METEOROLOGICAL APPLICATIONS
Meteorol. Appl. **23**: 1–13 (2016)
Published online 11 January 2016 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/met.1524

RMetS
Royal Meteorological Society

Review

Prediction of summer monsoon rainfall over India and its homogeneous regions

Santosh Kakade* and Ashwini Kulkarni
Indian Institute of Tropical Meteorology, Pune, India

ABSTRACT: The coherent regions for various fields such as sea level pressure, temperature, geopotential height and zonal wind anomalies at the surface, 850, 500 and 200 hPa levels in pre-monsoon months (January through May) and seasons (winter, spring) have been identified by applying the shared nearest neighbour algorithm. The fields over the corresponding cluster regions could be possible predictors for Indian summer monsoon rainfall as well as the rainfall over various homogeneous regions of India. The time series have been constructed by averaging the parameters over the respective clusters. The relationship between these time series and the summer monsoon rainfall over India and its well-defined homogeneous regions over India, (northwest India, central northeast India, northeast India, west central India and peninsular India), has been examined during the positive and negative phases of effective strength index tendency using the simple technique of multiple regression. Along with the linear relationship, the non-linear relationships between the cluster parameters and the seasonal rainfall have also been considered. Independent cluster parameters have been selected by cross-validation procedure and the performance of each predictive model is tested. The extreme yearly rainfall departures over India are qualitatively well predicted by the model. Also, the unprecedented droughts over India in 2002 and 2009, where all earlier models have failed to forecast, are well predicted by the present model. The performances of models for summer monsoon rainfall prediction over homogeneous regions of India are convincing.

KEY WORDS shared nearest neighbour (SNN); cross validation; effective strength index (ESI) tendency

Received 11 November 2014; Revised 17 April 2015; Accepted 14 May 2015

1. Introduction

Long range forecasting (LRF) of Indian summer monsoon rainfall (ISMR) has a long history. The first official seasonal monsoon forecast was issued by Blanford in 1886. Sir Gilbert Walker laid the basis for a statistical forecast and issued the forecast based on a regression equation in 1906. Since then, many researchers (Jagannathan, 1960; Rao and Rama Moorthy, 1960; Rao, 1965; Banerjee *et al.*, 1978; Kung and Sharif, 1982; Bhalme *et al.*, 1986; Parthasarathy *et al.*, 1988, 1991; Gowariker *et al.*, 1989, 1991; Krishna Kumar *et al.*, 1995, 1997; Rajeevan *et al.*, 2004) have contributed to seasonal forecasting of the ISMR. The LRF predictors are identified by analysing the relationships between all India summer monsoon rainfall and the regional/global fields of several surface/upper-air parameters. Banerjee *et al.* (1978), Mooley *et al.* (1986) and Shukla and Mooley (1987) have suggested the 500 hPa ridge along 75 °E in April as the best predictor for the ISMR. Verma and Kamte (1980) and Joseph *et al.* (1981) have identified the 200 hPa meridional wind component in May over five stations (Mumbai, Delhi, Chennai, Nagpur and Srinagar) as the predictor for the ISMR. Mooley and Paolino (1988) have considered May minimum temperatures over the western Indian region as a good potential predictor for the ISMR. Shukla and Paolino (1983) have obtained winter to spring tendency (March to May and December to February) of sea level pressure (SLP) as a reliable predictor for the ISMR. Verma *et al.* (1985) have identified the northern hemisphere winter (January + February) surface air temperature anomaly as an important predictor for LRF of the ISMR. Parthasarathy *et al.* (1988) have shown that winter to spring tendency in sea surface temperatures (SSTs) over the Niño4 region is inversely associated with ISMR and the relationship is significant at the 5% level. Parthasarathy *et al.* (1990, 1992) have suggested mean surface temperatures (STs) and SLPs at six stations (Jodhpur, Ahmedabad, Mumbai, Indore, Sagar and Akola) during spring (March–April–May) as LRF predictors. Parthasarathy *et al.* (1991) have considered the arithmetic average of the 200 hPa meridional component of wind for May at Mumbai, Delhi, Chennai, Nagpur and Srinagar as the predictor for the ISMR. Gowariker *et al.* (1991) have developed the operational LRF model with three predictors, minimum temperature over northern, central and east coastal areas of India. Krishna Kumar *et al.* (1992) have shown that April location minus March location of 500 hPa ridge along 75 °E can be used as a predictor for the ISMR. Krishna Kumar *et al.* (1995) identified two predictors based on the minimum temperatures during March over east peninsular India and during May over west central India. In all these studies, predictors are obtained by averaging a parameter field at the corresponding level over a region showing significant association between the parameter field and the overall ISMR.

* Correspondence: S.B. Kakade, Indian Institute of Tropical Meteorology, Dr. Homi Bhabha Road, Pashan, Pune 411008, Maharashtra, India.
E-mail: kakade@tropmet.res.in

The clustering of the parameter field at the corresponding level is not considered. In the present study, the shared nearest neighbour (SNN) clustering procedure was used to find clusters of temperature, pressure, geopotential height and U-wind field at surface, 500 hPa and 200 hPa levels. The SNN cluster procedure is discussed by Kakade and Kulkarni (2013) and for more details the reader is referred to Jarvis and Patrick (1973), Delsole and Shukla (2002), Sahai *et al.* (2002), Ertoz *et al.* (2003) and Boriah *et al.* (2004).

Kakade and Kulkarni (2012a, 2012b) have demonstrated that evolution of ST and SLP from winter to spring changes with the phase of effective strength index (ESI) tendency. Kakade and Kulkarni (2013) obtained two separate multiple regression equations for LRF of the ISMR during contrasting phases of the ESI tendency using independent cluster parameters of ST and SLP. In the present study, independent cluster parameters of temperature, geopotential height and zonal wind at the 850, 500 and 200 hPa levels along with the ST and SLP are considered.

For any regression model, the sample size is always limited in practice and the parameter values are adjusted to the peculiarities of the sample. Therefore, such models perform very well on the data set from which they were derived, but perform poorly on a new, independent data set. Hence, in this study, the parameters that perform well on an independent data set have been selected. The cross-validation procedure has been applied for selecting the cluster parameters in each phase of the ESI tendency. Two separate multiple regression equations are developed during contrasting phases of the ESI tendency for LRF of the ISMR. A similar procedure is adopted for obtaining multiple regression equations for LRF of the ISMR over the homogeneous regions of India.

The data used in the analysis and the methodology are described in Sections 2 and 3, respectively. Section 4 details the main results of prediction of the overall ISMR and the rainfall over the homogeneous regions of India, and conclusions are given in Section 5.

2. Data used

In this study, the following data have been used for the period 1951–2012.

- Monthly rainfall data for the monsoon season (June–September) for India and its homogeneous regions (Figure 1) have been taken from the website of the Indian Institute of Tropical Meteorology, Pune, http://www.tropmet.res.in. The seasonal rainfall series is prepared by accumulating the rainfall from June to September. The percentage departures from the long-term (1951–2012) mean are calculated and have been used in further analysis.
- The gridded $2.5^\circ \times 2.5^\circ$ latitude/longitude global monthly mean values for SLP, ST, geopotential height and zonal wind anomaly data at the surface, 850, 500 and 200 hPa levels have been taken from the NCEP/NCAR Reanalysis data set. The data have been interpolated on $5^\circ \times 5^\circ$ latitude/longitude to apply the SNN algorithm.
- Monthly North Atlantic Oscillation (NAO) and Southern Oscillation (SO) data have been taken from http://www.cpc.ncep.noaa.gov. The ESI is defined as the algebraic difference between monthly indices of NAO and SO. The anomalies from the annual mean have been calculated for each month and these anomaly series are then divided by the standard deviation (SD). These series are called as the ESI series of the respective month. The ESI tendency is the difference between April and January ESI values (April–January).
- Monthly Arctic Oscillation (AO) data have been taken from the website http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/monthly.ao.index.b50.current.ascii.table. The indices are constructed by projecting monthly mean 1000 hPa height anomalies onto the leading Empirical Orthogonal Function (EOF) mode. The time series is normalized by the SD of the monthly index (1979–2000 base period).

Seasonal means for all the above parameters are computed by averaging over the winter (December–January–February), spring (March–April–May), summer (June–July–August) and autumn (September–October–November).

3. Methodology

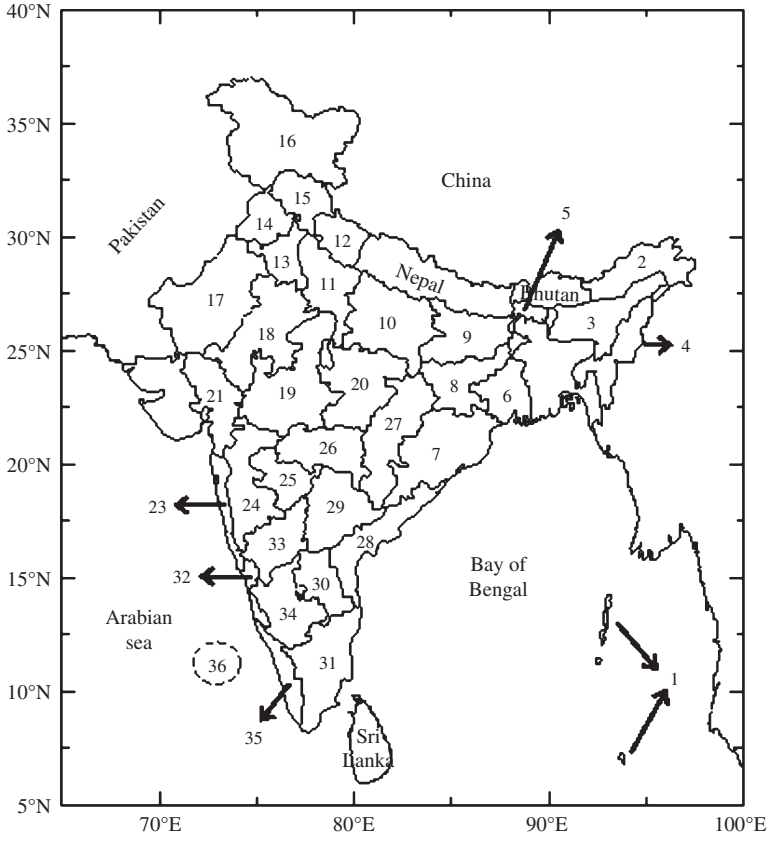
The following methods are used in this study.

3.1. SNN cluster analysis

Cluster analysis has been used to group the data points into useful and meaningful groups (clusters). Clustering has a long history and a large number of clustering techniques have been developed. For high-dimensional data such as meteorological data, traditional clustering methods such as K-means algorithm and agglomerative hierarchical technique do not perform well because the data contain outliers and clusters of different sizes, shapes and densities. In high-dimensional data sets, the traditional Euclidean notion of density, which is the number of points *per* unit volume, is meaningless. Therefore, a traditional density-based method cannot be used to identify the core points in the high-density regions. An alternative to this is to define the similarity (closeness) between a pair of points in terms of their SNNs. The technique was first developed by Ertoz *et al.* (2003) to find the clusters of different sizes, shapes and densities in noisy, high-dimensional data. Boriah *et al.* (2004) then applied this SNN technique to ocean temperatures and used the clusters to predict land temperatures.

The SNN technique is applied to pressure, temperature, zonal wind and geopotential height at the surface, mid-troposphere and upper troposphere to obtain cluster regions. The steps involved in the SNN clustering algorithms are described below.

1. To compute the similarity matrix. The correlation co-efficient (CC) between the time series over two grid points, for 1951–2012, is the typical measure of similarity.
2. To sparsify the similarity matrix by keeping only its k strongest links. Here, k is called the neighbourhood list size. It is the most important factor as it adjusts the focus of the clusters. If k is too small, even a uniform cluster will be broken up into pieces due to local variations in similarity, and the algorithm will tend to find many small, but tight clusters. On the other hand, if k is too large, then the algorithm will tend to find only a few large, well-separated clusters and small local variations in similarity will not have an impact. In the SNN, a point can be similar to at most k other points.
3. To construct the SNN list from the sparsified similarity matrix. At this point, one could apply a similarity threshold and find the connected components to obtain the clusters (Jarvis–Patrick algorithm). For each grid point, a neighbour list containing grid points that show significant (at the 1% level) CC is prepared. Jarvis and Patrick (1973) suggested



1. Andaman and Nicobar Islands
2. Arunachal Pradesh
3. Assam and Meghalaya
4. Nagaland, Manipur, Mizoram and Tripura
5. Sub-Him. West Bengal and Sikkim
6. Gangetic West Bengal
7. Orissa
8. Jharkhand
9. Bihar
10. East Uttar Pradesh
11. West Uttar Pradesh
12. Uttaranchal
13. Haryana, Chandigarh and Delhi
14. Punjab
15. Himachal Pradesh
16. Jammu and Kashmir
17. West Rajasthan
18. East Rajasthan
19. West Madhya Pradesh
20. East Madhya Pradesh
21. Gujarat
22. Saurashtra, KKutch and Diu
23. Konkan and Goa
24. Madhya Maharashtra
25. Marathwada
26. Vidarbha
27. Chattisgarh
28. Coastal Andhra Pradesh
29. Telangana
30. Rayalaseema
31. Tamil Nadu and Pondicherry
32. Coastal Karnataka
33. North Interior Karnataka
34. South Interior Karnataka
35. Kerala
36. Lakshadweep

Northwest India (NWI) 13, 14, 17, 18, 21, 22
Central northeast India (CNEI) 7, 8, 9, 10, 11
Northeast India (NEI) 3, 4, 5, 6
West central India (WCI) 19, 20, 23, 24, 25, 26, 27, 29, 33
Peninsular India (PI) 28, 30, 31, 32, 34, 35
Hilly regions not considered 2, 12, 15, 16

Figure 1. Homogeneous regions of India.

that a link is created between points i and j if and only if both i and j have each other in their closest k nearest neighbour lists, where k is the nearest neighbour size (here $k = 100$). Let i, j be two points. The strength of the link between i and j is then calculated as:

$\mathrm{Str}(i,j) = \Sigma(k+1-m) \times (k+1-n)$, where k is the nearest neighbour list size, and m and n are positions of the SNNs in the lists of i and j.

4. To find the SNN density of each point. Here, the sum of link strengths for every point is considered. The points having high total link strength will become candidates for representative points, while those having very low total link strength become candidates for noise points.
5. To find the core points. These are the points having SNN density greater than the threshold value. This value is to be decided by trial and error. (Here, we have taken it to be 100).
6. To form clusters from the core points by averaging the grids in the cluster.

3.2. Cross-validation scheme

In the cross-validation procedure, the data are divided into two disjoint sets: one is treated as the dependent sample for developing (or 'training') the model and the other as the independent sample for verifying (or 'testing') the estimated forecasts. Suppose the total independent sample is partitioned into K mutually exclusive sets of equal size; let the kth partition be I_k. Any one of these partitions can be considered the independent set, while the complement of this is considered as the dependent set (D_k). The quality of the resulting forecasts can be measured by the mean-square forecast errors and anomaly correlation co-efficients (ACCs) in the independent set.

The distinction between the dependent and independent error variances becomes clear for small sample sizes N. For large N, the two error estimates approach the same value. If P is the number of predictors, then as P/N increases, the expected error in the dependent data tends to be negligible while the expected error in the independent set approaches infinity. Davis (1976) derived similar results in the context of autoregressive models for moderately large N. Therefore, for a good predictive model, a minimum number of predictors should be used to achieve reasonably good forecasts in the independent set. Delsole and Shukla (2002) proposed a cross-validation scheme to select the best linear prediction model for the ISMR. The cross-validation procedure first screens out all models that are likely to perform poorly on independent data sets and then the prediction error of each model is compared with those of all other models to determine whether the difference in error variance exceeds some

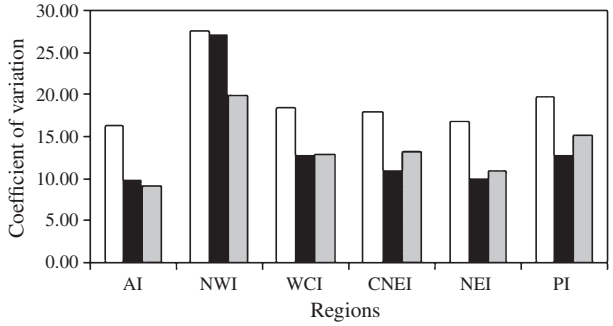


Figure 2. Co-efficient of variation (CV %) of summer monsoon rainfall (in mm) for total period 1951–2012 (white column), 30 positive ESI-tendency years (black column) and 32 negative ESI-tendency years (grey column) for all India and its five homogeneous regions.

threshold of significance. Sahai *et al.* (2002) used this scheme for selecting the best predictor set. The cross-validation scheme involves the following steps for selecting the best predictor set for predictions in the independent data sets.

1. Model development data (N number of years) are divided into two mutually exclusive sets, the independent set consisting of 1 year and the dependent set consisting of the remaining $N - 1$ years. In the present study, the data of model development period (1951–2012) have been divided into N (=62) mutually exclusive dependent and independent sets in which each independent set consists of 1 year and the remaining $N - 1$ years are in the dependent set.
2. For each predictor, N independent predicted values can be obtained along with the observed ones. The values of root mean square error (RMSE) are calculated for each predictor and the predictor having the minimum RMSE value is selected as the first predictor.
3. Let P be the number of predictors. After selecting the first predictor, trial regression equations are again constructed using the first selected predictor in combination with the remaining $P - 1$ predictors. The combination of two predictors having the minimum RMSE value is selected and this procedure is repeated for all P predictors.
4. The minimum RMSE value is plotted against the number of predictors and the set of predictors associated with the number of predictors having the least RMSE value is considered as the best predictor set.

4. Results and discussion

Among the 62 years (1951–2012), 30 years show positive ESI tendency and 32 years show negative ESI tendency. Figure 2 shows the co-efficient of variation (CV) of summer monsoon rainfall throughout India and its homogeneous regions for 30 positive ESI tendency years, 32 negative ESI tendency years and the total period of 62 years (1951–2012). It shows that the CV for the total period 1951–2012 is greater than the CVs for the positive and negative ESI tendency years during 1951–2012 throughout India and its homogeneous regions. Thus, summer monsoon rainfall variation over India and its homogeneous regions is reduced by classifying the rainfall years as positive and negative ESI tendency. In view of this and the methodology of predicting the ISMR using separate regression equations depending on the phase of the ESI tendency (Kakade and Kulkarni, 2013), the regression equations are modified by considering

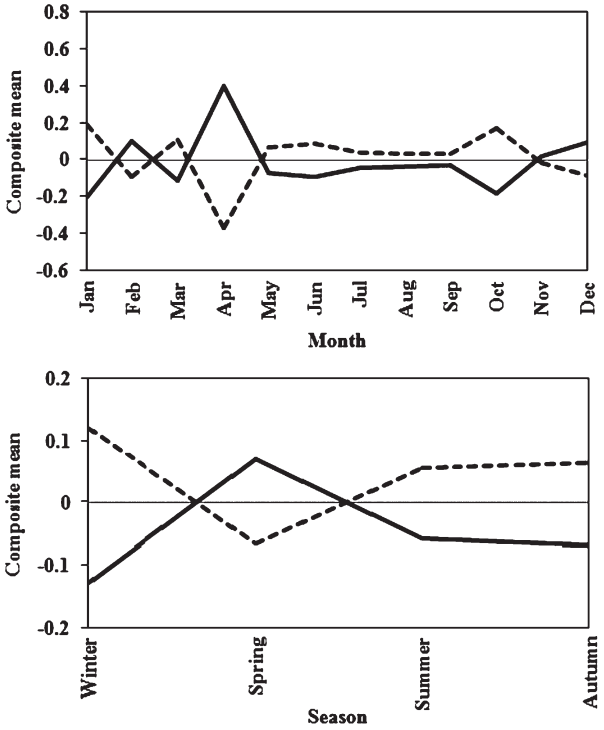


Figure 3. Composite mean anomalies of Arctic Oscillation (AO) from 1951–2012 during 30 positive ESI-tendency years (continuous line) and 32 negative ESI-tendency years (dashed line) on monthly (top panel) and seasonal (bottom panel) scale.

cluster parameters at different levels of various meteorological parameters. The study is further extended for prediction of summer monsoon rainfall over five homogeneous regions of India.

4.1. Monthly and seasonal composites of AO

The AO is a large-scale oscillation and is also referred to as the Northern Hemisphere annular mode. The AO is a pattern characterized by winds circulating counterclockwise around the Arctic at around 55 ° N latitude. During positive AO, a ring of strong winds (westerlies) around the North Pole confine colder air across the polar regions. This belt of winds becomes weaker and more distorted in the negative phase of the AO, which allows a penetrating colder Arctic air mass southward by increasing storminess into the mid-latitude. Figure 3 shows composite monthly and seasonal means of the AO anomaly from the 1951–2012 mean during 30 positive ESI tendency years and 32 negative ESI tendency years. It shows that during positive (negative) ESI tendency, the AO anomaly is negative (positive) from May to October. It further shows that during positive (negative) ESI tendency, the AO anomaly increases (decreases) from winter to spring and decreases (increases) from spring to autumn and becomes negative (positive) in summer.

4.2. Cluster parameters for LRF of the ISMR

Kakade and Kulkarni (2013) have discussed the SNN procedure to find clusters of SLP and ST anomaly. The cluster time series have been prepared by averaging the SLP and ST over each cluster region. When the relationship between the cluster parameter and the seasonal monsoon rainfall is non-linear, various powers of the parameters are considered. The power

Table 1. Number of cluster regions of meteorological parameter anomalies at different altitudes during pre-monsoon months and seasons.

	January	February	March	April	May	Winter	Spring	Total
SLP	12	12	6	8	7	20	9	74
STEMP	12	4	10	17	11	22	13	89
500T	7	7	6	16	9	20	21	86
500G	13	12	10	11	15	7	8	76
200T	9	9	13	14	14	14	13	86
200G	15	6	8	11	11	7	8	66
850U	0	0	0	0	0	9	9	18
500U	0	0	0	0	0	4	6	10
200U	0	0	0	0	0	13	10	23
850G	12	13	16	8	13	13	12	87
Total	80	63	69	85	80	129	109	615

SLP, sea level pressure; STEMP, surface temperature; 500T, 500 hPa temperature; 500G, 500 hPa geopotential height; 200T, 200 hPa temperature; 200G, 200 hPa geopotential height; 850U, 850 hPa U-wind; 500U, 500 hPa U-wind; 200U, 200 hPa U-wind; 850G, 850 hPa geopotential height.

showing the maximum correlation was considered for developing the prediction equation. They further showed that prediction skill is enhanced if separate multiple regression equations for positive and negative ESI tendency are developed. Since the atmosphere is heated by terrestrial radiation and atmospheric circulations are known to affect the ISMR, inclusion of cluster parameters at 850, 500 (mid-troposphere) and 200 hPa (upper troposphere) levels of temperature, geopotential height and zonal wind along with surface cluster parameters may improve the predictive skill in LRF of the ISMR. The present study has been extended to predict the summer monsoon rainfall over different homogeneous regions of India.

Table 1 shows the number of cluster regions of meteorological parameters at different altitude levels. There are a total of 615 cluster regions, out of which zonal wind anomaly clusters are observed on seasonal scale only. Cluster time series for all these 615 cluster regions, for 1951–2012, is prepared by averaging the corresponding meteorological parameter at all grids in the respective cluster region.

In order to understand the linear and non-linear relationship of the cluster time series for 615 cluster regions with the ISMR, the CCs between the ISMR and the different powers of cluster time series are computed during contrasting phases of the ESI tendency. The powered cluster time series showing significant (at 5% level) CC is considered as the cluster parameter for LRF of the ISMR. The set of independent cluster parameters for predicting the ISMR has been developed in each phase of the ESI tendency.

4.3. Selection of independent cluster parameters

During positive (negative) ESI tendency, we have $N = 30$ ($N = 32$) and $P = 8$ ($P = 10$). The cross-validation procedure has been applied to each phase of the ESI tendency and the RMSE values in the independent data set are computed for each group of predictors containing 1 to P number of predictors. Figure 4 shows the least RMSE, in the independent data set, in predicting the summer monsoon rainfall throughout India and its five homogeneous regions for each number of predictors during contrasting phases of the ESI tendency. Figure 4 (top panel left side) depicts the least RMSE in predicting the ISMR for each group of predictors during contrasting phases of the ESI tendency. It suggests that during positive and negative ESI tendency, each of the two disjoint sets containing eight independent cluster parameters can predict the ISMR. Table 2 shows selected independent cluster predictors by cross-validation procedure for LRF of the ISMR during positive and negative ESI tendency, respectively. Figure 5 shows the locations of cluster predictors during contrasting phases of the ESI tendency. It shows that during positive (negative) ESI tendency, cluster predictors are located up to mid-troposphere (upper troposphere). During positive (negative) ESI tendency, the strength of NAO (SO) tendency from winter to spring is relatively stronger than that of SO (NAO) tendency. This increased strength of north–south (east–west) oscillation in northern hemisphere (equatorial Pacific Ocean) may have changed the locations of cluster predictors.

4.4. LRF of summer monsoon rainfall throughout India

Two separate multiple regression equations are obtained for both the phases of ESI tendency. Equation (1) represents multiple regression equation for LRF of the ISMR during positive ESI tendency:

$$\begin{aligned}\text{Rainfall} = {} & 1.46 - 0.84\,(\text{AI-CPP1}) + 0.16\,(\text{AI-CPP2}) \\ & - 2.55\,(\text{AI-CPP3}) - 0.17\,(\text{AI-CPP4}) - 1.13\,(\text{AI-CPP5}) \\ & - 2.77\,(\text{AI-CPP6}) + 1.98\,(\text{AI-CPP7}) - 0.75\,(\text{AI-CPP8}) \quad (1)\end{aligned}$$

In the independent data set, RMSE is 5.9 and the ACC during positive ESI tendency is 0.85. The RMSE and ACC are computed by the following formula:

$$\text{RMSE} = \left((1/N)\left(\Sigma\left(R_{\text{actual}} - R_{\text{estimated}}\right)^2\right)\right)^{0.5}$$

$$\text{ACC} = \left(\Sigma\left(\left(R_{\text{actual}}\right)\left(R_{\text{estimated}}\right)\right)\right) / \times\left(\left(\left(\Sigma\left(R_{\text{actual}}\right)^2\right)\left(\Sigma\left(R_{\text{estimated}}\right)^2\right)\right)^{0.5}\right)$$

Equation (2) represents multiple regression equation for LRF of the ISMR during negative ESI tendency:

$$\begin{aligned}\text{Rainfall} = {} & 3.14 + 0.36\,(\text{AI-CPN1}) - 1.7\,(\text{AI-CPN2}) \\ & - 4.73\,(\text{AI-CPN3}) - 3.11\,(\text{AI-CPN4}) + 3.42\,(\text{AI-CPN5}) \\ & + 2.89\,(\text{AI-CPN6}) - 1.91\,(\text{AI-CPN7}) + 2.03\,(\text{AI-CPN8}) \quad (2)\end{aligned}$$

In the independent data set, the RMSE is 5.83 and the ACC during negative ESI tendency is 0.80. Since the phase of the ESI tendency is known by April, much before the beginning of monsoon season, we can choose the equation depending upon the phase of the ESI tendency for LRF of the ISMR. Figure 6(a) depicts the actual and estimated summer monsoon rainfall over India for 1951–2012. Mean error and RMSE on the domain 1951–2012 are 0.00 and 4.25, respectively, whereas CC between the actual and estimated rainfall departure is 0.90. When the monsoon rainfall departure over India or its homogeneous regions is greater than or equal to sum of the long-term mean and SD, then that year is considered as flood year for India or that homogeneous region. When the monsoon rainfall departure over India or its homogeneous regions is less than or equal to difference of the long-term mean and SD, then that year is considered as drought year for India or that homogeneous region. During 1951–2012, 13 years are drought years and 10 years are flood years. All estimated rainfall departure values in drought (flood) years are negative (positive) indicating that extreme rainfall departures are qualitatively predicted well. Where all the statistical as well as dynamical models failed to forecast the unprecedented droughts of 2002 and 2009, this model predicts them with reasonable skill.

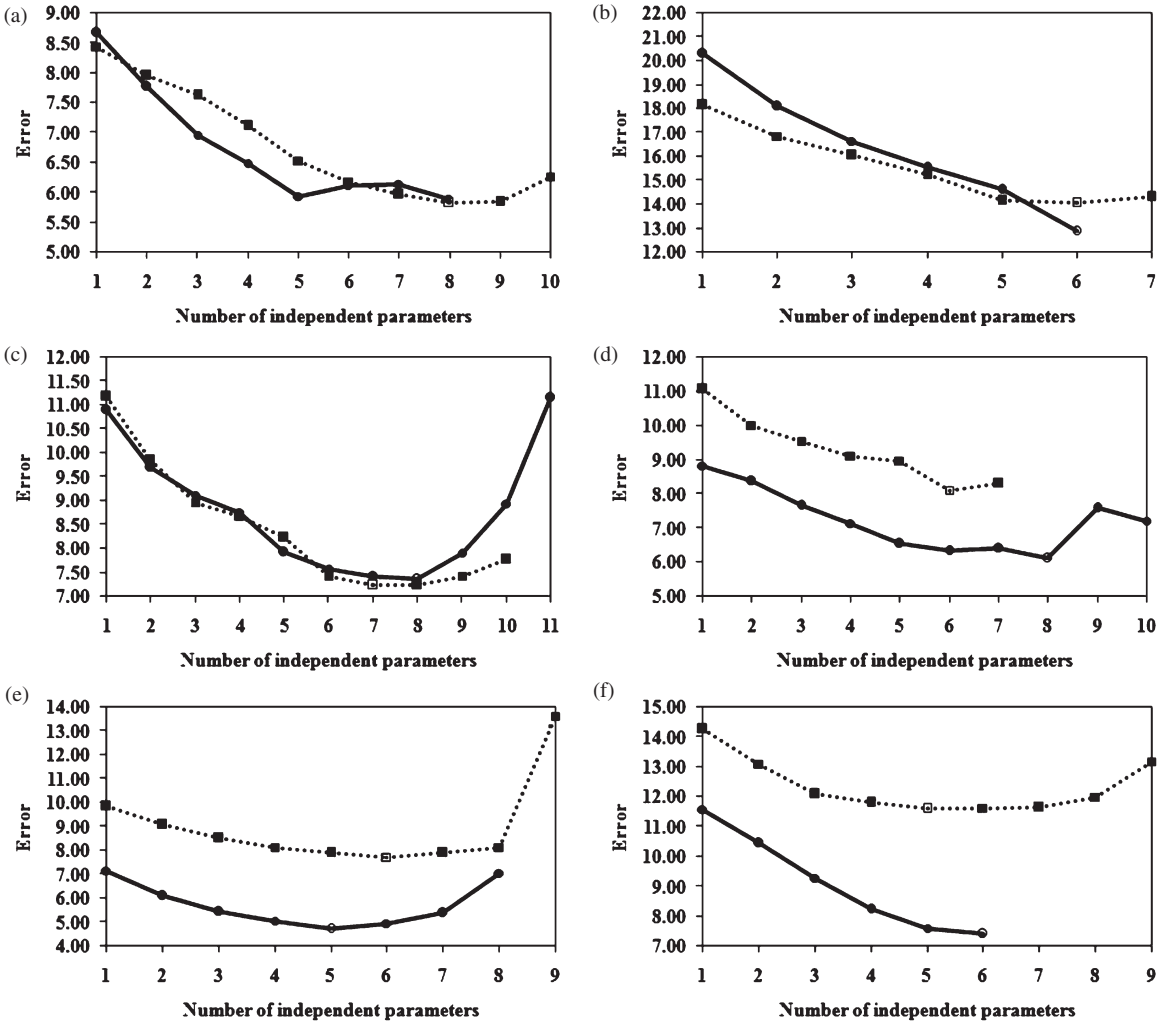


Figure 4. Least root mean square error (RMSE) in predicting summer monsoon rainfall over All India (top panel left), North West India (top panel right), West central India (middle panel left), Central north east India (middle panel right), North east India (bottom panel left) and Peninsular India (bottom panel right) for each number of predictors during positive (continuous line) and negative ESI-tendency (dotted line).

Table 2. Independent cluster parameters selected by cross validation for LRF of all India (AI) summer monsoon rainfall during (a) positive and (b) negative phase of ESI tendency.

Abbreviation	Description	CC
(a) Positive ESI tendency (30 years)		
AI-CPP1	4th power of averaged Spring 850 hPa U-wind anomaly for the region (60°–70° N; Equatorial –40° E)	−0.52
AI-CPP2	5th power of averaged Spring 850 hPa geopotential height anomaly for the region (15°–35° N; 90–60° W)	−0.36
AI-CPP3	Square of averaged Winter 850 hPa geopotential height anomaly for the region (20°–55° N; 85°–40° W)	−0.58
AI-CPP4	5th power of averaged May 500 hPa temperature anomaly for the region (80°–90° N; 70°–25° W)	−0.43
AI-CPP5	Square of averaged Winter 850 hPa U-wind anomaly for the region (60°–70° N; 65°–20° W)	−0.53
AI-CPP6	Square of averaged April surface temperature (ST) anomaly for the region (10°–20° N; 100°–115° E)	−0.41
AI-CPP7	5th power of averaged March 500 hPa temperature anomaly for the region (Equatorial –20°N; 0°–70° E)	0.37
AI-CPP8	5th power of averaged may ST anomaly for the region (Equatorial –30° N; 35°–90° E)	−0.39
(b) Negative ESI tendency (32 years)		
AI-CPN1	Averaged May sea level pressure anomaly for the region (Equatorial –10° N; 55°–75° E)	−0.52
AI-CPN2	Square of averaged May 850 hPa geopotential height anomaly for the region (50°–90° N; 0°–70° E)	−0.48
AI-CPN3	Square of averaged April 850 hPa geopotential height anomaly for the region (50°–75° N; 0°–40° E)	−0.38
AI-CPN4	Averaged Spring 500 hPa temperature anomaly for the region (65°–70° N; 0°–25° E)	−0.38
AI-CPN5	Square of averaged Winter 200 hPa temperature anomaly for the region (20°–35° N; 15°–5° W)	0.44
AI-CPN6	Square of averaged Winter ST anomaly for the region (5°–25° N; 170°–105° W)	0.43
AI-CPN7	Cube of averaged January 200 hPa geopotential height anomaly for the region (50°–60° N; 60°–15° W)	−0.42
AI-CPN8	Averaged April 500 hPa temperature anomaly for the region (Equatorial –20° N; 15°–35° E)	0.35

CC, correlation coefficient; ESI, effective strength index; LRF, long range forecasting.

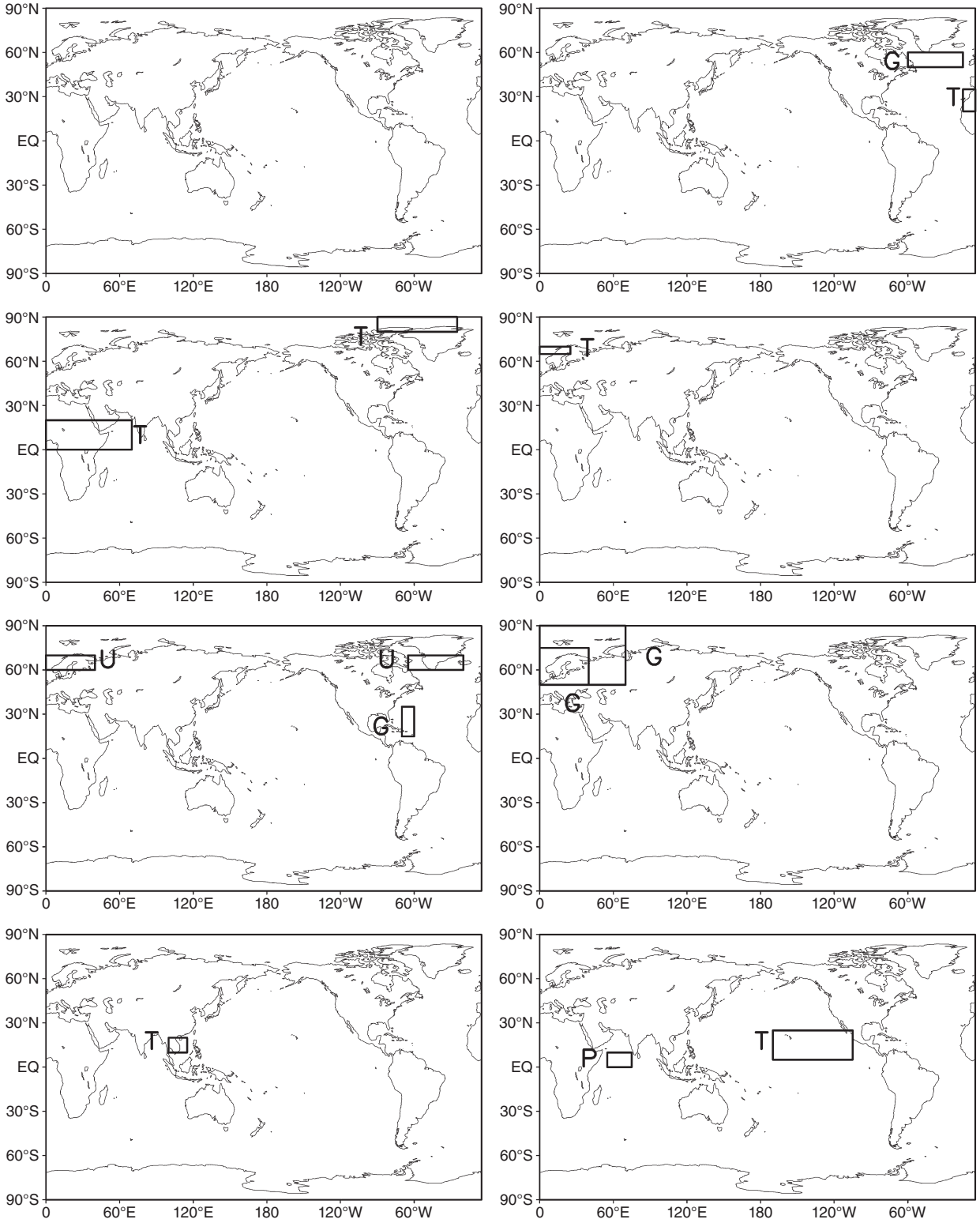


Figure 5. Locations of cluster parameters for temperature (T), pressure (P), geopotential height (G) and zonal wind (U) fields at 200 hPa (top panel) and sequentially followed at 500 hPa, 850 hPa and surface respectively; selected by cross validation procedure for predicting Indian summer monsoon rainfall during positive (left side) and negative (right side) phases of ESI-tendency.

4.4.1. Non-linear relationship between cluster parameters and ISMR during positive ESI tendency

During positive ESI tendency, eight independent cluster parameters are selected by cross-validation procedure, which shows the least error in predicting the ISMR on the independent data set (RMSE = 5.9). All of these cluster parameters show non-linear relationship with the ISMR (the power of the averaged meteorological parameter over the corresponding cluster region is greater than one). It is not so easy to give a simple explanation for the non-linear impact of meteorological parameters on

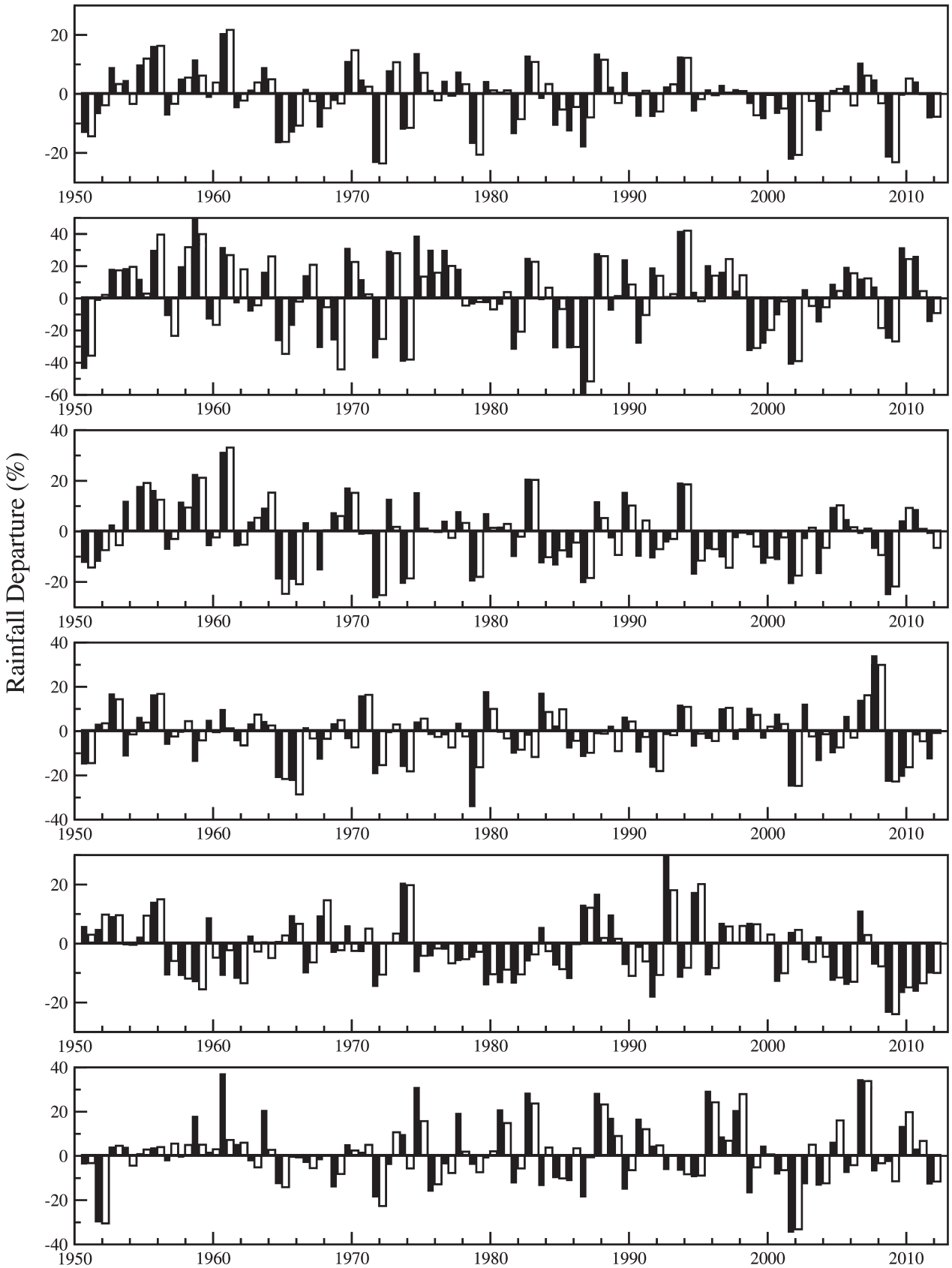


Figure 6. Estimated (white column) and actual (black column) summer monsoon rainfall departures in % over All India (top panel) and subsequently followed over North west India, West central India, Central north east India, North east India and Peninsular India respectively; using separate equations depending upon positive or negative phase of ESI-tendency for 1951–2012.

Table 3. Independent cluster parameters selected by cross validation for LRF of North West India (NWI) monsoon rainfall during (a) positive and (b) negative phase of ESI-tendency.

Abbreviation	Description	CC
(a) Positive ESI-tendency (30 years)		
NWI-CPP1	Square of averaged March surface temperature (ST) anomaly for the region (10°–25° N; 25°–100° E)	−0.65
NWI-CPP2	5th power of averaged May 500 hPa geopotential height anomaly for the region (80°–90° N; 140°E–145° W)	−0.51
NWI-CPP3	Averaged Spring 500 hPa temperature anomaly for the region (Equatorial –25° N; 160°–115° W)	−0.53
NWI-CPP4	4th power of averaged Winter 500 hPa geopotential height anomaly for the region (60°–80° N; 100°–170° E)	0.36
NWI-CPP5	Square of averaged March 500 hPa temperature anomaly for the region (Equatorial –20°N; 0°–70° E)	−0.40
NWI-CPP6	4th power of averaged Spring 850 hPa U-wind anomaly for the region (60°–70° N; 0°–40° E)	−0.39
(b) Negative ESI-tendency (30 years)		
NWI-CPN1	Averaged May sea level pressure (SLP) anomaly for the region (Equatorial –10°N; 55°–75° E)	−0.60
NWI-CPN2	Averaged April 500 hPa temperature anomaly for the region (Equatorial –20°N; 15°–35° E)	0.55
NWI-CPN3	Square of averaged Spring SLP anomaly for the region (10°N–10° S; 60°–75° E)	0.40
NWI-CPN4	Averaged Spring 200 hPa U-wind anomaly for the region (85°–90° N; 30°–5° W)	−0.49
NWI-CPN5	4th power of averaged February 500 hPa geopotential height anomaly for the region (80°–90° N; 50°E–150° W)	−0.38
NWI-CPN6	Square of averaged Winter ST anomaly for the region (20°–30° N; 40°–5° W)	0.43

CC, correlation coefficient; ESI, effective strength index; LRF, long range forecasting.

Table 4. Independent cluster parameters selected by cross validation for LRF of West Central India (WCI) monsoon rainfall during (a) positive and (b) negative phase of ESI tendency.

Abbreviation	Description	CC
(a) Positive ESI-tendency (30 years)		
WCI-CPP1	Square of averaged March surface temperature (ST) anomaly for the region (10°–20° N; 110°–120° E)	−0.55
WCI-CPP2	Cube of averaged April 850 hPa geopotential height anomaly for the region (10°S–25° N; 40°–5° W)	−0.42
WCI-CPP3	4th power of averaged March ST anomaly for the region (10°–25° N; 25°–100° E)	−0.45
WCI-CPP4	5th power of averaged May 500 hPa temperature anomaly for the region (80°–90° N; 70°–25° W)	−0.40
WCI-CPP5	4th power of averaged Spring 850 hPa geopotential height anomaly for the region (15°–35° N; 90°–60° W)	−0.40
WCI-CPP6	Square of averaged Winter 200 hPa U-wind anomaly for the region (85°–90° N; 75°–120° E)	−0.47
WCI-CPP7	5th power of averaged May 200 hPa temperature anomaly for the region (5°–20° N; 115°–150° E)	0.39
WCI-CPP8	Square of averaged Winter 850 hPa U-wind anomaly for the region (60°–70° N; 65°–20° W)	−0.39
(b) Negative ESI tendency (32 years)		
WCI-CPN1	4th power of averaged April sea level pressure (SLP) anomaly for the region (5°−−25° N; 0°–20° E)	0.59
WCI-CPN2	Averaged February 500 hPa temperature anomaly for the region (Equatorial –30°N; 130° E–130° W)	0.49
WCI-CPN3	Square of averaged March SLP anomaly for the region (70°–80° N; 50°–85° E)	−0.40
WCI-CPN4	5th power of averaged Spring 850 hPa geopotential height anomaly for the region (40°–50° N; 20°–80° E)	−0.37
WCI-CPN5	4th power of averaged March 850 hPa geopotential height anomaly for the region (Equatorial –30°N; 0° –30°E)	0.40
WCI-CPN6	Averaged Spring SLP anomaly for the region (Equatorial –20° N; 50°–40° W)	0.41
WCI-CPN7	Averaged May SLP anomaly for the region (15°–25° N; 55°–100° E)	−0.52

CC, correlation coefficient; ESI, effective strength index; LRF, long range forecasting.

the ISMR. The understanding of the high CC between the ISMR and the specific power of meteorological parameter needs further investigation.

During positive ESI tendency,

- The AO anomaly is negative from May to October, which is linked to the weak easterlies over the Arctic (making positive AI-CPP1 and AI-CPP5) and the anomalous warming over the Arctic (positive AI-CPP4). Due to the more southward penetration of the Arctic cold air mass, the Eurasian snow-cover may increase causing less rainfall over India (Bamzai and Shukla 1999).
- The NAO phase increases from winter to spring and decreases from spring to autumn. This is associated with intensification of the Azores high from winter to spring (positive AI-CPP3) and reduction of Azores high from spring to autumn (negative AI-CPP2). The negative phase of summer NAO may reduce the tropospheric temperature (TT) anomaly over Eurasia, which may weaken the meridional gradient of TT over the Indian monsoon region causing reduced rainfall activity over India (Goswami *et al.*, 2006).
- Anomalous warming over the equatorial Pacific starts from spring to autumn (positive AI-CPP6) and reduces the convection over India (negative AI-CPP7 and positive AI-CPP8). Weak Walker circulation reduces the convection over India and hence reduces the rainfall activity.

4.4.2. Linear and non-linear relationship between cluster parameters and ISMR during negative ESI tendency

During negative ESI tendency, eight independent cluster parameters are selected by cross-validation procedure, which show the least error in predicting the ISMR on the independent data set (RMSE = 5.83).

During negative ESI tendency, strong winds around the North Pole confine the cold Arctic air mass to the Arctic region causing surface cooling over this region, which is associated with above-normal rainfall activity over India (Prabhu *et al.*, 2012). Moreover, SO also restores positive phase before the beginning of monsoon, which is indicative of increased Indian rainfall activity.

During negative ESI tendency, the following patterns are observed.

- Strong convection over India and its surrounding is linked with reduced SLP (negative AI-CPN1). It will strengthen

Table 5. Independent cluster parameters selected by cross validation for LRF of Central North East India (CNEI) monsoon rainfall during (a) positive and (b) negative phase of ESI tendency.

Abbreviation	Description	CC
(a) Positive ESI tendency (30 years)		
CNEI-CPP1	4th power of averaged March surface temperature (ST) anomaly for the region (10°–20° N; 110°–120° E)	−0.61
CNEI-CPP2	Square of averaged May 500 hPa temperature anomaly for the region (Equatorial –15° N; 30°–60° E)	0.38
CNEI-CPP3	5th power of averaged Spring 200 hPa U-wind anomaly for the region (75°–80° N; 180°E–130° W)	0.45
CNEI-CPP4	Square of averaged January 200 hPa temperature anomaly for the region (30°–40° N; 35°–95° E)	−0.39
CNEI-CPP5	Square of averaged Spring 850 hPa U-wind anomaly for the region (60°–70° N; 0°–40° E)	−0.51
CNEI-CPP6	Square of averaged Winter 200 hPa U-wind anomaly for the region (85°–90° N; 75°–120° E)	−0.46
CNEI-CPP7	4th power of averaged March ST anomaly for the region (10°–25° N; 25°–100° E)	−0.38
CNEI-CPP8	5th power of averaged May 500 hPa temperature anomaly for the region (80°–90° N; 70°–25° W)	−0.37
(b) Negative ESI tendency (32 years)		
CNEI-CPN1	Square of averaged Winter 850 hPa geopotential height anomaly for the region (45°–55° S; 20°–5° W)	−0.62
CNEI-CPN2	Square of averaged January 500 hPa temperature anomaly for the region (50°–60° N; 95°–70° W)	0.50
CNEI-CPN3	4th power of averaged May ST anomaly for the region (30°–50° N; 90°–115° E)	0.37
CNEI-CPN4	Averaged Winter 200 hPa U-wind anomaly for the region (5°S–5° N; 35°–15° W)	−0.49
CNEI-CPN5	5th power of averaged April ST anomaly for the region (70°–80° N; 25°–75° E)	0.51
CNEI-CPN6	Square of averaged Winter sea level pressure anomaly for the region (Equatorial –15° N; 55°–75° E)	0.39

CC, correlation coefficient; ESI, effective strength index; LRF, long range forecasting.

Table 6. Independent cluster parameters selected by cross validation for LRF of North East India (NEI) monsoon rainfall during (a) positive and (b) negative phase of ESI tendency.

Abbreviation	Description	CC
(a) Positive ESI tendency (30 years)		
NEI-CPP1	Square of averaged February 850 hPa geopotential height anomaly for the region (Equatorial –30° N; 45°–5° W)	0.73
NEI-CPP2	4th power of averaged April 850 hPa geopotential height anomaly for the region (Equatorial –30° N; 120°–175° E)	0.61
NEI-CPP3	4th power of averaged January surface temperature (ST) anomaly for the region (10°–20° N; 75°–90° E)	0.43
NEI-CPP4	Cube of averaged January ST anomaly for the region (10°–25° N; 55°–70° E)	−0.51
NEI-CPP5	Cube of averaged Spring 500 hPa u-wind anomaly for the region (60°–70° N; 15°–5° W)	−0.43
(b) Negative ESI tendency (32 years)		
NEI-CPN1	Square of averaged Spring 500 hPa geopotential height anomaly for the region (60°–70° N; 50°–130° E)	0.54
NEI-CPN2	Square of averaged Spring sea level pressure anomaly for the region (20°–40° N; 110°–130° E)	0.49
NEI-CPN3	4th power of averaged Winter 500 hPa temperature anomaly for the region (70°–80° N; 60°–90° E)	0.47
NEI-CPN4	5th power of averaged Spring 500 hPa temperature anomaly for the region (25°–40° N; 35°–95° E)	−0.39
NEI-CPN5	Averaged Winter 200 hPa temperature anomaly for the region (55°–75° N; 0°–115° E)	−0.38
NEI-CPN6	Square of averaged March 850 hPa geopotential height anomaly for the region (40°–55° S; 130°–150° E)	−0.38

CC, correlation coefficient; ESI, effective strength index; LRF, long range forecasting.

Table 7. Independent cluster parameters selected by cross validation for LRF of Peninsular India (PI) monsoon rainfall during (a) positive and (b) negative phase of ESI tendency.

Abbreviation	Description	CC
(a) Positive ESI tendency (30 years)		
PI-CPP1	Square of averaged Spring 200 hPa U-wind anomaly for the region (15°–25° N; 100°–120° E)	0.47
PI-CPP2	4th power of averaged May surface temperature (ST) anomaly for the region (70°–80° N; 100°–120° E)	−0.46
PI-CPP3	5th power of averaged May 850 hPa geopotential height anomaly for the region (15°–30° N; 90°–65° W)	0.49
PI-CPP4	Square of averaged Spring 850 hPa U-wind anomaly for the region (70°–75° N; 75°–110° E)	0.40
PI-CPP5	Cube of averaged May 500 hPa temperature anomaly for the region (30°–40° N; 30°–10° W)	−0.51
PI-CPP6	4th power of averaged Spring ST anomaly for the region (10°N–5° S; 140°–175° E)	−0.38
(b) Negative ESI tendency (32 years)		
PI-CPN1	Square of averaged Winter 200 hPa temperature anomaly for the region (Equatorial –25°N; 10°–80° E)	0.52
PI-CPN2	4th power of averaged January sea level pressure anomaly for the region (80°–90° N; 105°–95° W)	−0.47
PI-CPN3	Averaged April 500 hPa temperature anomaly for the region (20°–35° N; 110°–150° E)	0.44
PI-CPN4	5th power of averaged Winter 500 hPa temperature anomaly for the region (5°–15° N; 55°–105° E)	0.42
PI-CPN5	Square of averaged Spring 850 hPa geopotential height anomaly for the region (40°–55° N; 160–180°E)	−0.36

CC, correlation coefficient; ESI, effective strength index; LRF, long range forecasting.

the cross-equatorial flow, which may increase the rainfall activity over India (Saha, 1974; Pisharoty, 1976; Cadet and Revardin, 1981; Shukla, 1984; Parthasarathy and Sontakke, 1988; Kakade and Dugam, 2008).

- The AO anomaly is positive in monsoon months. Strong winds are circulating counterclockwise around the Arctic at 55 ° N. This is linked with low pressure over the North Polar Region (negative AI-CPN2 and AI-CPN3) and cold air mass is confined to northern polar region (negative AI-CPN4).
- Restriction of southward penetration of the Arctic cold air masses and positive winter NAO suggests anomalous warming over Eurasia, which reduces the Eurasian snow-cover causing more rainfall over India (Bamzai and Shukla, 1999).
- Positive NAO phase in winter changes to negative phase in spring, which causes anomalous warming in winter (positive AI-CPN5) and intense Icelandic Low (negative AI-CPN7). The positive phase of summer NAO may rise the TT anomaly over Eurasia, which may strengthen the meridional gradient of TT over the Indian monsoon region causing enhanced rainfall activity over India (Goswami *et al.*, 2006).
- Warm winter SST anomalies over the Niño regions change to anomalous cold SSTs in spring and afterwards these cold anomalies persists till autumn. They cause anomalous warming over AI-CPN6 region in winter. The positive phase of the SO in summer indicates strengthening of the Walker circulation, which may increase the rainfall activity over India.

4.5. Summer monsoon rainfall prediction on homogeneous regions of India

During the contrasting phases of the ESI tendency, independent cluster parameters are obtained for predicting monsoon rainfall over different homogeneous regions of India. Figure 4 (remaining graphs) shows the least RMSE in predicting the summer monsoon rainfall over five homogeneous regions of India for each number of predictors during positive and negative ESI tendency. It suggests that for each homogeneous region of India, different numbers of cluster parameters are grouped for predicting the summer monsoon rainfall over it during contrasting phases of the ESI tendency. Tables 3–7 describe the independent cluster parameters obtained by cross-validation procedure, for predicting the summer monsoon rainfall over northwest India (NWI), west central India (WCI), central northeast India (CNEI), northeast India (NEI) and peninsular India (PI) during positive and negative ESI tendency. Table 8 shows multiple regression equations and their performances in predicting the summer monsoon rainfall over homogeneous regions of India during contrasting phases of the ESI tendency. The RMSE in predicting the summer monsoon rainfall over each homogeneous region is much less than the respective SD of rainfall, indicating better skill in prediction. The ACC during positive ESI tendency is either greater than or equal to the negative ESI tendency for all homogeneous regions. During positive (negative) ESI tendency, the predictive equations of rainfall over NEI (WCI) show the least RMSE. Figure 6 (remaining graphs) shows the actual and estimated rainfall departures (%) throughout India and its five homogeneous regions for 1951–2012. The frequencies of drought (flood) years for NWI, WCI, CNEI, NEI and PI are 15 (12), 11 (11), 10 (9), 11 (12) and 7 (13), respectively, and almost all extreme monsoon rainfall years of the five homogeneous regions of India are qualitatively well predicted. The CCs between the actual and estimated rainfall departures are 0.90, 0.91, 0.88, 0.87 and 0.82 for NWI, WCI, CNEI, NEI and PI, respectively.

Table 8. Multiple regression equations for predicting summer monsoon rainfall over homogeneous regions of India and their performance statistics in independent data set during contrasting phases of ESI tendency.

Regions	ESI tendency	Equations	RMSE (standard deviation)	ACC
NWI	Positive	$R = 6.91 - 10.23$ (NWI-CPP1) $- 0.62$ (NWI-CPP2) $- 16.69$ (NWI-CPP3) $+ 3.17$ (NWI-CPP4) $- 19.19$ (NWI-CPP5) $- 6.93$ (NWI-CPP6)	12.89 (25.63)	0.87
	Negative	$R = 2.3 - 6.66$ (NWI-CPN1) $+ 6.93$(NWI-CPN2) $+ 4.43$(NWI-CPN3) $- 6.54$ (NWI-CPN4) $- 1.26$ (NWI-CPN5) $+ 2.92$ (NWI-CPN6)	14.05 (21.48)	0.79
WCI	Positive	$R = 3.12 - 2.5$(WCI-CPP1) $- 1.86$ (WCI-CPP2) $- 2.27$WCI-CPP3) $- 0.28$ (WCI-CPP4) $- 0.46$(WCI-CPP5) $- 2.56$(WCI-CPP6) $+$ 0.67(WCI-CPP7) $- 1.74$(WCI-CPP8)	7.36 (12.41)	0.83
	Negative	$R = 0.3 + 0.84$ (WCI-CPN1) $+ 3.52$ (WCI-CPN2) $- 5.1$(WCI-CPN3) $- 0.39$ (WCI-CPN4) $+2.57$ (WCI-CPN5) $+ 4.82$ (WCI-CPN6) $- 1.84$ (WCI-CPN7)	7.24 (13.02)	0.84
CNEI	Positive	$R = -0.85 - 1.62$(CNEI-CPP1) $+ 4.92$ (CNEI-CPP2) $+ 1.55$ (CNEI-CPP3) $- 1.68$(CNEI-CPP4)-2.18 (CNEI-CPP5) $- 1.41$(CNEI-CPP6) $- 0.82$(CNEI-CPP7) $- 0.16$ (CNEI-CPP8)	6.11 (10.67)	0.85
	Negative	$R = -2.15 - 3.97$(CNEI-CPN1) $+ 1.92$ (CNEI-CPN2) $+ 2.18$ (CNEI-CPN3) $- 2.64$(CNEI-CPN4) $+ 0.68$(CNEI-CPN5) $+ 2.82$(CNEI-CPN6)	8.11 (13.31)	0.80
NEI	Positive	$R = -11.36 + 5.33$(NEI-CPP1) $+ 4.93$(NEI-CPP2) $+ 0.42$(NEI-CPP3) -0.89(NEI-CPP4) $- 0.57$(NEI-CPP5)	4.75 (9.61)	0.90
	Negative	$R = -1.24 + 1.41$(NEI-CPN1) $+ 4.01$ (NEI-CPN2) $+ 0.81$(NEI-CPN3) $- 1.7$(NEI-CPN4) $- 2.81$(NEI-CPN5) $- 2.7$(NEI-CPN6)	7.73 (10.90)	0.74
PI	Positive	$R = -7.98 + 5.02$(PI-CPP1)-2(PI-CPP2) $+0.36$(PI-CPP3) $+ 4.71$(PI-CPP4) $- 1.19$(PI-CPP5) $- 0.8$(PI-CPP6)	7.43 (12.51)	0.82
	Negative	$R = 5.19 + 4.63$(PI-CPN1) $- 1.58$(PI-CPN2) $+ 5.75$(PI-CPN3) $+ 0.4$ (PI-CPN4) $- 2.33$(PI-CPN5)	11.61 (15.92)	0.72

5. Conclusions

The following conclusions can be drawn from this study.

1. Seasonal rainfall variability is reduced when the rainfall years are classified as positive and negative ESI tendency years.
2. The seasonal mean monsoon rainfall throughout India and its homogeneous regions depends upon mutually exclusive sets of independent cluster parameters during contrasting phases of the ESI tendency.
3. The skill in predicting the monsoon rainfall over India will enhance if cluster parameters of various meteorological parameters at different levels are selected depending on the phase of the ESI tendency. Use of a cross-validation procedure in selecting independent cluster parameters improves the predictive skill. Separate multiple regression equations should be obtained for positive and negative ESI tendency. Unprecedented droughts in 2002 and 2009, where all models failed to forecast, are predicted well.
4. Separate equations depending on the phase of the ESI tendency and using independent cluster parameters can produce improved predictions of the summer monsoon rainfall over homogeneous regions of India. The strong association between the actual and estimated rainfall departure over 1951–2012 suggests its higher skill.

Acknowledgement

The authors wish to thank the Director, Indian Institute of Tropical Meteorology, for all the facilities provided. The authors also thank the anonymous reviewers for providing useful suggestions that helped to improve the manuscript.

Acronyms

ACCs	anomaly correlation co-efficients
AI	all India
AO	Arctic Oscillation
CC	correlation co-efficient
CNEI	Central North East India
CV	co-efficient of variation
EOF	Empirical Orthogonal Function
ESI	effective strength index
ISMR	Indian summer monsoon rainfall
LRF	Long range forecasting
NAO	North Atlantic Oscillation
NCEP/NCAR	National Centers for Environmental Prediction/National Center for Atmospheric Research
NEI	North East India
NWI	North West India
PI	Peninsular India
RMSE	root mean square error
SD	standard deviation
SLP	sea level pressure
SNN	shared nearest neighbor
SO	Southern Oscillation
SSTs	sea surface temperatures
STEMP	surface temperature
STs	surface temperatures
TT	tropospheric temperature
WCI	West Central India

References

Bamzai A, Shukla J. 1999. Relation between Eursian snow cover, snow depth, and the Indian summer monsoon: an observational study. *J. Clim.* **12**: 3117–3132.

Banerjee AK, Sen PN, Raman CRV. 1978. On foreshadowing southwest monsoon rainfall over India with mid-tropospheric circulation anomaly of April. *Indian J. Meteorol. Hydrol. Geophys.* **29**: 425–431.

Bhalme HN, Jadhav SK, Mooley DA, Ramana Murty BV. 1986. Forecasting of monsoon performance over India. *Int. J. Climatol.* **6**: 347–354.

Boriah S, Simon G, Naorem M, Steinbach M, Kumar V, Klooster S, et al. 2004. *Predicting Land Temperature Using Ocean Data*. KDD: Seattle, WA.

Cadet D, Revardin G. 1981. Water vapour transport over the Indian Ocean during summer 1975. *Tellus* **33**: 476–487.

Davis RE. 1976. Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.* **6**: 249–266.

Delsole T, Shukla J. 2002. Linear prediction of Indian monsoon rainfall. *J. Clim.* **15**: 3645–3658.

Ertoz L, Steinbach M, Kumar V. 2003. Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data. *Proceedings of Third SIAM International Conference on Data Mining*, 1–3 May 2003, San Francisco, CA, USA.

Goswami BN, Wu G, Yasunari T. 2006. Annual cycle, Intraseasonal oscillations and roadblock to seasonal predictability of the Asian summer monsoon. *J. Clim.* **19**: 5078–5099.

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

Gowariker V, Thapliyal V, Sarkar RP, Mandal GS, Sikka DR. 1989. Parametric and power regression models – new approach to long range forecasting. *Masaum* **40**: 115–122.

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

Jarvis RA, Patrick EA. 1973. Clustering using a similarity measure based on shared nearest neighbors. *IEEE Trans. Comput.* **22**(11): 1025–1034.

Joseph PV, Mukhopadhyaya RK, Dixit WV, Vaidya DV. 1981. Meridional wind index for long-range forecasting of Indian summer monsoon rainfall. *Mausam* **32**: 31–34.

Kakade SB, Dugam SS. 2008. Impact of cross-equatorial flow on intra-seasonal variability of Indian summer monsoon rainfall. *Geophys. Res. Lett.* **35** 1–5: L12805.

Kakade SB, Kulkarni A. 2012a. Relationship between ESI tendency and Indian monsoon rainfall: a possible mechanism. *Atmos. Sci. Lett.* **13**: 22–28, DOI: 10.1002/asl.357.

Kakade SB, Kulkarni A. 2012b. The changing relationship between surface temperatures and Indian monsoon rainfall with the phase of ESI tendency. *Adv. Meteorol.* **2012**: 1–8, DOI: 10.1155/2012/934624.

Kakade SB, Kulkarni A. 2013. Prediction of Indian summer monsoon rainfall using surface temperature and sea-level pressure cluster parameters. *Curr. Sci.* **105**: 964–970.

Krishna Kumar K, Rupa Kumar K, Pant GB. 1992. Pre-monsoon ridge location over India and its relation to monsoon rainfall. *J. Clim.* **5**: 979–986.

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

Krishna Kumar K, Soman MK, Rupa Kumar K. 1995. Seasonal forecasting of Indian summer monsoon: a review. *Weather* **12**: 449–467.

Kung EC, Sharif TA. 1982. Long-range forecasting of the Indian summer monsoon onset and rainfall with upper air conditions. *J. Meteorol. Soc. Jpn.* **60**: 672–681.

Mooley DA, Paolino DA. 1988. A predictive monsoon signal in the surface level thermal field over India. *Mon. Weather Rev.* **111**: 339–352.

Mooley DA, Parsarathy B, Pant GB. 1986. Relationship between Indian summer monsoon rainfall and location of the ridge at 500 hPa level along 75°E. *J. Clim. Appl. Meteorol.* **25**: 633–640.

Parthasarathy B, Diaz HZ, Eischeid JK. 1988. 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, DeshPande VR. 1991. Indian summer monsoon rainfall and 200mb meridional wind index: application for long range prediction. *Int. J. Climatol.* **11**: 165–176.

Parthasarathy B, Rupa Kumar K, Kothawale DR. 1992. Surface pressure and summer monsoon rainfall over India. *Adv. Atmos. Sci.* **9**: 359–366.

Parthasarathy B, Rupa Kumar K, Sontakke NA. 1990. Surface and upper air temperatures over India in relation to monsoon rainfall. *Theor. Appl. Climatol.* **42**: 93–110.

Parthasarathy B, Sontakke NA. 1988. El-Nino/SST of Puerto Chicama and Indian summer monsoon rainfall – statistical relationships. *Geofis. Int.* **27**: 37–59.

Pisharoty PR. 1976. Forecasting droughts in the sub-continent of India. *Proc. Indian Natl. Sci. Acad. A* **42**: 220–223.

Prabhu A, Mahajan PN, Khaladkar RM. 2012. Association of the Indian summer monsoon rainfall variability with the geophysical parameters over the Arctic region. *Int. J. Climatol.* **32**: 2042–2050.

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**(3): 422–431.

Rao KN. 1965. Seasonal forecasting-India. *Proceedings of Symposium on 'Research and Development Aspects of Long Range Forecasting'*, WMO-IUGG Tech. Note No. 66, WMO-No. 162-TP-79; 17–30.

Rao KN, Rama Moorthy KS. 1960. Seasonal (Monsoon) rainfall forecasting in India. *Proceedings of Symposium on 'Monsoon of the World'*, February 1958, held at New Delhi, Published by India Meteorological Department, New Delhi; 237–250.

Saha KR. 1974. Some aspects of the Arabian Sea summer monsoon. *Tellus* **26**: 464–476.

Sahai AK, Grimm AM, Satyan V, Pant GB. 2002. Prospects of prediction of Indian summer monsoon rainfall using global SST anomalies. Research Report-93, Indian Institute of Meteorology: Pune.

Shukla J. 1984. Predictability of time averages: Part II: the influence of the boundary forcing. In *Problems and Prospects in Long and Medium Range Weather Forecasting*, Burridge DM, Källén E (eds). Springer: Berlin; 155–206.

Shukla J, Mooley DA. 1987. Empirical prediction of the summer monsoon rainfall over India. *Mon. Weather Rev.* **115**: 695–703.

Shukla J, Paolino JA. 1983. The southern oscillation and the long-range forecasting of summer monsoon rainfall over India. *Mon. Weather Rev.* **111**: 1830–1853.

Verma RK, Kamte PP. 1980. Statistical technique for long-range forecasting of summer monsoon activity over India. *Proceedings of Symposium on The Probabilistic and Statistical Methods in Weather Forecasting*, 8–12 September 1980, Nice, WMO, Geneva; 303–307.

Verma RK, Subrmaniam K, Dugam SS. 1985. Interannual and long-term variability of the summer monsoon and its possible link with northern hemispheric surface air temperature. *Proc. Indian Acad. Sci.* **94**: 187–198.