

Seasonal prediction of summer monsoon rainfall over cluster regions of India

S B Kakade* and Ashwini Kulkarni

Indian Institute of Tropical Meteorology, Pune 411 008, India. *Corresponding author. e-mail: kakade@tropmet.res.in

Shared nearest neighbour (SNN) cluster algorithm has been applied to seasonal (June–September) rainfall departures over 30 sub-divisions of India to identify the contiguous homogeneous cluster regions over India. Five cluster regions are identified. Rainfall departure series for these cluster regions are prepared by area weighted average rainfall departures over respective sub-divisions in each cluster. The interannual and decadal variability in rainfall departures over five cluster regions is discussed. In order to consider the combined effect of North Atlantic Oscillation (NAO) and Southern Oscillation (SO), an index called effective strength index (ESI) has been defined. It has been observed that the circulation is drastically different in positive and negative phases of ESI-tendency from January to April. Hence, for each phase of ESI-tendency (positive and negative), separate prediction models have been developed for predicting summer monsoon rainfall over identified clusters. The performance of these models have been tested and found to be encouraging.

1. Introduction

India receives about 75–80% of its annual rainfall during the monsoon season (June–September). But the rainfall amount received during monsoon season depicts large variablity in time and space. This spatial and temporal variability of Indian summer monsoon rainfall (ISMR) increases the complexity in its prediction on smaller space—time domain. With the aim to achieve improved prediction skill of ISMR on smaller spatial scales, a large number of studies have been done to construct the contiguous regions over Indian landmass. Nicholson (1986) has shown that average over a homogeneous region reduces the data volume and takes care of small-scale variability. Also, the averaging helps to enhance the signal variation at larger spatial scales. A large number of techniques have been used to identify homogeneous rainfall regions over India, e.g., correlation analysis (Nicholson

1986; Kulkarni et al. 1992; Gadgil et al. 1993), principal component analysis – PCA (Kulkarni et al. 1992: Ivengar and Basak 1994), spectral analysis (Azad et al. 2010), cluster analysis (Matulla et al. 2003; Venkatesh and Jose 2007; Rao and Srinivas 2008; Malik et al. 2010), and PCA in association with cluster analysis [k-means cluster] (Kulkarni et al. 1992; Satyanarayana and Srinivas 2008) and fuzzy c-means cluster (Kulkarni and Kripalani 1998). Regionalization of summer monsoon rainfall over India has wide applications, in which agricultural planning is of high concern. Gadgil and Iyengar (1980) identified eight clusters by applying principal component analysis to the mean temporal profiles of rainfall at different stations of the Indian peninsula. Kulkarni et al. (1992) classified ISMR by objective methods like map-to-map correlation method and k-means clustering method. They suggested that the dominant patterns are similar, irrespective of the method used. They further

Keywords. SNN cluster; effective strength index; rainfall clusters over India.

showed that the monsoon rainfall for India as a whole and the spatial distribution of rainfall need not have any relationship. Iyengar and Basak (1994) identified four major regions over India with homogeneous variability of monsoon rainfall. They suggested that these four regional rainfall series are mutually uncorrelated. Azad et al. (2010) have identified 10 regions which have similar power spectral density. Parthasarathy et al. (1993) identified five homogeneous regions of India (depicted in figure 2.1) on the basis of similarity in rainfall characteristics and association of subdivisional monsoon rainfall with regional/global circulation parameters. Kulkarni and Kripalani (1998) obtained summer monsoon rainfall patterns over India with fuzzy c-means classification method. They have shown that this method is more flexible and better suited for classification of meteorological fields. Since 1999, India Meteorological Department (IMD) has been issuing long-range forecasts for northwest India, northeast India and the peninsula. These regions are identified by choosing regions in which the rainfall variation in each of the meteorological sub-divisions comprising the region is positively and significantly correlated with the area-weighted rainfall variation over the region as a whole (Rajeevan et al. 2004). Twenty-six homogeneous regions over India were identified using correlation analysis on daily rainfall over 1025 rain gauges over India (Saikranthi et al. 2012).

Since meteorological data is a high-dimensional data which may contain outliers and clusters of different sizes, shapes and densities, the traditional clustering methods may not be useful. Ertoz et al. (2003) suggested an alternative method as the similarity (closeness) between a pair of points in terms of their shared nearest neighbours (SNN). Boriah et al. (2004) obtained the clusters of ocean temperatures and used them to predict land temperatures. The SNN algorithm removes noise (low-density points) and connect non-noise or core points (highdensity points). This algorithm identifies the clusters of different shapes, sizes and densities. Kakade and Kulkarni (2016) described the SNN algorithm to obtain cluster regions of various meteorological fields. In this paper, an attempt has been made to identify the regions of India depending upon monsoon rainfall over sub-divisions of India. The rainfall variability, for each region, is discussed on interannual, decadal and climatological scales.

2. Data and methodology

2.1 Data

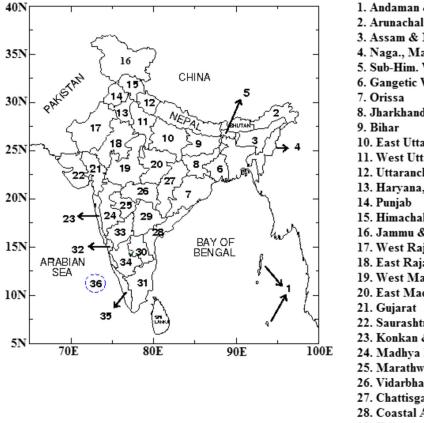
Indian summer monsoon rainfall: The sub-divisional monthly rainfall (figure 1) have been downloaded from the website of Indian Institute of Tropical Meteorology, Pune, www.tropmet.res.in. The seasonal (June–September) rainfall series is the total rainfall of the months June-September. The analysis involves percentage departures from long term (1951–2012) mean on each sub-division. Figure 1 depicts 36 sub-divisions over India as defined by IMD. The abbreviations for subdivisions are given in brackets. The sub-divisions in the hilly regions (Nos. 2, 12, 15 and 16), Andaman and Nicobar islands (No. 1) and the Lakshdweep islands (No. 36) are not considered in this study.

Effective strength index (ESI) tendency: Monthly North Atlantic Oscillation (NAO) and Southern Oscillation (SO) data for the period, 1951–2012 have been taken from www.cpc.ncep.noaa.gov. 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. These series are called as ESI series of respective month. ESI-tendency from winter to spring is defined as the difference, i.e., April minus January ESI values. Kakade and Kulkarni (2012) pointed out that the relationship between ISMR and ESI-tendency is statistically significant in negative phase and is insignificant in positive phase of ESI-tendency. Further, they have also shown that the evolution of surface temperature from winter to spring is exactly opposite during contrasting phases of ESI-tendency. During 1951–2012, there are 30 years (1951, 1954, 1955, 1957, 1960, 1962, 1963, 1965, 1967, 1969, 1972, 1974, 1976, 1977, 1980, 1981, 1982, 1985, 1987, 1990, 1991, 1992, 1994, 1996, 1997, 2001, 2002, 2004, 2006, 2009, 2011 and 2012) with positive ESItendency, while 32 years (1952, 1953, 1956, 1958, 1959, 1961, 1964, 1966, 1968, 1970, 1971, 1973, 1975, 1978, 1979, 1983, 1984, 1986, 1988, 1989, 1993, 1995, 1998, 1999, 2000, 2003, 2005, 2007, 2008 and 2010) with negative ESI-tendency.

Temperature, geo-potential height and zonal wind: The NCEP/NCAR reanalysis gridded $2.5^{\circ}\times$ 2.5° long./lat. global temperature (Kelvin), geopotential height (meter), zonal wind (ms⁻¹) data for all pre-monsoon months January-May), win-(December-January-February) and spring (March-April-May) at surface, 850, 500, 200 hPa for 1948–2012 have been taken from http:// www.esrl.noaa.gov/psd/data/gridded/data.ncep. reanalysis.html. The data have been interpolated on $5^{\circ} \times 5^{\circ}$ lat./long.

2.2 Shared nearest neighbour (SNN)

The meteorological data is high-dimensional data which contain outliers and clusters of different sizes, shapes and densities. Therefore, an alternative



- 1. Andaman & Nicobar Islands
- 2. Arunachal Pradesh
- 3. Assam & Meghalaya
- 4. Naga., Mani., Mizo. & Tripura
- 5. Sub-Him. W. Bengal & Sikkim
- 6. Gangetic West Bengal
- 8. Jharkhand
- 10. East Uttar Pradesh
- 11. West Uttar Pradesh
- 12. Uttaranchal
- 13. Haryana, Chandigarh & Delhi
- 15. Himachal Pradesh
- 16. Jammu & Kashmir
- 17. West Rajasthan
- 18. East Rajasthan
- 19. West Madhya Pradesh
- 20. East Madhya Pradesh
- 22. Saurashtra, Kutch & Diu
- 23. Konkan & Goa
- 24. Madhva Maharashtra
- 25. Marathwada
- 27. Chattisgarh
- 28. Coastal Andhra Pradesh
- 29. Telangana
- 30. Rayalaseema
- 31. Tamil Nadu & Pondicherry
- 32. Coastal Karnataka
- 33. North Interior Karnataka
- 34. South Interior Karnataka
- 35. Kerala
- 36. Lakshadweep

Figure 1. Map of 36 sub-divisions of India.

to the traditional density-based cluster methods was suggested as the closeness between a pair of points in terms of their SNN (Ertoz et al. 2003). New climate indices using meteorological parameters like sea surface temperature (SST), sea level pressures (SLP), precipitation are discovered by Steinbach et al. (2003). Boriah et al. (2004) have applied this SNN technique to ocean temperatures and used the clusters to predict land temperatures.

The steps involved in the SNN clustering algorithm are:

- 1. Preprocessing of data: Seasonality, trends and autocorrelation are removed from time series at each grid if it exists.
- 2. Compute the similarity matrix: The measure of similarity between two grid points is taken to be correlation coefficient (CC) between the time series over the two grid points for 1951–2012.

3. Sparsify the similarity matrix by keeping only its k strongest links.

Here k is called the neighbourhood list size. It adjusts the focus of the clusters. In case of its small (large) value, even a uniform cluster will be broken up into pieces and we may get many small, but tight clusters (we may get a few wellseparated clusters and small local variations in similarity, will not have an impact).

4. Construct the SNN list from the sparsified similarity matrix.

Jarvis-Patric algorithm is applied to find similarity threshold and the connected components. A neighbour list which contains grid points showing significant (at 1% level) CC, for each grid point, is prepared. Jarvis and Patrick (1973) suggested that points P and Q are connected if and only if both P and Q have each other in their closest k nearest neighbour lists,

where k is the nearest neighbour size (here k = 100). If i, j are two points, then strength of the link between i and j is calculated as:

$$Str(i,j) = \Sigma(k+1-m) \times (k+1-n),$$

where k is the nearest neighbour list size, and m and n are positions of SNNs in the lists of i and j.

- 5. Find total strength at each point: The points with high total link strength will become candidates for representative points, while those having very low total link strength become candidates for noise points.
- 6. Find the core points: These are the points with SNN density greater than the threshold value. This value is to be decided by trial and error (here, we have taken it as 100).
- 7. Form clusters from the core points by averaging the grids in the cluster: For detailed description of the method, please refer Kakade and Kulkarni (2016).

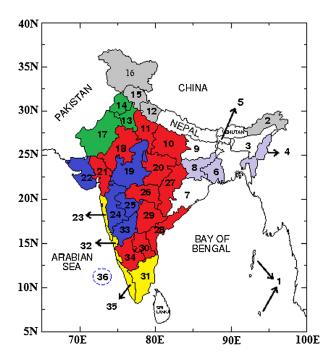
3. Results and discussion

SNN clustering algorithm is applied to summer monsoon rainfall departures over 30 subdivisions (figure 1) of India. Cluster regions are obtained by considering the points with high

(>75), medium (50-75) and low (<50) connectivity strengths. Figure 2 shows five rain-cluster (RC) regions of India, RC1 and RC2 consist of subdivisions having high and medium connectivity strength respectively. RC1 consists of four subdivisions, namely, West Madhya Pradesh, Madhya Maharashtra, Marathwada and North Interior Karnataka. RC2 consists of 11 sub-divisions of India, out of which two sub-divisions are from NWI and CNEI each, three are from PI and four are from WCI. RC3, RC4 and RC5 consist of subdivisions having low connectivity strength. West coast of India and Tamil Nadu delineate RC3, northwestern sub-divisions of NWI demarcate RC4 and three sub-divisions of NEI constitute RC5. The sub-divisions 3, 7 and 9 are not accommodated in any cluster regions (high, medium and low connectivities), which indicates lack of similarity in rainfall characteristics over these subdivisions with the remaining subdivisions and hence, a very low connectivity with any of the subdivisions. This may be due to the limitations of the SNN algorithm.

3.1 Summer monsoon rainfall variability over RC regions

The seasonal monsoon rainfall time series for each RC region is prepared by area weighted average



Region	Sub-division name (abbreviation)
RC1	WMPRA, MADMH, MARAT, NIKNT
RC2	EUPRA, WUPPL, ERJST, EMPRA, GUJRT, VDABH, CHHAT, COAPR, TELNG, RLSMA, SIKNT
RC3	KNGOA, COKNT, KERLA, TLNAD
RC4	HARYA, PUNJB, WRJST
RC5	NMAMT, GNWBL, JHKND

Figure 2. Five rain-cluster (RC) regions over India using SNN.

of summer monsoon rainfall over respective sub-divisions. Figure 3 shows the interannual variability of summer monsoon rainfall (mm) of five RC regions for 1951–2012. Mean, standard deviation (SD) and coefficient of variation (CV) for all five RC regions are also shown. The long period average

(LPA), for the period 1951–2012, is maximum for RC5 (LPA = 1167.71 mm) and minimum for RC4 (LPA = 342.84 mm). The cluster region RC5 consists of NMAMT, GNWBL and JHKND (northeastern part of India), which receives rainfall in monsoon season due to north or northwestward

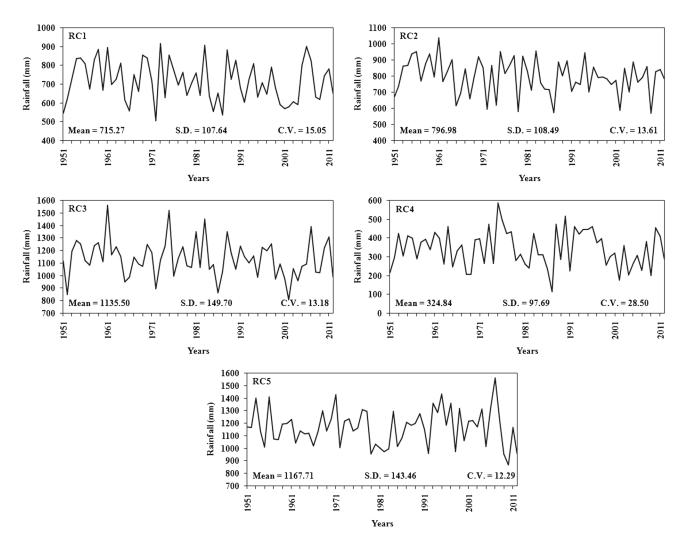


Figure 3. Interannual variability of summer monsoon rainfall for five rain clusters over India for the period 1951–2012.

Table 1. Extreme monsoon rainfall years.

	Excess monsoon years	Deficient monsoon years		
All India	1956, 1959, 1961, 1970, 1975, 1983, 1988,	1951, 1965, 1966, 1968, 1972, 1974, 1979,		
	1994, 2007	1982, 1985, 1986, 1987, 2002, 2004, 2009		
RC1	1954, 1955, 1958, 1959, 1961, 1969, 1970,	1951, 1966, 1972, 1985, 1987, 1992, 2000,		
	1973, 1975, 1983, 1988, 1990, 2006, 2007	2001, 2002, 2003, 2004		
RC2	1955, 1956, 1959, 1961, 1970, 1975, 1978,	1951, 1965, 1968, 1972, 1974, 1979, 1987,		
	1980, 1983, 1994	2002, 2009		
RC3	1961,1975,1981,1983,1988,2007,2011	1952, 1965, 1966, 1972, 1986, 1999, 2001,		
		2002, 2004		
RC4	1964, 1973, 1975, 1976, 1988, 1990, 1992,	1951, 1968, 1969, 1982, 1986, 1987, 1991,		
	1994, 1995, 1996, 2010	2002, 2004, 2007, 2009		
RC5	1953, 1956, 1971, 1993, 1995, 1997, 1999,	1955, 1966, 1972, 1979, 1981, 1982, 1983,		
	2004, 2006, 2007	1985, 1992, 1998, 2005, 2009, 2010, 2012		

movement of the low pressure systems occurred in head Bay of Bengal. The cluster region RC4 is northwestern part of India (HARYA, PUNJAB and WRJST) and WRJST is receiving less rainfall in monsoon season as compared with rainfall over other sub-divisions. The rainfall variation is highest for RC4 (CV = 28.5%) and lowest for RC5 (CV = 12.29%), which is in accordance with the high rainfall variability over northwest India and low rainfall variability over northeast India. Thus the cluster with high average rainfall (RC3 and RC5) has large variability also. The cluster RC5 has minimum mean rainfall and minimum variability. The cluster rainfall series for RC1–RC4 shows decreasing trend, but it is not significant at 5\% level (trend significance is tested by Mann Kendall trend test). The correlation coefficients between ISMR and regional cluster rainfall time series are 0.82, 0.94, 0.65, 0.71 and 0.19 for RC1, RC2, RC3, RC4 and RC5, respectively. It suggests that contribution to ISMR is maximum from RC1 to RC4 regions and is minimum from RC5 region; which may be because RC5 consists of subdivisions in northeast India and rainfall over this region is not necessarily associated with northward or northwestward movement of the systems originating over Bay of Bengal.

For each RC region, the excess (deficient) monsoon rainfalls are defined as the year's summer monsoon rainfall greater (less) than mean + standard deviation (mean - standard deviation) for that RC region. The extreme monsoon years for all India and its five RC regions are listed in table 1. It can be seen that extreme monsoon rainfall condition over India as a whole need not be extreme over cluster regions of India. Recently, in 2002, 2004 and 2009, India as a whole experienced severe drought conditions, but due to high spatial variability of Indian monsoon rainfall, all the five cluster regions do not have deficit conditions in all the three years. In 2002, the region RC5, in 2004 the regions RC2 and RC5 and in 2009, the regions

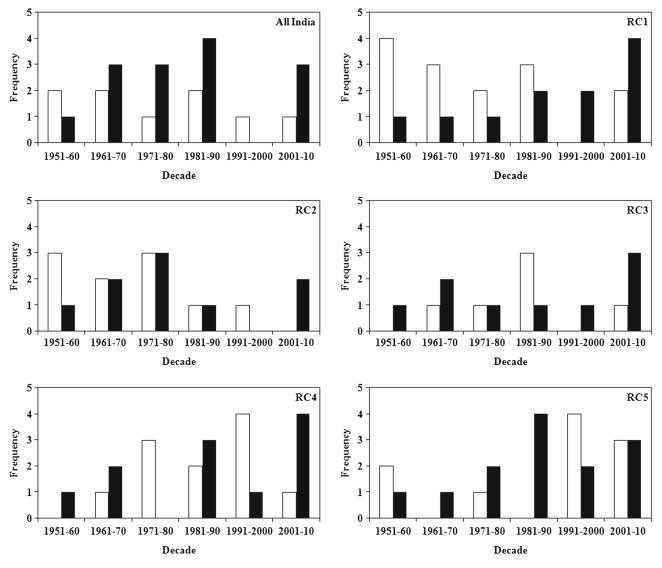


Figure 4. Decadal frequency of excess (white bars) and deficit (black bars) monsoons in six decades of 1951–2010.

RC1 and RC3 do not experience deficit monsoons. Only RC4 had all the three droughts. The distribution of excess/deficit monsoons in six decades 1951–2010 is shown in figure 4. The white (black) columns represent the decadal frequency of excess (deficient) monsoons. It is observed that, out of six decades, four decades (1961–1970, 1971–1980, 1981–1990 and 2001–2010) show large frequency of deficit monsoon years than frequency of excess monsoon years for all India. Similar frequency distribution of extreme monsoon years is observed for RC3, RC4 and RC5 with changes in decades. On decadal scale, the frequency of excess monsoon rainfall decreases and frequency of deficient monsoon rainfall years increases for RC1 region (these trends are also not significant at 5% level). RC1, RC3 and RC4 experienced maximum number of deficit monsoons in the recent decade 2001–2010. RC4 and RC5 were in better position in the decade of 1991–2000, having maximum number of excess monsoons. In this decade, RC1 and RC3 did not have any excess monsoon while RC2 did not have any deficit monsoon. RC5 did not have any excess monsoon in the decades 1961–1970 and 1981–1990. Thus extreme monsoon rainfall conditions over India depict a large spatial and temporal variability.

3.2 LRF of summer monsoon rainfall over RC regions

NAO and SO are the two large-scale atmospheric oscillations in northern and southern hemispheres, respectively. NAO and SO exist throughout the year simultaneously and are known to modulate

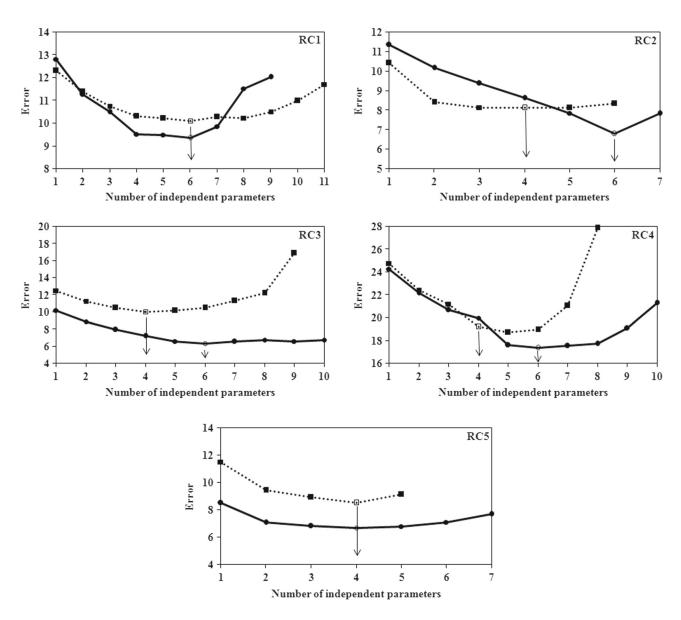


Figure 5. Least root mean square error (RMSE) against each number of predictors during positive (straight line) and negative (dotted line) phases of ESI-tendency for five RC regions of India.

 ${\it Table 2. Independent \ cluster \ parameters \ selected \ by \ cross \ validation \ for \ LRF \ of \ monsoon \ rainfall}$ over RC1 during positive and negative phases of ESI-tendency.

Abbreviation	Description	CC
Positive phase		
RC1 -CPP1	Square of averaged March surface temperature anomaly for the region ($10^{\circ}-20^{\circ}N$; $110^{\circ}-120^{\circ}E$).	-0.52
RC1–CPP2	Averaged winter sea level pressure anomaly for the region (Eq25°N; 5° -40°W).	-0.48
RC1-CPP3	Square of averaged winter 500-hPa temperature anomaly for the region ($5^{\circ}-15^{\circ}N$; $105^{\circ}-130^{\circ}W$).	-0.38
RC1-CPP4	Fourth power of averaged March surface temperature anomaly for the region $(10^{\circ}-25^{\circ}N; 25^{\circ}-100^{\circ}E)$.	-0.48
RC1-CPP5	Fifth power of averaged May 500-hPa geo-potential height anomaly for the region (75°-90°N; 5°-145°W).	-0.47
RC1-CPP6	Fifth power of averaged Spring 500-hPa temperature anomaly for the region $(5^{\circ}-20^{\circ}N; 50^{\circ}-70^{\circ}E)$.	0.37
Negative phase		
RC1-CPN1	Square of averaged winter 500-hPa temperature anomaly for the region (10°-20°N; 15°-55°E).	0.65
RC1-CPN2	Square of averaged winter 200-hPa temperature anomaly for the region $(20^{\circ}-35^{\circ}N; 5^{\circ}-15^{\circ}W)$.	0.43
RC1-CPN3	Square of averaged winter surface temperature anomaly for the region $(20^{\circ}-30^{\circ}\text{N}; 5^{\circ}-40^{\circ}\text{W})$.	0.47
RC1-CPN4	Cube of averaged spring 500-hPa temperature anomaly for the region ($65^{\circ}-75^{\circ}$ N; $0^{\circ}-25^{\circ}$ E).	-0.41
RC1-CPN5	Cube of averaged March 850-hPa geo-potential height anomaly for the region $(75^{\circ}-90^{\circ}\text{N}; 0^{\circ}-140^{\circ}\text{E})$.	-0.37
RC1-CPN6	Averaged May sea level pressure anomaly for the region $(15^{\circ}-30^{\circ}N; 20^{\circ}-40^{\circ}E)$.	-0.38

 ${\it Table 3. Independent \ cluster \ parameters \ selected \ by \ cross \ validation \ for \ LRF \ of \ monsoon \ rainfall}$ over RC2 during positive and negative phases of ESI-tendency.

Abbreviation	Description		
Positive phase			
RC2 -CPP1	Square of averaged March surface temperature anomaly for the region ($10^{\circ}-25^{\circ}N$; $25^{\circ}-100^{\circ}E$).	-0.59	
RC2-CPP2	Fifth power of averaged April surface temperature anomaly for the region $(45^{\circ}-55^{\circ}\text{N}; 135^{\circ}-155^{\circ}\text{W})$.	-0.36	
RC2-CPP3	Fourth power of averaged spring 850-hPa U-wind anomaly for the region $(60^{\circ}-70^{\circ}\text{N}; 0^{\circ}-40^{\circ}\text{E})$.	-0.49	
RC2-CPP4	Square of averaged February 500-hPa temperature anomaly for the region (Eq. -15° N; $5^{\circ}-20^{\circ}$ W).	-0.50	
RC2-CPP5	Fifth power of averaged May 850-hPa geo-potential height anomaly for the region (Eq25°N; 120°-165°E).	-0.42	
RC2-CPP6	Square of averaged April surface temperature anomaly for the region (10°-20°N; 100°-115°E).	-0.49	
Negative phase			
RC2-CPN1	Square of averaged April sea level pressure anomaly for the region $(5^{\circ}-25^{\circ}N; 0^{\circ}-20^{\circ}E)$.	0.59	
RC2-CPN2	Square of averaged May 850-hPa geo-potential height anomaly for the region $(50^{\circ}-90^{\circ}N; 0^{\circ}-70^{\circ}E)$.	-0.58	
RC2-CPN3	Averaged May 200-hPa temperature anomaly for the region $(25^{\circ}-55^{\circ}N; 5^{\circ}-35^{\circ}W)$.	-0.41	
RC2-CPN4	Square of averaged Winter surface temperature anomaly for the region (20°–30°N; 5°–40°W).	0.40	

Table 4. Independent cluster parameters selected by cross validation for LRF of monsoon rainfall over RC3 during positive and negative phases of ESI-tendency.

Abbreviation	Description	CC
Positive phase		
RC3-CPP1	Fifth power of averaged April surface temperature anomaly for the region $(70^{\circ}-80^{\circ}N; 25^{\circ}-75^{\circ}E)$.	-0.54
RC3-CPP2	Averaged winter sea level pressure anomaly for the region $(20^{\circ}-35^{\circ}N; 90^{\circ}-120^{\circ}W)$.	0.52
RC3-CPP3	Square of averaged spring 200-hPa U-wind anomaly for the region $(15^{\circ}-25^{\circ}N; 100^{\circ}-120^{\circ}E)$.	0.49
RC3-CPP4	Cube of averaged March surface temperature anomaly for the region $(20^{\circ}-35^{\circ}N; 70^{\circ}-90^{\circ}W)$.	-0.42
RC3-CPP5	Square of averaged winter surface temperature anomaly for the region (Eq10°; 135°-180°E).	-0.39
RC3-CPP6	Square of averaged winter 850-hPa geo-potential height anomaly for the region $(20^{\circ}-55^{\circ}N; 40^{\circ}-85^{\circ}W)$.	-0.54
Negative phase		
RC3-CPN1	Square of averaged winter 200-hPa temperature anomaly for the region $(20^{\circ}-35^{\circ}N; 5^{\circ}-15^{\circ}W)$.	0.58
RC3-CPN2	Fourth power of averaged February 500-hPa geo-potential height anomaly for the region (80°-90°N; 50°E-150°W).	-0.52
RC3-CPN3	Averaged April 500-hPa temperature anomaly for the region $(20^{\circ}-35^{\circ}N; 110^{\circ}-150^{\circ}E)$.	0.39
RC3-CPN4	Square of averaged spring 850-hPa geo-potential height anomaly for the region (45°–90° N; 65°–145° W).	-0.42

Table 5. Independent cluster parameters selected by cross validation for LRF of monsoon rainfall over RC4 during positive and negative phases of ESI-tendency.

Abbreviation	Description		
Positive phase			
RC4-CPP1	Averaged spring 500-hPa temperature anomaly for the region (Eq. -25° N; $115^{\circ}-160^{\circ}$ W).	-0.64	
RC4-CPP2	Square of averaged March 500-hPa temperature anomaly for the region (Eq15°N; 5°-15°W).	-0.41	
RC4-CPP3	Cube of averaged May surface temperature anomaly for the region $(65^{\circ}-80^{\circ}N; 5^{\circ}-80^{\circ}E)$.	-0.43	
RC4-CPP4	Square of averaged March surface temperature anomaly for the region $(10^{\circ}-25^{\circ}N; 25^{\circ}-100^{\circ}E)$.	-0.48	
RC4-CPP5	Averaged January 200-hPa geo-potential height anomaly for the region $(20^{\circ}-30^{\circ}N; 120^{\circ}-165^{\circ}W)$.	-0.40	
RC4-CPP6	Fourth power of averaged spring 500-hPa temperature anomaly for the region $(70^{\circ}-90^{\circ}N; 5^{\circ}-55^{\circ}W)$.	-0.36	
Negative phase			
RC4-CPN1	Fourth power of averaged May sea level pressure anomaly for the region $(80^{\circ}-90^{\circ}\text{N}; 105^{\circ}\text{E}-120^{\circ}\text{W})$.	-0.47	
RC4-CPN2	Fourth power of averaged January 200-hPa temperature anomaly for the region (15°-25°N; 40°-65°E).	0.47	
RC4-CPN3	Square of averaged spring sea level pressure anomaly for the region (10°N–10°S; 60°–75°E).	0.36	
RC4-CPN4	Fifth power of averaged spring 200-hPa temperature anomaly for the region $(5^{\circ}-25^{\circ}N; 125^{\circ}-150^{\circ}E)$.	0.42	
RC4-CPN5	Averaged spring 500-hPa temperature anomaly for the region (45°-60°N; 155°-175°E).	-0.39	

the inert-annual variability of the Indian summer monsoon rainfall (Sikka 1980; Pant and Parthasarathy 1981; Shukla and Paolino 1983; Rasmusson and Carpenter 1983; Verma 1990; Dugam et al. 1997; Kripalani and Kulkarni 1997; Goswami et al. 2006; etc.). The interactive mechanism between these two oscillations has been discussed by VanLoon and Rogers (1978), Chen

Table 6. Independent cluster parameters selected by cross validation for LRF of monsoon rainfall over RC5 during positive and negative phases of ESI-tendency.

Abbreviation	Description		
Positive phase			
RC5-CPP1	Cube of averaged Spring 500-hPa U-wind anomaly for the region $(85^{\circ}-90^{\circ}N; 75^{\circ}-145^{\circ}W)$.	-0.52	
RC5-CPP2	Fifth power of averaged January 850-hPa geo-potential height anomaly for the region $(55^{\circ}-75^{\circ}N; 105^{\circ}-165^{\circ}E)$.	0.49	
RC5-CPP3	Square of averaged January 200-hPa temperature anomaly for the region (30°-40°N; 35°-95°E).	-0.41	
RC5-CPP4	Averaged February 500-hPa geo-potential height anomaly for the region $(30^{\circ}-55^{\circ}N; 65^{\circ}-95^{\circ}W)$.	0.36	
Negative phase	- ,		
RC5-CPN1	Averaged Spring surface temperature anomaly for the region (Eq20°N; 100°-120°E).	-0.61	
RC5-CPN2	Fifth power of averaged Spring 850-hPa U-wind anomaly for the region (55°-80°N; 50°-75°E).	-0.47	
RC5-CPN3	Fourth power of averaged January surface temperature anomaly for the region (50°-60°N; 60°-80°E).	-0.47	
RC5-CPN4	Square of averaged Winter 500-hPa geo-potential height anomaly for the region $(30^{\circ}-60^{\circ} \text{ N}; 30^{\circ}-100^{\circ} \text{W})$.	0.36	

Table 7. Multiple regression equations for predicting summer monsoon rainfall over rain-culster regions of India and their performance statistics in independent dataset during contrasting phases of ESI-tendency.

Regions	ESI-tendency	Equations	RMSE (SD)	ACC
RC1	Positive	R = 3.43 - 3.93(RC1-CPP1) - 6.03(RC1-CPP2) - 4.51(RC1-CPP3) - 2.83 (RC1-CPP4) - 0.67(RC1-CPP5) + 0.74 (RC1-CPP6)	9.37 (14.22)	0.77
	Negative	R = -8.40 + 8.01(RC1-CPN1) + 4.06(RC1-CPN2) + 2.74(RC1-CPN3) - 2.12(RC1-CPN4) - 0.93 (RC1-CPN5) - 2.49 (RC1-CPN6)	10.12 (15.28)	0.76
RC2	Positive	R = 7.02 - 5.27(RC2-CPP1) - 0.28 (RC2-CPP2) - 1.15 (RC2-CPP3) - 5.73 (RC2-CPP4) - 0.97 (RC2-CPP5) - 6.94 (RC2-CPP6)	6.81 (13.46)	0.89
	Negative	R = 1.09 + 6.21 (RC2-CPN1) - 4.86 (RC2-CPN2) - 4.29 (RC2-CPN3) + 1.36 (RC2-CPN4)	8.14 (12.38)	0.76
RC3	Positive	R = 0.5 - 0.78 (RC3-CPP1) + 3.05 (RC3-CPP2) + 4.17 (RC3-CPP3) - 1.94 (RC3-CPP4) - 4.13 (RC3-CPP5) - 2.66 (RC3-CPP6)	6.35 (11.22)	0.84
	Negative	R = 3.55 + 6.71 (RC3-CPN1) - 0.93 (RC3-CPN2) + 6.11 (RC3-CPN3) - 5.28 (RC3 - CPN4)	10.04 (14.63)	0.74
RC4	Positive	R = 8.89 - 17.96 (RC4-CPP1) - 11.29 (RC4-CPP2) - 1.6 (RC4-CPP3) - 15.17 (RC4-CPP4) - 12.33 (RC4-CPP5) - 3.86(RC4 - CPP6)	17.37 (30.08)	0.83
	Negative	R = 6.2 - 1.95 (RC4-CPN1) + 2.4 (RC4-CPN2) + 6.62 (RC4-CPN3) + 1.05 (RC4-CPN4) - 6.04 (RC4-CPN5)	18.71 (26.35)	0.77
RC5	Positive	R = -0.36 - 2.3 (RC5-CPP1) + 0.96 (RC5-CPP2) - 2.12 (RC5-CPP3) + 2.62 (RC5-CPP4)	6.71 (9.58)	0.75
	Negative	R = 0.12 - 6.88 (RC5-CPN1) - 14.22 (RC5-CPN2) - 3.84 (RC5-CPN3) + 13.65 (RC5-CPN4)	8.51 (13.87)	0.79

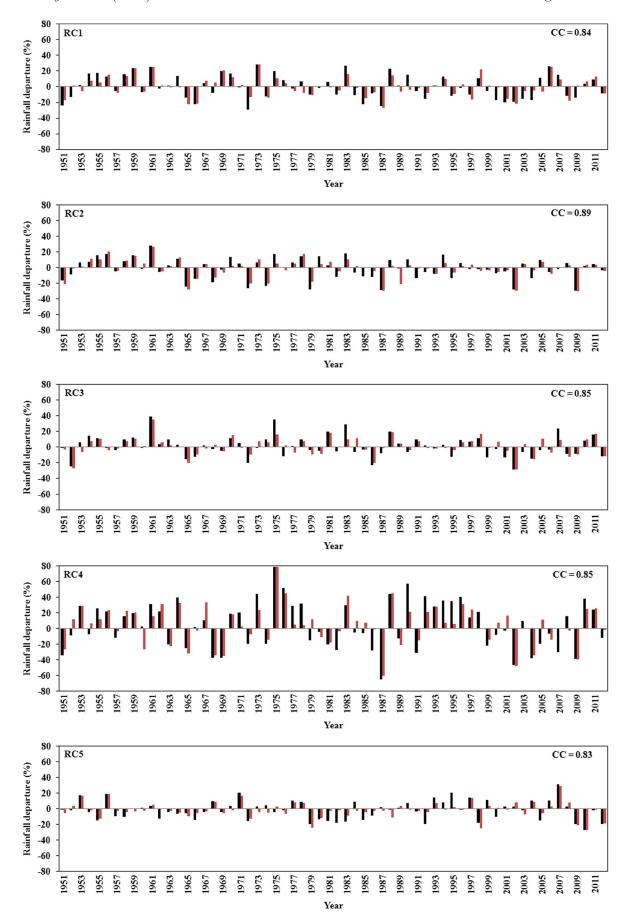


Figure 6. Observed (black) and estimated (red) summer monsoon rainfall departures (%) over five RC regions of India.

(1982), Rogers (1984), Alexander and Elena (1992), Hurrell (1996), Dugam and Kakade (1999). It has been observed that these two oscillations behave differently during the contrasting (excess/deficit) Indian monsoons, hence it is necessary to understand and consider the index for combined strength of NAO and SO. We coin it as ESI. It is observed that during excess monsoon years, ESI weakens from January to April, whereas during deficient monsoon years, there is a strengthening tendency of ESI from January to April (Kakade and Dugam 2000). Thus the behaviour of ESI from January to April is the critical factor between excess and deficit monsoon activity over India. Kakade and Kulkarni (2012) have shown that there is a clear line of demarcation in the relationship of ESI-tendency and Indian summer monsoon rainfall (ISMR) during contrasting phases of ESItendency. The relationship between ISMR and ESI-tendency is statistically significant in negative phase and is insignificant in positive phase of ESI-tendency. Further, they have also shown that the evolution of surface temperature over western Eurasia surrounding Moscow from winter to spring is exactly opposite during contrasting phases of ESI-tendency. Therefore in this section, we propose to understand the seasonal monsoon rainfall variability over RC regions during contrasting phases of ESI-tendency.

Kakade and Kulkarni (2014) have developed separate LRF equations for predicting summer monsoon rainfall departures (%) for all India and its homogeneous regions during contrasting phases of ESI-tendency. While obtaining independent predictors, first the cluster regions of various meteorological parameters have been identified by applying SNN algorithm to the parameter fields. Then by considering the linear and nonlinear impacts of

Table 8. Performance statistics of multiple regression equations for predicting summer monsoon rainfall over raincluster regions of India for 1951–2012.

Regions	RMSE	SD	CC	Chi-square
RC1 (high connectivity)	8.10	15.06	0.84	20.93
RC2 (medium connectivity)	6.09	13.37	0.89	44.06
RC3 (low connectivity)	7.10	13.31	0.85	18.76
RC4 (low connectivity)	18.80	29.67	0.85	18.80
RC5 (low connectivity)	6.79	12.07	0.83	18.07

different cluster series on summer monsoon rainfall over India and its homogeneous regions, the independent sets of predictors have been selected. The cross validation procedure is adopted over each homogeneous region, to obtain proper sets of independent cluster parameters in opposite phases of ESI-tendency.

Delsole and Shukla (2002) proposed crossvalidation scheme to select best linear prediction model for ISMR. The procedure first screens out all models which do not perform well on independent datasets, then the prediction error of each model is compared with those of all other models to examine whether the difference in error variance exceeds some significant threshold value. Sahai et al. (2002) used this scheme for selecting the best predictor set using 'leave one out' method, i.e., if there are P predictors and data is available for N years, the data is divided in two parts consisiting of 1 year as independent dataset and N-1 years as dependent dataset. The regression model is developed based on N-1 years and the value is predicted for the remaining year. This procedure is repeated for all N number of years. Values of root mean square error (RMSE) are calculated for each predictor and the first predictor is the one with the minimum RMSE. After selecting the first predictor, trial regression equations are again constructed using first selected predictor in combination with the remaining P-1 predictors. The combination of two predictors having minimum RMSE value is selected and this procedure is repeated for all P predictors. The plot of minimum RMSE value against number of predictors gives the set of best predictors, that with the least RMSE.

This procedure is applied to summer monsoon departure time series over each RC region. Figure 5 shows least RMSEs in independent datasets for each number of predictors during contrasting phases of ESI-tendency. Tables 2–6 show the best predictor sets over five RC regions of India during contrasting phases of ESI-tendency. Table 7 describes the corresponding LRF equations. It is clearly seen that by this method, for all RC regions of India, the RMSE is less than the standard deviation in both the phases of ESI-tendency.

Figure 6 shows observed and estimated summer monsoon rainfall departures (%) (1951–2012)

Table 9. Observed and estimated rainfall departures for 2013 and 2014 for all RC regions.

regrerie	•					
Year	Rainfall departure (%)	RC1	RC2	RC3	RC4	RC5
2013	Observed	31.89	17.54	20.40	10.37	-19.17
	Estimated	18.26	12.82	-11.49	-0.40	-23.43
2014	Observed	-14.18	-19.96	-1.54	-25.42	-15.36
	Estimated	-2.59	-29.04	-8.39	-13.25	-18.10

34

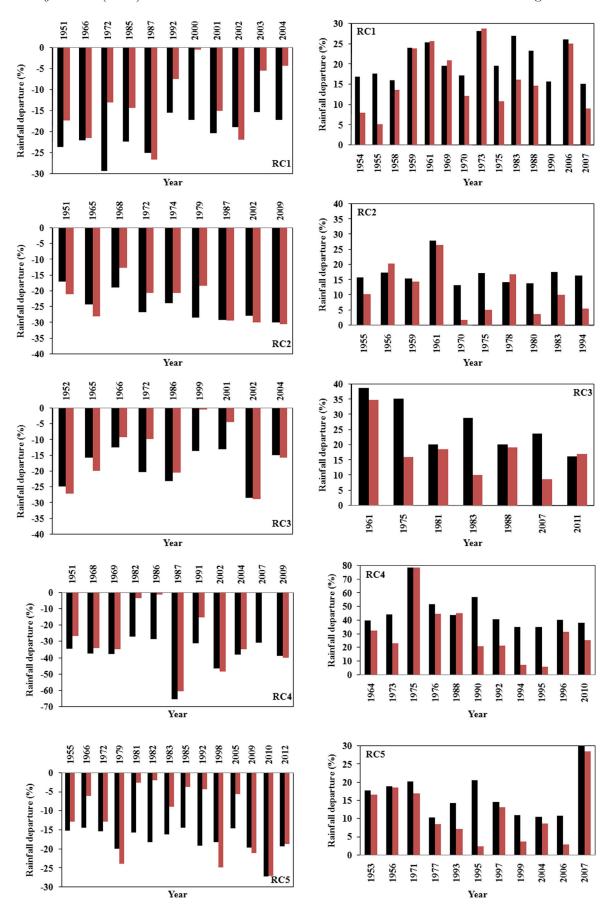


Figure 7. Observed (black) and estimated (red) summer monsoon rainfall departures (%) for deficit (left panels) and excess (right panels) monsoons over five rain clusters over the period 1951–2012.

over five RC regions of India using two separate equations for two phases of ESI-tendency. The performance statistics for all these LRF equations is shown in table 8. It suggests that RMSE is less than the standard deviation of summer monsoon rainfall departures over all these rain-cluster regions of India. The anomaly correlations are also more than 0.80 for all regions. Chi-square statistics is calculated for the contingency table showing qualitative agreement between observed and estimated rainfall departures. It also suggests significant relationship in qualitative prediction of monsoon rainfall over these rain-cluster regions. Table 9 shows observed and estimated rainfall departures for 2013 and 2014 for all RC regions. It suggests good agreement between observed and estimated rainfall departures for high connectivity regions (RC1 and RC2). The performance of these models in predicting extreme summer monsoons of each RC region of India is also examined. Figure 7 shows observed and estimated rainfall departures (%) for extreme monsoons over all RC regions of India. It can be seen that this method is able to predict almost all extreme monsoon rainfall years of five RC regions of India reasonably well.

4. Summary and conclusions

Five homogeneous zones have been constructed from the 30 subdivisions over India by applying the SNN technique to seasonal rainfall time series. Over each zone, two separate LRF models have been developed for two phases, positive and negative, of ESI-tendency. The cross validation procedure is applied to select the set of best predictors. Following are the main conclusions from this study:

- SNN gives five cluster regions over India
- Amongst all the cluster regions of India, the cluster region RC5 (RC4) receives maximum (minimum) rainfall and it shows minimum (maximum) variation with respect to mean rainfall.
- RMSE between the observed and predicted rainfall over all the five zones are well within the limits of observed standard deviation.
- The method well predicts the rainfall amount in excess/deficit monsoons over five zones of India.

Acknowledgements

Authors wish to thank, Director, Indian Institute of Tropical Meteorology for all the facilities provided.

References

Alexander B P and Elena N V 1992 The NAO and ENSO teleconnection; TOGA Notes 6 10-11.

- Azad S, Vignesh T S and Narasimha R 2010 Periodicities in Indian monsoon rainfall over spectrally homogeneous regions; Int. J. Climatol. 30 2289-2298.
- Boriah S, Simon G, Naorem M, Steinbach M, Kumar V, Klooster S and Potter C 2004 Predicting land temperature using ocean data; KDD Seattle, WA, USA.
- Chen W Y 1982 Fluctuations in Northern Hemisphere 700 Mb height field associated with the southern oscillation; Mon. Wea. Rev. 110 808-823.
- Delsole T and Shukla J K 2002 Linear prediction of Indian monsoon rainfall; J. Climate 15 3645–3658.
- Dugam S S, Kakade S B and Verma R K 1997 Interannual and long-term variability in north Atlantic oscillation and Indian summer monsoon rainfall; Theor. Appl. Climatol. **58** 21–29.
- Dugam S S and Kakade S B 1999 Interactive mechanism between ENSO and NAO and its relationship with Indian summer monsoon variability; In: Proc. National Symposium on 'Meteorology beyond 2000' TROPMET 99, pp. 53-58.
- Ertoz L, Steinbach M and Kumar V 2003 Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data; In: Proceedings of third SIAM international conference on data mining, San Francisco, CA, USA.
- Gadgil S and Iyengar R N 1980 Cluster analysis of rainfall stations of the Indian peninsula; Quart. J. Roy. Meteor. Soc. 106 873-886.
- Gadgil S, Yadumani and Joshi N V 1993 Coherent rainfall zones of the Indian region; Int. J. Climatol. 13 547–566.
- Goswami B N, Madhusoodanan M S, Neema C P and Sengupta D 2006 A physical mechanism for North Atlantic SST influence on the Indian summer monsoon; Geophys. Res. Lett. 33 L02706, doi: 10.1029/2005GL024803.
- Hurrell J W 1996 Influence of variations in extratropical winter time teleconnections on northern hemisphere temperature; Geophys. Res. Lett. 23(6) 665-668.
- Iyengar R N and Basak P 1994 Regionalization of Indian monsoon rainfall and long term variability signals; Int. J. $Climatol.\ 14\ 1095-1114.$
- Jarvis RA and Patrick EA 1973 Clustering using a similarity measure based on shared nearest neighbors; IEEE Trans. Comput. C-22 1025-1034.
- Kakade S B and Dugam S S 2000 The simultaneous effect of NAO and SO on the monsoon activity over India; Geophys. Res. Lett. 7(21) 3501–3504.
- Kakade S B and Kulkarni A 2012 Relationship between ESItendency and Indian monsoon rainfall: A possible mechanism; Atmos. Sci. Lett. 13 22–28, doi: 10.1002/asl.357.
- Kakade S B and Kulkarni A 2014 Convective activity over heat-low region and Indian summer monsoon rainfall during contrasting phases of ESI-tendency; Theor. Appl. Climatol. 115 591–597, doi: 10.1007/s00704-013-0920-2.
- Kakade S B and Kulkarni A 2016 Prediction of summer monsoon rainfall over India and its homogeneous regions; Meteorol. Appl. 23 1-13, doi: 10.1002/met.1524.
- Kulkarni A, Kripalani R H and Singh S V 1992 Classification of summer monsoon rainfall patterns over India; Int. J. Climatol. 12 269–280.
- Kulkarni A and Kripalani R H 1998 Rainfall patterns over India – Classification with Fuzzy C-means method; Theor. Appl. Climatol. 59 137-146.
- Kripalani R H and Kulkarni A 1997 Rainfall variability over southeast Asia: Connections with Indian monsoon and ENSO extremes: New perspectives; Int. J. Climatol. 17 1155 - 1168.
- Malik N, Marwan N and Kurths J 2010 Spatial structures and directionalities in monsoonal precipitation over south Asia nonlinear processes; Geophys. 17 371–381.
- Matulla C, Edouard K P, Haas P and Formayer H 2003 Comparative analysis of spatial and seasonal variability:

- Austrian precipitation during the 20th century; *Int. J. Climatol.* **23** 1577–1588.
- Nicholson S E 1986 The spatial coherence of African rainfall anomalies: Interhemisphere telecommunications; *J. Clim. Appl. Meteorol.* **25** 1365–1379.
- Pant G B and Parthasarathy B 1981 Some aspects of an association between the southern oscillation and Indian summer monsoon; Arch. Meteorol. Geophys. Bioklimatol. **B29** 245–251.
- Parthasarathy B, Rupa Kumar K and Munot A A 1993 Homogeneous Indian monsoon rainfall: Variability and prediction; *Proc. Indian Acad. Sci. (Earth Planet. Sci.)* 102 121–155.
- Rajeevan M, Pai D S, Dikshit S K and Kelkar R R 2004 IMD's new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003; *Curr. Sci.* 86(3) 422–431.
- Rao A R and Srinivas V V 2008 Regionalization of watersheds An approach based on cluster analysis; Springer, 241p.
- Rasmusson E M and Carpenter T H 1983 The relationship between eastern equatorial Pacific sea surface temperature and rainfall over India and Sri Lanka; *Mon. Wea. Rev.* 111 517–528.
- Rogers J C 1984 The association between the NAO and the southern oscillation in the Northern Hemisphere; *Mon. Wea. Rev.* **112** 1999–2015.
- Sahai A K, Grimm A M, Satyan V and Pant G B 2002 Prospects of prediction of Indian summer monsoon rainfall using global SST anomalies; *IITM Research Report*

- Saikranthi K, Narayana Rao T, Rajeevan M and Rao S V B 2012 Identification and validation of homogeneous rainfall zones in India using correlation analysis; *J. Hydrometeorol.* **14** 304–317.
- Satyanarayana P and Srinivas V V 2008 Regional frequency analysis of precipitation using large-scale atmospheric variables; *J. Geophys. Res.* **113** D24110, doi:10.1029/2008JD010412.
- Sikka D R 1980 Some aspects of the large-scale fluctuations of summer monsoon rainfall over India in relation to fluctuations in the planetary and regional scale circulation parameters; *Proc. Indian Acad. Sci. (Earth Planet. Sci.)* **89** 179–195.
- Shukla J and Paolino J A 1983 The southern oscillation and the long-range forecasting of summer monsoon rainfall over India; *Mon. Wea. Rev.* 111 1830–1853.
- Steinbach M, Klooster S, Potter C and Kumar V 2003 Discovery of climate indices using clustering; KDD, Washington DC, USA.
- Van Loon H and Rogers J C 1978 The seasaw in winter temperature between Greenland and Northern Europe. Part II: Some oceanic and atmospheric effects in middle and high latitude; *Mon. Wea. Rev.* **107** 509–519.
- Venkatesh B and Jose M K 2007 Identification of homogeneous rainfall regimes in parts of Western Ghats region of Karnataka; *J. Earth Syst. Sci.* **116** 321–329.
- Verma R K 1990 Recent monsoon variability in the global climate perspective; *Mausam* 41 315–320.

MS received 1 January 2016; revised 9 November 2016; accepted 12 January 2017

Corresponding editor: ASHOK KARUMURI