Theoretical and Applied Climatology (2021) 146:1007–1018
https://doi.org/10.1007/s00704-021-03781-z

ORIGINAL PAPER

Recent changes in Indian monsoon in light of regionalization based on various rain features

Chandrani Chatterjee[1] · **Saurabh Das**[2]

Received: 10 June 2020 / Accepted: 12 September 2021 / Published online: 20 September 2021

Abstract
The effect of climate change on precipitation pattern is emerging out as a serious global challenge. However, the quantification and assessment of the climate change effect on precipitation are a challenging issue given the spatio-temporal inhomogeneity of rainfall. Previous studies mostly used only rain intensity information for the identification of climate change signature by selecting homogeneous rain regions. Rain climatology, however, has several other important characteristics like occurrence of multiple rain type, occurrence of extreme rain events and their spatial distribution. The present study provides a novel approach to combine satellite data with ground measurements for addressing this issue. In this study, the Indian subcontinent has been divided into homogeneous regions based on six different rain features, obtained from ground observations and TRMM satellite measurements of 12 years, using K-medoid clustering technique. The rain features showed distinguishable patterns in the resulting nine homogeneous rain zones and distinct behavioural patterns in the inter-relationships between the various rain features. To verify the consistency of the proposed regionalization, the recent changes in mean monsoon rain rates are then studied using 40 years of rain gauge measurements for these regions. Increasing trends in monsoon rain are observed in two homogeneous regions whereas five of them have shown decreasing signature of rain intensity. Although the present study focused only on monsoon rain over India, the results indicate the suitability of the proposed method as an efficient technique of identifying homogeneous region for other seasons and locations as well.

Keywords Regionalization · Climate change · K-medoid · TRMM data · Multiple rain types · Monsoon rain trend

1 Introduction

Climate change as well as its impact on human society is one of the major concerns in recent times. This has been evidenced through noticeable changes in almost all the important climate parameters globally. The increase in global temperature is one of the major reasons as well as pathways of climate change.

A rise of 0.4 °C in global surface temperature has been reported during the last century (Hansen et al. 2010). This surface temperature rise, in turn, affects the global precipitation cycle. According to Clausius-Clapeyron, each 1 °C increase in temperature causes 7% increase in atmospheric water holding capacity (Pall et al. 2007; Wang et al. 2017). The water cycle intensification is supposed to correspond directly with the temperature increase (Bosilovich et al. 2005), which results in the formation of severe storm systems. This increases the amount of precipitation over storm areas. The distant places, on the contrary, experience very low rain which makes the turbulent zones wetter and quite areas drier. Various climate models indicate that rain extremes may increase under global temperature change scenario (Asadieh and Krakauer 2015;Wentz et al. 2007). However, the extreme rainfall intensity and surface temperature relation is modulated by several other factors as well (Panthou et al. 2014).

Increasing surface temperature and resulting precipitation variabilities as well weather extremities are also reported by several authors over the Indian subcontinent (De et al. 2005; Mishra and Srinivasan 2013; Mishra et al. 2018; Carvalho and Wang 2019). Significant decrease in

✉ Saurabh Das
das.saurabh01@gmail.com; saurabh.das@iiti.ac.in

Chandrani Chatterjee
chandrani.chatterjee9@gmail.com

[1] Indian Institute of Technology, Indore, India

[2] Department of Astronomy, Astrophysics and Space Engineering, Indian Institute of Technology Indore, Simrol, Indore, Madhya Pradesh 453552, India

monsoon rainfall is reported in different parts of the subcontinent during the last century (Thomasa and Prasannakumarb 2016), whereas significant delay in monsoon onset is noticed over the northern part of the country (Lacombe and Mccartney 2014), especially during the latter half of the last century. Changes have also been observed in rainfall frequency during the last century. Falling trend in frequency of wet days has been observed over Central India. Many portions of Northern India showed a significance decrease in heavy rainfall days and peninsular India showed an increasing trend during the last century. One-day extreme rainfall events seemed to increase in some parts of India like coastal Andhra Pradesh, Saurasthra, Kutch, Orissa and West Bengal, whereas places like Chhattisgarh and Jharkhand have experienced a fall in the frequency of extreme rainfall. Besides, flood risks over regions like eastern coast, West Bengal and East Uttar Pradesh have increased to a noticeable extent (Guhathakurta et al. 2011). In a country like India, whose economy is largely dependent on the monsoon rainfall, any changes introduced in the characteristics of monsoon due to climate change will be definitely of great concern.

Study of climate change effect on precipitation, however, seeks serious consideration of spatial variation of rainfall as the local topography and associated rain climate vary greatly for different regions. Hence, proper methodology for identification of homogeneous rain regions is a vital step for exploring the changes in rain pattern. This is especially true for a country like India. Different regions of India seem to have huge diversity in precipitation amount and nature due to topography and monsoon circulation (Varikoden et al. 2012). Several researchers have tried to resolve this inhomogeneity by subdividing the country into different homogeneous rainfall regions based on different approaches. Some of the approaches are heuristic; some are based on statistical analysis whereas some of them deployed pattern recognition techniques for this purpose. The Indian Institute of Tropical Meteorology used rain intensity thresholding to regionalize India in six homogeneous monsoon zones based on century-long rain intensity data. Ahuja and Dhanya (2012) used cluster ensemble for finding homogeneous rainfall zones over India. They subdivided India into different homogeneous rain zones using the RCDA cluster ensemble algorithm, based on discriminant analysis of 53 years (1951-2003) of rain data. In some cases, correlation analysis have been used for identifying homogeneous zones across the country (Saikranthi et al. 2013a). The country has been divided into 26 homogeneous regions based on seasonal and annual rain data from 1025 rain measuring stations spread all over the country. These zones are reported to have homogeneity over rain frequency, rain type and temporal scale. The zones, however, were reported to have shown better coherency during excess rainfall years than during deficit rain years. Kulkarni (2012) have studied the rainfall probability of century-long daily rain data over India and clustered India in five homogeneous zones. The study also focused on the changes introduced in the homogeneous zones due to recent warming. The authors studied two time epochs: 1901–1975 and 1976–2004. The homogeneous regions obtained in two time periods are found to be notably different. The latter showed much more dispersion in homogeneous rain zones. Kakade and Kulkarni (2017) reported similar clustering results. They divided India into five homogeneous rainfall zones using shared nearest neighbour clustering technique on summer monsoon rainfall data. The study attempted monsoon homogeneity in India over rainfall subdivisions. Cluster regions are obtained by considering the points with high, medium and low connectivity strengths.

It should be noted that all the previous studies focused mostly on a single rain parameter, i.e. rain accumulation. This is primarily because rain gauge can provide information on rain accumulation and frequency but not on the different types of rain. However, climate change impact will not be limited to only the rain amount, but will also be on several other rain features, such as rain type, occurrence, and spatial structures. The novelty of this study is to redefine the homogeneous monsoon zones over the country using six different rain features simultaneously instead of only rain intensity. This is achieved by combining satellite data with rain gauge measurements and using a clustering algorithm. The TRMM satellite data is used for this purpose as it provides various rain features as well as the spatial variation of the rainfall. Dense network of rain gauges as well weather radar can also provide information of spatial rain structure; however, such setups are extremely rare in India. Machine learning technique, particularly K-medoid algorithm, is then used here in order to identify the homogeneous region based on the six rain parameters. Finally, the trends of the monsoon rainfall are investigated for these newly clustered regions. One should note that the proposed method is applicable to any region; however, for the present study, we focused only on the Indian region.

2 Data

Three different precipitation parameters, namely convective rain intensity, stratiform rain intensity and rain frequency, are taken from TRMM satellite. Data product (Version 7.0) 3A25 from TRMM satellite is used in the present study (TRMM 2011). The TRMM Precipitation Radar (PR) is an electronically scanning radar, with an operating frequency of 13.8 GHz, which measures the 3-D rainfall distribution. Unlike the ground-based measurements, the time duration of this TRMM satellite covers only one and half decade. Even though the data period is not long enough, the satellite measurements are advantageous since areal coverage can be obtained with simultaneously measured multiple rain parameters. We have selected 12 years' time period of 2003–2014 for our study because of the changes introduced in the resolution of data due to the orbital boost up (Jeremy and Kenneth 2007). In August 2001, the NASA changed the TRMM orbit to 402.5 km from 350 km. This height was chosen as it was the next feasible altitude at which PR would work, given the designed pulse repetition rate. As a result, the near-nadir field of view of the PR sensor changed to 5 km from 4.3 km. High-resolution (0.5° × 0.5°) 3A25 data product from PR sensor has been used here over India (5°–40° N, 68°–95° E).

The 3A25 data product provides various rainfall statistics over a month. The output includes various variables like rainfall rate (mm/hour) profile at 2, 4, 6, 10 and 15 km; fractional rain; snow ice layer; and surface rain rate (mm/hour). The output statistics includes probabilities of occurrence, means and standard deviations, histograms, and correlation coefficients. These statistics are available at two different spatial resolutions: (1) 5° × 5° (latitude × longitude) cells and (2) 0.5° × 0.5° cells. The convective/stratiform rain intensities are estimated from the mean of non-zero near-surface convective/stratiform rain rate. The rain fraction is the number of raining pixel detected divided by the total pixel seen in a grid by TRMM satellite. All the TRMM parameters used here are monthly mean statistics. The temporal resolution of TRMM is 16 orbits/day. It has a revisit time of 11–12 h.

The other three rain features used, i.e. average monsoon rainfall (mm/day), rainy days/year and extreme rainfall days/year (> 40 mm/day), are obtained from high-resolution (0.5° × 0.5°) daily rain gauge measurements from the CPC (Climate Prediction Center) Global Unified Precipitation data provided by the NOAA/OAR/ESRL PSL, Boulder, CO, USA, in their website at https://psl.noaa.gov/in. Here, the daily rain products are obtained from CPC data because TRMM is reported to show excellent performance at the annual and monthly scales but CPC data are reported to be the more accurate option for daily scale (Dinh et al. 2020).

To verify the consistency of the regionalization, the monsoon rain trend analysis has been done on a completely separate dataset, i.e. GPCP (Adler et.al. 2003) monthly precipitation data provided by the NOAA/OAR/ESRL PSL, Boulder, CO, USA (https://psl.noaa.gov/in). Mean monthly monsoon rainfall measurements during 1979–2018 are used for this purpose. The said dataset is developed by combining satellite data with rain gauge measurements and sounding observations.

3 Methodology

3.1 Selection of rain features for the study

Six different rain features have been used in this study. The principal features of rainfall lie in its intensity and frequency. Here, instead of only generalized rain intensity, rainfall amount along with rain intensities for two broad classes of rain (i.e. convective and stratiform) have been taken into consideration because of clearly separable nature of these two types of precipitation which have distinct contributions towards the rain climatology of a region. Secondly, the rain fraction over the subcontinent is taken into account in order to look at the frequency of rain.

In addition to these rain intensity and frequency features, two other rain characteristics describing the number of rainy days and extreme rainfall days are considered as well for this study to provide an outlook towards the recent picture of extreme weather over the country. However, instead of only monsoon period, data of the entire year are considered for these two rain features. This is because the sole idea behind choosing these two features was to visualize how the precipitation cycle over different parts of the country has been affected by climate change. It was not possible to achieve the same with only monsoon rain measurements but has an obvious implied effect on monsoon rain.

3.2 Method of analysis

All the six rain features, i.e. convective rain intensity (mm/h), stratiform rain intensity (mm/h), rain frequency and average rainfall amount (mm/day) of monsoon period (JJAS), and number of rainy days/year and number of extreme rainfall days/year (> 40 mm/day), are averaged over the time period of 2003–2014. This has resulted in six different features for every single grid across the country. The features are then normalized in order to reduce the computational disproportion. As machine learning techniques are efficient to handle multiple numbers of features simultaneously, the Indian subcontinent is divided into several homogeneous regions using K-medoid (Kaufman and Rousseeuw 1987) clustering based on the similarity of these six parameters.

K-medoid is a simple clustering method which selects K-medoids and tries to minimize the distance of each data point from the medoids by reassigning them to different clusters. Using medoid instead of mean makes a process statistically more robust. This allows K-medoid to handle noise and outlier better than K-means algorithm does. In the present study, the K-medoid operates on the 6-dimentional dataset and forms the cluster by minimizing the 6-D distance measure between the medoid and the each point on that cluster. The visualization of the clusters in 6-D is not possible;

however, each point belonging to different clusters can be identified and studied. The behaviour of the cluster points can be visualized on a 2-D plot between any two variables. This provides information how the inter-relation between any two parameters varies among different clusters. Another important factor in K-medoid algorithm is the selection of the number of clusters. The cluster number optimization is done using a silhouette plot as discussed in Sect. 4.1.

A trend analysis of recent monsoon rainfall is then performed using the GPCP dataover the proposed monsoon zones. Again, only south-west monsoon (June–September) period is considered here for this study because of its pronounced effect on Indian annual rain budget and agriculture.

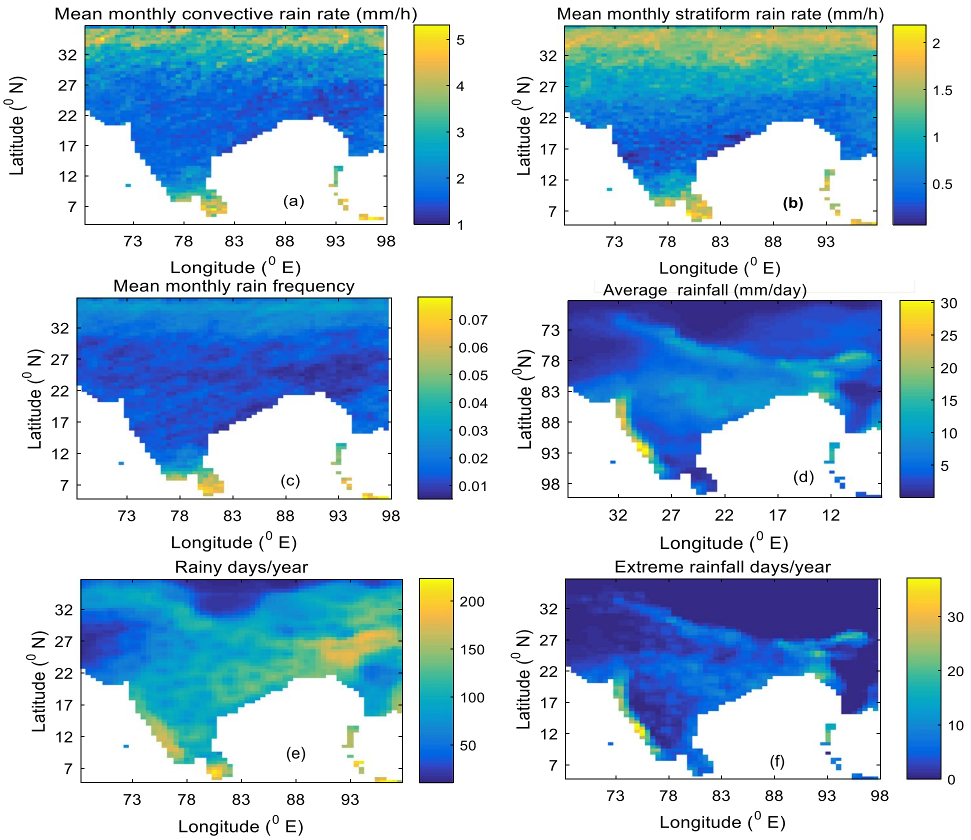
The non-parametric Mann–Kendall (MK) test is chosen for the trend analysis because of its ability to handle non-normal data distribution, unlike a parametric one. The ability of the MK test to handle outliers makes it suitable for trend analysis of climate variables.

4 Results

Rain has a noteworthy spatial variability over different parts of Indian subcontinent. Figure 1a–f shows the spatial variability of convective rain intensity, stratiform rain intensity, rain frequency and average rainfall amount of monsoon period (JJAS), number of rainy days/year and number of extreme rainfall days/year. The first three features are obtained from satellite data, whereas ground and satellite combined measurements are used for the subsequent three. As the time duration of the satellite observation is comparatively small, in the present study, covering only 12 years, such observation may not be useful for trend analysis but can indicate the rain climatology of a location. The spatial inhomogeneity of 12-year averaged rain characteristics based on two broad categories of rain, stratiform and convective, is shown in Fig. 1a–f.

If we look at the convective rain intensity (Fig. 1a), the northern part of the subcontinent shows the highest rain intensity and it decreases continuously as we move downwards. The nature of spatial change is more or less identical for stratiform rain intensity (Fig. 1b); however, the width of the similar intensity zones is totally different here. On the other hand, the frequency of rain is highest in the extreme northern and extreme southern parts of the subcontinent and decreases towards the middle part of the subcontinent (Fig. 1c). The rain amount seems to be significantly high along the Western Ghats mountain chain. The north eastern part of the country and Gangetic plains show a good amount of monsoon rainfall as well (Fig. 1d). The scenario changes completely if we focus on the number of rainy days/year. It can be observed from Fig. 1e that the north eastern part of

Fig. 1 Spatial variability of rain features over Indian subcontinent. (a) Mean monthly convective rain intensity (mm/h). (b) Mean monthly stratiform rain intensity(mm/h). (c) Mean monthly rain frequency during summer monsoon. (d) Average rainfall (mm/day) during summer monsoon. (e) Rainy days/year. (f) Extreme rainfall days/year



India gets the highest number of rainy days per year. The coastal parts of India seem to get a good number of rainy days, whereas the central part of India gets lesser number of rainy days. The number decreases significantly towards the desert, at the north western part of the country. The western coast of India gets the highest number of extreme rain fall days. The Indo-Gangetic Plain also shows values on the higher side for extreme rainfall days, whereas the other parts of the subcontinent seem to get lesser number of extreme rainfall days (Fig. 1f). All of these features show the different natures of variability in different parts of the subcontinent. Therefore, it is imperative to regionalize Indian subcontinent based on various rain features simultaneously in order to better understand the climate change effect on the monsoon rain. Accordingly, regionalization of monsoon rain zones over Indian subcontinent is opted using a clustering technique based on the six above-mentioned rain features.

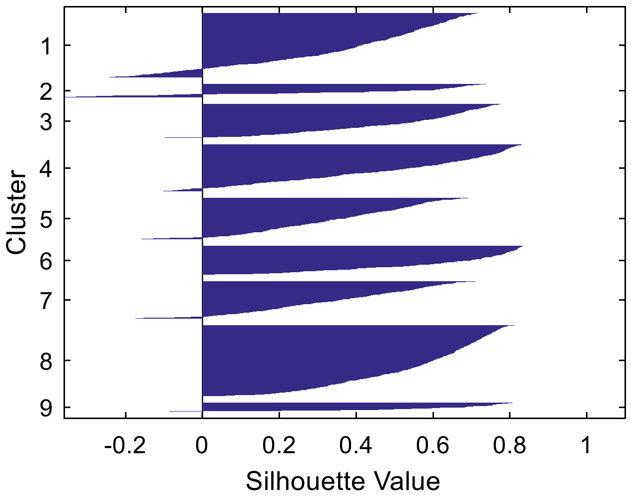


Fig. 2 Performance analysis of clustering using silhouette plot

4.1 Clustering performance analysis

The K-medoids clustering has divided the landmass of the Indian subcontinent into nine separable homogeneous regions based on the six rain features. The number of clusters is verified by a silhouette plot. A silhouette plot is a representation of how well the objects are clustered. A silhouette value can range between − 1 and + 1, and a positive silhouette value indicates that the object is well-matched to its assigned clustered, whereas a negative value represents a poorly matched object. The silhouette coefficient is calculated by two factors, namely the mean intra-cluster distance ‘*a*’ and the mean nearest-cluster distance (distance between a sample and the nearest cluster other than its own cluster) ‘*b*’ for each sample. The silhouette coefficient for a sample is given by

$$S(i) = (b - a)/\max(b, a) \qquad (1)$$

The mean silhouette value for all points in a specific cluster provides how tightly that cluster points are grouped. Therefore, it gives a measure of how well the data is clustered if we take a mean over the entire dataset. So, if too many or too few clusters are chosen, then the silhouette figure will show narrower columns for some clusters in comparison to others. Here, Fig. 2 indicates a positive silhouette value for 2289 sample points among 2420 points, which confirms consistency of the clusters.

In order to understand the separability among the nine clusters, the distribution of each of the rain features is then studied for different clusters. Figure 3a–f shows the fitted histograms for all the rain features. It is evident that the rain features are clearly separable in various clusters. As an example, the convective rain intensity (Fig. 3a) has the steepest peak towards a lower range in cluster 8 whereas it has a much wider spread in cluster 6. In cluster 4 and cluster 3, it has clear dominance of low and high values respectively. Similarly, the distribution of stratiform rain intensity (Fig. 3b) has a clear gradual right shift starting from cluster 8 while moving towards clusters 4, 5, 7, 6 and 3.

Next, the association among different rain features is investigated. The results indicate that the association between different features also varies in different parts of the subcontinent. Figure 4a–i shows some of the inter-relations between features in different clusters. For example, if we focus on the relationship between convective rain intensity and rainy days/year (Fig. 4a), then we can see that convective rain intensity does not vary much with rainy days/year in cluster 6, whereas it has almost a linear relationship in cluster 9. On the other hand, rain frequency and rainy days/year (Fig. 4c) has a negative correlation in cluster 8 but it has a positive correspondence in cluster 3. Stratiform and convective rain intensities seemed to have a strong linear dependence on each other in cluster 7 but they are merely relatable in cluster 1 (Fig. 4d). These clear variations in inter-relationships between the features, in the different clusters, indicate the presence of different rain climatologies over different regions of the subcontinent.

4.2 Homogeneous rain zones over Indian subcontinent

The proposed regionalization has clustered India in nine distinct homogeneous monsoon zones, and the locations of the regions are shown in Fig. 5. The different colours indicate different clusters numbered as C1, C2 etc. for the sake of readability. The data points of a single cluster are found to occupy more or less the same neighbourhood. However, it is also to be noted that some of the clusters, such as C5 and C3, appear at two disjoint regions. This simply indicates similar

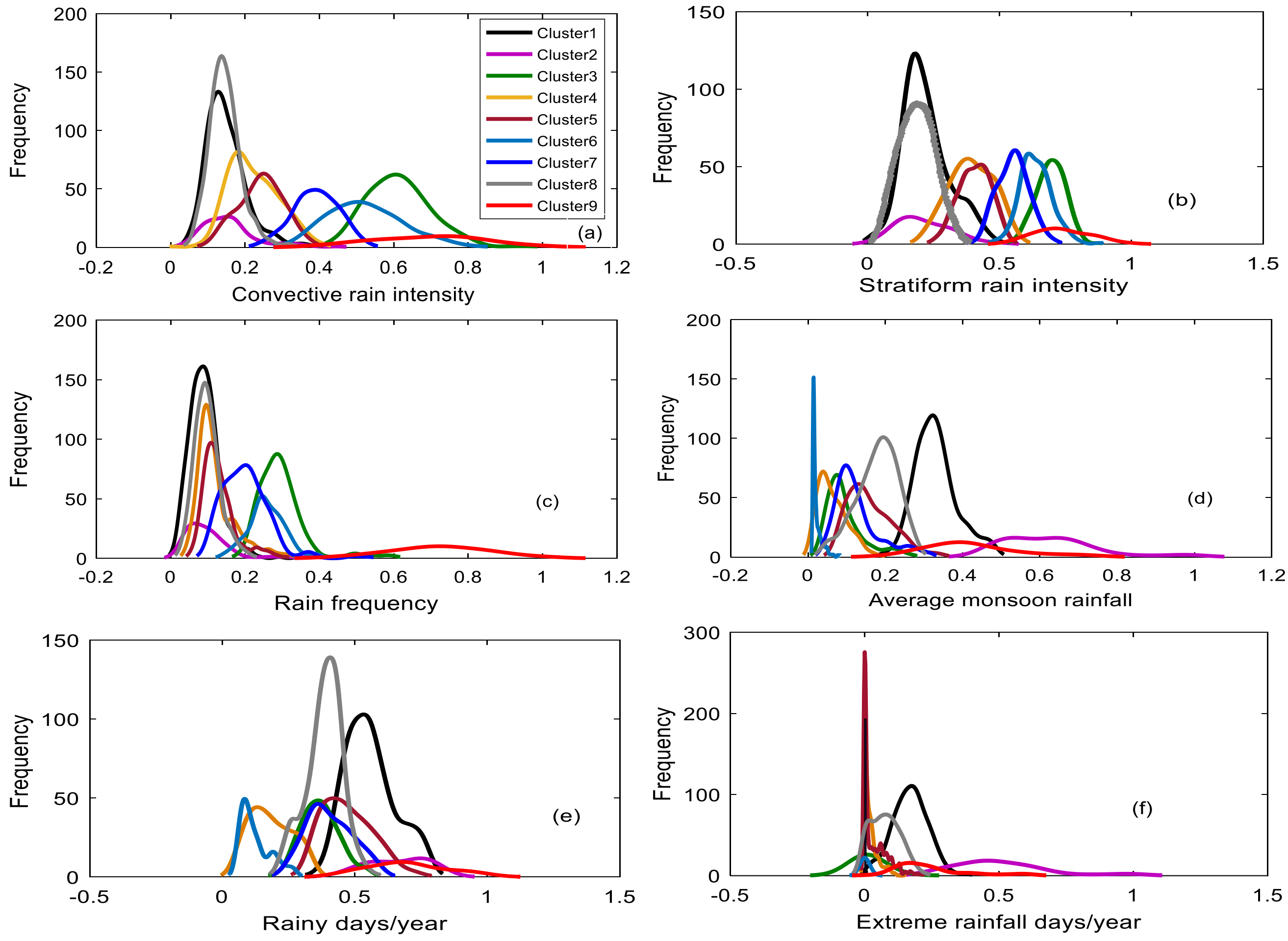


Fig. 3 The distribution of clustered data points for the studied features (in normalized unit). (a) Mean monthly convective rain intensity. (b) Mean monthly stratiform rain intensity. (c) Mean monthly rain frequency. (d) Average rainfall during summer monsoon. (e) Rainy days/year. (f) Extreme rainfall days/year

rain climatology at two different locations treated as part of the same homogeneous region for further analysis.

4.3 Study of recent rain trend over newly formed rain zones

The homogeneous regions are formed using the satellite and rain gauge data of 12 years. However, to understand the consistency of the proposed regionalization, monsoon rain trend is studied for each of these said zones. Since satellite data are of limited temporal coverage, the GPCP monthly precipitation data of 40 years (1979–2018) has been used for this purpose. Each of these proposed zones is studied separately to investigate if there is any noticeable trend in monsoon rain over the time and whether it is consistent with the previous studies (Fig. 6a–i).

4.3.1 Region C1

This rain zone is primarily spread across the eastern coast of India, i.e. along the Bay of Bengal. The trend of monsoon rain here is shown in Fig. 6i. A decreasing rate of 0.036mm/year is observed in average monsoon rain intensity. The result of the Mann–Kendall trend analysis presented in Table 1 also confirmed this signature. The H and p values observed are 1 and 0.012 respectively. This indicates the presence of a monotonic trend. The nature of this trend is confirmed as a downward one with the negative value of Mann–Kendall statistics Z_{MK} (−2.645). The state-wise trend reported by Kumar et al. (2010) also supports this nature of rainfall trend.

4.3.2 Region C2

This monsoon zone mainly covers the western coast of India, the coastal part of Myanmar and a small portion of Indian

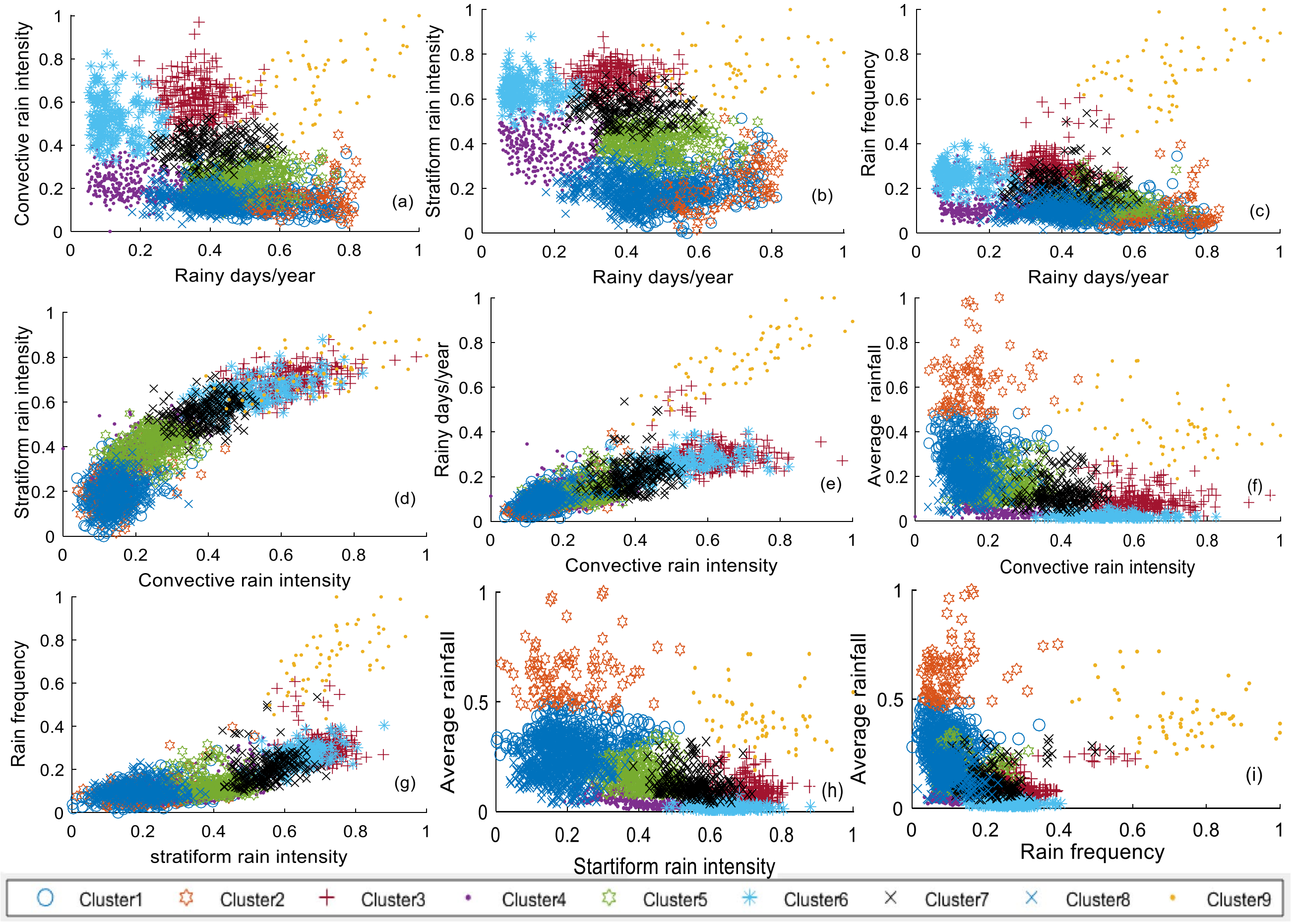


Fig. 4 The inter-relationships between rain features in different clusters

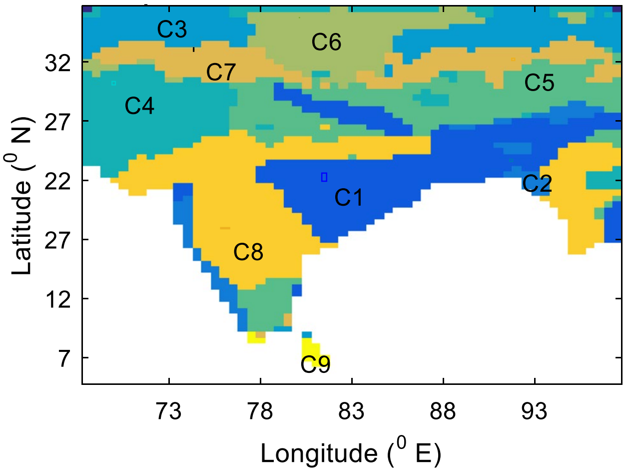


Fig. 5 Proposed regionalization

state Assam. The monsoon rain intensity of this zone during the last 40 years is shown in Fig. 6b. Mean monsoon rain intensity shows a steep fall over time at a rate of 0.056 mm/year. This trend is further validated with the Mann–Kendall trend analysis test (Table 1). The trend test rejects ($H = 1$) the null hypothesis of no trend with a p value of 0.046. The result implies a monotonic trend in the data. This region has showed a Mann–Kendall statistics Z_{MK} value of − 1.990. Z_{MK} reveals an important downward trend in rainfall during the last 40 years over this particular rain zone. The trend observed here finds good agreement with previously reported 100 years rain trend over western coastal part of India (Varikoden et al. 2013).

4.3.3 Region C3

This rain zone mostly covers two major parts in Himalayan belt, separately. In one part, it includes a portion of Afghanistan and some portion of upper Jammu and Kashmir State of India. In other part, a portion of China shows similar rain climatology. No strong rain trend (Table 1) in precipitation is observed in these regions (Fig. 6c).

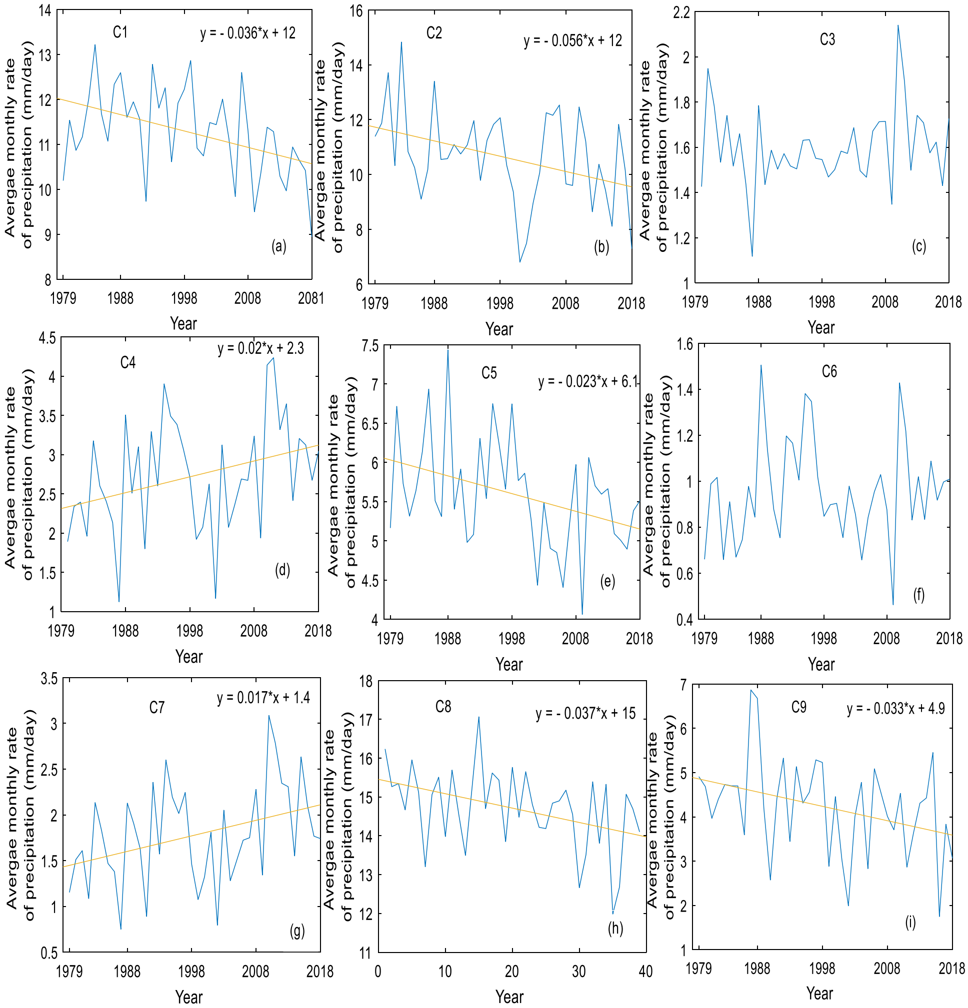


Fig. 6 Variation of mean monthly monsoon rain over the years in newly proposed homogeneous regions

Table 1 Trend of monsoon rainfall over the years in proposed rain zones with Mann–Kendall trend analysis results

Homogeneous Monsoon zone	Result of Mann-Kendall trend analysis			Conclusion drawn from the test on the trend of Monsoon rain over last 40 years
	H	*P*	*Z*	
	H	P	Z	
C1	1	0.012	-2.645	decreasing
C2	1	0.046	-1.990	decreasing
C3	0	0.380	0.870	No trend
C4	1	0.039	2.060	increasing
C5	1	0.031	-2.150	decreasing
C6	0	0.450	0.750	no trend
C7	1	0.043	2.015	increasing
C8	1	0.017	-2.395	decreasing
C9	1	0.030	-2.106	decreasing

4.3.4 Region C4

This region is located in the north west portion of India and mainly covers the Great Indian Desert along with a small portion of Pakistan. Figure 6d shows the monsoon rain intensity over this region during the last 40 years. The intensity plot points out an increasing trend of average monsoon rain intensity with a rate of 0.020 mm/

year. The Mann–Kendall trend analysis test (Table 1) is performed in order to verify the trend. The H value of test is 1, i.e. the trend test rejects the null hypothesis of no monotonic trend. The p value is obtained as 0.039. This indicates the presence of a monotonic trend. The high positive value (2.060) of Mann–Kendall statistics Z_{MK} indicates a significant increasing trend of rainfall. Climate change seemed to increase mean monsoon precipitation in the Indian desert. Similar results are also obtained by Kumar et al. (2010) which showed a significant increase in rain amount in this region during the last century using subdivisional rain data prepared by the Indian Institute of Tropical Meteorology.

4.3.5 Region C5

This homogeneous monsoon zone consists of a part of the Indo-Gangetic plains (IGP), some portion of north eastern India, a small portion of South India and a small portion of Southern China. The mean monsoon rain intensity shows a decreasing rate of 0.023 mm/year as shown in Fig. 6e. The Mann–Kendall trend analysis test (shown in Table 1) is performed on the data to confirm the trend. The H value of test is 1 and the p value of the test is 0.031. This indicates successful rejection of no monotonic trend assumption. The Mann–Kendall statistics Z_{MK} showed a significant negative value (-2.150) which concludes in favour of a falling rainfall trend in this region over time. Zhai et al. (2004) have shown similar rain trend over southern China.

The precipitation characteristics of IGP have been studied exhaustively by many researchers and the present study is consistent with the earlier reports. It is to be noted that it is observed in several studies that precipitation over IGP has a strong dependence on Northern sea oscillations and aerosol climatology and therefore possesses several periodicities. However, many of the studies are able to detect the weakening trend of the precipitation irrespective of these short-term trends. The weakening of monsoon rain in IGP is also reported previously by Bhatla et al. (2015).

4.3.6 Region C6

The eastern part of Jammu and Kashmir and a small part of western China come under this rain cluster. The region showed a weak increasing trend of rainfall; however, the trend observed here is not significant (Fig. 6f and Table 1).

4.3.7 Region C7

This rain zone covers the small part of Pakistan; Indian states Punjab, Himachal Pradesh and Uttarakhand (partially); and some parts of China. The average monsoon rain intensity over this region (Fig. 6g) shows an increase rate of 0.017 mm/year. The result of the Mann–Kendall trend test shown in Table 1 suggests that the null hypothesis of no trend is rejected with a p value of 0.043. A Z_{MK} value of 2.015 reflects a rising trend of rainfall in this region over the years. Similar trend is observed over Punjab by Krishan et al. (2015). Randhawa et al. (2015) have also performed a seasonal rainfall trend analysis and reported increase in monsoon rainfall over Himachal Pradesh.

4.3.8 Region C8

This region mostly covers a major portion of Central India. The monsoon rain intensity over this region during the last 40 years is shown in Fig. 6h. The time series of mean monsoon rain intensity clearly indicates a decreasing rate of 0.037 mm/year over the region and this trend is validated by the Mann–Kendall trend analysis test (Table 1). The trend test rejects the null hypothesis of no trend with a p value of 0.017 which is much lower than α (0.05). Hence, the test suggests the presence of a monotonic trend over the observation period. The Mann–Kendall statistics Z_{MK} is obtained as -2.395. This significant negative value of Z_{MK} supports a downward trend of rainfall over this region during the last 40 years. Varikoden et al. (2013) have also reported similar trend of rainfall over Central Indian landmasses using gridded rainfall data from IITM.

4.3.9 Region C9

The C3 rain cluster comprises the southernmost tip of the country along with the neighbouring island nation, Sri Lanka. This small region has shown a very significant decreasing trend (Fig. 6i) in monsoon rain intensity with a decreasing rate of 0.033 mm/year which is confirmed by the Mann–Kendall analysis (Table 1). The value of p (0.030) clearly depicts the successful rejection of no trend hypothesis. The Mann–Kendall coefficient is obtained as -2.106. This negative value proves the presence of a falling rain trend. The result finds good agreement with the result obtained by Jayawardene et al. (2005) which have also reported similar decreasing rainfall trends for 13 rain gauge stations in Sri Lanka.

4.4 Consistency and advantages of the proposed regionalization with respect to previous studies

Summer monsoon over the Indian subcontinent is principally a thermally driven process. High positive correlation is observed between surface temperature and monsoon rainfall over the north west and central parts of India (Parthasarathy et al. 1990). Significant impact of SST over the Arabian Sea and Bay of Bengal on Indian monsoon has been reported by several authors (Shukla 1975; Shukla and Mooley 1987; Kothawale et al. 2008). However, the Indian Ocean SST seemed to have a contrary impact on different parts of the Indian subcontinent. ENSO seemed to have a negative correlation with the Indian summer monsoon rainfall (Wang et al. 2006), whereas the formation of monsoon depression associated with occurrence of heavy rainfall is likely during El Nino years (Singh et al. 2000). The European surface temperature also supposed to have an impact on Indian monsoon (Hahn and Shukla 1976). Even though there are some common atmospheric factors affecting monsoon rainfall over the country, the variability of summer monsoon rainfall over different parts of the country has different sets of atmospheric variables to account for (Vecchi and Harrisson 2004). Therefore, proper identification of the set of influencing parameters requires identification of homogeneous rain zones. The present study is focused primarily on defining homogeneous regions having similar rain characteristics. The causal relation between monsoon rain of these regions and the influencing factors will be the subject of future investigation.

Rain regionalization in India has started with a geographic or political point of understanding. Connected regions in a broad scale were grouped as a rain zone. Based on 306 rain gauges spread all across the country, 36 meteorological subdivisions (MSD) have been formed which were initially considered as homogeneous because of their large numbers. However, negative correlations are reported between rainfalls of various MSDs (Parthasarathy 1984). In view, of these shortcomings, the IMD (Indian Meteorological Department) re-classified India in six homogeneous monsoon rainfall zones (north west, west central, north east, peninsular and central north east) based on their contribution in annual rainfall (Parthasarathy et al. 1995). Recent studies revealed heterogeneity of summer monsoon rainfall distribution in these regions (Satyanarayana and Srinivas 2008). Inter-site correlation analysis of rain gauge stations was adopted on seasonal and monthly scales to regionalize India in 31 homogeneous rainfall regions (Gadgil and Yadumani 1993). The study used a rain dataset of 25 years from 200 IMD stations. However, Satyanarayana (2009) reported seventeen of these regions as heterogeneous in terms of monsoon rainfall frequency distribution based on $1^0 \times 1^0$ gridded data provided by IMD. A subsequent study, however, identified 26 homogeneous rain regions with a 50-year-long dataset (1951–2000) from 1025 rain stations over India (Saikranthi et. al. 2013b). The main issue with the approach of correlation analysis lies in selection of the threshold value which changes the regionalization. Another very popular technique of rain regionalization involves PCA which has been successfully applied to various countries like the USA, Italy and Spain. However, the incredible complexity in spatial distribution of Indian rainfall makes it tough for first few leading PCs to account for significant percentage of the total variance in monsoon season (Iyengar and Basak 1994; Singh and Singh 1996). Conventional statistical analyses were replaced soon by emerging machine learning techniques. Several researchers have applied advanced clustering techniques for regionalizing Indian summer monsoon into more logical and stable homogeneous regions. However, all these regionalization approaches are not feasible for areas having poor numbers of rain gauges. So, a different pathway was formed to account for such regions which include other long-term atmospheric variables for regionalization. This approach using reanalysis gridded data divided India into seventeen homogeneous rainfall zones using K-means clustering method (Satyanarayana and Srinivas 2011).

The rain regionalization so far was principally based on rain gauge data with a single rain feature of rain intensity. Rain climatologies having similar rain intensities may have distinct rain features in terms of rain frequency and different rain categories. The current study has focused on six different rain features including extreme rainy days' number which finds its application in defining different climatologies in terms of several rain features along with the extreme weather conditions. Nine distinct rain zones have been formed after clustering the region using these rain features (Fig. 2). C2 and C5 were previously clustered as a single homogeneous region based on rain intensity (Kakade and Kulkarni 2017) but it can be noted that the rain frequency is significantly different for these two regions. Hence, considering all the rain features, logically, it should be clustered into two separate rain climatologies. Major portions of C7, C3 and C6 were also clustered as a single rain zone due to similar reasons. Most of the regions in C5 and C1 zones clustered as a single rain zone due to similar rain amount (Ahuja and Dhanya 2012) but interestingly, consideration of rain types divides them into two separate zones as shown in Fig. 5. Central Indian regions are observed to have the same rain climatology (C8) which was previously seemed to be two different zones. Mannan et al. (2018) have also identified the C2 region (Western Ghats) having different rain climatologies. C4, some portion of C5 and C8 were clustered as a rain zone by the authors. Similarly, some portion of C5, a part of C8 and C1 are clustered as a same rain zone. The present study has shown that these regions have distinction in rain frequency and convective and stratiform rain features in spite of having mostly similar rain intensity.

Rain climatology over a region depends on several parameters like rainfall statistics, topography, cloud type and rain formation process. Inclusion of different rain features related to types of rain and precipitation cycle is therefore advantageous in defining rain climatologies which can in turn lead to successful measurements, understanding and prediction of rain distribution over the region in future. This method, therefore, can be useful in successful modelling of rain over the subcontinent. The proposed method, however, can be applied to any other region in the globe.

5 Conclusion and future scope

Precipitation is an inhomogeneous process both in space and in time. It is influenced by several weather parameters and topography. Identification of homogeneous rain region is, therefore, important to study and quantify the climate change signature. The present study proposes a novel rain regionalization approach based on multiple rain features obtained from satellite and ground-based measurements. The K-medoid algorithm is used here to handle the multiple rain features. This resulted in nine homogeneous monsoon regions which show distinct behaviour as well varied inter-relationship between different rain features. The novelty of the present study is that the proposed technique provides a more compact clustering of different regions based on multiple rain features.

The study has found decreasing trend over five newly identified homogeneous regions out of nine homogeneous regions formed. The trend test results are significant in all these regions. Two rain zones have shown strong upward trends of monsoon rain. The results are in accordance with previously published results from these locations. The rain trends observed over newly formed monsoon zones indicate that majority of Indian regions are vulnerable under climate change. Even though this study focuses on Indian subcontinent, the use of satellite data for the regionalization makes this approach suitable for global rain climatologies. It should be noted that the homogeneous region will change if there is a substantial change in climate. The use of averaged value for a longer period may minimize the year-to-year variability, but long-term change will change the homogeneous regions. As the newly defined homogeneous regions are having distinct rain features, the causal relation between monsoon rain and other meteorological parameters can be explored separately for these locations. We expect this will help in better quantification of climate change effect on these regions and will be attempted in future.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00704-021-03781-z.

Acknowledgements The authors acknowledge the official websites (https://www.esrl.noaa.gov/psd/; https://www.esrl.noaa.gov) of the National Aeronautics and Space Administration (NASA) and National Oceanic and Atmospheric Administration (NOAA) for providing access to the rainfall datasets.

Author contribution Both the authors contributed to the study and have read and approved the final manuscript.

Funding This work received financial support from the Indian Space Research Organisation and Department of Science and Technology (DST), India.

Data availability Publicly available datasets are used for this article from official websites of the NASA and NOAA and the links are properly mentioned and cited.

Code availability Not applicable.

Declarations

Ethics approval The study is an atmospheric study and does not deal with any living being.

Consent to participate The study is based on publicly available data on atmospheric parameters and does not include any individual data or image.

Consent for publication The study has been performed on properly cited publicly available database and there are no restrictions in publishing the results.

Conflict of interest The authors declare no competing interests.

References

Ahuja S, Dhanya C (2012) Regionalization of rainfall using RCDA cluster ensemble algorithm in India. J Softw Eng Appl 5:568–573

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P (2003) The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present*)*. J Hydrometeor 4:1147–1167

Asadieh B, Krakauer NY (2015) Global trends in extreme precipitation: climate models versus observations. Hydrol Earth Syst Sci 19:877–891

Bhatla R, Singh M, Mall RK, Tripathi A, Raju PVS (2015) Variability of summer monsoon rainfall over Indo-Gangetic plains in relation to El-Nino/La-Nina. Nat Hazards 78:837–853

Bosilovich MG, Schubert SD, Walker GK (2005) Global changes of the water cycle intensity. Am Meteorol Soc 18:1591–1608

Carvalho KS, Wang S (2019) Characterizing the Indian Ocean sea level changes and potential coastal flooding impacts under global warming. J Hydrol 569:373–386

De U, Dube R, Prakasa G (2005) Extreme weather events over India in the last 100 years. J Ind Geophys Union 9:173–187

Dinh KD, Anh TN, Nguyen NY, Bui DD, Srinivasan R (2020) Evaluation of grid-based rainfall products and water balances over the mekong river basin. Remote Sens 12(11):1858. https://doi.org/10.3390/rs12111858

Guhathakurta P, Sreejith OP, Menon PA (2011) Impact of climate change on extreme rainfall events and flood risk in India. J Earth Syst Sci 120:359–373

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

Hahn DG, Shukla J (1976) An apparent relationship between Eurasian snow cover and Indian monsoon rainfall. J Atmos Sci 33:2461–2462

Hansen J, Ruedy R, Sato M, Lo K (2010) Global surface temperature change. Rev Geophys 48:RG4004. https://doi.org/10.1029/2010RG000345

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

Jeremy DD, Kenneth PB (2007) Changes in TRMM rainfall due to the orbit boost estimated from buoy rain gauge data. J Atmos Oceanic Tech 24:1598–1607

Jayawardene H, Sonnadara D, Jayewardene DR (2005) Trends of rainfall in Sri Lanka over the last century. Sri Lankan J Phys 6:7–17

Kulkarni A (2012) Weakening of Indian summer monsoon rainfall in warming environment. Theor Appl Climatol 109:447–459

Krishan G, Chandniha SK, Lohani AK (2015) Rainfall trend analysis of Punjab, India using statistical non-parametric test. Curr World Environ 10:792–800

Kumar V, Jain S, Singh Y (2010) Analysis of long-term rainfall trends in India. Hydrol Sci 55:484–496

Kakade S, Kulkarni A (2017) Seasonal prediction of summer monsoon rainfall over cluster regions of India. J Earth Syst Sci 126:34

Kothawale DR, Munot AA, Borgaonkar HP (2008) Temperature variability over the Indian Ocean and its relationship with Indian summer monsoon rainfall. Theoret Appl Climatol 92:31–45

Kaufman L, Rousseeuw, P (1987). Clustering by means of medoids. In: IDY and editor (ed.), North Holland / Elsevier, pp 405–416 .

Lacombe G, McCartney M (2014) Uncovering consistencies in Indian rainfall trends observed over the last half century. Clim Change 123:287–299

Mannan A, Chaudhary S, Dhanya CT, Swamy AK (2018) Regionalization of rainfall characteristics in India incorporating climatic variables and using self-organizing maps. ISH J Hydraul Eng 24:147–156

Mishra AK, Nagaraju V, Rafiq M, Chandra S (2018) Evidence of linkage of regional climate change with precipitation extremes over India. Weather 74(6). https://doi.org/10.1002/wea.3259

Mishra AK, Srinivasan J (2013) Did a cloud burst occur in Kedarnath during 16 and 17 June 2013? Curr Sci 105:1351–1352

Pall P, Allen MR, Stone DA (2007) Testing the Clausius-Clapeyron constraint on changes in extreme precipitation under CO2 warming. Clim Dyn 28:351–363. https://doi.org/10.1007/s00382-006-0180-2

Parthasarathy B (1984) Interannual and long-term variability of Indian summer monsoon rainfall. Proc. Indian Acad. Sci. (Earth Planet Sci) 93(4):371–385

Parthasarathy B, Kumar KR, Sontakke NA (1990) Surface and upper air temperatures over india in relation to monsoon rainfall. Theor Appl Climatol 42:93–110

Parthasarathy B, Munot AA, Kothwale DR (1995) Monthly and seasonal rainfall series for all-India homogeneous regions and meteorological subdivision: 1871–1994. Research Report No. RR-065, Indian Institute of Tropical Meterology, Pune. Available online https://www.tropmet.res.in/~lip/Publication/RR-pdf/RR-65.pdf

Panthou G, Mailhot A, Laurence E, Talbot G (2014) Relationship between surface temperature and extreme rainfalls: a multi-timescale and event-based analysis. J Hydrometeorol 15:1999–2011

Randhawa SS, Randhawa S, Rai I (2015) Seasonal monthly and annual rainfall trends in Himachal Pradesh during 1901–2002. State Centre On Climate Change (Himachal Pradesh). Available online: http://www.hpccc.gov.in/documents/Final%20Rainfall%20Document%20-%20Copy.pdf

Shukla J (1975) Effect of Arabian Sea surface temperature anomaly on Indian summer monsoon: a numerical experiment with the GFDL model. J Atmos Sci 32(3):503–511

Satyanarayana P (2009) Regional frequency analysis of hydrometeorological events—an approach based on climate information. PhD Thesis, Indian Institute of Science, Bangalore. Available online: https://etd.iisc.ac.in/

Shukla J, Mooley DA (1987) Empirical prediction of the summer monsoon rainfall over India. Mon Weather Rev 115(3):695–703

Singh O, Masood T, Khan A, Rahman S (2000) Impact of Southern Oscillation on the frequency of monsoon depressions in the Bay of Bengal. Nat Hazards 25:101–115

Saikranthi K, Rao TN, Rajeevan M (2013a) Identification and validation of homogeneous rainfall zones in India using correlation analysis. J Hydrometeorol 14:304–317

Saikranthi K, Rao T, Rajeevan M, Rao S (2013b) Identification and validation of homogeneous rainfall zones in India using correlation analysis. J Hydrometeorol 14:304–317. https://doi.org/10.1175/JHM-D-12-071.1

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(3):251–262

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

Satyanarayana P, Srinivas VV (2011) Regionalization of precipitation in data sparse areas using large scale atmospheric variables—a fuzzy clustering approach. J Hydrol 405(3–4):462–473

Tropical Rainfall Measuring Mission (TRMM) (2011) TRMM Radar Rainfall Statistics L3 1 month (5 x 5) and (0.5 x 0.5) degree V7, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: January, 2020. https://disc.gsfc.nasa.gov/datacollection/TRMM_3A25_7.html

Thomasa J, Prasannakumarb V (2016) Temporal analysis of rainfall (1871–2012) and drought characteristics over a tropical monsoon-dominated State (Kerala) of India. J Hydrol 534:266–280

Wang X, Li C, Zhou W (2006) Interdecadal variation of the relationship between Indian rainfall and SSTA modes in the Indian Ocean. Int J Climatol 26(5):595–606

Wentz FJ, Ricciardulli L, Hilburn K, Mears C (2007) How much more rain will global warming bring? Science 317(5835):233–235. https://doi.org/10.1126/science.1140746

Wang G, Wang D, Trenberth KE, Erfanian AYM, Bosilovich MG, Parr DT (2017) The peak structure and future changes of the relationships between extreme precipitation and temperature. Nat Clim Chang 7(4):268–274. https://doi.org/10.1038/nclimate3239

Vecchi G, Harisson D (2004) Interannual Indian rainfall variability and Indian Ocean sea surface temperature anomalies. Am Geo Union 147:247–260

Varikoden H, Kumar KK, Babu CA (2013) Long term trends of seasonal and monthly rainfall in different intensity ranges over Indian subcontinent. Mausam 64:481–488

Varikoden H, Preethi B, Revadekar JV (2012) Diurnal and spatial variation of Indian summer monsoon rainfall using tropical rainfall measuring mission rain rate. J Hydrol 475:248–258

Zhai P, Zhang X, Wan H, Pan X (2004) Trends in total precipitation and frequency of daily precipitation extremes over China. J Clim 18:1096–1107