ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/301594641

Homogeneous clusters over India using probability density function of daily rainfall

Article *in* Theoretical and Applied Climatology · July 2017

DOI: 10.1007/s00704-016-1808-8

CITATIONS
9

READS
293

1 author:

Ashwini Kulkarni
Indian Institute of Tropical Meteorology

73 PUBLICATIONS **4,563** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project

Short term Climate Variability and prediction View project

All content following this page was uploaded by Ashwini Kulkarni on 03 October 2018.

The user has requested enhancement of the downloaded file.

Theor Appl Climatol
DOI 10.1007/s00704-016-1808-8

ORIGINAL PAPER

Homogeneous clusters over India using probability density function of daily rainfall

Ashwini Kulkarni[1]

Received: 25 September 2014 / Accepted: 6 April 2016
© Springer-Verlag Wien 2016

Abstract The Indian landmass has been divided into homogeneous clusters by applying the cluster analysis to the probability density function of a century-long time series of daily summer monsoon (June through September) rainfall at 357 grids over India, each of approximately 100 km × 100 km. The analysis gives five clusters over Indian landmass; only cluster 5 happened to be the contiguous region and all other clusters are dispersed away which confirms the erratic behavior of daily rainfall over India. The area averaged seasonal rainfall over cluster 5 has a very strong relationship with Indian summer monsoon rainfall; also, the rainfall variability over this region is modulated by the most important mode of climate system, i.e., El Nino Southern Oscillation (ENSO). This cluster could be considered as the representative of the entire Indian landmass to examine monsoon variability. The two-sample Kolmogorov-Smirnov test supports that the cumulative distribution functions of daily rainfall over cluster 5 and India as a whole do not differ significantly. The clustering algorithm is also applied to two time epochs 1901–1975 and 1976–2010 to examine the possible changes in clusters in a recent warming period. The clusters are drastically different in two time periods. They are more dispersed in recent period implying the more erroneous distribution of daily rainfall in recent period.

✉ Ashwini Kulkarni
ashwini@tropmet.res.in

[1] Center for Climate Change Research, Indian Institute of Tropical Meteorology, Dr Homi Bhabha Road, NCL PO, Pashan, Pune 411008, India

1 Introduction

Indian summer monsoon rainfall (ISMR) exhibits a large spatial variability. Hence, the all-India rainfall index may not be a meaningful index in various applications. Also in need of the seasonal rainfall predictions on smaller spatial domains, a lot of efforts have been put in to divide the Indian landmass in homogeneous regions. The average over a homogeneous region not only reduces the data volume but also small-scale variability and enhances signal variation at larger spatial scales (Nicholson 1986). A variety of techniques have been used to identify homogeneous rainfall regions over India such as correlation analysis (Nicholson 1986; Kulkarni et al. 1992; Gadgil et al. 1993), principal component analysis (PCA; Kulkarni et al. 1992; Iyengar and Basak 1994; Singh and Singh 1996), 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; Satyanarayana and Srinivas 2011)). Each method has its own merits and demerits. Kulkarni et al. (1992) applied PCA to summer monsoon rainfall over 52 grids over India, each of size 2.5° lat/long while Gadgil and Iyengar (1980) applied PCA to seasonal rainfall at some stations over the Indian peninsula. Gadgil et al. (1993) clubbed together the stations whose seasonal rainfalls have maximum positive correlations with each other. Azad et al. (2010) have identified ten spectrally homogeneous regions, i.e., the regions which have similar power spectral density. Parthasarathy et al. (1993) have constructed five homogeneous regions. They have used sub-divisional rainfall over India and the subdivisions having similar rainfall characteristics, and similar relationships with 12 global/regional long-range forecasting parameters have been clubbed together.

Published online: 22 April 2016

The five homogeneous regions are northwest India, west central India, central northeast India, northeast India, and peninsular India. Rajeevan et al. (2004) divided the Indian landmass in three homogeneous regions using sub-divisional rainfall. The regions are thus chosen in which the rainfall variations in each of the meteorological subdivision comprising the region is positively significantly correlated with the area-weighted rainfall variations over the region as a whole. These are northwest India, northeast India, and south peninsular India. Saikranthi et al. (2012) identified 26 homogeneous zones over India using correlation analysis on daily rainfall over 1025 rain gauges over India.

It is a well-known fact that along with a large spatial variability, the Indian summer monsoon rainfall depicts temporal variability on inter-annual as well as intra-seasonal time scale. The one and half century-long time series of all-India summer monsoon rainfall has been prepared by area-weighted average of 306 stations well distributed over the Indian landmass and has been archived by the Indian Institute of Tropical Meteorology, Pune.(www.tropmet.res.in). The all-India summer monsoon rainfall series has long-term mean of 850 mm with standard deviation of 85 mm based on years 1871–2012. This time series is characterized by excess/deficient and normal monsoons. The monsoon in a particular year is defined to be excess (deficient) if the standardized rainfall for a particular year is greater (less) than + (−) 1. It has been observed that the spatial distribution of seasonal rainfall may be drastically different in two excess/deficient monsoons. Also, the day-to-day behavior of rainfall at a place may be different over the years though the seasonal total rainfall is the same. The intra-seasonal or day-to-day variability of Indian summer monsoon contributes substantially to its inter-annual variability. Thus, the distribution of daily rainfall decides the total strength of the summer monsoon rainfall. India, being an agrarian country, has an economy that depends on the annual crop production. Different crops require different amounts of water in different phases of life cycles. Hence, the distribution of daily rainfall during the season is very important. Here, we try to construct homogeneous regions over India using probability distribution of daily summer monsoon rainfall over 357 grids each of size 100 × 100 km^2. The method applied is the k-means clustering method. Thus, the method clubs together the grids which have similar probability distribution of daily rainfall. The clusters thus formed are similar within and dissimilar with each other.

It has been observed that the surface temperature has been rising unequivocally over the last century. The globe has been warmed by 0.74 °C in the last 100 years. The rate of warming happens to be more accelerated in the last 30 years. This has adverse effects on extreme events. Although the impact of global warming on changing climate has always been discussed in terms of mean climate, the potential changes in extremes have devastating effects on the society's and the ecosystem's adaptive capacity (Kharin et al. 2007; Trenberth et al. 2007). Also, a small change in the mean climate may be associated with the large changes in the extremes. The probability distribution of the climate variable gives idea about the extremes. The probability density function (pdf) of various climate parameters has been used in many studies. Hannachi (2006) has examined changes in pdfs and various quantiles of climate variables using order statistics. Donat and Alexander (2012) have used pdfs of daily maximum and minimum temperatures and showed that the distribution of both have shifted to hotter values in recent period. Ruff and Neelin (2012) examined long tails of pdfs of daily average, maximum and minimum temperature, and their implications for extremes under global warming. Loikith et al. (2013) have used pdf of daily temperature to construct the homogeneous clusters representing the climate regimes over USA. The pdfs help us to quantify the vulnerability to the risks associated with extreme events and to examine the projected changes in risks.

The data used are detailed in Sect. 2. Section 3 describes the methodology. Results are discussed in Sect. 4 and finally, conclusions are given in Sect. 5.

2 Data

The datasets used in this study are as follows:

1. High resolution (1° × 1° lat/long) daily gridded rainfall data set for the Indian region from 1901 to 2004 prepared by India Meteorological Department (Rajeevan et al. 2008) and updated up to 2010. For this analysis, the rainfall data over 1384 stations with minimum 70 % data availability were considered. The interpolation method proposed by Shepard (1968) had been used for interpolating station rainfall data into regular grids. The quality of this dataset was examined by comparing with the similar kind of data set developed for the period of 1951–2004 (Rajeevan et al. 2006). The data are available for 1 January to 31 December for every year; here, the data for Indian summer monsoon season, 1 June to 30 September, i.e., 122 days for each of these 110 years (1901–2010) have been used.
2. All India seasonal summer monsoon rainfall (June–September) for the period of 1871–2012 has been downloaded from the website of the Indian Institute of Tropical Meteorology, www.tropmet.res.in.

3 Methodology

To get the homogeneous clusters over Indian landmass, the k-means clustering algorithm has been applied to pdfs

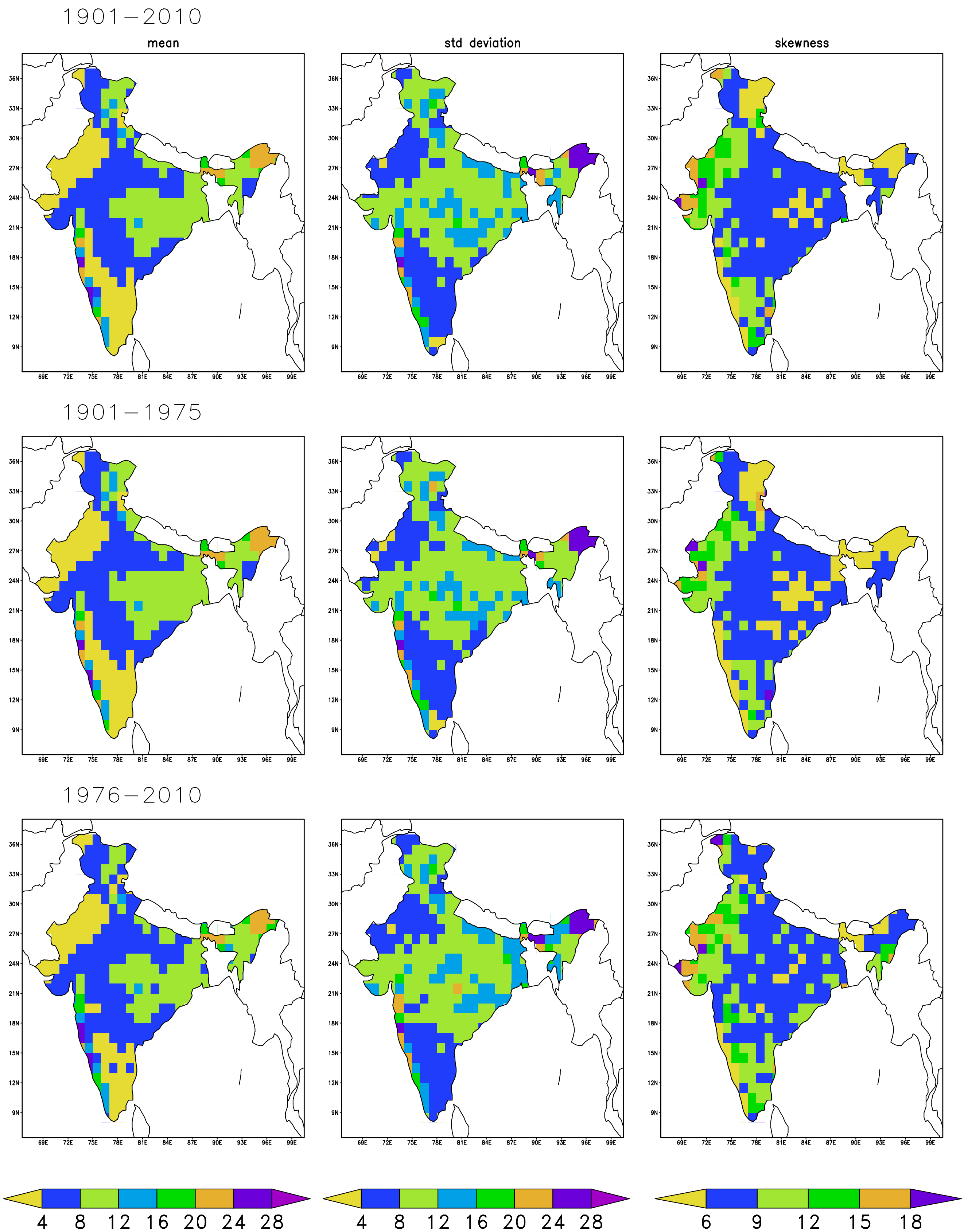


Fig. 1 Maps of mean (mm/day), standard deviation (mm/day), and skewness of daily summer monsoon rainfall (June through September) based on periods of 1901–2010 (*first row*), 1901–1975 (*second row*), and 1976–2010 (*third row*)

of daily rainfall at each grid point. The daily rainfall data is first deseasonalized by removing the daily climatology over the 110 year period (1901–2010). The linear trends are removed at each grid and the daily rainfall anomalies are computed which makes the mean 0 and helps in comparison of pdfs across the domain. Anomalies for 122 days × 110 years = 13,420 values are classified into bins of 5 mm at every grid. In all, there are 46 bins to span all values for all the grids, though at some grids, the frequencies for some bins may be zero. So, in all, we have 357 grids, each having a pdf of 46 bins. The k-means clustering algorithm is applied to {log10(fr(i)/13,420)}, i = 1,2…46 over a spatial domain of 357 grids to increase the weight of distribution tails in clustering (Loikith et al. 2013). Generally, daily rainfall has less frequency at right extreme end. The initial clusters were selected randomly. The k-means algorithm has been applied taking initial number of clusters 3 to 8; however, five clusters give reasonable picture. In other words, the clustering algorithm seeks 5 sets among these 357 vectors of length of 46 of log of probabilities which minimize within clusters, sum of squares, and maximize between cluster distances. Loikith et al. (2013) have defined clusters for the January temperature over North America.

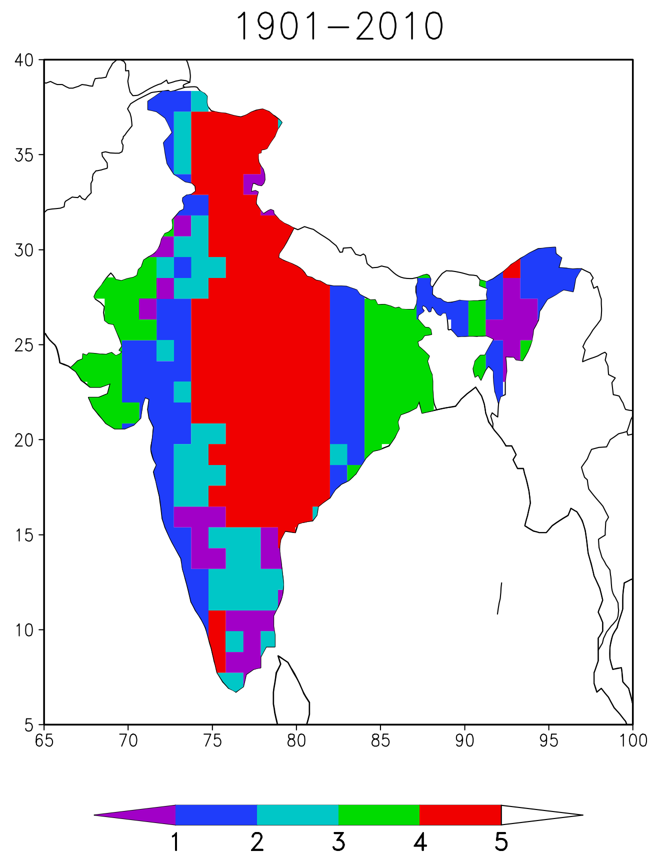


Fig. 2 Five clusters over India based on probability distribution function of daily rainfall over India for the period of 1901–2010

4 Results

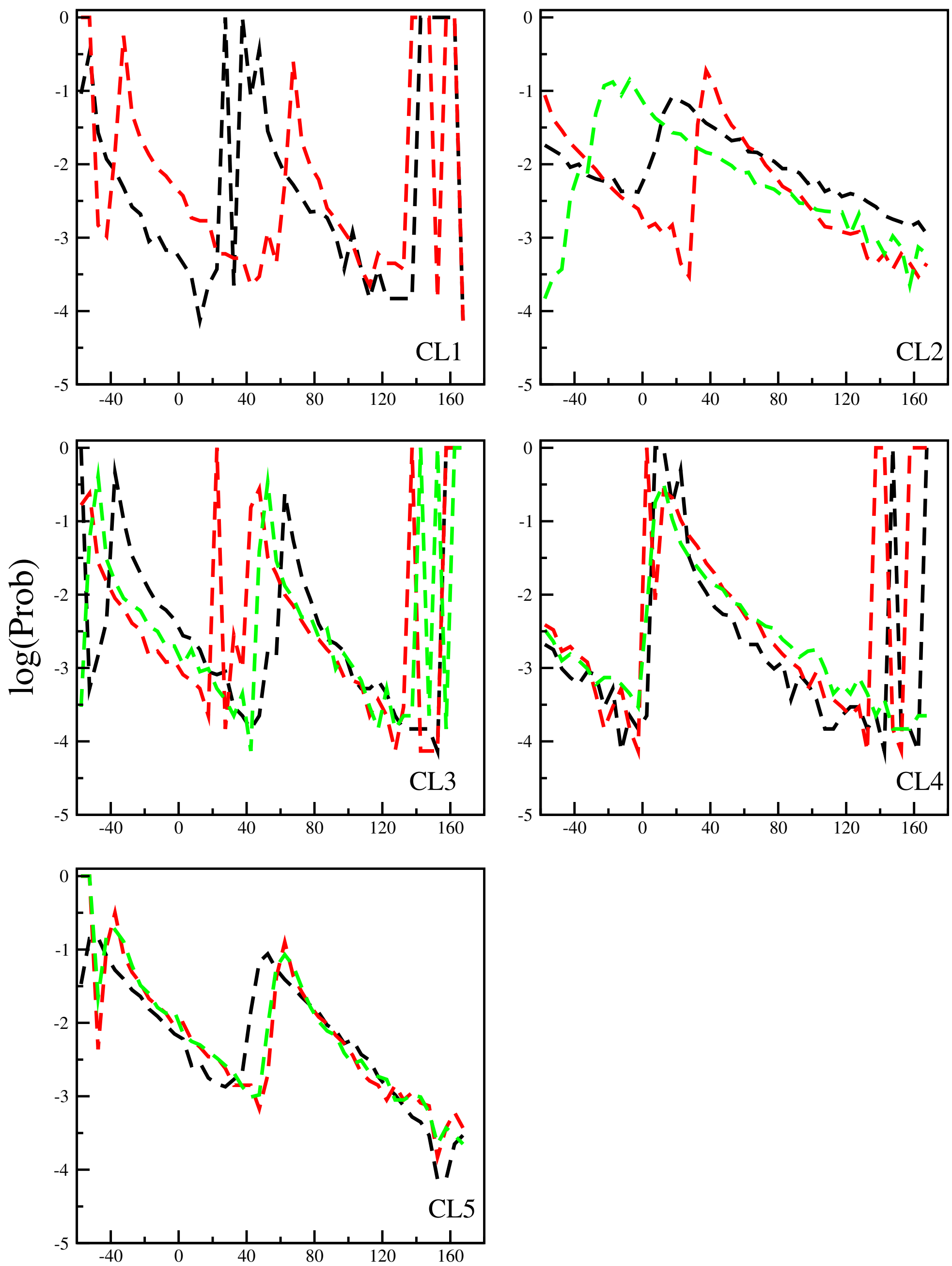
4.1 Moments of daily rainfall distribution

The important parameters of any distribution are its moments. Figure 1 depicts the mean, standard deviation, and skewness of the daily rainfall distribution based on the period of 1901–2010 (first row). Mean pattern shows that the northwest and southeastern parts of India are the dry regions with average daily rainfall of less than 4 mm. West coast and the northeastern tip are wet regions with mean rainfall of 20–24 mm/day. Over the central parts, there is a moderate rainfall of 4–12 mm/day. The middle panel shows the standard deviation (second moment) of the daily rainfall distribution. The regions of less (more) rainfall are also the regions of less (more) variability, which are very well brought out in this panel. Last panel shows the standard skewness coefficient defined as sqrt(μ_3^2/μ_2^3). The daily rainfall distributions are highly skewed over the regions of less rainfall, i.e., northwest and southeastern parts, whereas less skewed over the west coast and northeastern parts, comparatively wet regions. Over the central parts, the distributions are moderately skewed. For the simple sensitivity analysis, the total period is divided in two epochs: 1901–1975 and 1976–2010. In the second epoch, there is substantial warming over the globe as well as the Indian landmass. To examine whether there is any difference in these simple measures, the mean, standard deviation, and skewness coefficient have been computed separately over these two periods and presented here in Fig. 1 (second and third row). It is well seen that the general pattern of these three measures is not different in two epochs though the rainfall is reduced over the central parts in thelater period. The daily rainfall distributions remain equally asymmetric in the two time epochs which suggest that these patterns are stable with respect to time in the 20th and early 21st century.

4.2 Clusters using pdf

As discussed in Sect. 3, the k-means clustering algorithm is applied to 46 bins of daily pdf over 357 grid points spread over the Indian landmass. K-means algorithm forms the sub-regions (clusters) so that the within-group distances are at minimum and the between-group distances are at maximum. The algorithm has been applied with number of initial clusters to be three through eight. However, k = 5 gives reasonable groups. If we take k = 3, the northwest and west coast form one cluster which is known to be unreasonable since the northwest region is of scanty rainfall and west coast is a zone of maximum rainfall, and hence, the pdf of daily rainfall differ substantially in these two regions. If k is more than 5, clusters are quite distributed and it is hard to get a contiguous region. The initial 5 clusters are selected randomly from 1 to 357. The final clusters do not depend on the initial selection. Fig. 2 depicts the homogeneous clusters over India using the number of initial clusters to be five. The five clusters are formed over India as the vertical belts. Cluster 1 constitutes of some region on northwest and some region on southeast which are generally low rainfall regions and some region from northeast India. Cluster 1 does not give a contiguous region. Cluster 2 consists of two parts: one is along the entire west coast except Kerala and Gujarat, and southern parts of east Rajasthan and the other is along the western parts of Orissa, Jharkhand, and Bihar. Cluster 3 is again formed by discrete grids from some parts of Rayalseema, South Interior Karnataka, northern Tamilnadu, and Madhya Maharashtra which are again regions of low rainfall though not as low as northwest India, as also seen form Fig. 1. Cluster 4 is along two parts: one over the western side, Saurashtra, Kutch, Diu, and west Rajasthan and the other over the eastern side, West Bengal and eastern parts of Orissa, Jharkhand, and Bihar. Finally, cluster 5 is a contiguous region over the central parts of India. The subdivisions constituting

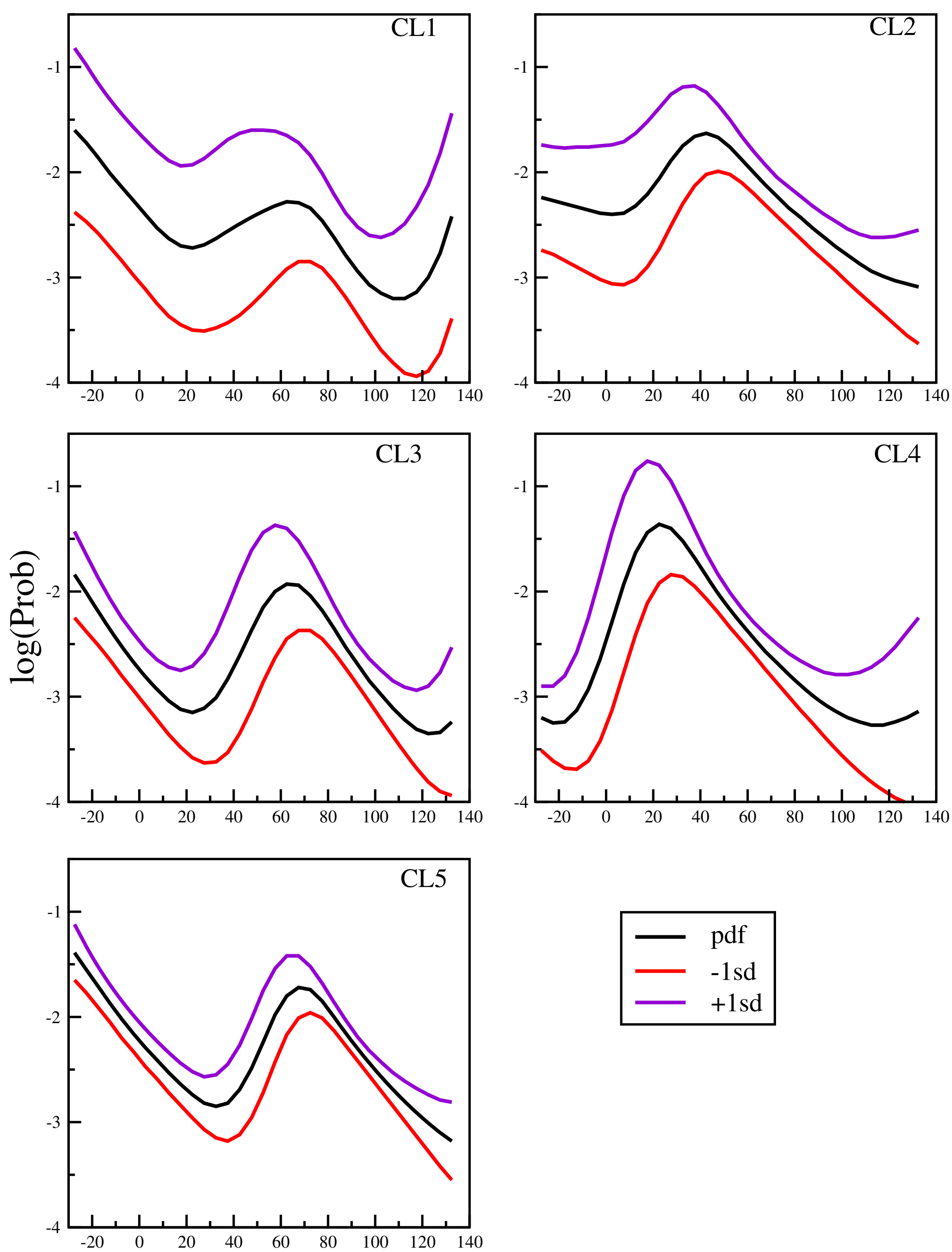
Fig. 3 The pdfs of representative grids in five clusters. Y-axis is $\log_{10}$ (probability) against x-axis daily rainfall anomalies



cluster 5 are Telangana, Vidarbha, Marathwada, Madhya Pradesh, Chattisgarh, Uttar Pradesh, Uttaranchal, Himachal Pradesh, Jammu and Kashmir, and southern parts of Kerala. Cluster 5 includes Uttaranchal, Himachal and Jammu, and Kashmir regions which are most vulnerable to extreme precipitation events. Why these clusters form along the longitudes is an interesting topic of research.

Since all the clusters except cluster 5 do not form the contiguous regions, we have examined the pdfs of grids in each cluster at various locations to show the similarity in their daily rainfall distributions. Fig. 3 shows these representative pdfs. The grids form each cluster have been selected from various locations in respective cluster. The logs (probability) are plotted against the daily rainfall anomalies. We select one grid on northwest and one on southeast from cluster 1, whose pdfs are plotted in panel a. Panel b represents pdfs of three grids from cluster 2: one from west coast, one from the central parts, and the third from the northeast. For cluster 3, the pdfs for three grids,, one each on southeast, middle part and northern parts, are given in panel c. In panel d, three representative pdfs for cluster 4 are shown: one grid on Gujarat, one on Bihar, and the third one on the northeastern region. Cluster 5 is a contiguous region, still to examine the pdfs, we have shown pdfs on three

Fig. 4 Average pdf of five clusters (*black*) with −1 standard deviation (*red*) and +1 standard deviation (*violet*) bounds. The pdfs have been plotted after applying 13-point binomial filter



grids in panel e: one on Kerala, one grid over the central region, and the third one on Jammu and Kashmir, the northernmost region. It can be clearly seen that though the grids are located over different parts of India, all the grids in a cluster have similar pattern of pdf, with slight difference in means. The overall pattern of pdfs in each cluster is similar and hence they constitute the cluster.

4.3 Average pdfs of clusters

To get the idea of pdfs of daily rainfall departures in each cluster, the average pdf has been computed for every cluster and is shown in Fig. 4. The black curve shows the average of the pdfs at all the grids in a cluster after smoothing using 13-point binomial filter. Red (violet) curve in each panel shows −1 (+1) average standard deviation for that cluster. In cluster 1, the extreme rainfall departures, i.e., very high or very low rainfall values, are more probable compared to moderate rainfall values. This cluster is formed by some grids from southeast, some from northwest which are low variability regions as seen from Fig. 1; however, it also includes the grids from the northeast region which are highly variable. The high rainfall variance is well evidenced by the wide spread of the distribution of cluster 1. Also the intra-cluster variability is very high which may be due to local variations in rainfall at northwest and northeast region.

The pdf of cluster 2 gives high probability to moderate values. The spread within the cluster is more toward lower rainfall values since this cluster includes two heavy rainfall regions of west coast and northeast; hence, its distribution is dominated by positive rainfall departures and thus is positively skewed. The pdf of cluster 3 depicts maximum probability to moderate values as compared to extremes, but the average rainfall is more for this cluster compared to that in cluster 2. The variability within the cluster is more toward too large or too small values and less toward the mean values. This cluster includes, in general, moderate rainfall grids.

The scanty rainfall regions of West Rajasthan and Sourashtra and Kutch and some parts of east central India constitute cluster 4. The pdf clearly shows more probability to lower rainfall values. The intra-cluster variability is more toward higher values of rainfall. Cluster 5, which includes a large number of grids, has a similar pattern of pdf as cluster 3; however, it has very less within cluster variability mainly because this is a contiguous region and all the grids in this cluster have moderate standard deviation as can be seen from Fig. 1.

It is well observed that the five clusters have drastically different pdfs hence the between-cluster variability is very large. The k-means clustering algorithm gives us distinct clusters over Indian landmass.

4.4 Relationship with all-India rainfall and ENSO

The central India region defined by Parthasarathy et al. (1993) which consists of 14 subdivisions over the central and northern parts of India depicts very high correlation with the entire Indian landmass. The variations in seasonal rainfall over this region are in phase with variations in all-India summer monsoon rainfall. It has been shown that the rainfall variations over northeast India are out of phase with all India rainfall. The first four clusters do not form contiguous regions and the correlations between seasonal rainfall for these four clusters and all-India summer monsoon rainfall are 0.18, 0.51, 0.57, and 0.44, respectively (figure not given). Cluster 5 gives the contiguous region and so to examine the association of rainfall variability over cluster 5 with that over all India, the scatter plots of cluster average rainfall with area-weighted all-India rainfall for summer monsoon season have been shown in Fig 5. Y-axis depicts the standardized ISMR against x-axis, the area-averaged standardized rainfall for cluster 5. The scatter is based on the period of 1901–2010. It is also seen that no cluster shows out-of-phase relationship with Indian summer monsoon rainfall. Cluster 5 which is the largest contiguous cluster shows a very strong relationship with Indian monsoon rainfall, 0.86, hence this cluster can be taken as representative of the entire Indian landmass.

The area-averaged standardized time series for all of cluster 5 has been shown in Fig. 6. This cluster clearly captures the excess/deficient rainfall of all India. The major deficits on all-India scale such as 1905, 1911, 1918, 1920, 1965, 1966, 1972, 1979, 1982, 1987, 2002, 2004, and 2009 happen to be large deficits on this cluster as well.

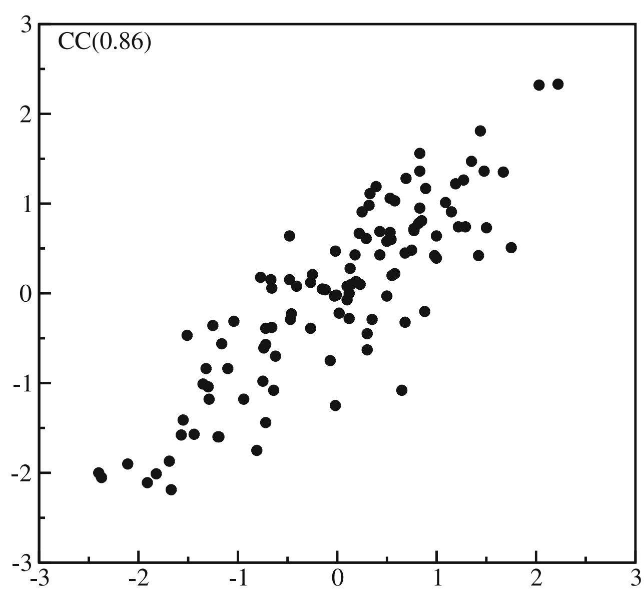


Fig. 5 Scatter plot of standardized anomalies of all-India summer monsoon rainfall against rainfall averaged over cluster 5. The *numbers in bracket* show correlation coefficient

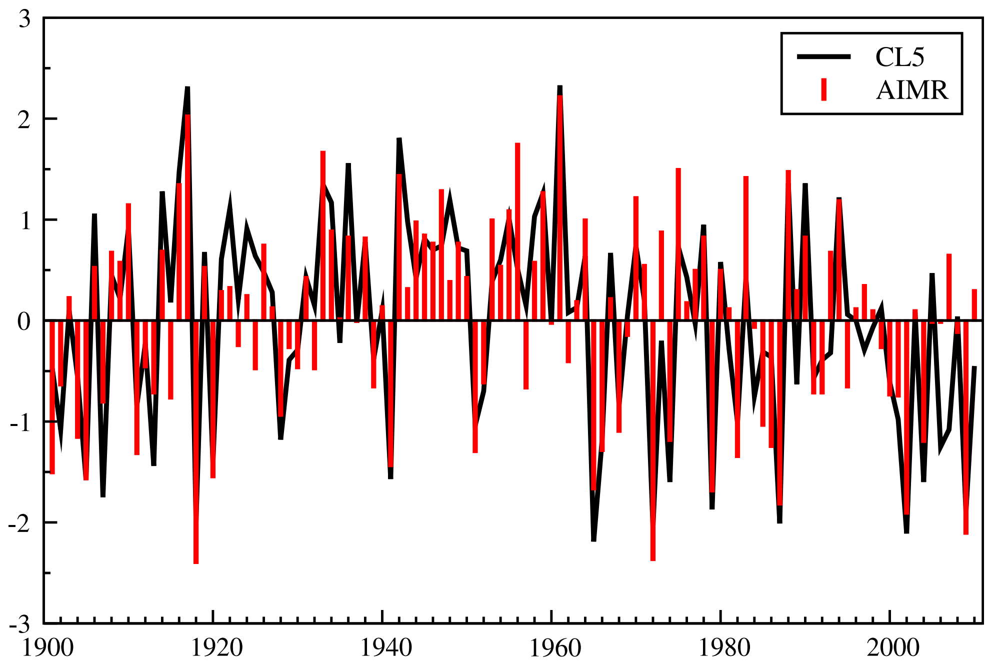


Fig. 6 Standardized summer monsoon rainfall anomalies for all-India and cluster 5 for the period of 1901–2010

The seasonal rainfall over cluster 5 does show highly significant relationship with ISMR; however, we examine the nature of relationship on daily scale. Figure 7 shows the magnitude of correlation coefficient between daily rainfall over cluster 5 and that averaged over all-India for 122 days in summer monsoon season. The lines stand for normal monsoon years, the red dots are for deficient ISMR, and green squares for excess ISMR. During normal monsoons, the correlations range from 0.57 to 0.9; except 4 years out of 74, the correlations are more than 0.7. During deficit ISMRs, the correlations range from 0.64 to 0.85 while for excess, they ranged from 0.72 to 0.85. The significant value of correlation is 0.233 (0.173) for 1 (5) % level, so all these correlations are highly significant which emphasizes that cluster 5 can be taken as representative for the all-India region. Also, the daily mean rainfall averaged over the period of 1901–2010 for cluster 5 and the all-India region depicts similar pattern and comparable values as shown in Fig. 8.

It is a well-known fact that El Nino Southern Oscillation (ENSO) is the most important mode of the earth's year-to-year variability. It has been shown that majority of the El Nino cases are associated with failure of monsoon over India (Sikka, 1980). However, this inverse relationship seems to be weakening in recent period (Kripalani and Kulkarni 1997; Krishna Kumar et al. 1999). It has been observed that in the recent 30 years, the Indian summer monsoon may not be deficient though it is a strong El Nino, e.g., 1997. One of the

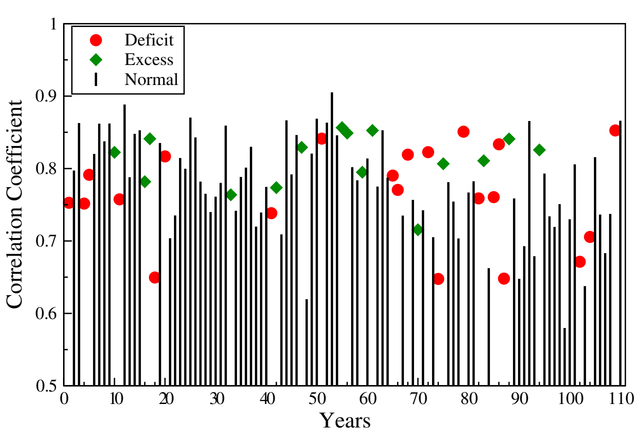


Fig. 7 Correlation coefficient between daily summer monsoon rainfall over cluster 5 and all-India for 110 years (1901–2010)

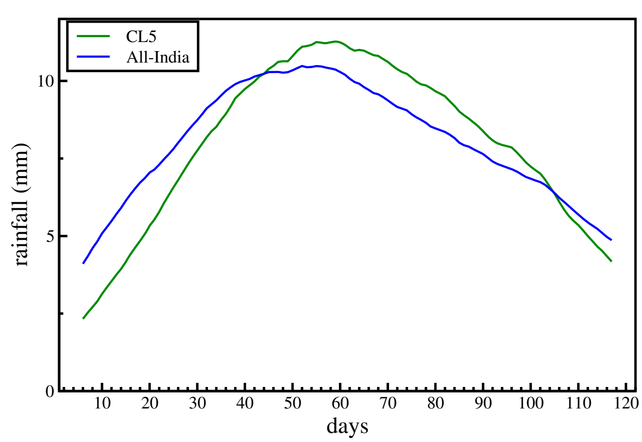


Fig. 8 Eleven-year running mean daily monsoon rainfall of cluster 5 and all-India based on the period of 1901–2010

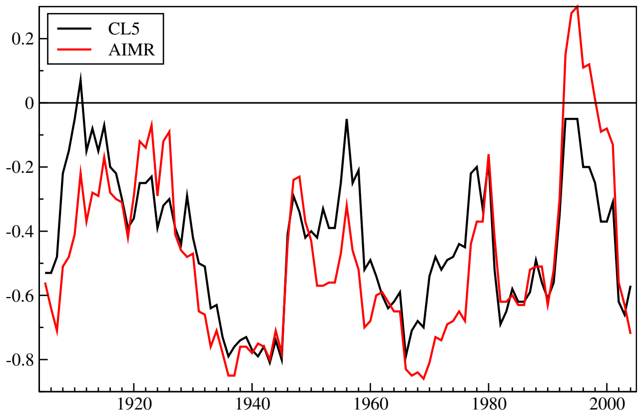


Fig. 9 Eleven-year sliding correlation coefficients of JJA NINO3.4 SST with summer monsoon rainfall over India and average rainfall of cluster 5

major cause of this is mainly contributed to the fact that in recent times, there are more frequent Modoki events rather than El Nino (Ashok et al. 2007) which are defined as anomalous warming in central Pacific rather than in east equatorial Pacific. The seasonal rainfalls associated with all the warm ENSO events happen to be large deficits over cluster 5. Even in 1997, a record ENSO event, when all-India summer monsoon rainfall was more than the long-term average rainfall, the seasonal rainfall over cluster 5 occur at the negative side (standardized rainfall −0.3), though not a strong deficit, as clearly seen from Fig. 9.

The relationship between an index of El Nino and seasonal rainfall over cluster 5 has been examined. Figure 9 shows 11-year sliding correlations of NINO3.4 SSTs with seasonal rainfall over cluster 5 as well as all-India. The sliding correlations of NINO3.4 SST do show epochal variability as shown by correlations with all-India rainfall (black curve). The relationship has been changed to positive at sometime during the 100 years; however, rainfall over cluster 5 has had a consistent inverse relationship with NINO3.4 SST since 1910. Hence, the relationship between ENSO and seasonal rainfall over cluster 5 has not been weakened for the last 100 years. Thus, ENSO may be a good predictor for seasonal rainfall over the CL5 region.

The Kolmogorov-Smirnov test for testing difference between two distributions has been applied to cumulative density functions of daily rainfall of cluster 5 and all-India. At the 0.025 level of significance, the difference between two distributions is statistically nonsignificant. This implies that not only for the teleconnections but even for the daily scale pdfs cluster 5 represents all-India daily rainfall reasonably well.

4.5 Epochal changes

Since 1976, the globe has experienced an unprecedented warming. Even the rate of warming is much accelerated in the post-1975 period over India. To examine whether there is any change in pdf of rainfall on daily scale so that the clusters formed by using pdfs are different in the two time epochs of 1901–1975 and 1976–2010, we applied the same

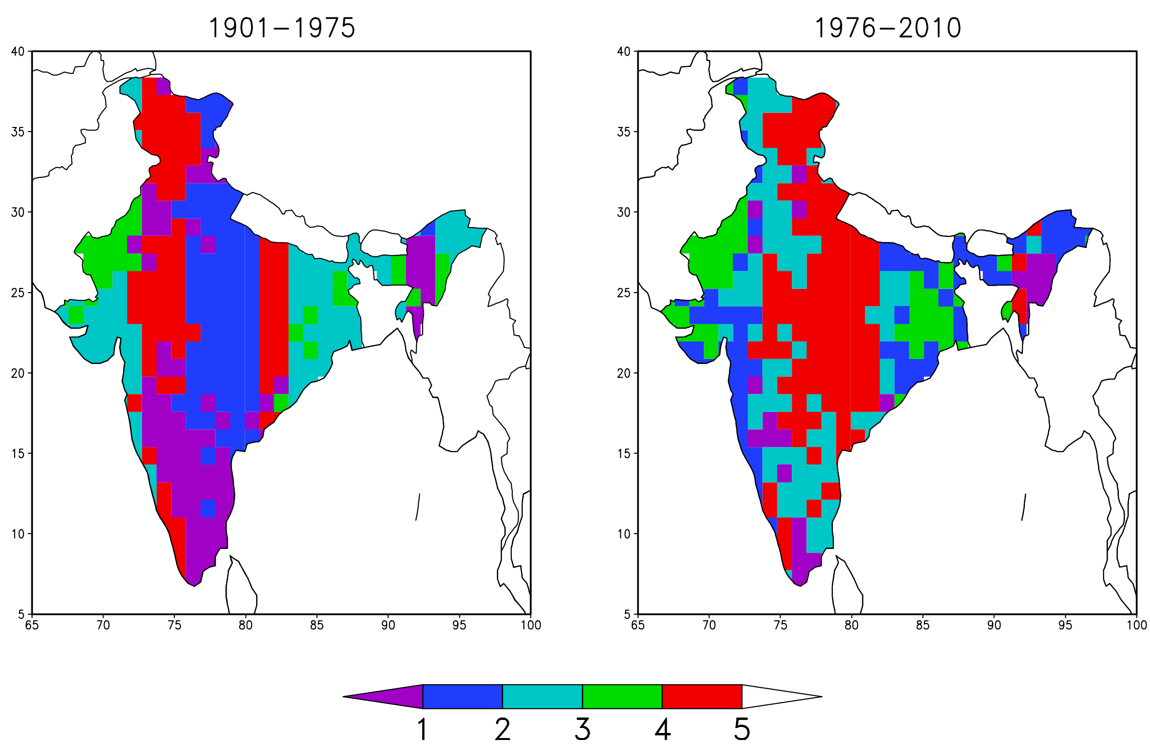


Fig. 10 Five clusters over India based on pdf of daily rainfall distribution in two epochs: 1901–1975 (*left*) and 1976–2010

k-means method to daily rainfall distributions in these two time epochs. Figure 10 shows the clusters obtained in these two time epochs, with same selection of initial clusters as for the period of 1901–2010. As shown in Fig. 1, there is no much change in mean and variability pattern as well as skewness of the daily rainfall distribution. However, the clusters formed are drastically different in two time epochs. It can be seen that during 1901–1975, the distributions are comparatively stable; the clusters formed do show contiguous regions. In the epoch of 1976–2010, the daily rainfall distribution is quite erratic, and the clusters are much dispersed. However, the clusters based on the entire period of 1901–2010 seem to be dominated by the period from 1976 to 2010. The clusters of years 1901–2010 resemble those obtained based on the years 1976–2010. Hence, the unequivocal warming of the post-1975 period has substantial impact on the daily rainfall distribution of India.

The average pdf of five clusters based on these two epochs have been compared with that of the period of 1901–2010 and is shown in Fig. 11. There is a large difference in pdfs in the two time epochs: 1901–1975 (red) and 1976–2010 (green) for clusters 1, 2, and 3. For CL4 and CL5, there is not much difference in pdfs based on the two time periods. Thus, the pdfs of daily rainfall do not change much over the regions of CL4 and CL5 even in warming period. Also, it is well observed that the average pdf based on the period of 1901–2010 (black) resembles that based on the period of 1976–2010 (green) for all the clusters implying that the pdfs in the period of 1901–2010 are dominated by those in the period of 1976–2010.

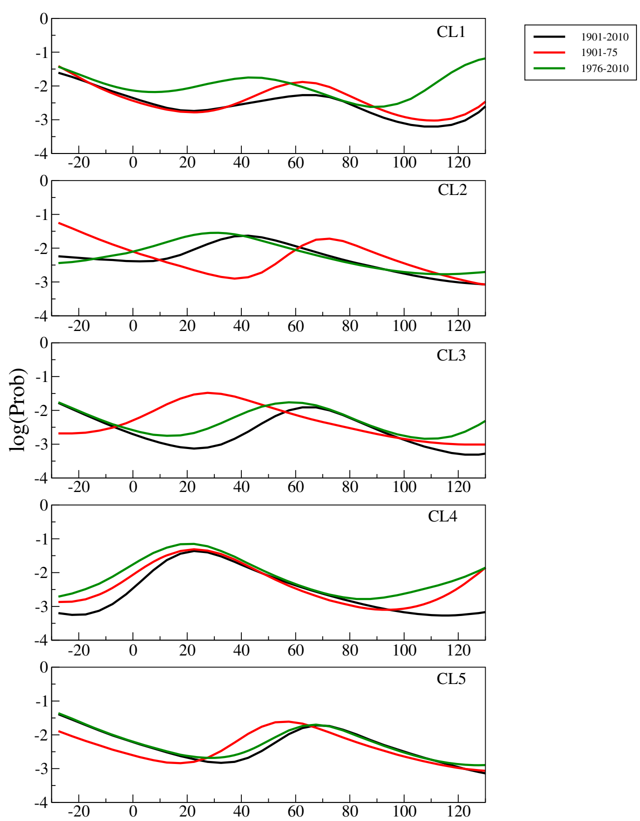


Fig. 11 Average pdf for five clusters in three time epochs: 1901–2010 (*black*), 1901–1975 (*red*), and 1976–2010 (*green*)

5 Conclusions

In this paper, the clusters over India have been constructed using the probability density function of daily rainfall over 357 grids each of approximately 100 km^2 based on century-long data of 1901–2010. By applying the k-means clustering algorithm, we obtain five clusters over India. The four out of five clusters are not contiguous regions and do not represent the rainfall regimes over India. Cluster 5 gives the largest contiguous region and has maximum positive relationship with ISMR on daily as well as seasonal scale. So, this cluster can be considered to be representative of the entire Indian landmass. Also, the seasonal rainfall variations over this cluster show consistent inverse relationship with the most important predictor of ISMR, the ENSO. The daily rainfall cumulative densities of cluster 5 and all-India do not depict the statistically significant difference. The clusters in two time epochs before and after global warming are drastically different. The clusters based on the period of 1976–2010 seem to dominate the entire period of 1901–2010.

Further studies will focus on developing the clusters over Indian landmass using pdfs of daily rainfall simulated by a suite of high resolution regional climate models from Coordinated Regional Downscaling Experiment (CORDEX) of WCRP and examining the projected changes in the patterns under various Representative Concentration Pathway (RCP) experiments. Perkins et al. (2007) used pdf skill scores to evaluate the model performance. This approach may help to identify regions over India where future changes in daily rainfall may be comparatively homogeneous. Since the method involves pdfs of daily rainfall, this may provide information in changes in climate extremes.

Acknowledgments The author thanks Director, IITM, and Chief Program Scientist, CCCR-IITM, Pune for their valuable support. Also, thanks are due to IMD for providing high resolution century-long daily rainfall data.

References

Ashok K, Weng H, Behera S, Rao SA, Yamagata T (2007) El Niño Modoki and its teleconnection. *J Geophys Res* 112:C11007. doi: 10.1029/2006JC003798

Azad S, Vignesh TS, Narasimha R (2010) Periodicities in Indian monsoon rainfall over spectrally homogeneous regions. Int J Climatol 30:2289–2298

Donat MG, Alexander LV (2012) The shifting probability distribution of global daytime and night-time temperatures. Geophys Res Lett 39: L14707. doi:10.1029/2012GL052459

Gadgil S, Iyengar RN (1980) Cluster analysis of rainfall stations of the Indian peninsula. Q J Roy Met Soc 106:873–886

Gadgil S, Yadumani JNV (1993) Coherent rainfall zones of the Indian region. Int J Climatol 13:547–566

Hannachi A (2006) Quantifying changes and their uncertainties in probability distribution of climate variables using robust statistics. Clim. Dyn 27: 301–317.

Iyengar RN, Basak P (1994) Regionalization of Indian monsoon rainfall and long term variability signals. Int J Climatol 14:1095–1114

Kharin VV, Zwiers FW, Zhang GCHX (2007) Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. J Clim 20:1419–1444

Kripalani RH, Kulkarni A (1997) Climatic impact of el Nino/La Nina on the Indian monsoon: a new perspective. Weather 52:39–46

Krishna Kumar K, Rajgopalan B, Cane MK (1999) On the weakening relationship between the Indian monsoon and ENSO. Science 284: 2156–2159

Kulakarni A, Kripalani RH, Singh SV (1992) Classification of summer monsoon rainfall patterns over India. Int J Climatol 12:269–280

Kulkarni A, Kripalani RH (1998) Rainfall patterns over India: classification with fuzzy c-means method. Theor Appld Climtol 59:137–146

Loikith PC, Lintner BR, Kim J, Lee H, Neelin JD, Waliser DE (2013) Classifying reanalysis surface temperature probability density functions (PDFs) over North America with cluster analysis. Geophys Res Lett 40:3710–3714. doi:10.1002/grl.50688

Malik N, Marwan N, Kurths J (2010) Spatial structures and directionalities in monsoonal precipitation over South Asia Nonlinear Processes. Geophys 17:371–381

Matulla C, Edouard KP, Haas P, Formayer H (2003) Comparative analysis of spatial and seasonal variability: Austrian precipitation during the 20th century. Int J Climatol 23:1577–1588

Nicholson SE (1986) The spatial coherence of African rainfall anomalies: Interhemisphere telecommunications. J Climate Appl Meteor 25: 1365–1379

Parthasarathy B, Munot AA, Rupa Kumar K (1993) Homogeneous Indian monsoon rainfall: variability and prediction. Proc Ind Acad Sci (Earth and Planet Sci) 102:121–155

Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20:4356–4376

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

Rajeevan M, Bhate J, Kale JD, Lal B (2006) A high resolution daily gridded rainfall for Indian region: analysis of break and active monsoon spells. Current Science 91:296–306

Rajeevan M, Bhate J, Jaswal AK (2008) Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. Geophys Res Lett 35:L18707

Rao A R, Srinivas VV (2008) Regionalization of watersheds—an approach based on cluster analysis. Springer, 241 pp

Ruff TW, Neelin JD (2012) Long tails in regional surface temperature probability distributions with implications for extremes under global warming. Geophys Res Lett 39:L04704. doi:10.1029/2011GL050610

Saikranthi K, Rao TN, Rajeevan M (2012) Identifiaction and validation of homogeneous rainfall zones in India using correlation analysis. J Hydrometeorology 14:304–317

Satyanarayana P, Srinivas VV (2008) Regional frequency analysis of precipitation using large-scale atmospheric variables. J Geophys Res 113:D24110. doi:10.1029/2008JD010412

Satyanarayana P, Srinivas VV (2011) Regionalization of precipitation in data sparse areas using large scale atmospheric variables—a fuzzy clustering approach. J Hydrol 405:462–473

Shepard D (1968) A two-dimensional interpolation function for irregularly- spaced data Proceedings of the 1968 23rd ACM National Conference pp 517–524 Assoc. Comput. Mach., New York.

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

Singh KK, Singh SV (1996) Space-time variation and regionalization of seasonal and monthly summer monsoon rainfall of the sub-Himalayan region and Gangetic plains of India. Climate Res 6: 251–162

Trenberth KE et al (2007) Observations: surface and atmospheric climate change, in Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change, edited by S. Solomon et al., pp. 235–336, Cambridge Univ. Press, Cambridge, U. K.

Venkatesh B, Jose MK (2007) Identification of homogeneous rainfall regimes in parts of Western Ghats region of Karnataka. J Earth Syst Sci 116:321–329

View publication stats