the subtropical westerly jet stream to the north of the Tibetan Plateau and the westward shift of the quasi-stationary trough at 500 hpa, from about 90°E to about 80°E. Yin (1949) was the first to link the process of monsoon onset to the displacement of westerly troughs in the circumpolar westerlies and shift of the subtropical jet (STJ) to the north of the Himalayan periphery. A tropical easterly jet stream also appears over south India in association with the monsoon onset (Koteswaram 1958, 1960). Murakami and Ding (1982) have suggested that the onset is related to the warming of the Eurasian region by diabatic heating. Thus, the onset of monsoon over India is linked to a combination of regional and planetary scale changes over the entire Indian monsoon region. There exists a variety in the linkages of the onset process with the seasonal developments of transitions in the regional and planetary-scale features. Pearce and Mohanty (1984) found that the period prior to monsoon onset consists of two main phases:

- a moisture buildup phase over the Arabian Sea during which synoptic and mesoscale transient disturbances develop and
- a rapid intensification of the Arabian Sea winds and a substantial increase in latent-heat release, essentially a large scale feedback process.

Soman and Krishna Kumar (1993) studied the climatological features of atmospheric circulation associated with the monsoon onset. The relative humidity builds up suddenly in the vertical, a few days before the onset at the respective stations. The vertically integrated zonal moisture transport at individual stations over the peninsula increases sharply with respect to the south Kerala onset, with appropriate lag in time. The composite outgoing long wave radiation fields over the north Indian Ocean (figure not shown) show rapid buildup of convective activity over the southeast Arabian Sea and east Bay of Bengal with the approach of the monsoon. Krishnamurti and Ramanathan (1982) examined observational aspects of the evolution of energy exchanges and differential heating during the GARP Monsoon Experiment MONEX.

2. Definition of monsoon onset over Kerala (MOK)

The onset of Asian monsoon can be considered as having two phases, one with a rainfall surge over South China Sea and the other with increased rainfall over India (Wang and LinHo 2002). Onset of South China Sea Summer Monsoon (SCSSM) is the first transition of Asian summer monsoon causing major changes in both convection and winds (Hsu *et al* 1999). The seasonal monsoon transition can occur in a variety of ways with abrupt, gradual or multiple transitions. Though at the surface, the monsoon transitions are first revealed by variability in rainfall, a variety of dynamic and thermodynamic precursors are known to exist (Ananthakrishnan and Soman 1991). The onset and dates of the monsoon season can be defined using a wide range of criteria that include rainfall, surface and upper level winds, outgoing long-wave radiation (OLR) indices, upper-tropospheric water vapour brightness temperature, etc. Zeng and Lu (2004) have suggested criteria for globally unified summer monsoon onset (or retreat) dates based only on the global daily $1° \times 1°$ normalized precipitable water data with the threshold value of 0.618. For defining the onset of SCSSM, Tanaka (1992) used satellite-derived high cloud amount, Wang and Wu (1997) used zonal wind and OLR and Wang *et al* (2004) used objective criteria based on 850-hPa zonal winds averaged over the central South China Sea (5°–15°N and 110°–120°E).

The India Meteorological Department (IMD) has determined the date of MOK operational every year, for more than 100 years. On an operational mode, the date of MOK is based on the synoptic conditions as given by Forecasting Manual Unit (FMU) Report No. IV–18.2 by Ananthakrishnan *et al* (1968). On real time mode, declaration of the date of MOK was based on rainfall (Ananthakrishnan *et al* 1967). This criteria (old criteria) state that after 10 May, if any five stations out of the following seven stations, viz., Colombo, Minicoy, Thiruvananthapuram, Alapuzha, Kochi, Kozhikode, and Mangalore receive rainfall of 1 mm in 24 h for two consecutive days, the MOK may be announced on the second day. Accompanying such rainfall, the lower tropospheric westerly wind over Kerala is strong and deep and the relative humidity of the air is high from the surface to at least 500 hpa (Rao 1976). IMD has been taking these factors into consideration in a subjective way to determine the onset date. Fasullo and Webster (2003) proposed a hydrological definition of Indian monsoon onset and withdrawal. To diagnose the onset and withdrawal, vertically integrated moisture transport is used instead of rainfall. They argued that using rainfall over Kerala may be susceptible to 'false' or 'bogus' monsoon onsets, which are associated with propagating tropical intraseasonal disturbances unrelated to the monsoon onset (Flatau *et al* 2001). The disturbances are characterized by an enhancement of convection and westerly surface winds similar to the MOK but occurring over a smaller scale and lasting for a smaller duration (a week or less). Often bogus onsets are followed immediately by extended periods of weak winds and clear skies that result in heat waves and droughts in India.

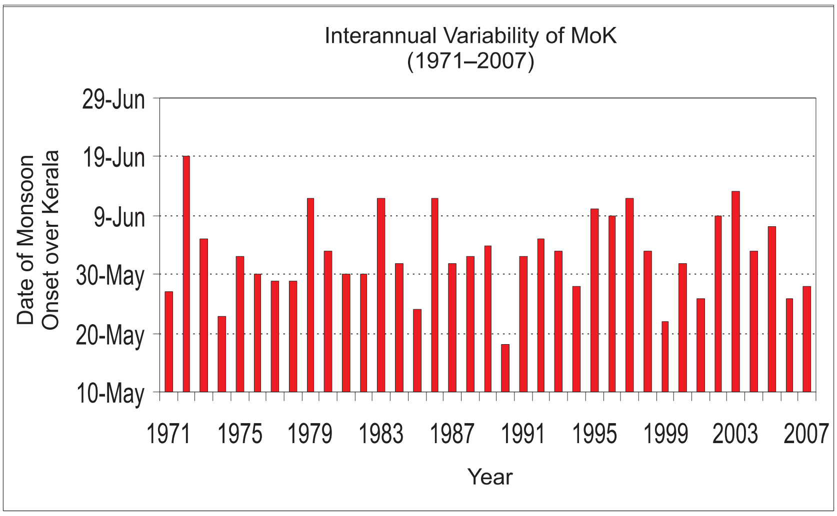

Figure 3. Interannual variation of date of monsoon onset over Kerala (MOK) during the period 1971–2007.

Flatau *et al* (2003) and Joseph *et al* (2006) discussed the bogus onset of 2002. A similar event happened in 2004 also associated with a severe cyclonic storm over Bay of Bengal. Joseph *et al* (2006) proposed a 3-tier strategy to determine the MOK objectively based on OLR and wind data in addition to rainfall realized around Kerala.

In 2006, India Meteorological Department adopted new criteria for declaring MOK operationally. These criteria use the information on rainfall and large scale circulation patterns as in Joseph *et al* (2006). The criteria for declaring the date of MOK are based on rainfall, wind field and OLR. They are given below:

- If after 10 May, 60% of the available 14 stations enlisted, viz., Minicoy, Amini, Thiruvananthapuram, Punalur, Kollam, Allapuzha, Kottayam, Kochin, Trissur, Kozhikode, Talassery, Cannur, Kasargode and Mangalore report rainfall of 2.5 mm or more for two consecutive days, the MOK may be declared on the second day, provided the following criteria are also satisfied in concurrence.
- Depth of westerlies should be maintained up to 600 hpa, in the box equator to latitude 10°N and longitude 55°–80°E. The zonal wind speed over the area bounded by latitude 5°–10°N, longitude 70°–80°E should be of the order of 15–20 knots at 925 hpa. The source of data can be RSMC New Delhi wind analysis/Satellite derived winds.
- INSAT derived OLR value should be below 200 Wm^{-2} in the box confined by latitude 5°–10°N and longitude 70°–75°E.

In these criteria, the emphasis has been given to the sharp increase in rainfall over Kerala. However, setting up of the large scale monsoon flow and extent of westerlies up to 600 hpa are also confirmed before declaring the monsoon onset over Kerala.

3. Scope of the present study and data

By April, general public as well as Government officials become curious to know the arrival of monsoon over Kerala. The arrival of monsoon is crucial for farmers to plan their crop strategy during the season. A delay in the MOK does not necessarily mean a delay in monsoon onset over NW India. However, a delay in the MOK is generally associated with a delay in onset at least over the southern states including the city of Mumbai. In spite of its importance, there are not many studies attempting to predict the date of MOK. Reddy (1977) proposed the May 50 hpa zonal wind component over Singapore with westerlies presaging an early and easterlies a late onset date. Kung and Shariff (1980, 1982) developed regression methods for predicting the onset date in Kerala based on April upper air patterns in the India–Australian region and sea surface temperature (SST) around India in the pre-monsoon season. Rajeevan and Dubey (1995) developed a regression model for long range prediction of monsoon onset over Kerala using April mean surface temperature and winter snow cover over Eurasia.

In this study, we have analyzed the variability of the date of MOK and an attempt has been made to predict the date of MOK using predictive signals derived from the thermal, convective and circulation patterns evolving over the Asia–Pacific region in association with the MOK, with the ultimate aim of developing a prediction tool for the operational use.

Table 1. *Dates of the MOK operationally derived by IMD for the period 1971–2007. For operationally declaring the date of MOK, IMD used old criteria till 2005 and new criteria from 2006 onwards. For the period 1971–2005, the dates of MOK were reworked using new criteria and are given in the last column of this table.*

Year	Date of MOK based on Old criteria	Date of MOK based on New criteria
1971	27 May	27 May
1972	18 June	19 June
1973	4 June	5 June
1974	26 May	23 May
1975	31 May	2 June
1976	31 May	30 May
1977	30 May	29 May
1978	28 May	29 May
1979	11 June	12 June
1980	1 June	3 June
1981	30 May	30 May
1982	30 May	30 May
1983	13 June	12 June
1984	31 May	1 June
1985	28 May	24 May
1986	4 June	12 June
1987	2 June	1 June
1988	26 May	2 June
1989	3 June	4 June
1990	19 May	18 May
1991	2 June	2 June
1992	5 June	5 June
1993	27 May	3 June
1994	28 May	28 May
1995	8 June	10 June
1996	3 June	9 June
1997	9 June	12 June
1998	2 June	3 June
1999	25 May	22 May
2000	1 June	1 June
2001	23 May	26 May
2002	29 May	9 June
2003	8 June	13 June
2004	18 May	3 June
2005	7 June	7 June
2006		26 May
2007		28 May

For this study, the data of date of MOK was derived based on the IMD criteria adopted in 2006 (new criteria) as mentioned in the section 2. The time series of date of MOK during the period 1971–2007 is shown in figure 3. The dates of MOK for the period 1971–2007 derived using the new criteria are also given in table 1 along with that derived using the old criteria. As seen in table 1, during 1971–2007, the extreme dates of monsoon onset over Kerala derived based on new criteria were 18 May 1990 and 19 June 1972.

The NCEP/NCAR (Kalnay *et al* 1996) daily data of surface mean sea level pressure, OLR and zonal winds at 925 hpa and 200 hpa were obtained from the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA (http://www.cdc.noaa.gov/). The spatial resolution of these data is $2.5° \times 2.5°$ (latitude × longitude). The other data used were daily minimum surface air temperatures of six stations (Deesa, Rajkot, Guna, Bikaner, Barmer and Akola) over northwest India and daily rainfall data of all available rainguage stations in the latitude region of 8°–13°N over south peninsula obtained from IMD's National Data Center. All the data were used for the period 1971–2007.

4. Methodology

For identification of the predictors, daily data of wind, mean sea level pressure, land surface minimum temperature, rainfall over south India and OLR over the Asia–Pacific region were analyzed. The monsoon onset process over Kerala is influenced by both intraseasonal oscillations (ISOs) as well as large scale circulation patterns. Hence in this study, the development of the model for the prediction of date of MOK has been designed on the basis of predictors indicating the large scale circulation patterns and ISOs. We have derived the data of predictors averaged over two short periods 16–30 April and 1–15 May. For establishing the relationship of some of the climate variables such as mean sea level pressure, zonal wind, OLR, etc., with MOK, we have prepared spatial maps of correlation of date of MOK with these climate fields. The correlation maps were prepared using data for the period 1975–2000. As seen in figure 3, the MOK has generally occurred after 15 May but in many years, the onset occurred before 31 May. Therefore, to predict the date of MOK around 15 days in advance, two prediction models have been developed. The first model (Model-1) uses the data up to 30 April and the second model (Model-2) makes use of data up to 15 May. With this, the first prediction can be made available by the end of April and the second prediction can be made available by 15 May. Prediction using Model-1 is particularly important in years when the MOK occurs in the middle of May. The Model-2 provides an update prediction, which will be of particular interest in years when the MOK occurs around the normal date or later.

The technique used here to develop the prediction models was the principal component regression (PCR). In this method, the principal components of the predictors are used in regression analysis to develop the prediction algorithm. The PCR technique is recommended when there is significant inter-correlation among the independent variables. The PCR model avoids the inter-correlation and helps to reduce the degrees of freedom by restricting the number of independent variables (Rao 1964). PCR model has been used for the prediction of all India summer monsoon season (June–September) rainfall for the country as a whole based on predictors from the Indian Ocean only (Singh and Pai 1996). PCR model has also been used for the prediction of seasonal summer monsoon rainfall over two homogeneous regions of India based on predictors from various observed climatic fields (Rajeevan *et al* 2000). The general mathematical formulation of PCR model is given below in brief.

A set of time series of 'm' inter-correlated standardised variables for 'n' years can be represented by $(n \times m)$ matrix, $\boldsymbol{Z} = [Z_{ij} : i = 1, \ldots, n;\ j = 1, \ldots, m]$. $\boldsymbol{Z}$ is transformed into $\boldsymbol{F}$ *via* the matrix transformation, $F_{ij} = Z_{ij}L_{ji}$, using principal component analysis (PCA). In matrix form above equation can be written as: $\boldsymbol{F} = \boldsymbol{ZL}$. Here $\boldsymbol{F}$ is the $(n \times m)$ matrix of PC scores each having zero mean and unit variance (i.e., the PC scores are standardized); $\boldsymbol{L}$ is $(m \times m)$ matrix of PC loadings. If we select any 'p' modes of the PC scores $(p < m)$, we can write the PCR model as $\boldsymbol{R} = \boldsymbol{BF'} + \varepsilon$, where $\boldsymbol{R}$ is the $(n \times 1)$ predictand matrix, $\boldsymbol{B}$ is the $(1 \times p)$ matrix of regression coefficients, $\boldsymbol{F'}$ is the $(n \times p)$ matrix of selected PC scores and ε is the $(n \times 1)$ error matrix.

For generating independent predictions, a sliding but fixed training period was used. In this technique, a fixed window of training period is moved across the data period and the prediction is made for one year just following the training period. This means that the prediction model gets updated regularly with the addition of latest data and at the same time the model training period remains the same. Such regular updating of the prediction models is necessary for better predictions (Kung and Sharif 1982; McBride and Nicholls 1983; Nicholls 1984) as the time series of meteorological parameters are statistically non-stationary. Kaiser's criterion (Kaiser 1958) was used for retaining the first few PCs for further analysis. As per this criterion, PCs with Eigen values equal to or more than one only are to be retained. Rationale behind this criterion is that the proportion of the variance explained by each of the retained PC should be at least equal to the contribution of variance (which is equal to one) by each of the standardized variables used as the input. The retained PCs were then used for training the MR model. Using the PC loadings of the retained PCs, PC scores were calculated for the reference year and the same were then used for the prediction for the reference year.

The skill of the PCR models is measured by calculating simple and well known model statistics such as correlation coefficient between the actual and predicted values and root mean square error (RMSE) of the model predictions.

5. Results

5.1 *Correlation maps*

The spatial maps of correlation coefficient (CC) between date of MOK and different climate variables from which some of the predictors were derived for the present study are shown in the figure 4(a–f). The climate variables considered for preparing the correlation maps are OLR, surface mean sea level pressure and zonal wind at 925 hpa and 200 hpa levels. In figure 4(a–f), the areas of positive (negative) CC significant at 5% significance level are shown shaded dark (light). In figure 4(a), the significant positive CC over the subtropical areas of Asia–Pacific region indicates below (above) normal surface mean sea level pressures over the region during earlier (later) than normal MOK years. Climatologically, during the second half of April, the surface pressure over the subtropical region of Asia is low mainly due to intense solar land heating. Whereas over the Pacific region, there is a climatological subtropical high. The intraseasonal oscillation leading to MOK is first noticed over the Pacific and its signal can be noticed over the subtropical northwest Pacific almost 20–25 days (Joseph *et al* 2006) before MOK. Thus, when the surface pressure over subtropical Asia–Pacific region is below normal during the second half of the April, it is indicative of the influence of intraseasonal oscillation over the region and earlier than normal MOK. Associated with the changes in the surface pressure pattern, changes also occur in the flow pattern. As the monsoon approach, the easterly trade winds over equatorial Indian Pacific region weaken and the westerly zonal winds set in. The significant negative CC between zonal wind at 925 hpa and MOK (figure 4b) over this region indicates stronger (weaker) than normal westerly winds associated with earlier (later) than normal MOK. Similarly, the significant positive CC areas over the equatorial east Indian Ocean and neighbouring Indonesian region (figure 4c) indicate stronger (weaker) than normal climatological easterly winds at 200 hpa associated with earlier (later) than normal MOK. The climatological low level westerly

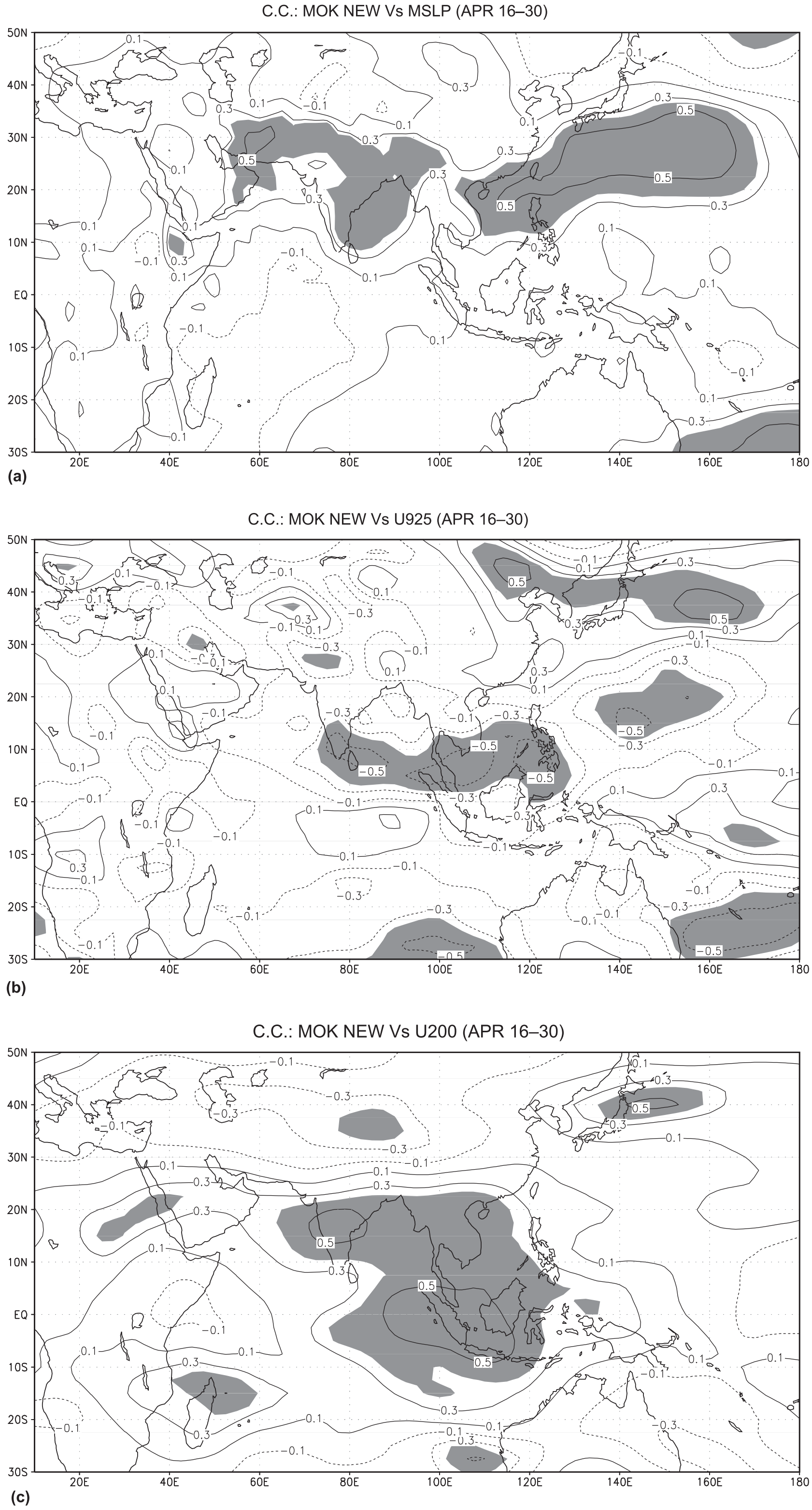

Figure 4. **(a)** Contour map of correlation coefficient (CC) between the surface mean sea level pressure (averaged over April 16 to 30) over Asia–Pacific region and date of MOK. The CC was computed using data for the period 1975–2000. Solid (dotted) contours are used for positive (negative) CC. The contour interval is 0.2. The areas of CC significant at and above the 95% significant level are shaded. **(b)** Same as figure 4(a) but for CC between zonal wind at 925 hpa (averaged over April 16 to 30) and the date of MOK. **(c)** Same as figure 4(a) but for CC between zonal wind at 200 hpa (averaged over April 16 to 30) and the date of MOK. **(d)** Same as figure 4(a) but for CC between outgoing long wave radiation (averaged over April 16 to 30) and the date of MOK. **(e)** Same as figure 4(a) but for CC between zonal wind at the 925 hpa (averaged over May 1 to 15) and the date MOK. **(f)** Same as figure 4(a) but for CC between outgoing long wave radiation (averaged over May 1 to 15) and the date of MOK.

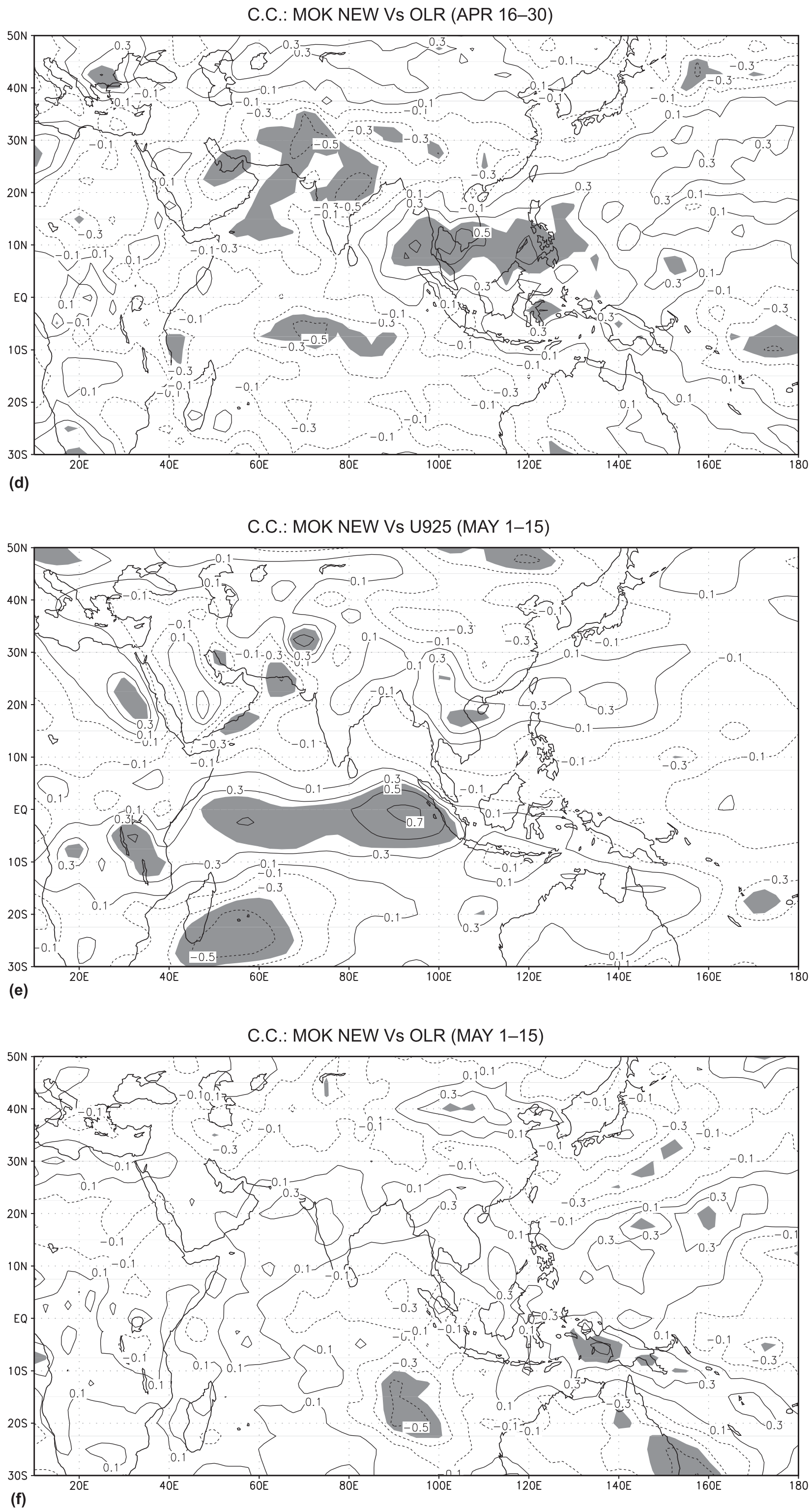

Figure 4. (*Continued*).

winds and upper level easterly winds over the equatorial east Indian Ocean during the second half of April is accompanied by climatological deep convection over the region. In fact, the intraseasonal variability leading to MOK is more prominent in the convective pattern. This is actually reflected in the spatial map of CC between OLR and MOK date (figure 4d), where significant positive CCs are observed in the latitude zone of 5°–15°N extending from the east Indian Ocean to Philippines. The positive CC indicates below (above) normal OLR over the region associated with earlier (later)

Table 2. *Details of the predictors used in the models (Model-1 and Model-2) for the prediction of date of MOK. First five predictors listed in this table were used in the Model-1 and last six predictors were used in the Model-2. All the correlations are significant at and above 95% significance level.*

No.	Name of predictor	Temporal domain	Geographical domain	Correlation coefficient (CC) 1975–2000
1	Minimum surface air temperature over NW India (P1)	16–30 April	1. Deesa 2. Rajkot 3. Guna 4. Bikaner 5. Akola 6. Barmer	−0.38
2	Surface mean sea level pressure over subtropical NW Pacific (P2)	16–30 April	20°–30°N, 130°–160°E	0.57
3	Zonal wind at 925 hpa over north-east Indian Ocean (P3)	16–30 April	5°–10°N, 80°–110°E	−0.52
4	Zonal wind at 200 hpa over Indonesian region (P4)	16–30 April	5°S–5°N, 90°–120°E	0.48
5	OLR over South China Sea (P5)	16–30 April	5°–15°N, 100°–120°E	0.40
6	Pre-monsoon rainfall peak date (P6)	Pre-monsoon April–May	South peninsula (8°–13°N, 74°–78°E)	0.48
7	Minimum surface air temperature over NW India (P7)	1–15 May	1. Deesa 2. Rajkot 3. Guna 4. Bikaner 5. Akola 6. Barmer	−0.37
8	Zonal wind at 925 hpa over equatorial south Indian Ocean (P8)	1–15 May	10°S–0°, 80°–100°E	0.52
9	OLR over southwest Pacific (P9)	1–15 May	30°–20°S, 145°–160°E	−0.53

than normal MOK. The OLR anomalies are caused by the northward movement of the convective region associated with the intraseasonal oscillation. The significant CC areas observed in the spatial maps of CC of date of MOK with 925 hpa zonal wind during 1–15 May (figure 4e) and that with OLR during 1–15 May can also be explained as the link between the intraseasonal oscillation in these climate fields and MOK event.

5.2 *Predictor dataset*

Table 2 shows the details of the nine predictors used for the two models. The geographical locations of the predictors are given in figure 5. OLR, surface mean sea level pressure and zonal wind predictors were derived as the simple arithmetic average of the actual values over the respective geographical region whose domain is given in the fourth column of table 2. These geographical regions are derived from the correlation maps (figure 4) and correspond to the areas where the CC between the respective climate variable and MOK was at least significant at 5% level. The time

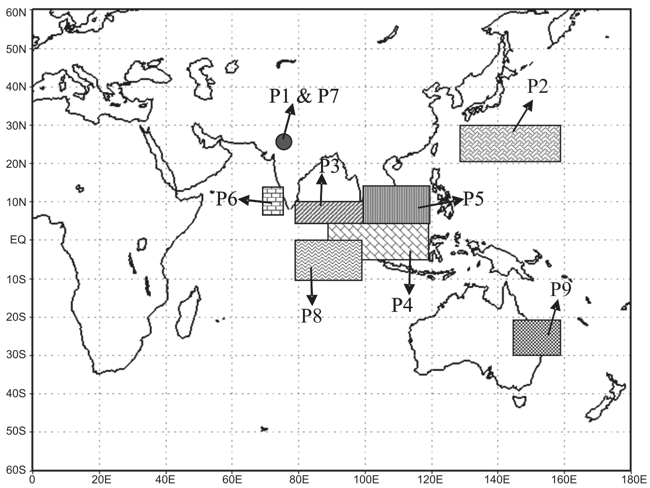

Figure 5. Geographical locations of the 9 predictors used in the PCR models. The 9 predictors (P1–P9) are listed in table 1.

periods used for the averaging are given in the third column of table 2. In 1978, due to the problem in the satellite, OLR data was not available. Therefore, this data gap was filled using normal

Table 3. *Intercorrelation coefficient among the nine predictors used for the forecasting of date of MOK. The intercorrelation coefficient was computed using data for the period 1975–2000. The intercorrelation coefficients between different predictors significant at and above 5% significance level are shown using bold letters.*

	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**
P1	1.00	−0.24	0.16	−0.22	−0.21	**−0.40**	**0.51**	−0.20	0.20
P2		1.00	**−0.75**	**0.55**	**0.69**	**0.46**	−0.16	0.25	**−0.58**
P3			1.00	**−0.75**	**−0.79**	−0.25	0.31	−0.27	**0.41**
P4				1.00	**0.66**	0.28	**−0.50**	**0.57**	**−0.52**
P5					1.00	0.24	−0.31	0.31	**−0.49**
P6						1.00	**−0.37**	**0.39**	−0.34
P7							1.00	−0.32	0.14
P8								1.00	**−0.53**
P9									1.00

value (base period 1975–2000) of the OLR over the relevant geographical region.

Other predictors in table 2 are date of Pre-Monsoon Rain Peak (PMRP) and surface air minimum temperature over northwest India. The PMRP (Joseph and Pillai 1988; Joseph *et al* 2006) is an event that occurs every year about 6–8 pentads earlier to the MOK, during which a cloud band passes through Kerala very similar to that occur during the monsoon onset phase with widespread rainfall over the extreme south Peninsula. The main difference between PMRP and MOK is that in association with the monsoon onset, a large area of deep convection can be seen over the south-east Arabian Sea. This feature is absent at PMRP and the large area of deep convection is, instead, seen over the south Bay of Bengal. Ramesh Kumar (2004) derived date of PMRP as the center date of the pentad with the maximum rainfall and that pertains to the period between 1 April to 10 May. For deriving date of PMRP, Ramesh Kumar (2004) used the rainfall derived from the Global Precipitation Climatology Project and *in situ* rain gauge data for the region bounded by 8°–13°N and 70°–95°E. In this study, the date of PMRP was derived from the daily rain gauge data averaged over the land region bounded between 8°–13°N as the date corresponding to the centre of the first peak rainfall envelop over the region.

The land surface air minimum temperature over the northwest India was computed as the average of surface air minimum temperatures of six surface observatories (Deesa, Rajkot, Guna, Bikaner, Barmer and Akola). The negative correlation between the minimum temperature index over northwest India and date of MOK indicates above (below) normal heating of the region associated with earlier (later) than normal MOK.

All the predictor time series were normalized using 1975–2000 base period before used in the model development. As shown in table 2, all the predictors have significant (> 95% level) correlation with the date of MOK. The cross correlation among the nine predictors is shown in table 3. As seen in this table, there is significant intercorrelation among some of the predictors. This is expected because the variations in most of the predictors are part of the evolution of the same physical mechanism leading to the monsoon onset. The first five predictors were used in Model-1 and the last six predictors were used in Model-2. As seen in table 2, all the five predictors used in Model-1 pertain to second half of April.

5.3 *Performance of the PCR models*

For predicting the date of MOK each year, predictor data for the 22 years just prior to the reference year were used for training the models. For example, for predicting the date of MOK in 2001, data for the period 1979–2000 was used for training the models. Similarly data for the period 1985 to 2006 were used for the prediction of date of MOK in 2007. Thus, with the available data, independent predictions were prepared for the 11-year period (1997–2007) for each of the two models (Models 1 and 2). As mentioned in the previous section, due to the significant intercorrelation among the predictors, PCA analysis was carried over each of the predictor sets. In case of Model-1, the PCA analysis of predictor series for each of the 11 independent prediction cases showed that eigen values of the only first two PCs were ≥ 1. Therefore, the first two PCs explaining about 86% of the total variability of the predictor set containing five predictors were retained for the regression analysis. In the case of Model-2 also the PCA analysis of predictor series resulted in the selection, the first two PCs having eigen value ≥ 1. These two PCs together explained about 72% of the total

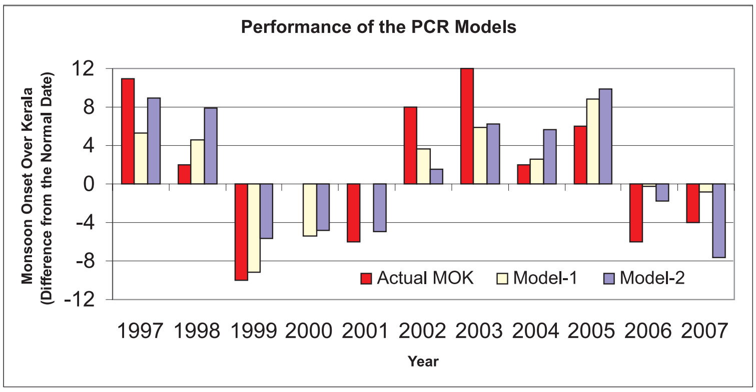

Figure 6. Actual dates of monsoon onset over Kerala (MOK) and forecasts from the PCR models for the period 1997–2007.

variability of the predictor set (containing 6 predictors) and were used for the regression analysis. We found that any addition of other PCs having eigen value less than one does not make any improvement in the model performance. Figure 6 shows the performance of the prediction by Models 1 and 2 for the period 1997–2007. As seen in figure 6, both the model predictions have shown good performance during the independent training period. During all the years both the models were able to correctly predict whether the MOK was earlier or later to the normal date (1 June as computed from dates of MOK for the period 1971–2007). The root mean square errors of model predictions during the independent test period of 11 years (1997–2007) for both the models was about 4 days (4.34 days for Model-1 and 4.30 days for Model-2) which is less than the standard deviation of the MOK (7 days). The CC between the actual and predicted MOK for both the models was 0.78. During the years like 1997, 2001, 2003 and 2006 when the MOK was earlier or later than the normal MOK date by more than or equal to 5 days, the prediction from Model-2 was closer to the actual MOK date than the prediction from Model-1. On the other hand, during years like 1999, 2002 and 2005, the prediction from Model-1 was closer to the actual MOK date than the prediction from Model-2. Though the average difference between the predictions from these two models during the period of 1997–2007 was 2.8 days, as a whole, the performance of both the models was nearly equal. The predictions from the Models 1 and 2 were also compared with that of a climatology-based model. The prediction from the climatology-based model for each year was computed as the mean of the actual MOK values during the training period (i.e., 22 years prior to the reference prediction year). The RMSE of the climatology-based model during the independent test period was 7.32, which is more than the RMSE of the predictions from PCR models discussed in this study. Thus, it is clear that the PCR models developed in this study have performed better than the climatology-based model.

6. Summary and conclusions

The summer monsoon over the Indian subcontinent first arrives over Kerala situated at the southern tip of the Indian Peninsula around 1 June with a standard deviation of about 7 days. The arrival of the monsoon over the region is signaled by widespread, persistent and heavy rainfall replacing the occasional pre-monsoon rains. In some years, the MOK occurs by the middle of May itself. Therefore, for the operational forecasts, we may need a model which can predict the date of MOK around 15 days ahead of the event. This means a model should predict the event at least by the end of the April. Fortunately, researchers have noticed that the evolution of atmospheric and oceanic process associated with the MOK starts almost 60–70 days ahead of the event. In this study, efforts were made to make use of predictive signals available in the convective, thermal and circulation patterns over the Asia–Pacific region associated with the event to predict the date of MOK well ahead of the event. For this purpose, two principal component regression (PCR) models using predictors were developed. Due to the possibility of the occurrence of MOK in the middle of May itself, one model (Model-1) was constructed using the April predictors only so that the prediction of MOK can be prepared in the end of April itself. The second model (Model-2) that can provide prediction for the date of MOK in the end of first half of May was prepared using four additional predictors pertaining to the first half of May along with two predictors used in the Model-1. Both the models showed good skill in the prediction of the date of MOK during the independent test period of 1997–2007. The RMSE of the predictions from both the models during the independent test period was about 4 days which was comparatively very less compared to RMSE of the predictions from a climatology-based model (7.32) during the same period and standard deviation (8 days) of the MOK date. This study demonstrates that the prediction of the date of MOK with satisfactory accuracy can be made by the end of April itself. In case of possibility of a late onset, the prediction can be updated by making use of predictive signals from May.

Acknowledgements

We express our sincere thanks to Dr Ajit Tyagi, DGM, Dr H R Hatwar ADGM (R) and Shri Thakur Prasad, DDGM (C) for their guidance, encouragement and support in carrying out this research work. We also thank Shri O P Sreejith, Shri S G Nargund, Shri J D Kale and all the staff members of the National Climate Centre for their assistance in the collection and compilation of the data required for this study.

References

Ananthakrishnan R and Soman M K 1988 The onset of southwest monsoon over Kerala for the period 1870–1990; *Int. J. Climatol.* **9** 283–296.

Ananthakrishnan R and Soman M K 1991 Onset of southwest monsoon in 1990; *Curr. Sci.* **61** 447–453.

Ananthakrishnan R, Acharya U R and Ramakrishnan A R 1967 On the criteria for declaring the onset of the southwest monsoon over Kerala; Forecasting Manual. FMU Report No. IV-18.1: 52, India Meteorological Department, Pune, India.

Ananthakrishnan R, Pathan J M and Aralikatti S S 1983 The onset phase of the southwest monsoon; *Curr. Sci.* **52** 155–164.

Ananthakrishnan R, Srinivasan V, Ramakrishnan A R and Jambunathan R 1968 Synoptic features associated with onset of southwest monsoon over Kerala; Forecasting Manual Report No. IV-18.2, India Meteorological Department, Pune, India.

Fasullo J and Webster P J 2003 A hydrological definition of Indian monsoon onset and withdrawal; *J. Climate* **16** 3200–3211.

Flatau M, Flatau P and Rudnick D 2001 The dynamics of double monsoon onsets; *J. Climate* **14** 4130–4146.

Flatau M, Flatau P, Schmidt J and Kiladys G 2003 Delayed onset of the 2002 Indian monsoon; *Geophys. Res. Lett.* **30** 1768.

Hsu H H, Terng C T and Chen C T 1999 Evolution of large-scale circulation heating during the first transition of Asian summer monsoon; *J. Climate* **12** 793–810.

Joseph P V and Pillai P V 1988 40-day mode of equatorial trough for long range forecasting of Indian summer monsoon onset; *Curr. Sci.* **57** 951–954.

Joseph P V, Eishcheid J K and Pyle R J 1994 Interannual variability of the onset of the Indian summer monsoon and its association with atmospheric features, El Niño and sea surface temperature anomalies; *J. Climate* **7** 81–105.

Joseph P V, Sooraj K P and Rajan C K 2006 The summer monsoon onset over process over south Asia and an objective method for the date of monsoon onset over Kerala; *Int. J. Climatol.* **26** 1871–1893.

Kaiser H F 1958 The varimax criterion for analytic rotation in factor analysis; *Psychometrika* **23** 187–200.

Kalnay E *et al* 1996 The NCEP/NCAR re-analysis project; *Bull. Am. Meteorol. Soc.* **77** 437–471.

Koteswaram P 1958 The Easterly Jet Stream in the Tropics; *Tellus* **10** 43–57.

Koteswaram P 1960 The Asian summer monsoon and the general circulation over the tropics; In: *Monsoons of the world*, India Meteorological Department, New Delhi, 105–110.

Krishnamurti T N and Ramanathan Y 1982 Sensitivity of the Monsoon Onset to Differential Heating; *J. Atmos. Sci.* **39** 1290–1306.

Kung E C and Sharif T A 1980 Regression forecasting of the onset of the India summer monsoon with antecedent upper air conditions; *J. Appl. Meteorol.* **19** 370–380.

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

McBride J L and Nicholls N 1983 Seasonal relationships between Australian rainfall and the southern oscillation; *Mon. Wea. Rev.* **111** 1998–2004.

Murakami T and Ding Y 1982 Wind and temperature changes over Eurasia during the early summer of 1979; *J. Meteorol. Soc. Japan* **60** 183–196.

Nicholls N 1984 The stability of empirical long range forecast techniques: A case study; *J. Appl. Meteorol.* **23** 143–147.

Pearce R P and Mohanty U C 1984 Onsets of the Asian Summer Monsoon 1979–82; *J. Atmos. Sci.* **41** 1620–1639.

Rajeevan M, Guhathakurta P and 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 and Dubey D P 1995 Long range prediction of monsoon onset over Kerala; *Mausam* **46** 287–290.

Ramesh Kumar M R 2004 Forecasting of onset of southwest monsoon over Kerala coast using satellite data; *IEEE Geosci. Remote Sensing Lett.* **1(4)** 265–267.

Rao C R 1964 The use and interpretation of principal component analysis in applied research; *Sankhya* (*Indian Journal of Statistics*) **26** 329–358.

Rao Y P 1976 Southwest monsoon; Meteorological Monograph, Synoptic Meteorology No. 1/1976. India Meteorological Department, New Delhi.

Reddy S J 1977 Forecasting the onset of southwest monsoon over Kerala; *Indian J. Meteorol. Hydrol. Geophys.* **28** 113–114.

Sikka D R and Gadgil S 1980 On the maximum cloud zone and the ITCZ over Indian longitudes during the southwest monsoon; *Mon. Wea. Rev.* **108** 1840–1853.

Singh O P and Pai D S 1996 An oceanic model for the prediction of SW monsoon rainfall over India; *Mausam* **47** 91–98.

Soman M K and Krishna Kumar K 1993 Space-time evolution of meteorological features associated with the onset of the Indian summer monsoon; *Mon. Wea. Rev.* **121** 1177–1194.

Tanaka M 1992 Intraseasonal oscillation and the onset and retreat dates of the summer monsoon east, southeast Asia and the western Pacific region using GMS high cloud amount data; *J. Meteorol. Soc. Japan* **70** 613–628.

Wang B and LinHo 2002 Rainy season of the Asian–Pacific summer monsoon; *J. Climate* **15** 386–396.

Wang B and Wu R 1997 Peculiar temporal structure of the South China Sea summer monsoon; *Adv. Atmos. Sci.* **14** 177–194.

Wang B, LinHo, Zhang Y and Lu M M 2004 Definition of South China Sea Monsoon Onset and Commencement of the East Asia Summer Monsoon; *J. Climate* **17** 699–710.

Webster P J, Magana V O, Palmer T N, Shukla J, Tomas R A, Yanai M and Yasunari T 1998 Monsoons: Processes, predictability, and the prospects for prediction; *J. Geophys. Res.* **103** 14,451–14,510.

Yin M T 1949 A synoptic-aerologic study of the onset of the summer monsoon over India and Burma; *J. Meteorol.* **6** 364–400.

Zeng X and Lu E 2004 Globally unified monsoon onset and retreat indexes; *J. Climate* **17** 2241–2248.

MS received 1 May 2008; revised 24 November 2008; accepted 25 November 2008