



# Spatial homogeneity of extreme precipitation indices using fuzzy clustering over northeast India

Manish Kumar Goyal<sup>1,2</sup> · Gupta Shivam<sup>1</sup> · Arup K. Sarma<sup>1</sup>

Received: 15 October 2016 / Accepted: 1 August 2019 / Published online: 13 August 2019  
© Springer Nature B.V. 2019

## Abstract

Regionalization on the basis of the properties of hydro-meteorological data helps in identifying the regions reflecting the similar characteristics which could be useful in designing hydrological structures as well as planning and management of water resources of the region. In this study, rainfall data of northeast India were utilized for calculation of extreme precipitation indices as suggested by expert team on climate change detection and monitoring. Trend analysis of the indices was carried out using Mann–Kendall nonparametric test, and Sen's slope estimator was used for calculating the magnitude of trend. Further fuzzy c-means method was used for clustering of the selected stations on the basis of six parameters of all these precipitation indices, i.e., latitude, longitude, mean, standard deviation, minimum value and maximum value. Three cluster validity indices, namely fuzzy performance index, modified partition entropy and cluster separation index were used for selecting the optimum cluster numbers. Analysis shows insignificant trend for the indices like consecutive dry days and consecutive wet days, whereas maximum 1-day precipitation (R1 day) and maximum 5-day precipitation (R5 day) are not showing any clear trend. It is observed that the number of rainy days is decreasing followed by increasing 1-day precipitation. Cluster analysis of the precipitation indices shows five major clusters for most of the indices.

**Keywords** Precipitation indices · Extreme events · ETCCDI · Fuzzy clustering · Fuzzy c-means

## 1 Introduction

Changes in the occurrence and magnitude of climatic parameter because of the global warming are very obvious and are being observed (IPCC 2007), but extreme events like heavy rainfall days, dry spells and wet spells were not considered for the climate change studies and in the recent years, some extreme events in different parts of the world forced

---

✉ Manish Kumar Goyal  
vipmkgoyal@gmail.com

<sup>1</sup> Department of Civil Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, India

<sup>2</sup> Discipline of Civil Engineering, Indian Institute of Technology Indore, Indore 453552, India

the climate change researchers to study about these extreme events. Expert team on climate change detection and monitoring (ETCCDI) has worked on extreme precipitation events and suggested some rainfall and temperature indices to monitor these extreme events (Zhang and Yang 2004). Manton et al. (2001) studied the trend in extreme daily rainfall and temperature in South East Asia for the period of 1961–1998 for 13 stations and found significant increasing trend at some stations for very wet days while decreasing trend observed in rainy days at some stations. Significant increase in very hot days and decrease in cold days were also observed for the different parts of the South East Asia. De Lima et al. (2014) selected 57 stations of mainland Portugal (Southwestern Europe) and precipitation data of 67 years were used to investigate the intensity and frequency of extreme precipitation events, study revealed that the consecutive dry day spells are increasing significantly at 17% of the total stations while decreasing trend was observed for 1-day and 5-day maximum rainfall. Heavy precipitation days and very heavy precipitation day indices were also representing a significant decreasing trend at several stations. Abbas et al. (2014) carried out a study in the semiarid and sub-humid region of Pakistan for identifying the pattern of extreme precipitation indices at five stations of Punjab, and it was observed that no significant temporal trend is found for annual precipitation but the spatial trend was observed according to variation in elevation. Roy and Balling (2004) conducted study on 129 stations across India to analyze the precipitation extreme events trend and concluded that Himalayan region and Deccan Plateau show increasing trends in precipitation extreme events, whereas parts of Gangetic Plain show decreasing trends. Goswami et al. (2006) found significant increase in heavy precipitation events over central India in last century.

Several studies related to rainfall, temperature, evapotranspiration and its temporal and spatial variation in northeast India have been carried out by several researchers, but main focus was on the recognition of trend and its variability for these climatic parameters and to the best of authors knowledge, no study of regionalization on the basis of precipitation indices has been carried out for this region. Goyal (2014) reported based on the study carried out for 68 stations of Assam that there is no trend in annual rainfall of this region during last century, but the seasonal variation in rainfall is significant for agriculture and other sectors of the state. Deka et al. (2013) in their study analyzed rainfall data for more than a century in Brahmaputra and Barak River basin at monthly, seasonal and annual basis and found significant change in annual and monsoon rainfall. Many other researchers studied the climatic variation within the northeast India (Jain et al. 2013; Jhahharia et al. 2009, 2012) and found significant change in seasonal rainfall at some stations. Shivam et al. (2017) conducted a study on a river basin of northeast India to analyze the changes in extreme rainfall events and found that the daily rainfall intensity is likely to increase for climate change scenarios, whereas the number of rainy days tends to decrease.

Apart from the trend analysis methods, another technique which is very helpful in identifying the trend and behavior of hydro-meteorological data is regionalization of the area on the basis of statistical properties of meteorological data. Clustering approach is used for regionalization, and in recent time, fuzzy clustering approach has gained much popularity because of its flexibility to attach a particular dataset with several clusters. In hydrological studies, unlike hard clustering, fuzzy clustering provides an idea of a particular station being associated with different clusters. Goyal and Sharma (2016) studied the meteorological drought pattern in three states of western India and found that the region reflects different numbers of clusters for different monthly timescales of drought analysis. Goyal and Gupta (2014) used cluster approach to identify the homogeneous region of precipitation events in northeast India using five different validity indices and found that the method provides sufficient homogenized cluster for regional frequency analysis. Rao and Srinivas

(2006a, b) used fuzzy cluster approach in regionalization of the watersheds to analyze the flood frequency. Rajsekhar et al. (2012) identified eight homogeneous regions for hydrological drought situation based on severity and duration by using hydrological simulation and fuzzy clustering approach.

Northeast India is one of the most vulnerable regions toward climate change, and the largest threat to its biodiversity and hydrological system is from changing climate (Ravindrath et al. 2011). Recent extreme climatic phenomena in this region have caused loss of life, agriculture and socioeconomic losses. Increasing temperature and changing pattern of rainfall in the region as observed by several researchers have compelled planners and policy makers to think and plan according to these results which would reduce the further losses. Taking into consideration all these facts, this study was carried out for following objectives (a) extreme rainfall events occurred in the previous decades in northeastern region of India, (b) analysis of the rainfall indices for the region, (c) identifying the existing trend of different precipitation indices and (d) clustering of the northeast India region on the basis of precipitation indices.

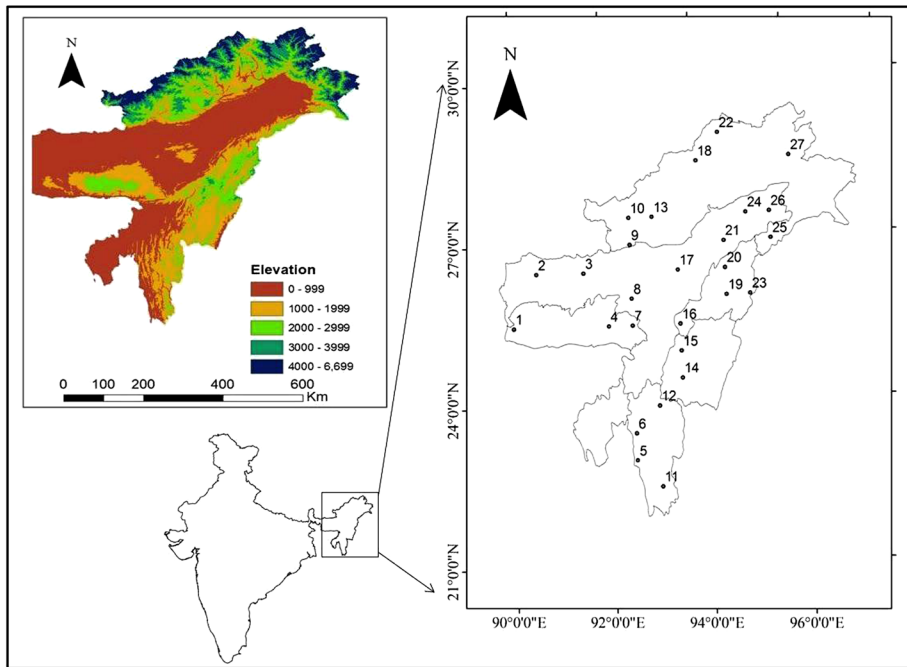
## 2 Study area and data collection

Northeast region of India which consists of the eight states Assam, Meghalaya, Arunachal Pradesh, Tripura, Nagaland, Mizoram, Manipur and Sikkim is selected for the study. This region extends from latitude 22.4°N to 28.7°N and longitude 88.2°E to 96.5°E. Elevation varies from 28 m above mean sea level to 7000 m above mean sea level. Region is very diverse in nature, and plains mainly comprise Brahmaputra and Barak valleys. Onset of monsoon occurs from middle of May and continues till October. On an average, the NE region receives about 2450 mm of rainfall. The Cherrapunjee–Mawsynram range receives rainfall as high as 11,500 mm, about 60% area under forest with Arunachal Pradesh having about 80% of its area under different kinds of forest, while Assam has the minimum percentage of forest area (30%). Large altitude difference and varying physical features are the main reason of diverse climate from near tropical to temperate. The annual rainfall in the region is received mainly from southwest monsoon, annually (Das et al. 2009; Dash et al. 2012; Goyal 2014). The temperature varies from 15 to 32 °C in summer and 0–26 °C in winter (Das et al. 2009). For studying the extreme precipitation indices and its trend, 27 stations from 0.5° × 0.5° gridded precipitation data from India Meteorological Department (IMD) were selected for the period of 1971–2005. These stations are shown in Fig. 1, and their details are given in Table 1.

## 3 Methodology

### 3.1 Precipitation indices and trend analysis

Precipitation indices are calculated according to the methodology suggested by the experts team on climate change detection, monitoring and indices (ETCCDMI). Details of these rainfall indices are given in Table 2. Mann–Kendall nonparametric test (Mann 1945; Kendall 1948) was used for calculating the trends of precipitation indices. As Mann–Kendall test is applicable on the data which is not auto-correlated so the auto correlation of the precipitation time series was checked at 1-lag, 2-lag, 3-lag and 4-lag, it was found that precipitation at any



**Fig. 1** Study area and location of selected stations

station is not auto-correlated significantly at 95%. The magnitude of trends in time series was determined using Sen's estimator (Sen 1968).

### 3.2 Fuzzy c-means clustering

Fuzzy c-means is a soft clustering approach where each vector belongs to every cluster with certain degree of membership (Dunn 1973; Bezdek et al. 1984). If a dataset of  $M$  objects which can be classified into  $c$  classes and  $Y_k$  is the data vector belongs to  $k_{th}$  object where  $k = 1, 2, 3 \dots M$ , then fuzzy c-means clustering algorithm is used to minimize the objective function represented as:

$$J(U, C) = \sum_{j=1}^M \sum_{i=1}^c u_{ik}^{\phi} \|Y_k - C_i\|^2 \quad (1)$$

where  $u_{ik}$  is the membership value of  $k_{th}$  data point in  $i_{th}$  cluster,  $\|Y_k - C_i\|^2$  is squared Euclidean distance between data vector  $k$  and center of  $i_{th}$  cluster center,  $C_i$  is the center of  $i_{th}$  cluster and  $\phi$  is generally called as fuzzifier or fuzziness index, which can have any value greater than 1. Fuzzifier value near to 1 denotes hard cluster, and in general, its value is set between 1 and 2.5 (Bezdek and Pal 1998). In fuzzy c-means clustering algorithm, initially the number of cluster and data vector of cluster center is selected randomly and then membership matrix is calculated using formula

**Table 1** Details of the selected stations

S. no.	Longitude	Latitude	Alt. (amsl)	Name of nearby station
1	90.0	25.5	87.0	West Garo Hills
2	90.5	26.5	55.0	Bongaigaon
3	91.5	26.5	56.0	Nalbari
4	92.0	25.5	1288.0	East Khasi Hills
5	92.5	23.0	138.0	Lunglei
6	92.5	23.5	662.0	Mamit
7	92.5	25.5	1145.0	Jaintiya Hills
8	92.5	26.0	181.0	Karbi Anglong
9	92.5	27.0	792.0	West Kameng-1
10	92.5	27.5	2113.0	West Kameng-2
11	93.0	22.5	755.0	Saiha
12	93.0	24.0	318.0	Aizawl
13	93.0	27.5	753.0	East Kameng
14	93.5	24.5	999.0	Churachandpur
15	93.5	25.0	975.0	Tamenglong
16	93.5	25.5	588.0	Kohima
17	93.5	26.5	439.0	Golaghat
18	94.0	28.5	3108.0	Upper Subansiri
19	94.5	26.0	1410.0	Zunheboto
20	94.5	26.5	381.0	Mokokchung
21	94.5	27.0	91.0	Sibsagar
22	94.5	29.0	1667.0	Siang
23	95.0	26.0	1041.0	Tuensang
24	95.0	27.5	112.0	Dibrugarh
25	95.5	27.0	632.0	Tirap
26	95.5	27.5	136.0	Tinsukia
27	96.0	28.5	3719.0	Dibang Valley

**Table 2** Details of rainfall indices

Indices	Description
CDD	Consecutive dry days
CWD	Consecutive wet days
PRCPTOT	Annual total wet day precipitation
R10 mm	Number of heavy precipitation days
R20 mm	Number of very heavy precipitation days
R25 mm	Number of days above 25 mm
R95P	Annual total precipitation when RR > 95th percentile
R99P	Annual total precipitation when RR > 99th percentile
R1 day	Monthly maximum 1-day precipitation
R5 day	Monthly maximum 5-day precipitation
SDII	Annual total precipitation divided by wet days in the year

$$u_{i < k}^{t+1} = \left[ \sum_{j=1}^c \left( \frac{\|y_k - c_i\|}{\|y_k - c_j\|} \right)^{\frac{2}{\phi-1}} \right]^{-1} \quad (2)$$

where  $i = 1, 2, 3 \dots c$  and  $k = 1, 2, 3 \dots M$ . By using updated membership function and Eq. (6), new cluster center can be calculated as follows:

$$c_i = \frac{\sum_{k=1}^M u_{ik}^{\phi} y_k}{\sum_{k=1}^M u_{ik}^{\phi}} \quad (3)$$

In clustering analysis, the number of clusters is very important as it represents the compactness of clustering. Bezdek (1981) and Bensaid et al. (1996) presented some validity checks for optimality of number of clusters. Optimal number of clusters refers to the compactness of the data points in a cluster and its separation from other clusters, and optimal value of number of clusters also decreases with increase in number of clusters (Pal and Bezdek 1995; Rao and Srinivas 2006b). In present study, three validity indices have been used to check the performance of clustering which are as follows:

- (a) Xie and Beni (1991) suggested compactness and separation validity function (S) which is measure of the ratio of variance within the clusters to the variance between the clusters, and smaller value of S indicates the better separation of clusters.

Separation index

$$S(U, V : X) = \frac{\sum_{i=1}^c \sum_k^M (u_{ik}^a) \|c_i - y_k\|^2}{M \min_{i \neq k} \|v_i - y_k\|^2} \quad (4)$$

- (b) The fuzziness performance index (FPI) is a measure of the degree to which different classes share membership (fuzziness) (Bezdek 1974a).

Fuzziness partition index

$$FPI(U) = \frac{1}{m} \sum_{i=1}^c \sum_{k=1}^M u_{ik}^2 \quad (5)$$

- (c) The modified partition entropy (MPE) estimates the degree of disorganization created by a specified number of classes (Bezdek 1974b).

Modified partition entropy

$$MPE(U) = \frac{1}{m} \left[ \sum_{i=1}^c \sum_{k=1}^M u_{ik} \log_a (u_{ik}) \right] \quad (6)$$

Minimizing the value for MPE, FPI and S gives the optimum number of clusters. These indices have been used in several studies (Minasny and McBratney 2002; Odeh et al. 1992; Halkidi et al. 2001; McBratney and Moore 1985).

Variables with different units affect the clustering results, so it is suggested in several studies (Lim and Voeller 2009; Dikbas et al. 2012) to normalize the data. In this study, six attributes of each station were used for identifying the clusters, i.e., latitude, longitude, average

value of the precipitation index, standard deviation, maximum value and minimum value. These data were normalized using the following transfer function:

$$x_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{7}$$

where  $x_n$  is the normalized data attribute,  $x$  is the original data attribute and  $x_{\min}$  and  $x_{\max}$  are the minimum and maximum data attributes of all stations.

## 4 Results and discussion

### 4.1 Analysis of precipitation indices

Consecutive dry days (CDD) are the number of consecutive days when rainfall is less than 1 mm, and this index is helpful in identifying the drought spell, scarcity of the water resources for the particular spell for agricultural requirement. Northeast India is humid region where major rainfall occurs in the monsoon season of June to October and on an average area receives high rainfall of 2450 mm annual, but major portion of rainfall occurs in the monsoon season only. According to the analysis of 35 years of rainfall data, it was found that average dry spell is approximately 126 days and the longest dry spell occurred

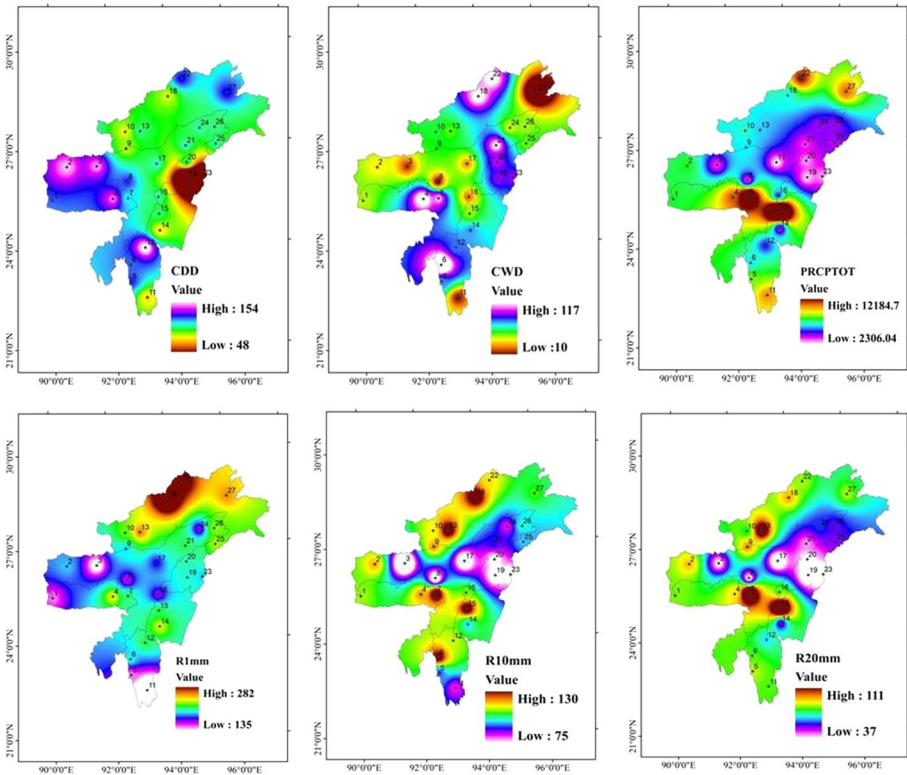


Fig. 2 Spatial variation in the precipitation indices

at West Garo Hills for 150 days in 1999 while the shortest dry spell was at upper Subansiri for 48 days in 1979. As the elevation of the station from mean sea level varies from minimum 58 m (Bongaigaon) to maximum 3719 m (Dibang Valley), variation in dry spells also varies according to the elevation. Length of dry days spells is lesser at higher elevation as compared to lower elevation stations, and average dry spell length at the stations having elevation less than 500 m was 51 days while stations above 1000 m have the average dry spell of 41 days. Reason behind this would be the orographic precipitation at the higher elevation. Figure 2 shows the spatial variation in dry days spells for the region. A significant negative trend was observed at Lunglei and Aizawl stations while at Tirap significant positive trend was found, and the rest of all the stations show non-significant increasing or decreasing trend at 95% confidence interval. Dry days spell is an indicator of drought situation in region, and its analysis is also necessary for the area like northeast India where major portion of the land is covered with forest; many studies have correlated the dry days spells to forest fire incidents (Chen et al. 2014) as during the dry spells dead biomass can get fired easily. There have been so many incidents of forest fire in northeast India and it is very vulnerable to it (Puri et al. 2011).

Similarly, the consecutive wet days spells are also important for detecting the climate change footsteps; as the threshold value for a rainfall to consider as significant was decided as 1 mm, it was assumed that the rainfall below this amount does not contribute in any hydrological process; many studies have suggested different methods to select the threshold value for the particular region (Manton et al. 2001; Groisman et al. 1999). The longest wet spell was observed at Mamit for 118 days, whereas minimum 29 days at Karbi Anglong. Karbi Anglong is the only station which shows a significant positive change, and remaining stations reflect insignificant positive or negative trend. Interpolation map of maximum wet days spells within the duration of 1971–2005 is shown in Fig. 2.

## 4.2 Changes in annual precipitation

Highest annual rainfall received at Tamenglong in 1971 which was 12,195.9 mm, whereas minimum rainfall was recorded in West Kameng-2 in 1972 that was 741.9 mm. As far as the average annual rainfall is concerned, it was observed at Dibang Valley which has the highest elevation among selected stations, and average annual rainfall was recorded to be 3358.3 mm. Trend analysis of annual precipitation shows that no significant trend was observed at any of the stations, whereas non-significant increasing or decreasing trend is existing at several stations. Only six stations (Mamit, Tamenglong, Sibsagar, Tinsukia, Upper Subansiri and Dibang Valley) are there with decreasing (statistically insignificant) annual precipitation trends, and remaining stations are showing insignificant increasing trend.

## 4.3 Analysis of rainy days (R1 mm), heavy precipitation days (R10 mm) and very heavy precipitation days (R20 mm)

As per the ETCCDI, it was considered that days receiving precipitation of 1 mm are rainy day; however, this threshold could be different for different areas. Similarly, days with precipitation more than or equal to 10 mm are termed as heavy precipitation days and days with 20 mm or more precipitation are termed as very heavy precipitation days. Analysis of these indices is important for hydrological studies as well as climate change studies (Aguilar et al. 2005). As far as the rainy days are concerned, maximum number of rainy days was



observed at Sibsagar in 1981 of 283 days and minimum 89 days at Mokokchung in 1971. Average annual precipitation days calculated for the region were 170 days, and spatial distribution of R1 mm, R10 mm and R20 mm is shown in Fig. 2. Trend of rainy days from 1971 to 2005 has not changed at any station significantly, although it is positive at some stations (Nalbari, East Khasi Hills, West Kameng-1 and 2, Aizawl, Tirap and Tinsukia) and decreasing at most of the stations, and the trend is non-significant at 95% confidence interval.

Maximum heavy precipitation days observed were 131 days at Jaintiya Hills in 1986, whereas minimum 19 days at Tuensang. This region has average 72 days of heavy precipitation days annually, and trend at these stations is not significant. Similarly, the very heavy precipitation (precipitation > 20 mm) days are also very often in northeast India. Maximum 71 days of very heavy precipitation observed at Tamenglong in 1971 and minimum 2 days at Tuensang in 1986. On average, 36 days in a year receives the very heavy rainfall, and this very heavy rainfall events cause the flash flood, landslide and other hazardous phenomena which is quite often in hilly terrain of northeast India. East Kameng and Churachandpur show the significant increasing trend, whereas in Dibang Valley it is significantly decreasing. Similarly, very wet days (R95p) and extremely wet days (R99p) indices suggest significant increasing trend at West Kameng-1 and West Kameng-2 with a magnitude of 28.71 mm and 22.36 mm, respectively, and extremely wet days index represents significant positive change at East Kameng and Churachandpur while at Karbi Anglong and Kohima this index is decreasing significantly.

#### 4.4 Changes in R1 day, R5 day and SDII index

Maximum 1-day precipitation and maximum 5-day precipitation are the indication of temporal distribution of precipitation. For the places like northeast India, this becomes more important to analyze these indices as the region receives rainfall more than 2400 mm annually, and the maximum amount of rainfall occurs in monsoon season, that too with a certain period of very high intensity rainfall. Urban hydrology has the certain significance of these indices; recently occurred 1-day heavy rainfall at Mumbai in 2005 was recorded 94.4 cm, and heavy rain in Kedarnath in Uttarakhand state of India caused severe destructions. Such incidents encourage to study and monitor these extreme precipitation indices. In the study, it was observed that stations in northeast India have received very heavy precipitation in one day during the 1971–2005 periods. It has been observed that a very much variation in 1-day maximum precipitation exists in this region depending on the spatial variation. Tamenglong (Manipur) has received maximum 1-day rainfall among all the selected stations which was 597.4 mm on June 1976, whereas Tuensang (Nagaland) received maximum 110 mm rainfall in August 1976. Trend analysis of the index R1 day represents that out of 27 stations 15 stations, namely Bongaigaon, Nalbari, West Kameng-1, West Kameng-2, Aizawl, East Kameng, Churachandpur, Tamenglong, Kohima, Golaghat, Upper Subansiri, Mokokchung, Sibsagar, Tirap and Dibang Valley depict significant positive or negative trend for maximum 1-day precipitation (Table 3). Five-day maximum precipitation shows significant increasing trend in Bongaigaon (June month with magnitude 1.69 mm), West Kameng (May month with magnitude 1.05 mm), East Kameng (0.71 mm in May month) and Churachandpur (August month with 1.16 mm), whereas significant negative trend in Nalbari, Tamenglong, Kohima, Siang and Tinsukia for the month of July, June, May, May and August, respectively.

**Table 3** Magnitude of trend (Sen's slope) for maximum 1-day (R1 day) and maximum 5-day (R5 day) rainfall

Index	Stations	Magnitude of trend (mm/year)
R1 day	Bongaigaon	1.69 (Jun), – 0.28 (Nov)
	Nalbari	– 1.06 (Jul)
	West Kameng-1	2.12 (Aug), 1.8 (Sep)
	West Kameng-2	1.49 (Aug)
	Aizawl	0.49 (Oct)
	East Kameng	1.16 (Aug), 1 (Sep), 0.92 (Oct)
	Churachandpur	0.83 (Sep)
	Tamenglong	– 0.89 (Jun)
	Kohima	– 0.62 (Jun)
	Golaghat	– 0.15 (Dec)
	Upper Subansiri	– 0.44 (Dec)
	Mokokchung	– 0.24 (Dec)
	Sibsagar	– 0.28 (Dec)
	Tirap	– 0.16 (Dec)
	Dibang Valley	– 0.42 (Dec)
R5 day	Bongaigaon	1.69 (Jun)
	Nalbari	– 1.06 (Jul)
	W. Kameng-1	1.05 (May), 2.12 (Aug), 1.8 (Sep)
	W. Kameng-2	1.49 (Aug)
	Aizawl	0.49 (Aug)
	East Kameng	0.71 (May), 1.16 (Aug), 1 (Sep), 0.92 (Oct)
	Churachandpur	1.16 (Aug), 1 (Sep), 0.92 (Oct)
	Tamenglong	– 0.89 (Jun)
	Kohima	– 0.63 (May), – 0.62 (Jun)
	Siang	– 1.1 (May)
	Tinsukia	– 0.77 (Aug)

Maximum 1-day precipitation has significant trend on monthly basis, but no significant trend was observed at annual basis. Daily intensity index (SDII) is defined as annual precipitation divided by the number of wet days (rainfall > 1 mm). Daily intensity index is important for understanding the distribution of rainfall, and northeastern region is a high rainfall receiving area but most of the rainfall occurs in short duration of monsoon period, so it is important to analyze intensity of precipitation throughout the year. Maximum intensity within the study period was observed 80.8 mm/day at Tamenglong in 1995, and minimum 5.1 mm/day was observed at Upper Subansiri in 1982.

#### 4.5 Clustering of rainfall indices

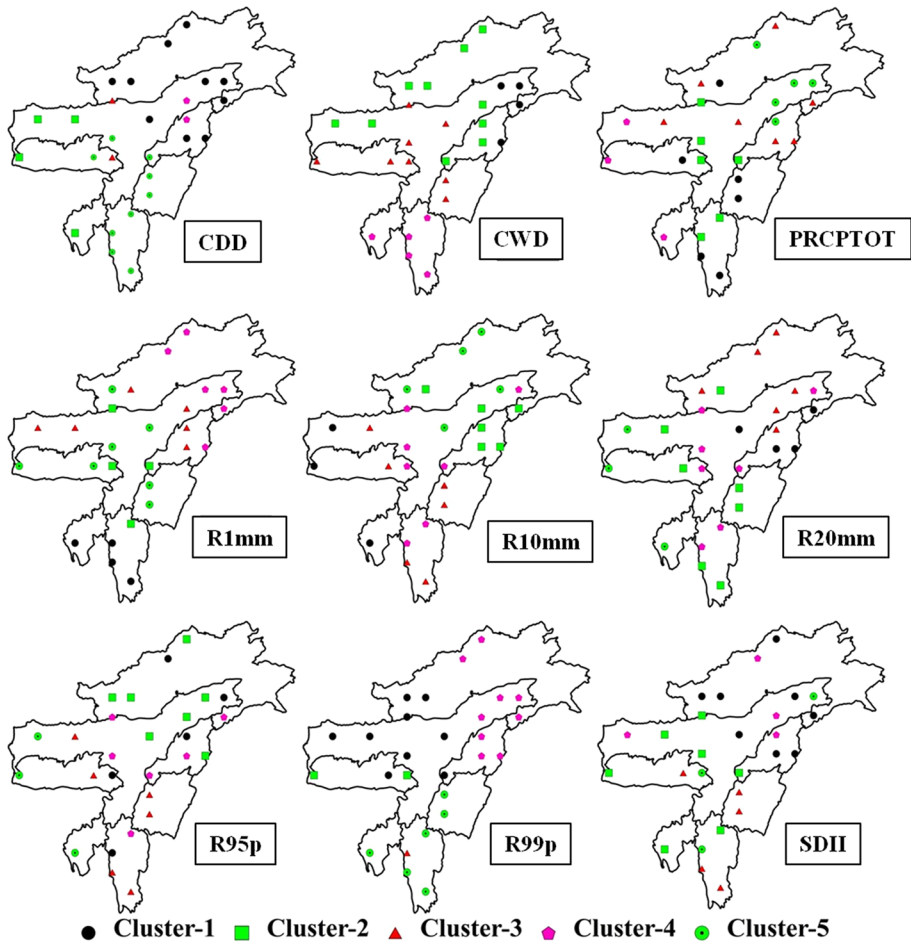
Fuzzy c-means clustering algorithm was applied on the matrix containing normalized value of the attributes of the stations, i.e., latitude, longitude, average, standard deviation, minimum value and maximum value of each index. As discussed in the methodology section, fuzziness index ( $\phi$ ) is taken as 2. Urcid and Ritter (2012) suggested the upper bound of the number of cluster as  $\sqrt{n}$  (where  $n$  is the number of stations), and minimum number

**Table 4** Values of validity indices for different precipitation indices

Index	Cluster	OFV	FPI	MPE	S
CDD	2	76.66	0.83	0.87	1.15
	3	50.22	0.81	0.84	0.53
	4	37.20	0.80	0.81	1.10
	5	24.43	0.67	0.66	0.32
CWD	2	44.66	0.78	0.83	0.56
	3	28.27	0.73	0.77	0.37
	4	19.88	0.70	0.72	0.22
	5	15.36	0.71	0.72	0.36
PRCPTOT	2	76.01	0.82	0.86	0.81
	3	48.58	0.75	0.78	0.46
	4	35.30	0.70	0.72	0.26
	5	27.27	0.69	0.70	0.51
R1 mm	2	80.66	71.93	0.85	0.88
	3	53.21	92.86	0.84	0.86
	4	38.73	93.93	0.77	0.78
	5	29.66	89.13	0.73	0.74
R10 mm	2	74.4	64.34	0.75	0.8
	3	44.8	68.27	0.65	0.69
	4	32.64	74.14	0.66	0.68
	5	25.04	70.21	0.64	0.65
R20 mm	2	76.16	67.62	0.81	0.86
	3	45.24	69.37	0.64	0.69
	4	33.18	72.63	0.67	0.69
	5	25.82	74.71	0.69	0.7

of cluster can be 2, so the number of cluster classes for the analysis was taken from 2 to 5. Further, optimum number of clusters was selected with help of certain validity indices discussed in methodology section. Objective function value (OFV) and validity indices values for indices CDD, CWD and PRCPTOT are given in Table 4. As in the table, FPI, MPE and S values are minimum at five numbers of clusters for CDD, whereas FPI and S values are minimum at four numbers of clusters and MPE value is minimum value at five number of clusters in case of CWD index. Therefore, optimum cluster for CWD is assigned as 4. FCM results for clustering of annual precipitation index (PRCPTOT) checked with validity indices, FPI and MPE value minimize at five numbers of clusters, whereas separation index (S) minimizes at four numbers of clusters (Table 4); hence for PRCPTOT, five groups were finalized and distribution of the stations can be seen in Fig. 3. Precipitation days (R1 mm), heavy precipitation days (R10 mm) and very heavy precipitation days (R20 mm) clustering results and validity indices values can be seen in Table 4. As FPI and MPE and S values minimize at five numbers of clusters and therefore for R1 mm and R10 mm cluster numbers are 5, whereas for R20 mm, FPI and MPE values minimize at three numbers of clusters, this is selected as maximum number of clusters for R20 mm. Stations falling under each cluster can be seen in Fig. 3.

Similarly, cluster analysis for the indices R95p, R99p and SDII also shows five clusters which can be seen in Fig. 3. Further, FCM was applied on 1-day maximum rainfall index



**Fig. 3** Clusters identified by fuzzy c-means for indices

(R1 day) and 5-day maximum rainfall (R5 day), as these two rainfall indices are calculated for each month, so for the analysis only May, June, July, August and September months were selected as these are the highly rainfall-receiving periods. Analysis reveals that for June and August months, there is not very much variation among the stations and thus the number of clusters is only 2 and 3, respectively, whereas in May, July and September months five clusters were found based on the values of validity indices. Reason behind this monthly variation could be the northeast monsoon winds of Bay of Bengal which dominates the region in summer months and southwest monsoon winds which reach to this region in monsoon months. Distribution of the stations in each cluster for R1 day and R5 day can be seen in Fig. 4.

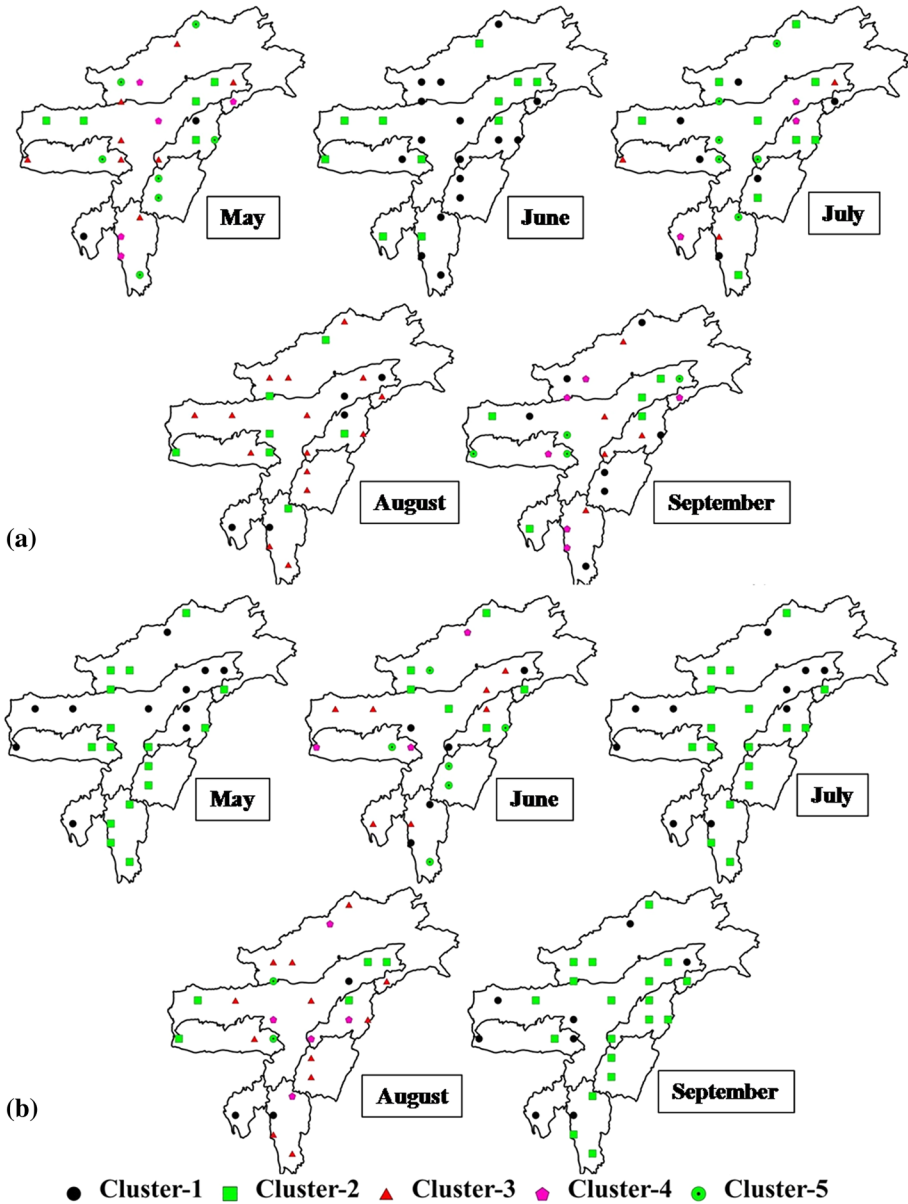


Fig. 4 Clusters identified by fuzzy c-means for index R1 day and R5 day

### 5 Discussion and conclusion

Northeast India has a very irregular terrain, and elevation ranges from 55 m (Bongaigaon) to 3719 m at Dibang Valley and this causes the high variation in climatic pattern of the region. Results of extreme precipitation indices analysis of the region show high variation in dry days spells (CDD) and wet spells (CWD) in stations. Dry day spells are quite often

at the stations situated at lower elevation, whereas prolonged wet days spells observed at high elevation zone. Maximum 1-day rainfall is good indicator for the flood analysis in the hilly region, and the study shows increasing trends in maximum 1-day rainfall at the stations of Arunachal Pradesh state (East and West Kameng) which has the several implications for the region as catchment of the Brahmaputra river drains run off entirely from states of Arunachal Pradesh and Assam. Extreme rainfall events in the upper catchment of Brahmaputra River cause flash flood in the downstream catchment.

Although the trend analysis of annual precipitation does not suggest statically significant trend, the decrease in rainy days and subsequent increase in intensity are the prime concern for the region. Changing magnitude of extreme rainfall events of the northeast region is matter of concern as the area is very much vulnerable to extreme events; because of its hilly terrain, even a small-duration heavy rainfall at upper part causes flood in plain areas in downstream of Brahmaputra River. Contrary to this, despite receiving high annual rainfall, some parts face drinking water and irrigation water scarcity in the non-monsoon period because of the draining out of large portion of rainfall water.

This study classifies the northeast India region in several homogeneous clusters according to their statistical properties for different precipitation indices using fuzzy clustering approach, which provides an insight into the rainfall distribution in the region. Optimum number of clusters from fuzzy clustering approach was validated using five validity indices, which assures the compactness of the clusters. Findings of the study could be helpful for frequency analysis and watershed management planning of the region. Homogeneous regions identification based on the different extreme events indices would be helpful for the preparedness and effective mitigation planning for natural hazard situation.

## References

- Abbas F, Ahmad A, Safeeq M, Ali S, Saleem F, Hammad HM, Farhad W (2014) Changes in precipitation extremes over arid to semiarid and subhumid Punjab, Pakistan. *Theoret Appl Climatol* 116(4):671–680
- Aguilar E, Peterson TC, Obando PR, Frutos R, Retana JA, et al (2005) Changes in precipitation and temperature extremes in Central America and northern South America, 1961–2003. *J Geophys Res Atmos* 1984–2012, 110D23
- Bensaid AM, Hall LO, Bezdek JC, Clarke LP, Silbiger ML, Arrington JA, Murtagh RF (1996) Validity-guided re clustering with applications to image segmentation. *IEEE Trans Fuzzy Syst* 42:112–123
- Bezdek JC (1974a) Numerical taxonomy with fuzzy sets. *J Math Biol* 1:57–71
- Bezdek JC (1974b) Cluster validity with fuzzy sets. *J Cybern* 3(3):58–72
- Bezdek JC (1981) *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers, Norwell
- Bezdek JC, Pal NR (1998) Some new indexes of cluster validity. *IEEE Trans Syst Man Cybern B Cybern* 28(3):301–315
- Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. *Comput Geosci* 102:191–203
- Chen F, Fan Z, Niu S, Zheng J (2014) The influence of precipitation and consecutive dry days on burned areas in Yunnan Province, Southwestern China. *Adv Meteorol*, pp 1–11
- Das A, Ghosh PK, Choudhury BU, Patel DP, Munda GC, Ngachan SV, Chowdhury P (2009) December climate change in North East India: recent facts and events—worry for agricultural management. In: *Proceedings of the workshop on impact of climate change on agriculture*, pp 32–37
- Dash SK, Sharma N, Pattanayak KC, Gao XJ, Shi Y (2012) Temperature and precipitation changes in the north-east India and their future projections. *Global Planet Change* 98:31–44
- De Lima MIP, Santo FE, Ramos AM, Trigo RM (2014) Trends and correlations in annual extreme precipitation indices for mainland Portugal, 1941–2007. *Theoret Appl Climatol* 119(1–2):55–75

- Deka RL, Mahanta C, Pathak H, Nath KK, Das S (2013) Trends and fluctuations of rainfall regime in the Brahmaputra and Barak basins of Assam, India. *Theoret Appl Climatol* 114(1–2):61–71
- Dikbas F, Mahmut F, Cem K, Gungor M (2012) Classification of precipitation series using fuzzy cluster method. *Int J Climatol* 32:1596–1603
- Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters
- Goswami BN, Venugopal V, Sengupta D, Madhusoodanan MS, Xavier PK (2006) Increasing trend of extreme rain events over India in a warming. *Environment* 314:1442–1445
- Goyal MK (2014) Statistical analysis of long term trends of rainfall during 1901–2002 at Assam, India. *Water Resour Manage* 286:1501–1515
- Goyal MK, Gupta V (2014) Identification of homogeneous rainfall regimes in northeast region of india using fuzzy cluster analysis. *Water Resour Manage* 28:4491–4511
- Goyal MK, Sharma A (2016) A fuzzy c-means approach regionalization for analysis of meteorological drought homogeneous regions in western India. *Nat Hazards* 84:1831–1847
- Groisman PY, Karl TR, Easterling DR, Knight RW, Jamason PF et al (1999) Changes in the probability of heavy precipitation: important indicators of climatic change. *Clim Change* 42:243–283
- Halkidi M, Batistakis Y, Vazirgiannis M (2001) On clustering validation techniques. *J Intell Inf Syst* 172(3):107–145
- IPCC (2007) Climate change 2007: impacts, adaptation, and vulnerability: working group II contribution to the intergovernmental panel on climate change fourth assessment report, summary for policymakers IPCC Secretariat, Geneva
- Jain SK, Kumar V, Saharia M (2013) Analysis of rainfall and temperature trends in northeast India. *Int J Climatol* 334:968–978
- Jhajharia D, Shrivastava SK, Sarkar D, Sarkar S (2009) Temporal characteristics of pan evaporation trends under the humid conditions of northeast India. *Agric For Meteorol* 149:763–770
- Jhajharia D, Dinpashoh Y, Kahya E, Singh VP, Fakheri-Fard A (2012) Trends in reference evapotranspiration in the humid region of northeast India. *Hydrol Process* 26:421–435
- Kendall MG (1955) Rank correlation methods. Griffin, London
- Lim YH, Voeller DL (2009) Regional flood estimations in Red river using L-moment-based index-flood and bulletin 17B procedures. *J Hydrol Eng* 14(9):1002–1016
- Mann HB (1945) Nonparametric tests against trend. *Econometrica* 13:245–259
- Manton MJ, Della-Marta PM, Haylock MR, Hennessy KJ, Nicholls N et al (2001) Trends in extreme daily rainfall and temperature in Southeast Asia and the South Pacific: 1961–1998. *Int J Climatol* 21:269–284
- McBratney AB, Moore AW (1985) Application of fuzzy sets to climatic classification. *Agric For Meteorol* 35:165–185
- Minasny B, McBratney AB (2002) FuzME version 21 Australian Center for Precision Agriculture The University of Sydney
- Odeh IOA, Chittleborough DJ, McBratney AB (1992) Soil pattern recognition with fuzzy-c-means: application to classification and soil-landform interrelationships. *Soil Sci Soc Am J* 56:505–516
- Pal NR, Bezdek JC (1995) On cluster validity for the fuzzy c-means model. *IEEE Trans Fuzzy Syst* 3(3):370–379
- Puri K, Areendran G, Raj K, Mazumdar S, Joshi PK (2011) Forest fire risk assessment in parts of Northeast India using geospatial tools. *J For Res* 224:641–647
- Rajsekhar D, Mishra AK, Singh VP (2012) Regionalization of drought characteristics using an entropy approach. *J Hydrol Eng* 18(7):870–887
- Rao AR, Srinivas VV (2006a) Regionalization of watersheds by fuzzy cluster analysis. *J Hydrol* 318(4):37–56
- Rao AR, Srinivas VV (2006b) Regionalization of watersheds by fuzzy cluster analysis. *J Hydrol* 318(4):57–79
- Ravindranath NH, Rao S, Sharma N, Nair M, Gopalakrishnan R et al (2011) Climate change vulnerability profiles for North East India. *Curr Sci* 101(3):384–394
- Roy SS, Balling JRRC (2004) Trends in extreme daily precipitation indices in India. *Int J Climatol* 24:457–466
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. *J Am Stat As* 63(324):1379–1389
- Shivam G, Goyal MK, Sarma AK (2017) Index-based study of future precipitation changes over Subansiri river catchment under changing climate. *J Environ Inform* 34:1–14

- Urcid G, Ritter GX (2012) Advances in knowledge-based and intelligent information and engineering systems. In: Graña M, Toro C, Posada J, Howlett RJ, Jain LC (eds) Advances in knowledge-based and intelligent information and engineering systems. IOS Press, Amsterdam, pp 2140–2149
- Xie XL, Beni G (1991) A validity measure for fuzzy clustering. *IEEE Trans Pattern Anal Mach Intell* 138:841–847
- Zhang X, Yang F (2004) RClmDex 10 user manual Climate Research Branch Environment Canada

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.