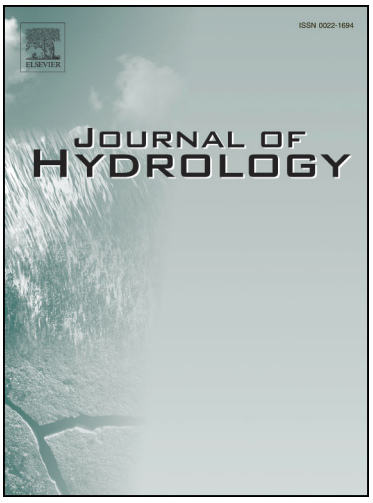
Journal Pre-proofs



Research papers

Accounting for temporal variability for improved precipitation regionalization based on self-organizing map coupled with information theory

Ravi Kumar Guntu, Rathinasamy Maheswaran, Ankit Agarwal, Vijay P. Singh

PII: S0022-1694(20)30696-X
DOI: https://doi.org/10.1016/j.jhydrol.2020.125236
Reference: HYDROL 125236

To appear in: *Journal of Hydrology*

Received Date: 18 May 2020
Revised Date: 20 June 2020
Accepted Date: 23 June 2020

Please cite this article as: Kumar Guntu, R., Maheswaran, R., Agarwal, A., Singh, V.P., Accounting for temporal variability for improved precipitation regionalization based on self-organizing map coupled with information theory, *Journal of Hydrology* (2020), doi: https://doi.org/10.1016/j.jhydrol.2020.125236

Accounting for temporal variability for improved precipitation regionalization based on self-organizing map coupled with information theory

Ravi Kumar Guntu[1], Rathinasamy Maheswaran[2], Ankit Agarwal[1,*] and Vijay P. Singh[3]

[1]Department of Hydrology, Indian Institute of Technology Roorkee, 247667, India

[2]Department of Civil Engineering, MVGR College of Engineering, Vizianagaram, 535005, India

[3]Department of Biological and Agricultural Engineering & Zachry Department of Civil & Environmental Engineering, Texas A&M University, 321 Scoates Hall, 2117 TAMU, College Station, TX 77843-2117, USA; National Water Center, UAE University, Al Ain, UAE

**Correspondence to:* Ankit Agarwal (ankit.agarwal@hy.iitr.ac.in)

Highlights

- Accounted for temporal variability and magnitude to identify homogeneous regions of precipitation in India
- Coupling of self-organizing map with standardized variability index reveals unique seasonal distribution of precipitation for each cluster
- South-central India; South-eastern coastlines; Konkan Coast exhibit stable clusters independent of temporal variability
- The temporal evolution of clusters unravels a new emerging pattern of Indian Summer Monsoon across Central India

Abstract

Precipitation regionalization deals with an investigation of the seasonality and its temporal variability and is useful for a wide variety of applications in hydro-meteorology. The d homogeneous regions can be used as a basis for transforming the information from gauged to ungauged sites and can reduce the uncertainty in estimating the seasonal characteristics of precipitation across India. Despite several studies stressing the importance of seasonality and temporal variability to the environment, there is a lack of studies on accounting for temporal variability in regionalization. Precipitation regionalization must account for both the precipitation magnitude and its temporal variability at multiple time-scales to extract the seasonality of a region representing coherent local and inter-annual variability. Therefore, in this study, we propose a framework for precipitation regionalization, considering both precipitation magnitude and its temporal variability. High resolution (0.25^0 x 0.25^0) gridded daily precipitation time series over the period 1901–2013 from Indian Meteorological Department (IMD) was used for the evaluation of the framework. First, the historical daily time series was transformed into multiple time scales, i.e., annual, seasonal, and monthly time scales. Entropy-based standardized variability index was used to measure the inter-annual variability of precipitation at each time scale. Regionalization of grid points was performed using self-organizing maps, an artificial neural network. Ten distinct regions were identified that can be tied back to two general categories, such as climate characteristics and physical characteristics. Coupling of the self-organizing map with standardized variability index reveals unique seasonal distribution of precipitation for each region. The temporal evolution of clusters unravels a new emerging pattern across Central India. Consideration of temporal variability plays an insignificant role in the shape, size and stability of south-central India, south-eastern coastlines, and Konkan Coast. Intriguingly, separate Rain-belt and Rain-shadow Western Himalayas are formed due to the difference in topography and seasonal characteristics of precipitation. The temporal evolution of clusters unravels a significant change in the occurrence of the 50^{th} percentile monsoon after the 1940s across the north-western region; a significant increase in the 50^{th} percentile monsoon after the 1940s across western India, and decrease in the 50^{th} percentile monsoon after the 1980s in the north-central Region.

Keywords: Precipitation regionalization, Spatiotemporal variability, Indian summer monsoon, Standardized variability index, Self-Organizing maps, Information theory

1 Introduction

Regionalization, a technique to identify homogenous regions in a particular aspect of a climate variable(s) of interest, is crucial for numerous hydro-meteorological applications (Brown et al., 2010). These applications include regional flood frequency analysis (Sugahara et al., 2009; Yang et al., 2010; Zhang and Singh, 2006), precipitation forecasting and downscaling (Sehgal et al., 2018), potential water resources availability (Maruyama et al., 2005), climate dynamics (Deng et al., 2019; Geen et al., 2018; Jiménez-Esteve and Domeisen, 2019; Zhao et al., 2019), agricultural planning, design of hydraulic structures (Adamowski, 2000), among others. Regionalization can be used as a method to extend information beyond the gauge network and can provide planners and designers with a better alternative for the assessment of precipitation characteristics at ungauged sites than can traditional single-site analysis (Alila, 1999). However, in many developing countries, like India, rain gauges are scarce. Information on the spatial distribution of precipitation as well as its temporal variability at ungauged sites can be gained by forming groups having similar precipitation characteristics (Hosking and Wallis, 1993; Modarres, 2008; Saf, 2009).

There have been several attempts to delineate homogenous precipitation regions that include numerous statistical and conceptual techniques (Caliński and Harabasz, 1974; Fraley, 1998; Rousseeuw, 1987). These include correlation analysis (Gadgil et al., 1993; Saikranthi et al., 2013), principal component analysis (Domroes et al., 1998; Iyengar and Basak, 1994), common factor analysis (Dinpashoh et al., 2004), spectral analysis (Azad et al., 2010; Türkeş and Tatlı, 2011), hierarchical approach (Santos et al., 2019), principal component analysis with hierarchical approach (Darand and Mansouri Daneshvar, 2014; Fazel et al., 2018), K-means (Carvalho et al., 2016), and self-Organizing maps (Hsu and Li, 2010). Among these, the Self-Organizing Map (SOM), proposed by Kohonen, (2001), is found to be a robust clustering technique in creating soft boundaries and in handling large amounts of complex data (Lin and Chen, 2006). It is extensively used due to its competence in reducing high dimensional data to low dimensional data (preferably to one or two) while preserving the topology (Chen et al., 2010). In the past, SOM was used for the regionalization of catchments with similar hydrologic characteristics (Agarwal et al., 2016a), and grid points with similar precipitation and runoff events (Dilmi et al., 2017; Nourani et al., 2013; Wolski et al., 2018).

The key to all these methods is to define regions that need not be spatially contiguous but homogenous concerning the variable of interest and are sufficiently distinct from other regions. The variable of interest could be a high intraregional correlation or low interregional correlation between members (Gadgil et al., 1993; Saikranthi et al., 2013), statistical characteristics of daily precipitation such as standard deviation and skewness (Kulkarni, 2017), geographical location, principal components extracted from large scale atmospheric predictor variables (Satyanarayana and Srinivas, 2011), synchronicity of extremes (Agarwal et al., 2018), periodicities in precipitation (Azad et al., 2010), and variability at multiple time scales and their signature (Roushangar et al., 2019). All these methods have contributed significantly to

our understanding; however, there exists an enormous scope for advancement by accounting for the temporal variability of precipitation in delineating homogenous regions. For instance, homogenous regions developed by prior studies (Bharath and Srinivas, 2015; Fukushima et al., 2019; Mannan et al., 2018; Saikranthi et al., 2013; Satyanarayana and Srinivas, 2008) used either the mean or total precipitation for identifying homogeneous regions. On the contrary, a few studies (Krstanovic and Singh, 1992; Maruyama et al., 2005; Roushangar et al., 2019) considered only the variability associated with precipitation parameters, while disregarding the precipitation amount. They were thus resulting in zones with similar variability but different precipitation magnitudes. In recent decades, a multitude of studies have reported a significant increase in the complexity of Indian precipitation in terms of temporal variability as well as its uncertainty (Ghosh et al., 2012; Goswami et al., 2006; Paul et al., 2018; Roxy et al., 2017; Vinnarasi and Dhanya, 2016). Besides, there has been a significant increase in the magnitude of total annual precipitation (Rajeevan et al., 2008; Vul and Mishra, 2019; Yadav and Roxy, 2019). Motivated by these realizations, we argue that regionalization studies must account for both the temporal variability and precipitation magnitude for robust identification of homogeneous regions. Therefore, this study attempted to develop a regionalization framework accounting for both temporal variability and magnitude of precipitation.

In general, temporal variability is defined as a measure of the unevenness of a random variable over different class intervals (Sang et al., 2013). There is a plethora of metrics for assessing the variability of a time series, including variance, diversity indices, and various measures based on Shannon entropy (Guntu et al., 2020a; Mishra et al., 2009). In the past decade, Shannon entropy has gained significant attention in surface and sub-surface hydrology (Koutsoyiannis, 2005; Molini et al., 2006; Singh, 2013, 2015). One of the main advantages of using Shannon entropy is that it does not make any assumption on the probability distribution or statistical properties of data. Besides, it can be applied to any type of distribution whether it is known or unknown, or if the nature of the underlying system is dynamic (Guntu et al., 2020b; Pechlivanidis et al., 2016). Kawachi et al. (2001) proposed Shannon entropy as precipitation entropy for assessing the temporal precipitation apportionment. Later, Maruyama et al. (2005) modified precipitation entropy into intensity entropy and apportionment entropy for quantifying the intra-annual variability of precipitation. Further, these measures are applied by Mishra et al. (2009) to assess the intra-annual variability of monthly precipitation and proposed marginal entropy to investigate the inter-annual variability. Furthermore, to have a single measure to represent both inter-annual and intra-annual variability, Mishra et al. (2009) proposed a disorder index based on the difference between maximum possible entropy and entropy obtained for a time series. In continuation, Cheng et al. (2017); Rodrigues da Silva et al. (2016); Roushangar et al. (2019); Zhang et al. (2016); Zhao et al. (2011) used the disorder index to assess precipitation variability for different regions around the globe.

Very recently, (Guntu et al., 2020a) highlighted that though the disorder index is indeed beneficial; however, it is limited to the time scale of evaluation as well as the length of the data. They modified the disorder index and proposed a new standardized variability Index (SVI) with capabilities to overcome the limitations of the disorder index. It is found robust in comparing the inter-annual and intra-annual variability of precipitation at different time scales and for regions with different temporal records.

Using SVI, this study developed a framework for precipitation regionalization based on precipitation magnitude and its temporal variability. Besides, we aim to analyse the effects of non-stationarity on regionalization from an abruptly changing climate point of view.

This paper is organized as follows. Section 2 describes the proposed methodology with a brief description of Marginal Entropy, Standardized Variability Index, and SOM clustering technique. Details of the study area and gridded precipitation data are presented in Section 3. Section 4 presents the application of the proposed methodology, followed by the discussion of results. Finally, Section 5 presents some of the essential conclusions and scope for further research.

2 Study area and Data

The present study focuses on improving the precipitation regionalization by accounting for temporal variability and precipitation magnitude in the Indian subcontinent which lies approximately between 8–37° N latitude and 68–98° E longitude, covering an area of 3,287,590 km^2. The climate of India comprises a wide range of weather conditions across a vast geographic scale and varied topography, making generalization indeed tricky.

In the present study, we used the gridded data of mean daily precipitation in mm (developed by Pai et al., (2014)) with a high spatial resolution of 0.25° × 0.25° for a spatial domain of 66.5°E to 100°E and 6.5°N to 38.5°N covering the mainland region of India (see Fig. S1). The gridded data was generated from a diverse network of 6995 gauging stations across India using Inverse distance weighted (IDW) interpolation scheme proposed by Shepard, (1968). The daily data were obtained for a period of 113 years (1901–2013), which was available in the archive of National Data Centre, India Meteorological Department (IMD), Pune (link to data is given in data sources). In the past, several studies have used the same dataset which includes downscaling (Sehgal et al., 2018), spatiotemporal variability of precipitation (Guntu et al., 2020a; Sahany et al., 2018), extreme precipitation analysis (Agarwal et al., 2018; Malik et al., 2016; Vinnarasi and Dhanya, 2016), intrinsic predictability of Indian precipitation (Guntu et al., 2020b), and spatial diversity of Indian precipitation teleconnections (Kurths et al., 2019), among others. The application of IMD gridded data in various applications shows that the data is highly accurate and capable of capturing the spatial distribution of precipitation over the country.

3 Methods and methodology

In this section, first, we explain the data preparation and computation of time series at different scales. Following that, the *Standardized Variability Index (SVI)* and Kohonen *Self-Organizing Maps (SOM)* are discussed. Furthermore, we illustrate the combination of SVI and SOM to regionalize precipitation.

3.1 Data preparation and computing time series at different time scales

This choice of clustering variable, i.e. time series of gridded precipitation at different time scales means that homogeneous regions are defined based on coherent climate variability at different time-scales (monthly, seasonal and annual) rather than mean climate conditions. This is important, as Seneviratne et al. (2012) say seasonality is a characteristic of a time series in which the data experiences regular and predictable changes that recur every calendar year due to local variability. Also, Kurths et al. (2019) suggest that coherent inter-annual variability shows a common response to large-scale climate drivers. When regionalization is applied to observations at different time-scales, then the approach can help us identify regions of common seasonality, climate sensitivity, and mechanisms that drive variability for each region.

Therefore, in the present study, climate data was analyzed for four seasons. (1) Summer season: March–April–May (MAM), (2) South-west monsoon season form June–July–August (JJA), (3) Autumn (Fall) season during which most of the northeast monsoon occurs: September–October–November (SON), and (4) Winter season: December–January–February (DJF).

We investigate the inter-annual variability of precipitation at monthly (Jan-Dec), seasonal (Spring, Summer, Fall, Winter), and annual time-scales. Variability of seasonal precipitation (e.g., summer) over the 113 year period (in this study) provides an understanding of how summer precipitation varies over time. Similarly, monthly precipitation (e.g., June) over the years helps quantify the precipitation variability in that month over the 113 years. This detailed investigation of the precipitation patterns would provide a comprehensive picture of the spatiotemporal precipitation variability. The method for creating the time series at multiple time-scales is presented next.

1. ***Formation of monthly time-series (January to December)***

   To form monthly time-series, sum all days in a particular month for the year 1901 and the process is repeated for all the years. Monthly time-series consists of 113 values; each value represents total monthly precipitation for that particular year (1st to 12th row in Table 1).

2. ***Formation of seasonal time-series (Spring, Summer, Fall, Winter)***

To form seasonal time-series, sum all days in a particular season for the year 1901 and repeat the process for all the years. Seasonal time-series consists of 113 values; each value represents the total seasonal precipitation for that particular year (13th -16th row in Table 1).

3. Formation of annual time-series (Annual)

The gridded data is available at daily time-step. To get annual values, sum all the 365 days in the year 1901 and repeat the process for all the years. Annual time-series consists of 113 values; each value represents the total annual precipitation for that particular year (17th row in Table.1).

Table 1: Formation of time-series at multiple time-scales

Sl No		Years (1901-2013)									
	Timescale	1901	1902	.	.	.	.	.	.	.	2013
	Monthly scale	Precipitation amount (mm)*									
1	Jan	20	25	.	.	.	.	.	.	.	22
2	Feb	30	26	.	.	.	.	.	.	.	28
3	Mar	100	89	.	.	.	.	.	.	.	95
.	.	.	.	.	.	.	.	.	.	.	.
12	Dec	30	32	.	.	.	.	.	.	.	31
	Seasonal scale	Precipitation amount (mm)*									
13	Spring	470	474	.	.	.	.	.	.	.	465
14	Summer	4560	4750	.	.	.	.	.	.	.	4520
15	Fall	310	292	.	.	.	.	.	.	.	299
16	Winter	80	83	.	.	.	.	.	.	.	81
	Annual scale	Precipitation amount (mm)*									
17	Annual	5420	5599	.	.	.	.	.	.	.	5365

*precipitation amount (mm) are random values showed for demonstration purpose

3.2 Standardized variability index

To introduce standardized variability Index (SVI), a background on marginal entropy and Shannon entropy is necessary.

Shannon Entropy

Shannon (1948) formulated the entropy concerning a concept called a surprise, i.e. information regarding the uncertainty of the random variable or its probability distribution. Singh, (2015, 2013, 1997) discussed the strengths of Shannon entropy (SE) and its limitations and accredited it to be a decision-making tool in surface and sub-surface hydrology. The SE can be defined as Eq. (1)

$$SE(X) = -\sum_{i=1}^{n_c} P(x_i) log_2[P(x_i)] \quad (1)$$

where $SE(X)$ represents the average information content of the random variable X, with probability distribution function $P(x_i)$; $-log_2P(x_i)$ directly represents the average surprise about the value x_i occurring with probability $P(x_i)$ (Gong et al., 2014); and n_c is the number of class intervals. The unit of SE depends on the base of the logarithm. For example, a unit of SE is bits when the base is 2, Napier for base e, and decibels for base 10. Here, we used the logarithm with base 2 and SE was measured in bits. For a deterministic single outcome system, the probability for a specific event that will take on value is one, and for any other value is zero. In such a case, the average surprise is zero, and this case corresponds to absolute certainty. On the contrary, for uniform distribution, the average surprise yields no information about the possible outcome, since every outcome, has the same probability of occurrence and therefore corresponds to absolute uncertainty. Thus, the value of $SE(X)$ varies from zero to log_2n_c i.e., from complete information to no information about the system's state. Viewed in this manner, SE captures the spread of data and can be used for measuring the variability associated with the precipitation time series (Kawachi et al., 2001). Gong et al., (2014) discussed the four problems of distribution-based estimation in hydrological time series that could result in biased entropy estimation. These are: "optimal bin width," "zero values," "measurement error," and "skewness effect.". Weijs et al., (2013) suggested that the optimal bin width consideration for evaluating the probability density function depends on the question one wishes to answer.

Marginal Entropy

The marginal entropy (ME), used by Mishra et al. (2009), as a measure of the randomness of a random variable *X* with the probability distribution *P(x)*, is useful in calculating the variability of the distribution and can be extended to any type of dataset. For instance, we considered total annual, seasonal, and monthly precipitation time series and applied ME to assess the inter-annual variability associated with individual time series. Individual seasons are compared to examine the interannual variability of seasonal precipitation within the year and help understand the season(s) responsible for the annual variability. The mathematical expression for evaluating ME of precipitation amount can be given as

$$ME = -\sum_{i=1}^{n_c} \frac{q_i}{Q} log_2\left[\frac{q_i}{Q}\right] \quad (2)$$

where q_i denotes the precipitation amount for the ith year; Q denotes the total precipitation for the period from 1901-2013; and n_c is the number of class intervals, depending on the length of data. For a deterministic event (hypothetical), the probability $p(\frac{q_i}{Q})$ is one and, correspondingly to this, ME is nearer to zero, and this case refers to absolute certainty. On the other hand, if the distribution is uniform, then Eq. (2) yields maximum uncertainty since every outcome has the same probability of occurrence. Thus, the value of ME varies from zero to log_2n_c i.e., from absolute certainty to maximum

uncertainty (or) complete information to no information about the system's state. Viewed in this manner, intuitively, ME can be used for measuring the spatiotemporal variability associated with any geophysical time series.

Standardized variability index

In this paper, we employed SVI proposed by Guntu et al., (2020a) to calculate the variability associated with the individual time series, defined as

$$SVI = \frac{ME_{max} - ME}{ME_{max}} \tag{3}$$

where ME_{max} is the maximum ME that can be obtained for a given distribution, and ME is the entropy obtained for the given time series. From this description, SVI takes on a value within a finite range of 0 to 1, where zero corresponds to no variability, and one represents high variability, i.e. minimum uncertainty to maximum uncertainty. It has an advantage in comparing the variability of precipitation series at multiple time scales and therefore in relatively addressing the uncertainty. As SVI increases the variability increases and vice versa. Since the range of SVI is finite, it has the competence of inter-comparison of results for datasets with different lengths in regionalization. We applied SVI to 17 time-series (12 monthly time series, 4 seasonal and 1 annual time series) to evaluate the inter-annual variability at each time scale and use their signatures in the regionalization of precipitation. The detailed procedure on how to calculate SVI is given in Appendix A.

3.3 Necessity for using both magnitude and variability

To capture the precipitation characteristics of a region satisfactorily, we argue that considering both the magnitude and variability is essential. We illustrate this concept by considering four grid locations A, B, C, and D (see Fig.1a) of IMD gridded precipitation dataset. Location A and Location B have different mean precipitation (mm/year) but the same variability (measured in terms of SVI_a); on the contrary, Location C and Location D have the same mean precipitation (mm/year) but different variability (see Fig. 1b). Thus, if variability alone is considered in the regionalization, it delineates the regions having similar precipitation variability patterns. On the contrary, if only mean precipitation is considered, regions will have similar mean precipitation, and the intra-cluster heterogeneity will be in terms of precipitation variability.

This study proposes to include both precipitation magnitude and its variability (in terms of SVI) to improve precipitation regionalization. Besides, precipitation time-series is nonlinear and exhibits multi-scale phenomenon (Rathinasamy et al., 2014). Thus, mean precipitation and SVI at multiple time scales (monthly, seasonal, annual) were used for precipitation regionalization.

a) **b)**

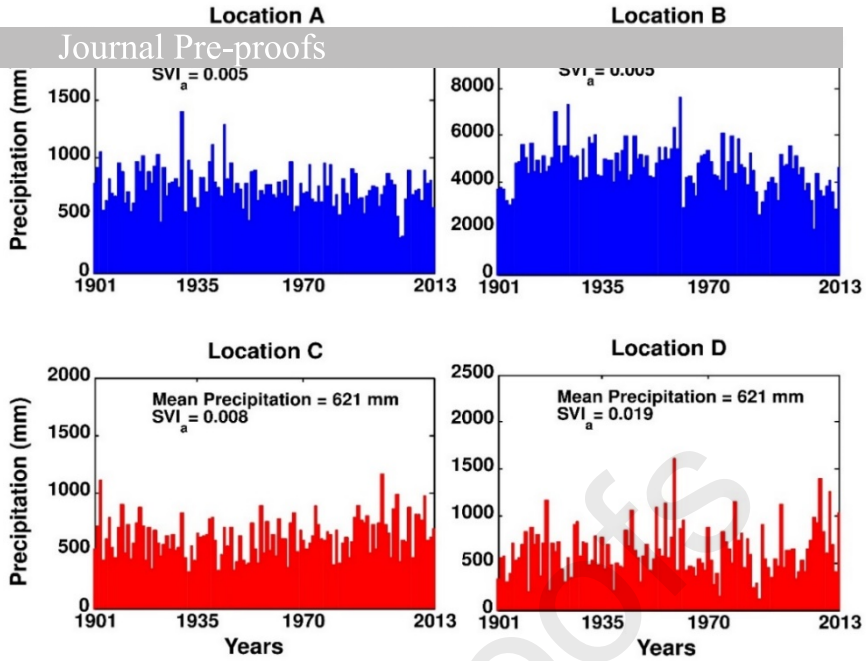


Figure 1: Geographical locations of grid points (a) considered for illustration of the necessity for using both magnitude as well as its variability. (b) Locations A and B describes the downside of only using SVI in regionalization, Locations C and D describes the downside of using only the mean precipitation in regionalization.

3.4 Kohonen Self Organizing Map

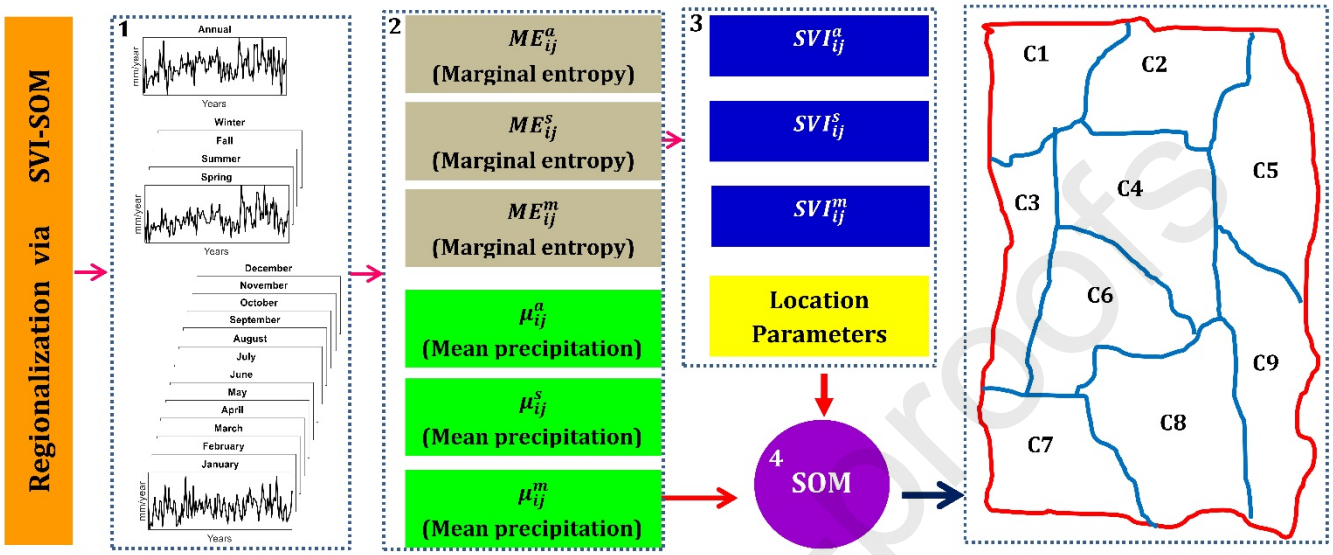
Kohonen Self Organizing Map (SOM) is based on artificial neural network and has been successfully applied in hydro-meteorological applications (Agarwal et al., 2016a; Hsu and Li, 2010; Mannan et al., 2018; Ohba et al., 2016). SOM is a widely used tool for clustering, visualization of high dimensional data and dimension reduction in various applications (Kohonen, 2001). SOM is unsupervised, and collection of neurons which are regularly arranged in a low-dimensional (one or two) grid pattern in the output layer. One of the advantages of SOM is to extract the inter-relationship between inputs of a high-dimensional dataset and classify the extracted pattern into a low-dimensional (one or two) grid pattern, where similar inputs stay close together while maintaining data structure (Vesanto and Alhoniemi, 2000). Thus, SOM can be interpreted as topology and can provide an insight into the system under transformation.

In topology, the neurons can be connected based on the number of clusters using either hexagonal or rectangular lattice. For visual display, Kohonen, (2001) pointed out the hexagonal lattice is to be preferred, because it does not favor horizontal and vertical directions as much as the rectangular lattice. Next, a network is constructed, trained in two phases using weight learning function (detailed information is provided in Vesanto and Alhoniemi, (2000)). For every input, weights will be calculated for each neuron correspondingly. From the obtained weight distance from each input, a weight map is obtained with bright and light colours. The bright colour represents a smoother distribution of samples from cluster to cluster and on the other side the dark colour represents a high variation between the samples from cluster to

cluster, and only a few samples will be assigned to that particular cluster. Finally, from the weight distances of each input, sample hits will be obtained, and it displays the numbers of samples assigned to each cluster, respectively.

3.5 Proposed methodology

The schematic of the proposed methodology is shown in Fig. 2, and the step by step procedure is described as follows.



Superscript refers to time scale of evaluation (a- annual, s- seasonal and m-monthly) and in subscript 'i' refers to years 1901 to 2013, 'j' = 1, 1-4 and 1-12 for annual, seasonal (Spring, summer, fall, winter) and monthly time-scales (January, February, March, April, May, June, July, August, September, October, November, December)

Figure 2: Flow diagram of the framework to regionalize precipitation.

Step-1: Daily data is converted to annual, seasonal (spring, summer, fall, winter), and monthly (January to December) time-series (as described in section 3.1). A total of 17 time-series with a data length of 113 would be formed, and data encloses the total precipitation of the particular time scale during the year and over the years (see Table.1).

Step-2: Calculate the mean precipitation (i.e., an average of the precipitation of 113 years is reported as mean precipitation) and ME for each time series respectively.

Step-3: From the obtained ME, SVI is calculated correspondingly for each time series. For each grid point, there are 37 inputs (mean precipitation (17) +SVI (17) +location parameters (3)). Different combinations of 37 inputs are tried (not shown here), and it is disclosed that all the 37 inputs play a useful role in regionalization.

Step-4: Regionalization using self-Organizing maps. Formation of the matrix with samples as rows and inputs as columns. In the present study, 4443 samples (total number of grid points) with 37 input vectors (described in step 3) for each sample are considered. Thus, the size of the matrix is 4443 X 37. Based on the number of clusters, the network is

constructed, trained and sample hits will be obtained for each cluster, respectively. Further, the obtained clusters are to be validated for homogeneity and thereby, an optimum number of clusters is considered for further analysis.

3.6 Cluster Validation index

Silhouette coefficient (SC) is employed to decide the optimum cluster number, as it is robust in portioning the data into clusters by combing two techniques, i.e. separation and cohesion. The advantages and limitations of SC are nicely presented in Rousseeuw (1987). Silhouette Coefficient of a cluster can specify the degree of similarity of grid points inside a cluster (cohesion) compared to other clusters (separation). Cohesion measures the intra-distance ($X(i)$) from one data point (i) to remaining data points within the cluster, and separation measures the inter-distance ($Y(i)$ from the data point (i) to data points falling under other clusters (for calculation procedure of X and Y see supplementary information S1). For each data point (i), SC is calculated as

$$SC(i) = \begin{cases} 1 - \frac{X(i)}{Y(i)}, & if\ X(i) < Y(i) \\ 0, & if\ X(i) = Y(i) \\ \frac{Y(i)}{X(i)} - 1, & if\ X(i) > Y(i) \end{cases} \quad (5)$$

From the above equation, for every data point (i), the range of SC(i) is -1 to +1. It says that intra-cluster distance should be less, and inter-cluster distance should be high to get well-separated clusters (positive values). On the other hand, more the intra-cluster distance and less the inter-cluster distance which leads to the formation of an incorrect cluster (negative values, i.e. presence of similar kinds of data points in different clusters). SC for the cluster is represented by an average of all SC's obtained within the cluster and is denoted as average Silhouette width. The average Silhouette width provides an evaluation of clustering validity and helpful in selecting the optimum number of clusters. As the homogeneous regions are not known initially, we applied SOM by changing the cluster number (k) from 3 to 25 following Mannan et al. (2018) and Satyanarayana and Srinivas (2008).

3.7 Assessment of Clusters

This section deals with the assessment of clusters in terms of seasonality, timing, and magnitude of 50th percentile pentad precipitation which can be applied in understanding the temporal evolution of clusters.

Seasonality

The seasonality was explored in terms of pentad precipitation distribution throughout the year following Shanmugasundaram and Lee (2018). Pentad time-scale is used as an alternative of daily time-scale to justify the formation of clusters. Since Ruiz-Barradas and Nigam (2013) pointed out that pentad time-scale can suppress high daily variations and thereby highlight some key intra-seasonal characteristics. The pentad is the summation of 5 day's precipitation, the 1st pentad corresponding to 1st to 5th January and 73rd pentad corresponding to 27th to 31st December,

in total 73 data points (a.k.a. pentads) for a year (dates corresponding to each pentad number is given in Table S1). The calculation procedure for the formation of pentad time series over the years for all the grid points is given in Appendix B.

Timing and magnitude of peak precipitation

Feng et al. (2013) developed the concept of the first moment of area to estimate the magnitude and timing of peak precipitation (used by Pascale et al. (2015) and Sahany et al. (2018)). In order to understand the precipitation distribution and magnitudes of precipitation in a cluster, we used the concept of the centroid of the pentad. The centroid was computed using Eq. (6) and it would indicate the timing of occurrence of the 50th percentile precipitation (in terms of pentad number) and the corresponding precipitation amount (mm/pentad), respectively (Fig.3).

$$C_i = \frac{\sum_{p=1}^{73} p r_p}{\sum_{p=1}^{73} r_p} \qquad (6)$$

In Equation 3, p represents the pentad number from 1 to 73, and r_p represents the corresponding amount of precipitation during the pentad number.

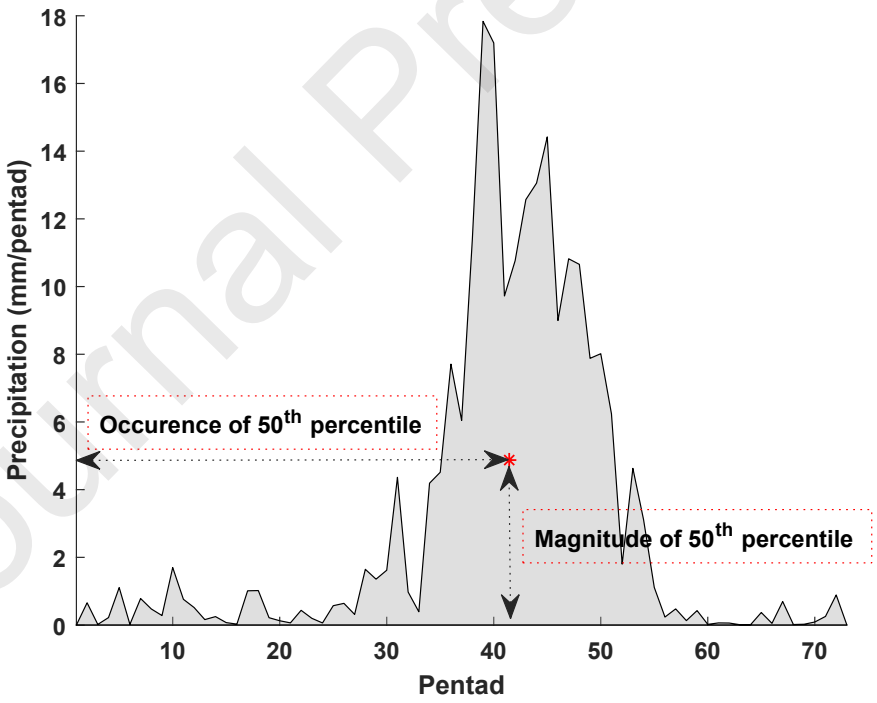


Figure 3: Computation of the centroid associated with the area of precipitation. Red asterisk marker (centroid) represents the occurrence of 50th percentile of precipitation and its corresponding magnitude.

4 Results

First, we demonstrate the potential of SVI to unravel the precipitation variability at different time-scales. Following that the identification of an optimal number of clusters is presented. We further analyze each identified cluster for its geographical and climatological characteristics. Besides, seasonality and temporal variability of the clusters are highlighted. Lastly, the temporal evolution of clusters is presented.

4.1 SVI revealed precipitation variability at different time-scales

To quantify the inter-annual variability of annual, seasonal, and monthly precipitation, the marginal entropy and SVI, respectively, are calculated based on the amount of precipitation over the period 1901–2013 for every single grid location. SVI provides an easy inter-comparison of variability across different time-scales. The spatial distribution of mean precipitation is plotted in Figs. S2-S6 and a brief description is provided in Supplementary Information S2. Fig. 4 (a) shows the spatial representation of ten randomly selected grid points scattered all over the Indian mainland to compare the SVI values (Fig. 4b) at annual, seasonal, and monthly time-scales for each station. We observed that the variability at annual time-scale was comparatively low and was almost similar for all the locations. At seasonal time-scale, the variability was more than annual and less than the monthly time-scale, and it was in congruence with the general understanding that the variability dampens from finer to coarser time-scales.

On comparing individual seasons, SVI reveals that the precipitation of summer season contributed less and that of winter contributed the highest to the annual variability. This is an interesting observation to quantify the annual variability of Indian precipitation in its seasonal components. The lowest contribution to annual variability is because summer precipitation contributes 75 % to the annual precipitation (Malik et al., 2016; Yadav and Roxy, 2019) and winter being the driest season of the year makes a low contribution to the annual precipitation.

In the winter season, except for the southern region (example grid point 10 in Fig.4) and Himalayan region (example grid point 1 in Fig.4), the rest of the country has high variability. This is also in congruence with the general understanding. The variability is low in the southern part, since it receives regular precipitation from the northeast monsoon (Rajeevan et al., 2012; Sreekala et al., 2012) and in the case of the Himalayan region, trade winds travel from the Mediterranean Sea (Western disturbances) causing significant precipitation during the winter season (Shukla et al., 2019). However, the other parts of the country (grid point 2 to 8 in Fig.4) irregularly receive unseasonal rains with no uniform pattern causing high variability (Rajeevan et al., 2012; Sreekala et al., 2012).

Further, intra-variability of months within season helps in understanding the months responsible for the variability of that particular season. There is a distinct variation of variability for the same month (except JJAS) at different locations, due to the intricate climatological pattern of precipitation.

a)

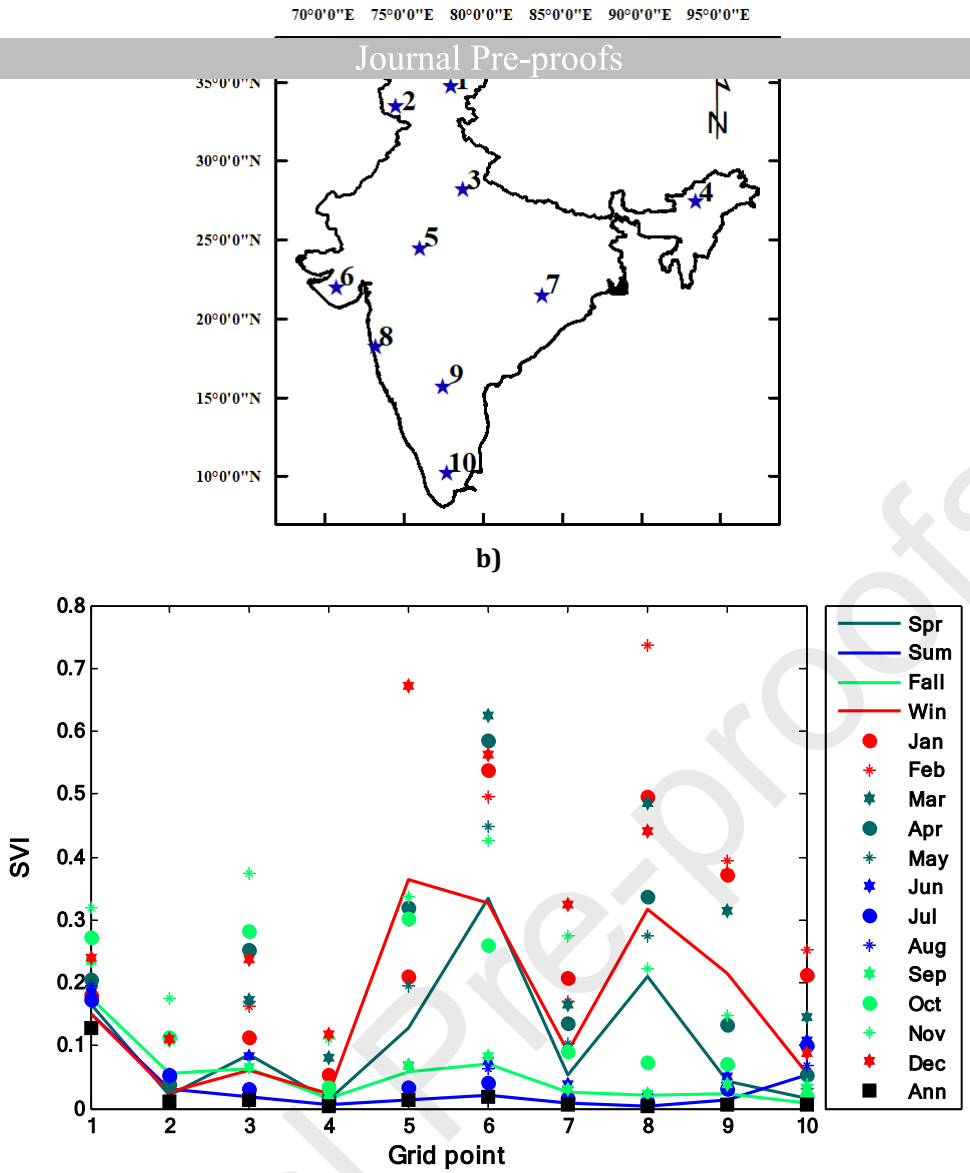


Figure 4: (a) Spatial representation of ten randomly selected grid points scattering all over the Indian mainland to illustrate the (b) variation of SVI at multiple time scales (annual, seasonal, and monthly). Annual time-scale results are indicated with 'square' marker, seasonal time-scale results with lines. Similarly, corresponding months of each seasonal time-scale are indicated with similar colors.

Comparison of SVI in the box plot discloses that the variability of seasonal time series is always lower than that of its constituent months (see Fig. 5). For instance, the mean of SVI for the spring season (0.09) is lower than that of its constituent months, i.e. March (0.232), April (0.193), May (0.143). A similar pattern is seen in other seasons; variability dampens from finer to coarser time-scales. Hence, considering finer time scales, variability information can be captured in clustering and forms well-separated clusters.

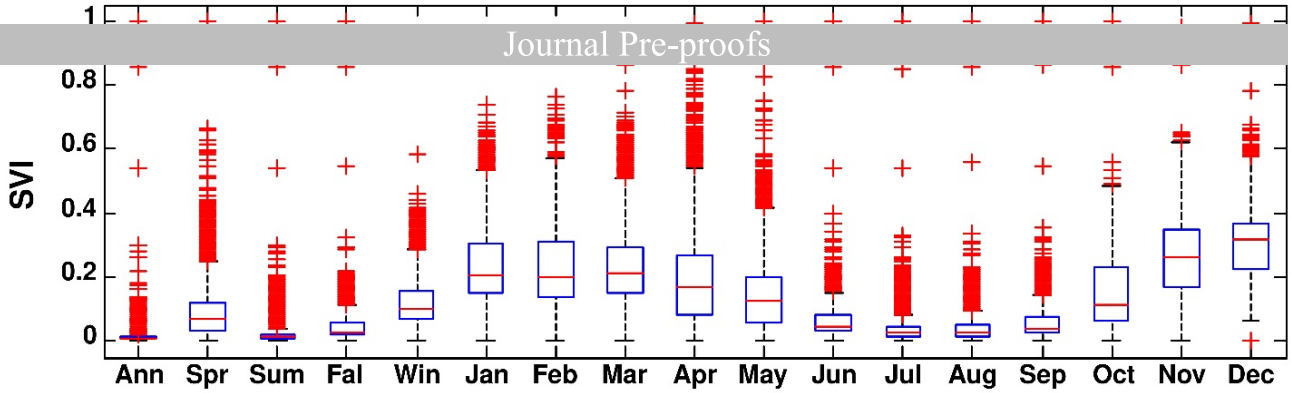


Figure 5: Box plot of the SVI for all the grid points at multiple time scales (annual, seasonal, and monthly)

4.2 Optimal number of clusters

For spatial clustering of gridded data over India, the mean precipitation and corresponding SVI values at multiple time scales were used as a basis for the clustering of grid points. The average Silhouette width revealed (see Fig.6) that probable cluster number was either 6, 10 or 14 with the corresponding values 0.4589, 0.4647, and 0.4499, respectively. It is observed that on working with a total of 14 clusters, some of the clusters have very few (one or two) stations. This can be one of the reasons that make the average Silhouette value fall to a low value suddenly. Also, on running the algorithm multiple times, the clusters were highly unstable. Now, the remaining two likelihoods for the optimal number of clusters was thoroughly investigated for seasonality and SVI pattern. In the case of selecting total 6 clusters, we noticed that the interregional heterogeneity was missing, i.e. properties of grid points within one cluster were matching the grid points in other clusters. On selecting 10 clusters, the intraregional homogeneity and interregional heterogeneity of grid points were maintained, and in earlier studies (Fukushima et al., 2019; Mannan et al., 2018) also the chosen cluster number was nearer to ten. Besides, the 10 clusters represent different climate regimes that exist in India. Therefore, we chose ten as the optimum number of clusters. Also, this choice was having the highest Silhouette coefficient, as well as the clusters were highly stable.

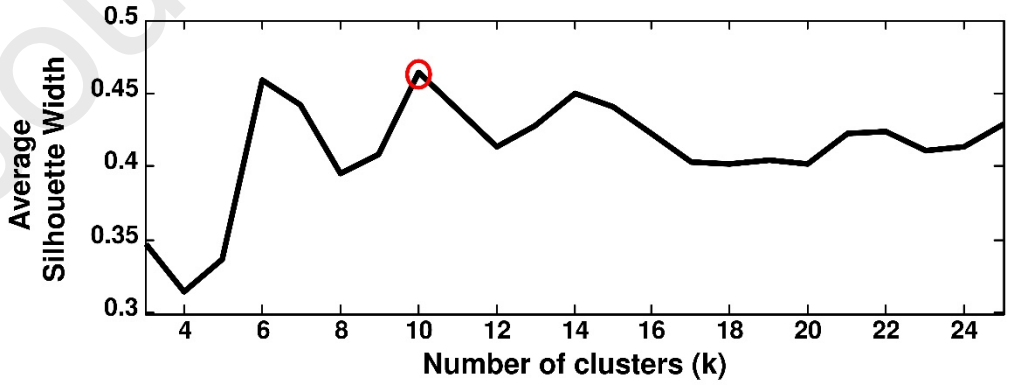


Figure 6: Silhouette analysis for SVI-SOM approach with a number of clusters (k) =3 to 25, and a circular mark at 10 represents the optimum cluster number.

4.3 Geographical and climatological characteristics of clusters

Based on the 37 input vectors (as described in the step-3 of section 3.5), grid points were segregated into 10 clusters (Fig. 7). The cluster structure, identified by precipitation magnitude and its temporal variability analysis, shared specific elements which can be tied back to two general categories, such as climate characteristics and physical characteristics. Therefore, the number of clusters reflects the climatological diversity of Indian subcontinent, and the number of grid points per cluster sets the extent to which each distinct climatology "family" is sampled. The spatial extent of the grid points in the regions prominently shows the ability of the method to capture the underlying driving forces. For instance, we do not impose any spatial constrains on the cluster classification, there is no guarantee that the region will be geographically cohesive, but as shown in Fig. 7, this is often the case.

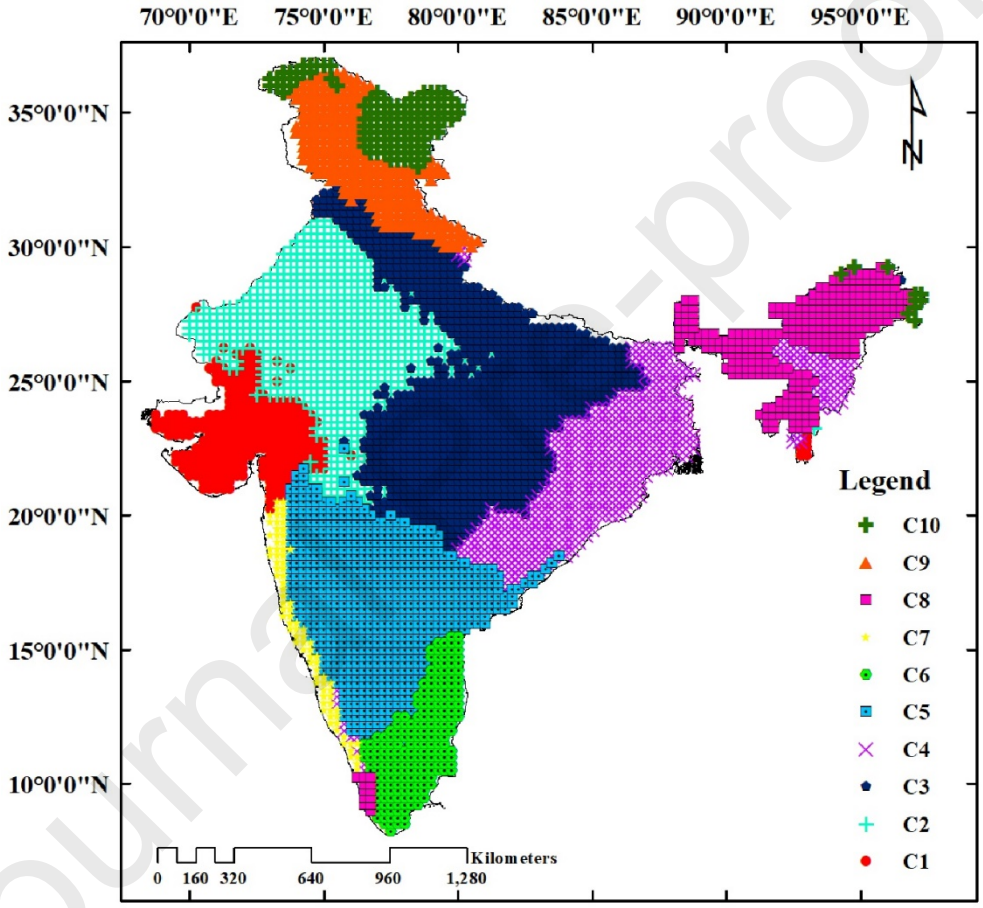


Figure 7: Geographical representation of grid points belonging to each region.

The geographical representation of cluster structure (Fig.7) confirms some similar patterns to those provided by earlier studies (Fukushima et al., 2019; Mannan et al., 2018). It is also essential to emphasize that the formation of clusters using self-Organizing maps is based on a similar kind of neurons representing precipitation magnitude and its temporal variability, rather than on our traditional measures like nearest neighbor's, and linear correlations. Precipitation grid points belonging to clusters 2, 3, 4, 5, 6, 7, and 9 (see Fig.6) are concentrated, i.e. exhibit geographical contiguity. On the other hand, certain grid points specifically belonging to clusters 1, 4, 8, and 10 are geographically separated,

demonstrating that apart from geographical contiguity the approach accounts for other additional governing mechanisms, temporal variability of precipitation or its magnitude at multiple time scales, local variability and large-scale climate drivers which might influence the precipitation pattern.

Table 2 shows the number of grid points, topographical, climate regime and statistical analysis of resultant clusters that include the mean and standard deviation of annual precipitation for each cluster. As per Koppen climate classification, the climate regimes existing in India are Warm hot desert, Semi-arid, Humid Subtropical, Tropical savannah, Semi-arid, Tropical monsoon and Alpine. The type of elevation and climate regime are interpreted as per the literature (Agarwal et al., 2018; Beck et al., 2018; Dimitrova and Bora, 2020). High mean precipitation illustrates the total amount of precipitation and higher standard deviation indicates greater variation among the grid points in the cluster (Hsu and Li, 2010).

Table 2: Number of grid points, a **summary of topographical and statistical analysis for each cluster.**

C. No.	Number of grid points	Mean annual precipitation (mm)	Std. Dev (mm)	Topographical analysis		
				Type of elevation	Climate regime	Cluster name
1	298	679	306	Low	Warm desert and semi-arid	Western India
2	722	562	238	Moderate	Warm desert and semi-arid	North-western India
3	1058	1135	216	Moderate	Humid subtropical	North-central India
4	580	1470	190	Moderate	Tropical savannah	Eastern India
5	734	795	226	Moderate	Semi-arid	South-central India
6	223	1003	296	Low	Warm and humid subtropical	South-eastern coastline
7	84	3169	712	low	Tropical monsoon	Konkan Coast
8	312	2775	586	High	Humid subtropical	North-eastern India
9	282	1081	376	Very high	Alpine	Rain-belt Western Himalayan
10	150	460	269	Highest	Alpine	Rain-shadow Western Himalayan

From statistical analysis, Clusters 1 and 2 (in Fig. 7) have comparable precipitation characteristics but are differentiated by elevation and coastlines features, have warm desert and semi-arid climate regime (Western and North-western India). Cluster-3 (in Fig.7) is the largest cluster enclosing 1058 grid points in that region and receives mean annual precipitation (1135 mm) almost nearer to all-India average precipitation (1119 mm) with a variation of 216 mm among the grid points in that cluster, has moderate elevation covering humid subtropical climate regime (North-central India). Cluster-4 (in Fig.7) receives a high amount of mean annual precipitation (1470 mm) with the lowest variation among the clusters (190 mm), and has a tropical savannah climate regime (Eastern India). Cluster-5 (in Fig.7) receives low mean annual precipitation (795 mm) with low variation (226 mm), has moderate elevation and semi-arid type of climate regime (South-central India). Cluster-6 (in Fig.7) receives high mean annual precipitation (1003 mm) with a variation of 296 mm, has low elevation, and warm and humid subtropical climate regime (South-eastern coastline). Cluster 7 (in Fig.7), is the smallest cluster having only 84 grid points, has low elevation, and has tropical monsoon climate regime (Konkan coast), has the highest mean annual precipitation (3169 mm) with highest variation (712 mm) in that cluster (Table 2). One likely reason for high variation could be that this cluster is positioned nearer to coastlines and lee side of Western Ghats. Cluster 8 (in Fig.7) receives 2nd highest mean annual precipitation (2775 mm) with a variation of 586 mm among the grid points in that cluster and covers humid subtropical climate regime (North-eastern India). Meanwhile, cluster 10 (in Fig. 7), covering alpine climate regime (Rain-shadow Western Himalayan), shows the lowest mean annual precipitation (460 mm) with a variation of 269 mm. Comparing the mean annual precipitation of clusters 9 and 10 (in Fig.7), Rain-belt Western Himalayan receives high mean annual precipitation than Rain-shadow Western Himalayas. In general, most of the regionalization approaches (Agarwal et al., 2018; Fukushima et al., 2019) classified the entire Western Himalayas as one, however in the present study, Western Himalayas become two distinct regions (Rain-belt and Rain-shadow) due to the differences in the mean annual precipitation.

4.4 Seasonality and temporal variability of the clusters

The grid points within the individual clusters were further analyzed for any common characteristics in terms of seasonality and temporal variability they may have among themselves following Agarwal et al., (2016b). Fig. 8 shows the pentad precipitation for 50th percentile and interquartile (25th-75th percentile) range for each cluster. It is interesting to note that the shape of seasonality is unique for each cluster. The precipitation pattern of every cluster during the Indian summer monsoon (ISM, i.e. in between pentad numbers 31 (31st May-4th June) and 54 (23rd-27th September)) and retreat phase of monsoon; northeast monsoon (NEM, i.e. in between pentad numbers 55 (23rd-27th September), and 69 (7th-11th Dec)) is discussed further in detail.

The ISM influence is visible in all the regions (Fig. 8), the peak points of pentad precipitation can be seen in all the regions except in the south-eastern coastlines (C6) and highest elevation Rain-shadow Western Himalayas (C10). In continuation,

the influence is strongest and phenomenal for the Konkan Coast (C7). This is in line with a general understanding that the inflow of the south-west monsoon into India brings about a total change in the weather. Early in this season, the windward side of the Konkan Coast receives very high precipitation during the ISM (June-September) with mean precipitation over some parts exceeding 2500 mm. Extreme events with precipitation more than 150 mm per day at one or more grid points along this region occur frequently and cause considerable damage. During the onset phase of ISM over the coastal area of Arabian Sea, there will be heavy precipitation when compared to other regions, and it is one of the prominent characteristics of ISM (Ananthakrishnan and Soman, 1988).

Similarly, Western India (C1), North-western India (C2), North-central India (C3), and Eastern India (C4)) receives significant precipitation during the ISM (Malik et al., 2016). Also, the magnitudes of the pentad are similar in the three regions, except North-western India (C2) where magnitudes of the pentad in that region are almost half of North-Central India (C3) but the seasonality pattern of both the regions is nearly similar.

Further, Rain-belt Western Himalayas (C9) and Rain-shadow Western Himalayas (C10) receive high precipitation during January to March when comparing with ISM months, Yadav et al. (2007) reported, western disturbances travelling from Mediterranean sea towards Tibetan plateau obstructed by Himalayas arc causing precipitation in this region. Furthermore, we observe that NEM dominates only in the south-eastern coastlines (C6), and this is the unique region where the contribution of NEM to annual budget is more than ISM. The average quantity of precipitation of NEM in south-eastern coastlines is nearly one-fifth of ISM in Konkan Coast. South-central India (C5) which has moderate elevation and semi-arid climate regime also experienced some part of NEM in addition to SWM.

There is a unique region, i.e. north-eastern India (C8), that receives precipitation in nine months (March to November) in a year. The seasonality pattern of the region is similar to the shape of a normal distribution, having a high magnitude in July and covers pre-monsoon, ISM and NEM.

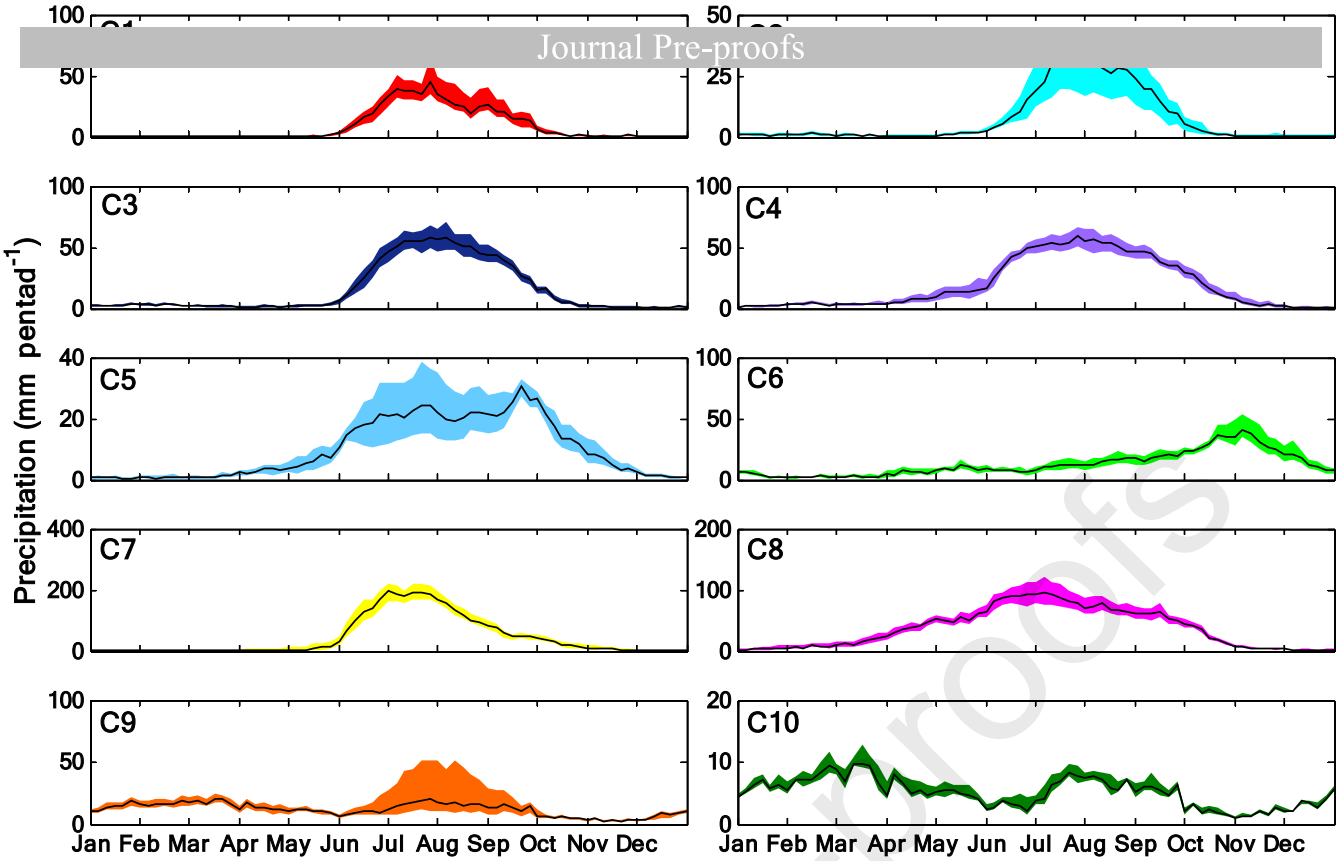


Figure 8: Seasonality of the pentad precipitation for each cluster. The black line represents the 50th percentile pentad precipitation distribution throughout the year for each cluster. The color band represents Inter-Quartile Range (25 to 75 percentile) and is colored in the same way as Fig.7 to aid comparison.

To further understand the clusters, mean monthly precipitation (mm) of January to December and corresponding SVI from 1901 to 2013 was calculated for all the regions (see Tables. S2 and S3). Fig. 9 illustrates the average monthly (bars) precipitation and along with the corresponding SVI (asterisk marker) and these can be considered as typical values for that region. The distribution of monthly precipitation is unique, and explicitly regional characteristics are captured for each region. During January, February, and March; rain-belt Western Himalayas (Fig .9) receives the highest precipitation from western disturbances with low variability. In April, May, and September; north-eastern India (Fig. 9) is dominant in receiving the highest precipitation in conjunction with low variability during monsoon and its retreat phase. In the case of June, July, and August (monsoon months); Konkan Coast (Fig.9) is the dominant cluster among others and in particular, in July, this region receives the highest precipitation of 1125 mm with SVI = 0.017. During the post-monsoon months (i.e., in October, November, and December), the csouth-eastern coastline (Fig.9) receives the highest amount of precipitation concurrently with low variability.

In Fig.9, north-western India and north-central India have similar SVI patterns across months but different magnitudes of monthly precipitation, making the two regions distinct. Similarly, the SVI pattern of eastern India and north-eastern

India; south-central India and Konkan Coast, and Rain-belt and Rain-shadow Western Himalayas are similar, but their corresponding monthly precipitation magnitudes are different forming well-separated clusters. On the other side, western India and north-western India have similar monthly precipitation patterns and different SVI magnitudes, and thereby, it says, considering SVI in regionalization also plays a vital role in forming homogeneous regions.

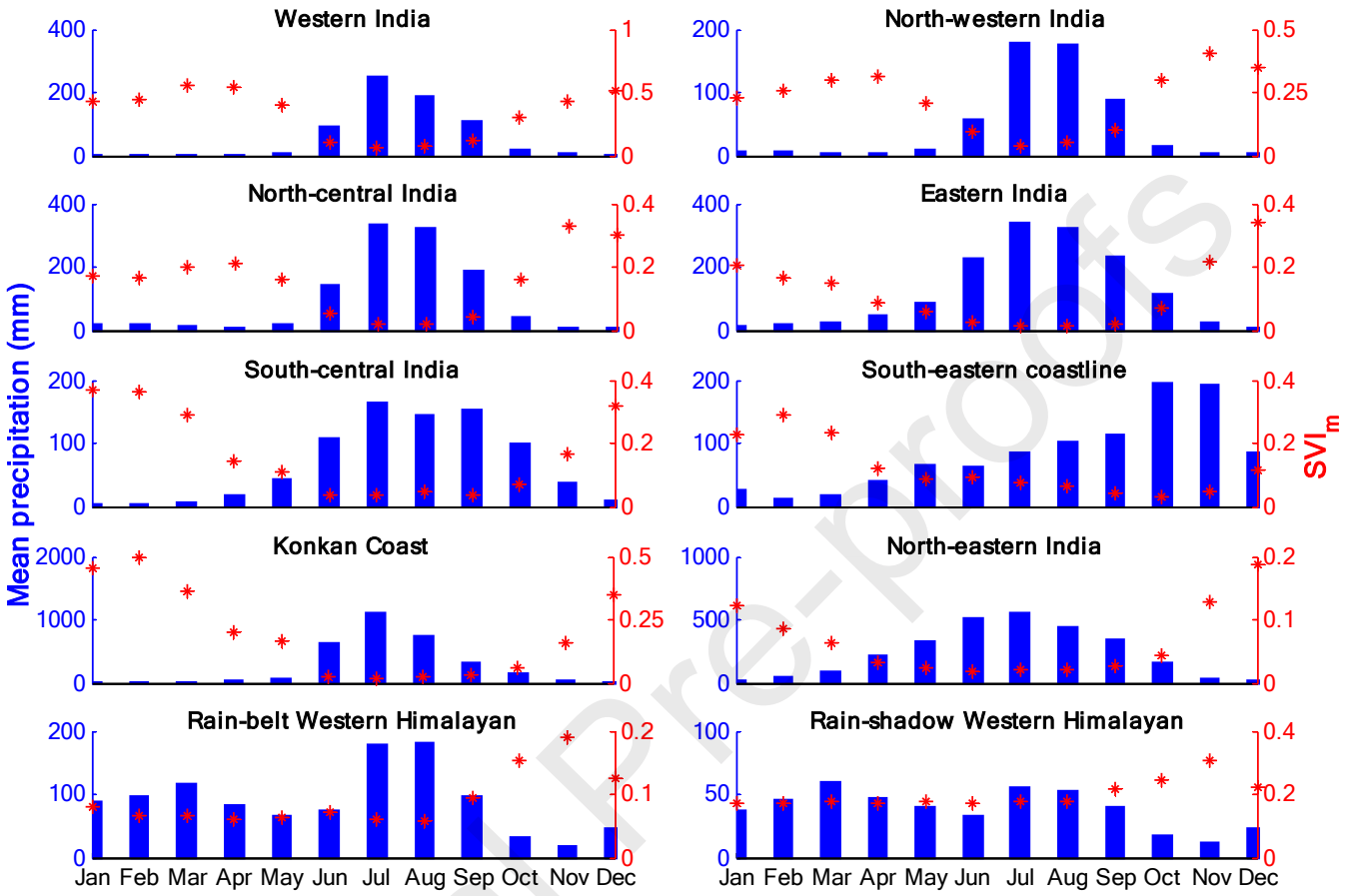
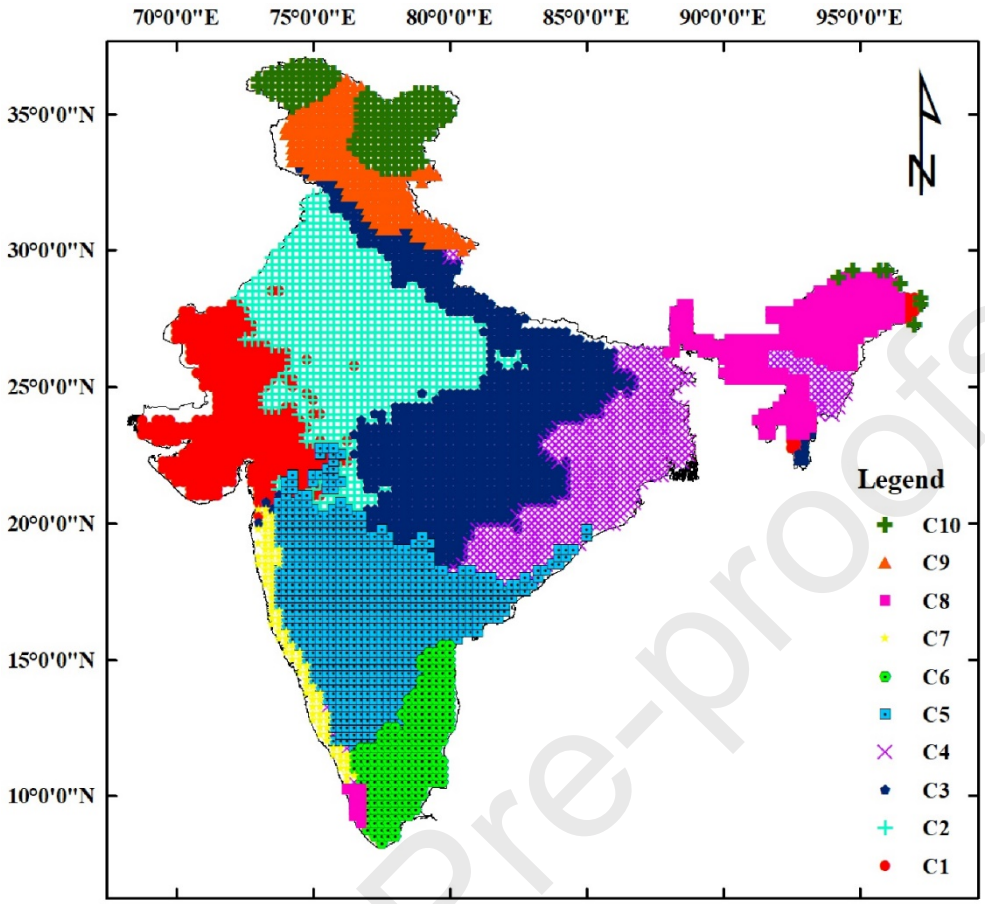


Figure 9: Mean annual, seasonal and monthly precipitation along with its corresponding SVI values for the 10 clusters. The abscissa represents the time scale, and the corresponding mean precipitation (primary axis-blue color) and SVI (Secondary axis-red color) are denoted by the ordinate.

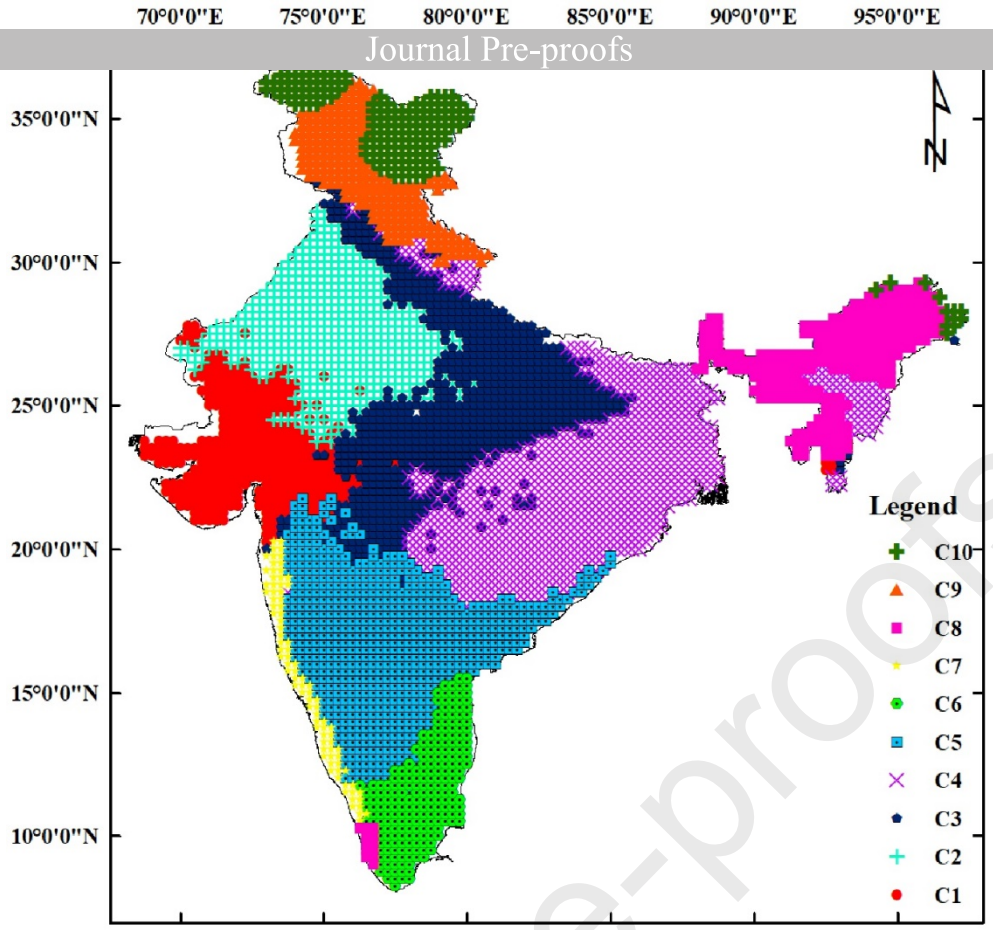
4.5 The temporal evolution of clusters

The complete data set, i.e. 1901-2013 was divided into three time windows (the span of the first two time windows was 40 years each and the remaining 33 years was the third time window) following Agarwal et al., (2016a) and Mannan et al., (2018). The same clustering approach was followed for each window, and results were compared to unravel the time-varying characteristics of precipitation and the stability of clusters. Fig. 10 compares the spatial representation of regions formed in the clustering analysis. Interestingly, there was no change in the cluster characteristics for south-eastern coastlines (cluster 6), north-eastern India (cluster 8), rain-belt and rain-shadow Western Himalayas (clusters 9 and10) in the three temporal windows, while there was a movement of grid points across south-central India (cluster 5), western

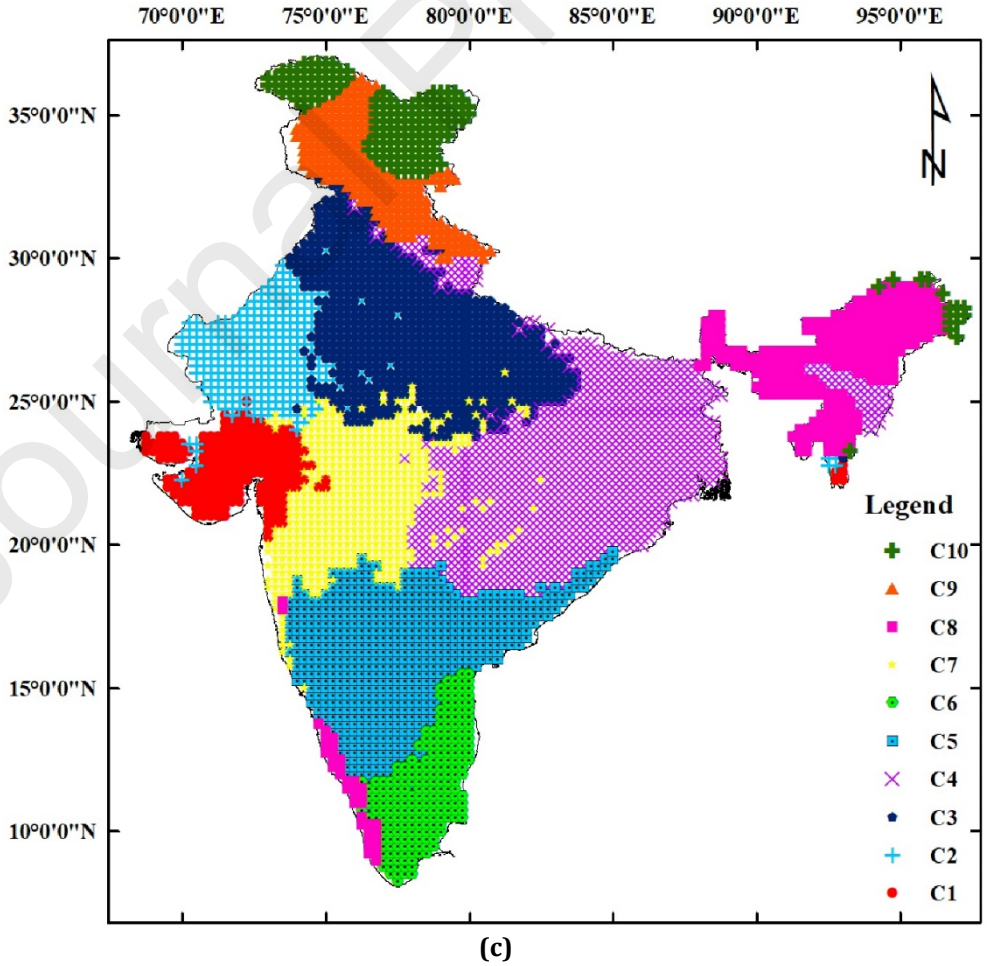
India (cluster 1), north-western India (cluster 2), north-central India (cluster 3), eastern India (cluster 4), and Konkan Coast (cluster 7).



(a)



(b)



(c)

Fig ent temporal windows a) 1901-1940, b) 1941-1980, and c) 1981-2013. The clusters formed in the regions of southern India, Himalayan regions, and north-eastern are similar in all the three temporal windows. On the contrary, the region bounded between 20^0N-30^0N and 70^0E-80^0E is dynamic; the majority of grid points are fluctuating among the regions. In the third temporal window, a new cluster is being formed due to the large spatial heterogeneity in precipitation in recent decades.

To examine the stability of clusters across western India, north-western India, north-central India, eastern India, three hotspots (A, B, and C) were identified (see Fig.11), based on the shifting of grids between the clusters from 1st time window to 3rd time window in Figs. 9a to 9c. Further investigation was for all the grids which shifted; however, the results are shown for 12 grid points in each hotspot. For every grid point in the identified hotspot the pentad distribution during the Indian summer monsoon months was plotted for the three temporal windows.

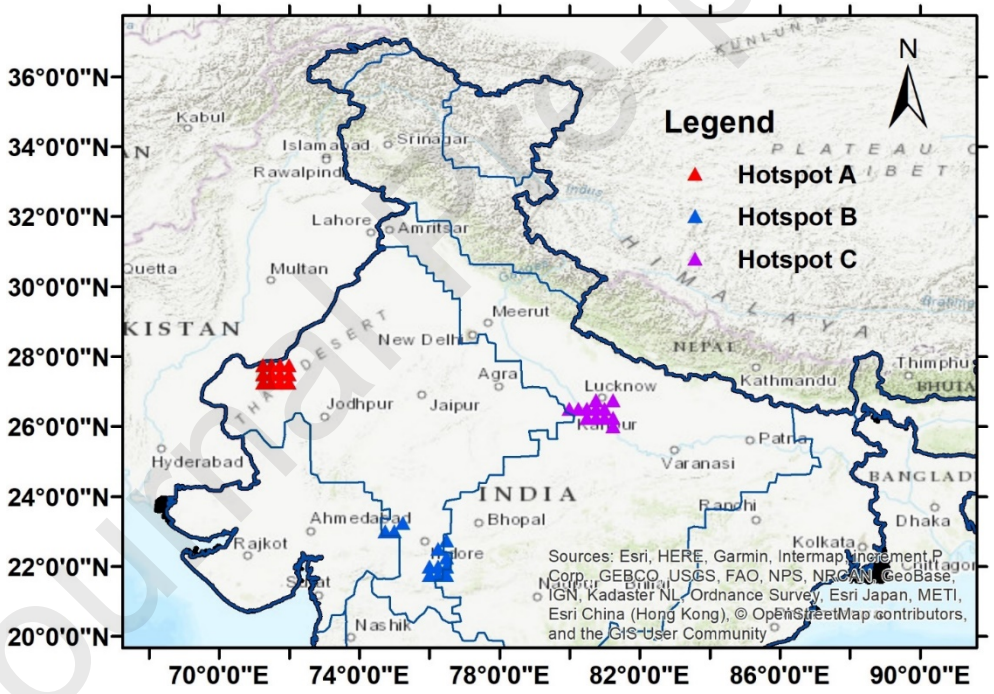


Figure 11: Geographical locations of the hotspots A, B and C. Each Hotspot contains 12 grid points, starts from Bottom left and ends at top right. Hotspot means a small area which is shown for movement among the regions from 1st time window to 3rd time window in Figs. 9a to 9c

Figs.12 to 14 illustrate the shifting and fluctuating pentad distribution in the three temporal windows (1901-1940 (black line), 1941-1980 (blue line) and 1981-2013 (red line)) and it revealed an essential facet of the Indian summer monsoon.

In hotspot A (Fig.12), there was a shift in the pentad distribution from 1st temporal window to 3rd temporal window. In the 1st temporal window, the highest pentad precipitaiton was observed in August, but in the 2nd and 3rd temporal window, the highest shifted towards July, disclosing changes in intra-annual variability. In hotspot B (Fig. 13), there was an increase in the strength of monsoon from 1st temporal window to last. On the contrary, there was a decrease in the strength of monsoon in hotspot C (Fig.14).

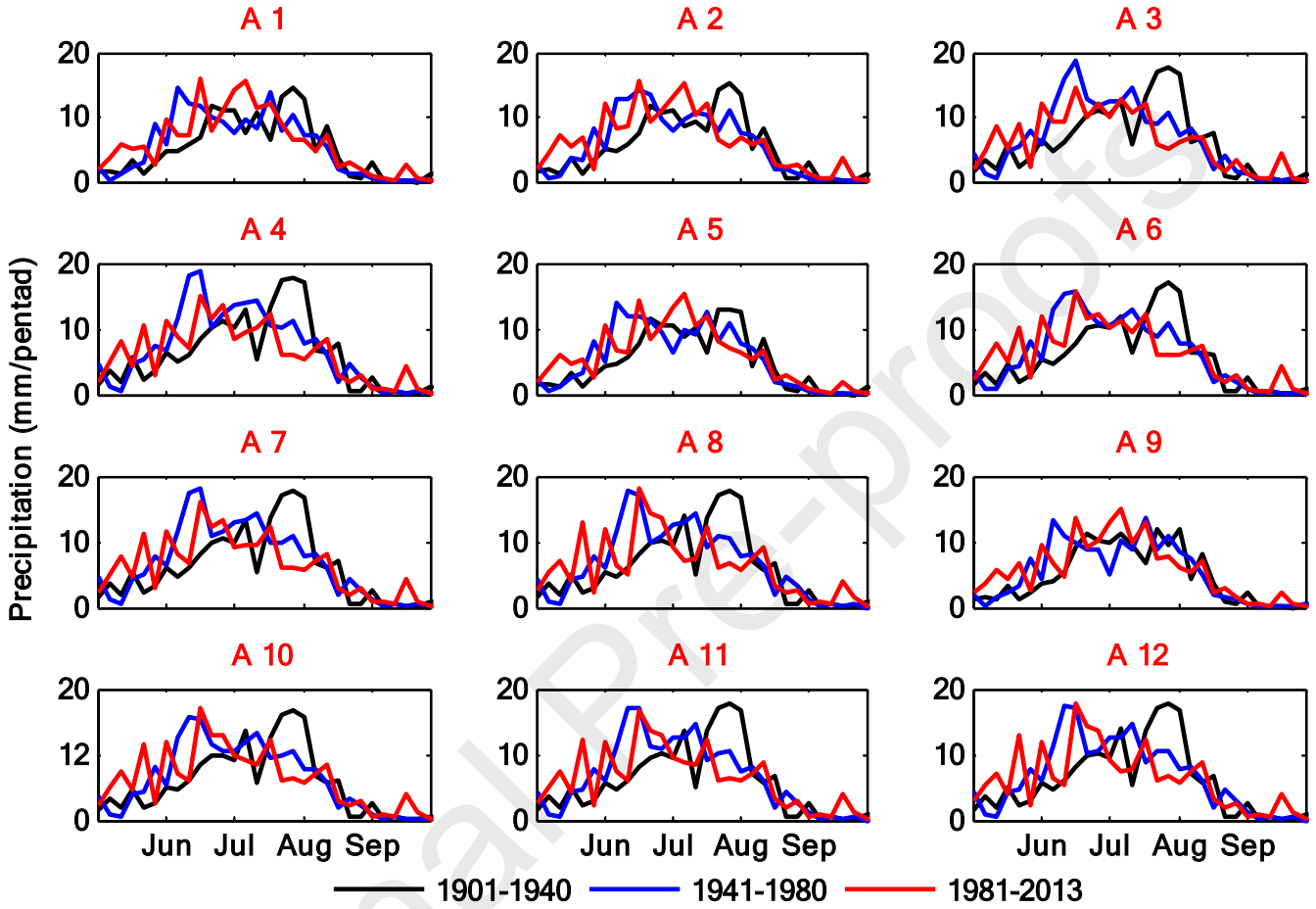


Figure 12: Seasonality of pentad distribution for hotspot-A (12 grid points) for the three temporal windows during the months of the Indian summer monsoon.

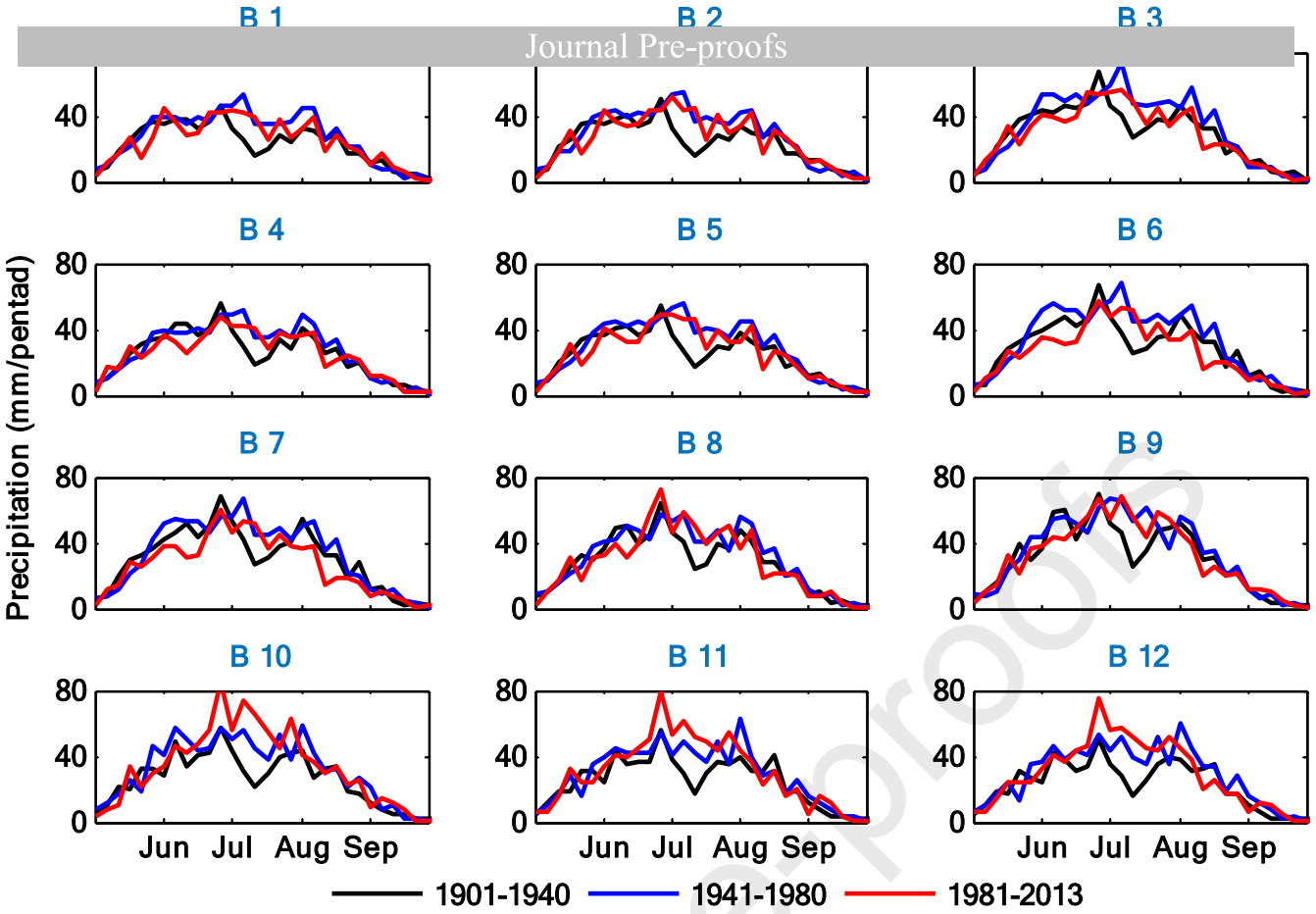


Figure 13: Seasonality of pentad distribution for hotspot-B (12 grid points) for the three temporal windows during the months of the Indian summer monsoon.

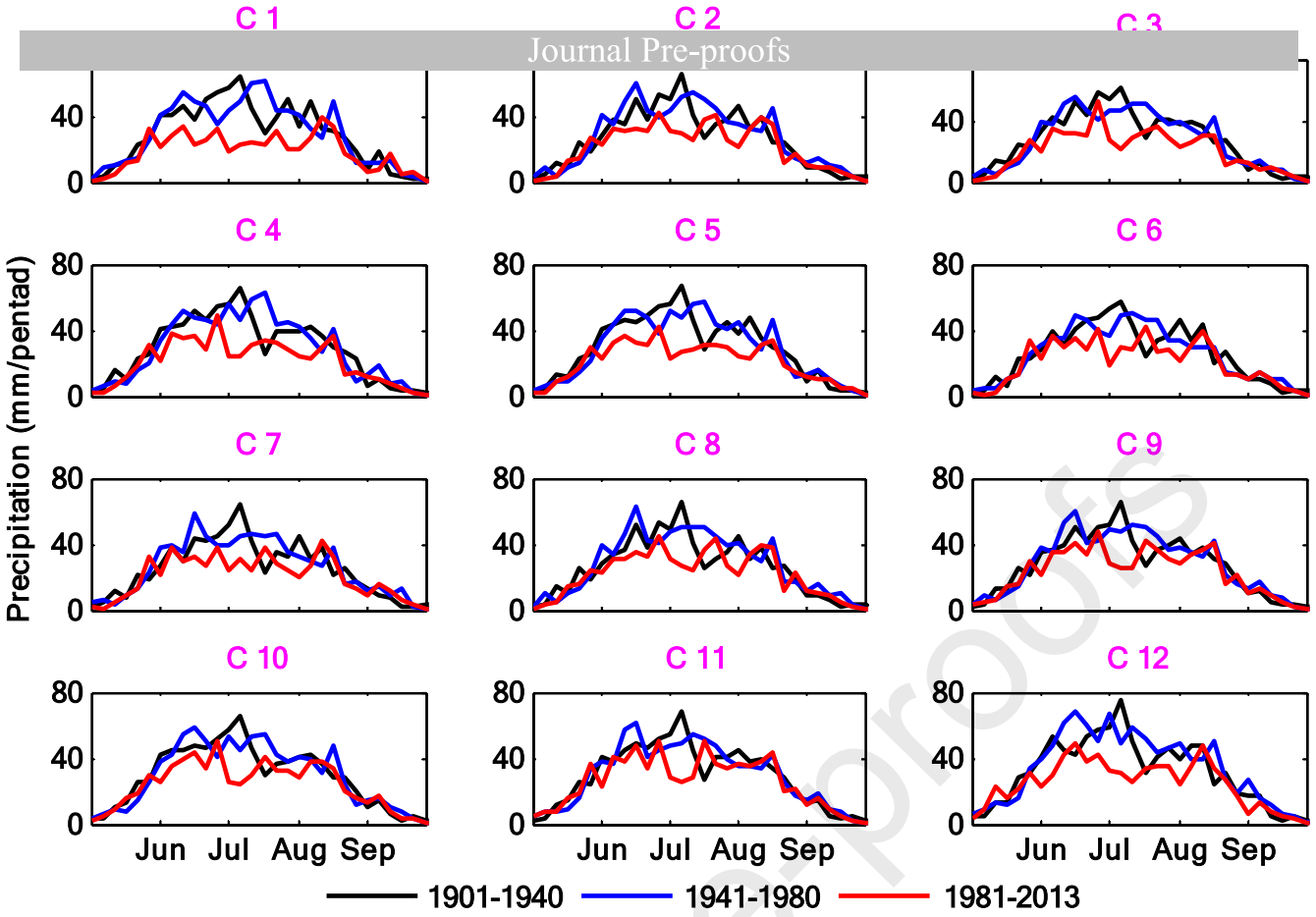


Figure 14: Seasonality of pentad distribution for hotspot-C (12 grid points) for the three temporal windows during the months of the Indian summer monsoon.

Further, the first moment of area (centroid) was calculated for every grid point following Feng et al., (2013). Fig. 15 shows the plot of the average values (over each time window) of time of occurrence of the 50th percentile precipitation and the corresponding precipitation amount for all the three hotspots.

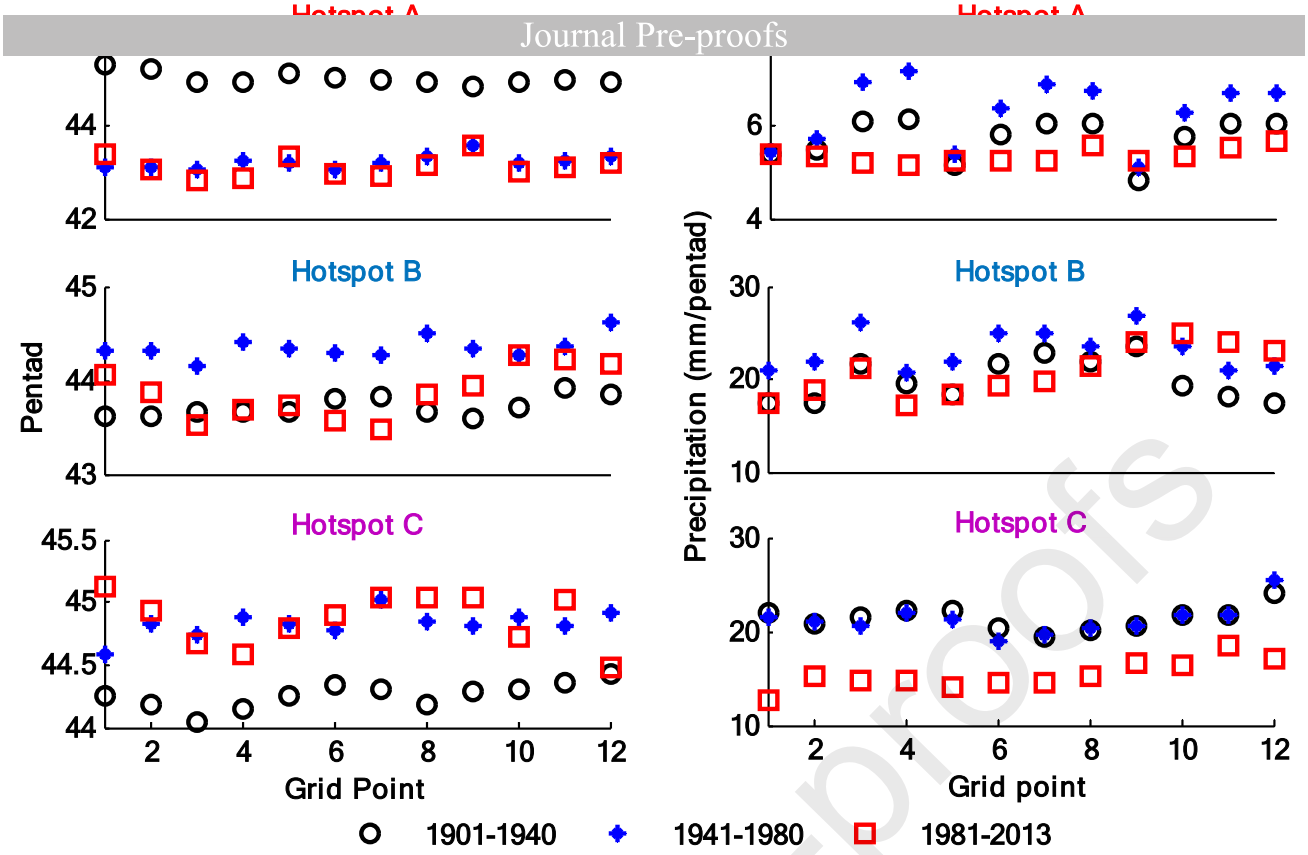


Figure 15: Centroid (Pentad, Precipitation(mm/pentad) of the pentad distribution for every grid point in Hotspot A, B and C, respectively for the three temporal windows.

In hotspot A, the time of occurrence of the 50th percentile shifted 10 days (i.e. from pentad number 45 to 43) from 1st to 2nd time window, and it was consistent in the 3rd time window. The significance testing was done using the two-tailed t-test for testing the difference in the mean values obtained from the three time windows. The results in terms of the p-values from the t-test (Table.3) revealed that the difference in the mean values from 1st to 2nd time window for all points was significant ($p<0.05$). On the other side, the p-values for the magnitude of precipitation (given in brackets) indicated no statistically significant change. In hotspot B (Fig.15 and Table 3), there was no significant change in the time of occurrence of the 50th percentile over the three windows. However, the magnitude of 50th percentile showed a significant increase in the precipitation amount at all grid points from 1st to 2nd time window. In hotspot C, a different behavior was observed wherein there was no significant shift in the precipitation timing, but a statistically significant decrease in the magnitude of 50th percentile in the 3rd time window (1980-2013). The observations on changes in the precipitation pattern were corroborating with the findings of Guntu et al. (2020a) and Sahany et al. (2018) who showed a significant trend in the variability of precipitation both in terms of quantity and timing in central India, These observations showed

that the proposed approach was sensitive in capturing the temporal changes in the precipitation magnitude and its variability, possibly due to changes in climate variability.

Table 3: Summary of t-test for change in the occurrence of 50th percentile precipitation and its magnitude (in bracket) from 1st to 2nd time window, 2nd to 3rd time window for each hotspot. Bold values represent the change is statistically significant at a 95% confidence interval.

Grid Point	Hotspot-A		Hotspot-B		Hotspot-C	
	1st - 2nd time	2nd - 3rd time	1st - 2nd time	2nd - 3rd time	1st - 2nd time	2nd - 3rd time
1	**0.004**(0.569)	0.895(0.490)	0.435(**0.017**)	0.801(0.315)	0.886(0.512)	0.874(**0.005**)
2	**0.003**(0.401)	0.612(0.627)	0.427(**0.033**)	0.647(0.329)	0.972(0.843)	0.504(**0.006**)
3	**0.011**(0.591)	0.647(0.554)	0.451(**0.019**)	0.545(**0.013**)	0.816(0.678)	0.958(**0.009**)
4	**0.025**(0.645)	0.539(0.569)	0.404(0.397)	0.906(**0.050**)	0.351(0.954)	0.671(**0.001**)
5	**0.008**(0.534)	0.708(0.527)	0.379(**0.046**)	0.928(0.190)	0.565(0.704)	0.771(**0.001**)
6	**0.002**(0.629)	0.575(0.862)	0.780(**0.042**)	0.651(**0.004**)	0.571(0.142)	0.979(0.262)
7	**0.017**(0.761)	0.488(0.885)	0.925(0.253)	0.511(**0.018**)	0.807(0.902)	0.808(0.077)
8	**0.040**(0.400)	0.352(0.400)	0.485(0.282)	0.730(0.515)	0.972(0.858)	0.798(**0.039**)
9	**0.050**(0.758)	0.370(0.462)	0.662(**0.054**)	0.975(0.348)	0.864(0.604)	0.933(**0.005**)
10	**0.006**(0.958)	0.452(0.520)	0.464(**0.019**)	0.387(0.250)	0.407(0.965)	0.785(**0.026**)
11	**0.026**(0.896)	0.477(0.649)	0.669(**0.039**)	0.906(0.279)	0.979(0.808)	0.978(0.080)
12	**0.035**(0.855)	0.393(0.481)	0.410(**0.022**)	0.757(0.759)	0.961(0.417)	0.418(**0.016**)

5 Discussion

This study developed a regionalization framework accounting for both temporal variability and magnitude of precipitation. We investigated the inter-annual variability of precipitation at monthly (Jan-Dec), seasonal (Spring, Summer, Fall, Winter), and annual time scales. The time series of gridded precipitation at different time scales was selected as a clustering variable. This means that homogeneous regions were defined, based on coherent climate variability at different time scales (monthly, seasonal and annual) rather than mean climate conditions. This is an essential step as an estimation of key variables, and grouping of the precipitation time series (Fig.7) captures the local and inter-annual variability at individual grid points, respectively, and allows regionalization irrespective of the contiguity of the grid points and the assumptions involved therein. Standardized Variability Index (SVI) was used to quantify the precipitation variability at different scales and self-Organizing maps were employed to delineate homogeneous precipitation regions. These were capable of identifying regions of common seasonality, climate sensitivity, and global mechanisms that drive variability for different regions.

The mean precipitation magnitude and its temporal variability (using SVI) were assessed over 113 years (1901-2013) for the Indian mainland. It was found that the temporal variability increased from annual time scale to monthly timescale. This is in congruence with the general understanding that the temporal variability dampens (Guntu et al., 2020a; Mishra et al., 2009) on coarser scales. Besides, at each time scale, the precipitation magnitude and its temporal variability were spatially varying (Fig.4) owing to different climatic conditions, and forms the basis for clustering. Based on this observation, ten different clusters (Fig.7) were identified that could be tied back to two general categories, such as climate characteristics and physical characteristics. The number of clusters reflects the climatological diversity of the Indian mainland, and the number of grid points per cluster sets the extent to which each distinct climatology "family" is sampled. The spatial extent of the grid points in the communities prominently shows the ability of the method to capture the underlying driving forces.

Fukushima et al. (2019) and Mannan et al. (2018) developed seven (ten) homogenous regions by considering the only magnitude of pentad precipitation (only daily precipitation) and employing hierarchical clustering and self-Organizing maps.

These studies pointed out the developed regions and well represented the seasonal characteristics of ISM and NEM.

On comparing clustered regions with the present study, three regions (i.e., south-central India (Cluster 5); south-eastern coastlines (Cluster 6); Konkan Coast (Cluster 7)) were in almost good agreement. However, the remaining seven regions turned out to be spatially different. A multitude of studies conducted in the past along these three regions did not show any indication for changes in precipitation characteristics (Ghosh et al., 2012; Goswami et al., 2006; Guntu et al., 2020a; Roxy et al., 2017). This is an interesting observation, i.e., consideration of temporal variability plays an insignificant role in the shape, size, and stability of these three homogeneous regions. Further, the discussion is limited to the remaining seven regions, which are new and different from the previous studies.

Intriguingly, the separate Rain-belt (cluster 9) and Rain-shadow (cluster 10) Himalayan cluster formed in the present study are different from the previous studies (Agarwal et al., 2018; Fukushima et al., 2019; Malik et al., 2016) where the entire Western Himalayas was classified as a single homogeneous region. However, this finding is supported by Shukla et al., (2019) highlighting that the topography plays an important role in differentiating precipitation over Rain-shadow and Rain-belt Western Himalayas. The orographic barrier (Pir Panjan range) divides the Rain-belt Western Himalayan, and the Rain-shadow Western Himalayan, i.e. the Rain-shadow region stands as a barrier to the moisture carrying monsoon winds crossing over to Tibetan plateau, thereby causing heavy to very heavy precipitation of short (3–4 h) to long (10–14 h) duration in the Rain-belt region (Nandargi and Dhar, 2011). The Pir Panjal Range of the Western Himalayas successfully intercepts the moisture from the monsoon currents; it acts as a rain belt region while the opposite side, i.e., the Rain-shadow Western Himalayan region, is a dry region. The monsoon is extremely weak over this region.

Hence, the Leh station located in Rain-shadow Western Himalayan merely receives 93 mm of precpitation in a whole year (Nandargi and Dhar, 2011).

The temporal evolution of clusters (Figs 10a to 10c) discloses an interesting observation. For instance, grid points belonging to the central India region as a whole (western India (cluster 1), north-western India (cluster 2), north-central India (cluster 3), and eastern India (cluster 4)) are time-variant during the three temporal windows. Figs. 12 to 14 indicate the changes in the precipitation characteristics of central India with time in terms of changing seasonality and magnitude of ISM (Goswami et al., 2006; Guhathakurta et al., 2015). Malik et al. (2016) found that the grid points belonging to central India are contributing 90~100% of ISM to the annual precipitation budget. Therefore Fig.15 highlights that in central India the significant changes in precipitation characteristics can be attributed only to the increasing heterogeneity in ISM in the last 50 years that has been reported in earlier studies (Ghosh et al., 2012; Vinnarasi and Dhanya, 2016). Mannan et al. (2018) pointed out that spatial heterogeneity of climatic variables over central India is also one of the causes for changes in precipitation with time. Also, Roxy et al. (2017) reported changes in frequency and magnitude of precipitation over these regions due to the increase in global warming. Recently Kurths et al. (2019) confirmed the emerging patterns and teleconnection in central Indian due to North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO) and Atlantic Multidecadal Oscillation (AMO).

The decrease of 10 days in the occurrence of 50th percentile monsoon in north-western India (Cluster 2) after the 1940s; an increase of precipitation by 5mm corresponding to the 50th percentile monsoon in western India (Cluster 1) after the 1940s; a decrease of 5mm precipitation corresponding to the 50th percentile monsoon in north-central India (Cluster 3) region after the 1980s were never reported in the previous studies. There is an effect of non-stationarity on the clusters in terms of varying magnitude and frequency with time. Interestingly, our approach is capable of capturing that. Therefore, as described in section 3.3, there is a necessity for using both precipitation magnitude and its inter-annual variability for robust homogenous regions.

Inclusion of precipitation magnitude and accounting for its temporal variability in the proposed regionalization offered more robust homogenous regions. The method was tested on a gridded precipitation dataset for two reasons: i) the availability and the access to rain gauge data is limited, and ii) gridded datasets provide an effective platform to understand the precipitation dynamics. The gridding process used to build the dataset might affect the relationships between grid points, owing to the assumptions underlying the spatial interpolation. However, these effects can be neglected considering the extent of the study area. Besides, the method is applicable to all sort of data, found to be highly effective in capturing the seasonality.

6 Conclusions

The present study has proposed a novel method for precipitation regionalization using self-organizing map coupled with multi-scale standardized variability index. The time series of gridded precipitation at different time scales is selected as a clustering variable. This means that homogeneous regions are defined based on coherent climate variability at different time scales (monthly, seasonal and annual) rather than mean climate conditions. This is important, as seasonality is a characteristic of a time series in which the data experiences regular and predictable changes that recur every calendar year due to local variability. Besides, coherent inter-annual variability suggests a common response to large-scale climate drivers. When regionalization is applied to observations at different time scales, then the approach can help identify regions of common seasonality, climate sensitivity and identify mechanisms that drive variability for each region. Application of the method to Indian precipitation offers promising results for regionalization. The following concluding remarks can be made from the study:

a) The Multi-Scale Standardized Variability Index measure captures the variability of precipitation dynamics at each grid point independently and then allows the formation of a homogeneous cluster that is not based on any prior assumptions.
b) The SVI-SOM clustering approach for precipitation regionalization is found to be robust and overcomes the limitations of existing approaches by including uncertainties of precipitation magnitude in a multi-scale approach and forms new clusters across central and northern India.
c) The temporal evolution of clusters unravels a significant change in the occurrence of 50^{th} percentile monsoon after the 1940s across north-western region (Cluster 2); a significant increase of 5 mm precipitation of the 50^{th} percentile monsoon after the 1940s across western India (Cluster 1) and decrease of 5 mm of 50^{th} percentile monsoon after the 1980s in north-central Region (Cluster 3). On the contrary, south-central India (Cluster 5), south-eastern coastlines (Cluster 6), and Konkan Coast (Cluster 7) are stable, independent of temporal variability and intriguingly, separate eastern (cluster 10) and western Himalayas (cluster 9) clusters formed due to the difference in the physiographic features.
d) The atmospheric dynamics across Central India regions could be forecasted by using a robust dynamical model to unravel large-scale drivers to local climate variability and has important implications for high performing seasonal prediction system in ungauged sites.

Applications of the proposed methodology to data from other parts of the world, with different climatic and environmental characteristics, would also help verify, and possibly strengthen and confirm, the conclusions drawn here.

Credit author statement

Ravi Kumar Guntu- preparing the data, developing codes, running the model, writing the manuscript.

R. Maheswaran and Ankit Agarwal- initial conceptualization, results interpretation, editing and reviewing

V P Singh- editing and reviewing.

Acknowledgements

Dr. RM gratefully acknowledges the funding from SERB, GOI for the project ECR/0016/1721. Dr. AA acknowledges the funding support provided by Indian Institute of Technology Roorkee through Faculty Initiation Grant number IITR/SRIC/1808/F.IG. All the authors sincerely thank the two anonymous reviewers and the editor for their constructive reviews and helpful suggestions on an earlier version of the paper.

Competing Interests

The authors declare no competing interests.

Data sources

http://www.imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html

Appendix A: Calculation of ME and ME_{max} at annual, seasonal and monthly time-scale for the period 1901-2013 for one grid point

1. x_i is the precipitation amount during the considered time-scale and form a time series as per the length of the data.

   $x_1, x_2, x_3 \ldots\ldots x_{113}$ for annual time-scale

   $x_1, x_2, x_3 \ldots\ldots x_{113}$ for seasonal time-scale

   $x_1, x_2, x_3 \ldots\ldots x_{113}$ for monthly time-scale

2. X is the total precipitation considered during the time scale of evaluation

   $X = \sum_{i=1}^{113} x_i$

3. Calculate ME

   $ME = -\sum_{i=1}^{113} \frac{x_i}{X} log_2\left[\frac{x_i}{X}\right]$

4. Calculation of ME_{max}

   $ME_{max} = log_2 113$

5. Calculation of SVI

   $SVI = \frac{ME_{max} - ME}{ME_{max}}$

Appendix B: Formation of Pentad time series

For one grid point (GP1)

First, the 73 Pentad's given in Table.S1 are calculated for each year, respectively. Second, for every individual pentad, the mean is calculated considering 113 years.

Table A2: Construction of pentad time series

Pentad	1901	1902	1903	.	.	.	.	.	.	.	2011	2012	2013	Mean
1	X_1	X_2	X_3	.	.	.	.	.	.	.	X_{111}	X_{112}	X_{113}	$x_1 = \frac{\sum_{i=1}^{113} X_i}{113}$
2	Y_1	Y_2	Y_3	.	.	.	.	.	.	.	Y_{111}	Y_{112}	Y_{113}	$y_1 = \frac{\sum_{i=1}^{113} Y_i}{113}$
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
73	Z_1	Z_2	Z_3	.	.	.	.	.	.	.	Z_{111}	Z_{112}	Z_{113}	$z_1 = \frac{\sum_{i=1}^{113} Z_i}{113}$

For Cluster 1:

In each cluster, the obtained mean pentad time series from every grid point respectively for further analysis. Next 25th, 50th,75th percentiles are calculated for each pentad.

Table A3: Construction of seasonality distribution

Pentad	GP1	GP2	GP3	.	.	.	.	.	.	GP296	GP297	GP298	25th percentile	50th percentile	75th percentile
1	x_1	x_2	x_3	.	.	.	.	.	.	x_{296}	x_{297}	x_{298}	0.1169	0.2150	0.3472
2	y_1	y_2	y_3	.	.	.	.	.	.	y_{296}	y_{297}	y_{298}	0.1596	0.2967	0.6132
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
73	z_1	z_2	z_3	.	.	.	.	.	.	z_{296}	z_{297}	z_{298}	0.0019	0.0236	0.0634

References

Adamowski, K., 2000. Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. J. Hydrol. 229, 219–231. https://doi.org/10.1016/S0022-1694(00)00156-6

Agarwal, A., Maheswaran, R., Kurths, J., Khosa, R., 2016a. Wavelet Spectrum and Self-Organizing Maps-Based Approach for Hydrologic Regionalization -a Case Study in the Western United States. Water Resour. Manag. 30, 4399–4413. https://doi.org/10.1007/s11269-016-1428-1

Agarwal, A., Maheswaran, R., Sehgal, V., Khosa, R., Sivakumar, B., Bernhofer, C., 2016b. Hydrologic regionalization using wavelet-based multiscale entropy method. J. Hydrol. 538, 22–32. https://doi.org/10.1016/j.jhydrol.2016.03.023

Agarwal, A., Marwan, N., Maheswaran, R., Merz, B., Kurths, J., 2018. Quantifying the roles of single stations within homogeneous regions using complex network analysis. J. Hydrol. 563, 802–810. https://doi.org/10.1016/j.jhydrol.2018.06.050

Alila, Y., 1999. A hierarchical approach for the regionalization of precipitation annual maxima in Canada. J. Geophys. Res. Atmos. 104, 31645–31655. https://doi.org/10.1029/1999JD900764

Ananthakrishnan, R., Soman, M.K., 1988. The onset of the southwest monsoon over Kerala: 1901–1980. J. Climatol. 8, 283–296. https://doi.org/10.1002/joc.3370080305

Azad, S., Vignesh, T.S., Narasimha, R., 2010. Periodicities in Indian monsoon rainfall over spectrally homogeneous regions. Int. J. Climatol. 30, 2289–2298. https://doi.org/10.1002/joc.2045

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future köppen-geiger climate classification maps at 1-km resolution. Sci. Data 5, 1–12. https://doi.org/10.1038/sdata.2018.214

Bharath, R., Srinivas, V. V., 2015. Regionalization of extreme rainfall in India. Int. J. Climatol. 35, 1142–1156. https://doi.org/10.1002/joc.4044

Brown, J.R., Jakob, C., Haynes, J.M., 2010. An evaluation of rainfall frequency and intensity over the Australian region in a global climate model. J. Clim. 23, 6504–6525. https://doi.org/10.1175/2010JCLI3571.1

Caliński, T., Harabasz, J., 1974. Communications in Statistics - Theory and Methods. Commun. Stat. 3, 1–27. https://doi.org/10.1080/03610927408827101

Carvalho, M.J., Melo-Gonçalves, P., Teixeira, J.C., Rocha, A., 2016. Regionalization of Europe based on a K-Means Cluster Analysis of the climate change of temperatures and precipitation. Phys. Chem. Earth, Parts A/B/C 94, 22–28. https://doi.org/10.1016/j.pce.2016.05.001

Chen, Y., Qin, B., Liu, T., Liu, Y., Li, S., 2010. The Comparison of SOM and K-means for Text Clustering. Comput. Inf. Sci. 3, 268–274. https://doi.org/10.5539/cis.v3n2p268

Cheng, L., Niu, J., Liao, D., 2017. Entropy-Based Investigation on the Precipitation Variability over the Hexi Corridor in China. Entropy 19, 660. https://doi.org/10.3390/e19120660

Darand, M., Mansouri Daneshvar, M.R., 2014. Regionalization of Precipitation Regimes in Iran Using Principal Component Analysis and Hierarchical Clustering Analysis. Environ. Process. 1, 517–532. https://doi.org/10.1007/s40710-014-0039-1

Deng, S., Yang, N., Li, M., Cheng, L., Chen, Z., Chen, Y., Chen, T., Liu, X., 2019. Rainfall seasonality changes and its possible teleconnections with global climate events in China. Clim. Dyn. 53, 3529–3546. https://doi.org/10.1007/s00382-019-04722-3

Dilmi, M.D., Mallet, C., Barthes, L., Chazottes, A., 2017. Data-driven clustering of rain events: microphysics information derived from macro-scale observations. Atmos. Meas. Tech. 10, 1557–1574. https://doi.org/10.5194/amt-10-1557-2017

Dimitrova, A., Bora, J.K., 2020. Monsoon weather and early childhood health in India. PLoS One 15, 1–22. https://doi.org/10.1371/journal.pone.0231479

Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S., Mirnia, M., 2004. Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. J. Hydrol. 297, 109–123. https://doi.org/10.1016/j.jhydrol.2004.04.009

Domroes, M., Kaviani, M., Schaefer, D., 1998. An Analysis of Regional and Intra-annual Precipitation Variability over Iran https://doi.org/10.1007/s007040050060

Fazel, N., Berndtsson, R., Uvo, C.B., Madani, K., Kløve, B., 2018. Regionalization of precipitation characteristics in Iran's Lake Urmia basin. Theor. Appl. Climatol. 132, 363–373. https://doi.org/10.1007/s00704-017-2090-0

Feng, X., Porporato, A., Rodriguez-Iturbe, I., 2013. Changes in rainfall seasonality in the tropics. Nat. Clim. Chang. 3, 811–815. https://doi.org/10.1038/nclimate1907

Fraley, C., 1998. How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. Comput. J. 41, 578–588. https://doi.org/10.1093/comjnl/41.8.578

Fukushima, A., Kanamori, H., Matsumoto, J., 2019. Regionality of long-term trends and interannual variation of seasonal precipitation over India. Prog. Earth Planet. Sci. 6, 20. https://doi.org/10.1186/s40645-019-0255-4

Gadgil, S., Yadumani, Joshi, N. V., 1993. Coherent rainfall zones of the Indian region. Int. J. Climatol. 13, 547–566. https://doi.org/10.1002/joc.3370130506

Geen, R., Lambert, F.H., Vallis, G.K., 2018. Regime Change Behavior during Asian Monsoon Onset. J. Clim. 31, 3327–3348. https://doi.org/10.1175/JCLI-D-17-0118.1

Ghosh, S., Das, D., Kao, S.-C., Ganguly, A.R., 2012. Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. Nat. Clim. Chang. 2, 86–91. https://doi.org/10.1038/nclimate1327

Gong, W., Yang, D., Gupta, H. V., Nearing, G., 2014. Estimating information entropy for hydrological data: One-dimensional case. Water Resour. Res. 50, 5003–5018. https://doi.org/10.1002/2014WR015874

Goswami, B.N., Venugopal, V., Sengupta, D., Madhusoodanan, M.S., Xavier, P.K., 2006. Increasing Trend of Extreme Rain Events Over India in a Warming Environment. Science (80-.). 314, 1442–1445. https://doi.org/10.1126/science.1132027

Guhathakurta, P., Rajeevan, M., Sikka, D.R., Tyagi, A., 2015. Observed changes in southwest monsoon rainfall over India during 1901-2011. Int. J. Climatol. 35, 1881–1898. https://doi.org/10.1002/joc.4095

Guntu, R.K., Rathinasamy, M., Agarwal, A., Sivakumar, B., 2020a. Spatiotemporal variability of Indian rainfall using multiscale entropy. J. Hydrol. 587, 124916. https://doi.org/10.1016/j.jhydrol.2020.124916

Guntu, R.K., Yeditha, P.K., Rathinasamy, M., Perc, M., Marwan, N., Kurths, J., Agarwal, A., 2020b. Wavelet entropy-based evaluation of intrinsic predictability of time series. Chaos An Interdiscip. J. Nonlinear Sci. 30, 033117. https://doi.org/10.1063/1.5145005

Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. Water Resour. Res. 29, 271–281. https://doi.org/10.1029/92WR01980

Hsu, K.-C., Li, S.-T., 2010. Clustering spatial–temporal precipitation data using wavelet transform and self-organizing map neural network. Adv. Water Resour. 33, 190–200. https://doi.org/10.1016/j.advwatres.2009.11.005

Iyengar, R.N., Basak, P., 1994. Regionalization of Indian monsoon rainfall and long-term variability signals. Int. J. Climatol. 14, 1095–1114. https://doi.org/10.1002/joc.3370141003

Jiménez-Esteve, B., Domeisen, D.I. V., 2019. Nonlinearity in the North Pacific Atmospheric Response to a Linear ENSO Forcing. Geophys. Res. Lett. 46, 2271–2281. https://doi.org/10.1029/2018GL081226

Kawachi, T., Maruyama, T., Singh, V.P., 2001. Rainfall entropy for delineation of water resources zones in Japan. J. Hydrol. 246, 36–44. https://doi.org/10.1016/S0022-1694(01)00355-9

Kohonen, T., 2001. Self-Organazing Maps. https://doi.org/10.1111/j.1530-0277.2007.00437.x

Koutsoyiannis, D., 2005. Uncertainty, entropy, scaling and hydrological stochastics. 1. Marginal distributional properties of hydrological processes and state scaling. Hydrol. Sci. J. 50, 381–404. https://doi.org/10.1623/hysj.50.3.381.65031

Krstanovic, P.F., Singh, V.P., 1992. Evaluation of rainfall networks using entropy: II. Application. Water Resour. Manag. 6, 295–314. https://doi.org/10.1007/BF00872282

Kurths, J., Agarwal, A., Shukla, R., Marwan, N., Rathinasamy, M., Caesar, L., Krishnan, R., Merz, B., 2019. Unravelling the spatial diversity of Indian precipitation teleconnections via a non-linear multi-scale approach. Nonlinear Process. Geophys. 26, 251–266. https://doi.org/10.5194/npg-26-251-2019

Lin, G.-F., Chen, L.-H., 2006. Identification of homogeneous regions for regional frequency analysis using the self-organizing map. J. Hydrol. 324, 1–9. https://doi.org/10.1016/j.jhydrol.2005.09.009

Malik, N., Bookhagen, B., Mucha, P.J., 2016. Spatiotemporal patterns and trends of Indian monsoonal rainfall extremes. Geophys. Res. Lett. 43, 1710–1717. https://doi.org/10.1002/2016GL067841

Mannan, A., Chaudhary, S., Dhanya, C.T., Swamy, A.K., 2018. Regionalization of rainfall characteristics in India incorporating climatic variables and using self-organizing maps. ISH J. Hydraul. Eng. 24, 147–156. https://doi.org/10.1080/09715010.2017.1400409

Maruyama, T., Kawachi, T., Singh, V.P., 2005. Entropy-based assessment and clustering of potential water resources availability. J. Hydrol. 309, 104–113. https://doi.org/10.1016/j.jhydrol.2004.11.020

Mishra, A.K., Özger, M., Singh, V.P., 2009. An entropy-based investigation into the variability of precipitation. J. Hydrol. 370, 139–154. https://doi.org/10.1016/j.jhydrol.2009.03.006

Modarres, R., 2008. Regional frequency distribution type of low flow in North of Iran by L-moments. Water Resour. Manag. 22, 823–841. https://doi.org/10.1007/s11269-007-9194-8

Molini, A., La Barbera, P., Lanza, L.G., 2006. Correlation patterns and information flows in rainfall fields. J. Hydrol. 322, 89–104. https://doi.org/10.1016/j.jhydrol.2005.02.041

Nandargi, S., Dhar, O.N., 2011. Extreme rainfall events over the Himalayas between 1871 and 2007. Hydrol. Sci. J. 56, 930–945. https://doi.org/10.1080/02626667.2011.595373

Nourani, V., Baghanam, A.H., Adamowski, J., Gebremichael, M., 2013. Using self-organizing maps and wavelet transforms for space-time pre-processing of satellite precipitation and runoff data in neural network based rainfall-runoff modeling. J. Hydrol. 476, 228–243. https://doi.org/10.1016/j.jhydrol.2012.10.054

Ohba, M., Kadokura, S., Nohara, D., Toyoda, Y., 2016. Rainfall downscaling of weekly ensemble forecasts using self-organising maps. Tellus A Dyn. Meteorol. Oceanogr. 68, 29293. https://doi.org/10.3402/tellusa.v68.29293

Pai, D.S., Sridhar, L., Rajeevan, M., Sreejith, O.P., Satbhai, N.S., Mukhopadhyay, B., 2014. Development of a new high spatial resolution (0.25° × 0.25°) long period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. Mausam 65, 1–18.

Pascale, S., Lucarini, V., Feng, X., Porporato, A., Hasson, S. ul, 2015. Analysis of rainfall seasonality from observations and climate models. Clim. Dyn. 44, 3281–3301. https://doi.org/10.1007/s00382-014-2278-2

Paul, S., Ghosh, S., Mathew, M., Devanand, A., Karmakar, S., Niyogi, D., 2018. Increased Spatial Variability and Intensification of Extreme Monsoon Rainfall due to Urbanization. Sci. Rep. 8, 3918. https://doi.org/10.1038/s41598-018-22322-9

Pechlivanidis, I.G., Jackson, B., Mcmillan, H., Gupta, H. V., 2016. Robust informational entropy-based descriptors of flow in catchment hydrology. Hydrol. Sci. J. 61, 1–18. https://doi.org/10.1080/02626667.2014.983516

Rajeevan, M., Bhate, J., Jaswal, A.K., 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. https://doi.org/10.1029/2008GL035143

Rajeevan, M., Unnikrishnan, C.K., Bhate, J., Niranjan Kumar, K., Sreekala, P.P., 2012. Northeast monsoon over India: Variability and prediction. Meteorol. Appl. 19, 226–236. https://doi.org/10.1002/met.1322

Rathinasamy, M., Khosa, R., Adamowski, J., Ch, S., Partheepan, G., Anand, J., Narsimlu, B., 2014. Wavelet-based multiscale performance analysis: An approach to assess and improve hydrological models. Water Resour. Res. 50, 9721–9737. https://doi.org/10.1002/2013WR014650

Rodrigues da Silva, V. de P., Belo Filho, A.F., Rodrigues Almeida, R.S., de Holanda, R.M., da Cunha Campos, J.H.B., 2016. Shannon information entropy for assessing space–time variability of rainfall and streamflow in semiarid region. Sci. Total Environ. 544, 330–338. https://doi.org/10.1016/j.scitotenv.2015.11.082

Roushangar, K., Alizadeh, F., Adamowski, J., Saghebian, S.M., 2019. Exploring the multiscale changeability of https://doi.org/10.2166/wcc.2019.097

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Roxy, M.K., Ghosh, S., Pathak, A., Athulya, R., Mujumdar, M., Murtugudde, R., Terray, P., Rajeevan, M., 2017. A threefold rise in widespread extreme rain events over central India. Nat. Commun. 8, 1–11. https://doi.org/10.1038/s41467-017-00744-9

Ruiz-Barradas, A., Nigam, S., 2013. Atmosphere–Land Surface Interactions over the Southern Great Plains: Characterization from Pentad Analysis of DOE ARM Field Observations and NARR. J. Clim. 26, 875–886. https://doi.org/10.1175/JCLI-D-11-00380.1

Saf, B., 2009. Regional Flood Frequency Analysis Using L-Moments for the West Mediterranean Region of Turkey. Water Resour. Manag. 23, 531–551. https://doi.org/10.1007/s11269-008-9287-z

Sahany, S., Mishra, S.K., Pathak, R., Rajagopalan, B., 2018. Spatiotemporal Variability of Seasonality of Rainfall Over India. Geophys. Res. Lett. 45, 7140–7147. https://doi.org/10.1029/2018GL077932

Saikranthi, K., Narayana Rao, T., Rajeevan, M., Vijaya Bhaskara Rao, S., 2013. 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

Sang, Y.-F., Wang, Z., Li, Z., Liu, C., Liu, X., 2013. Investigation into the daily precipitation variability in the Yangtze River Delta, China. Hydrol. Process. 27, 175–185. https://doi.org/10.1002/hyp.9202

Santos, C., Brasil Neto, R., da Silva, R., Costa, S., 2019. Cluster Analysis Applied to Spatiotemporal Variability of Monthly Precipitation over Paraíba State Using Tropical Rainfall Measuring Mission (TRMM) Data. Remote Sens. 11, 637. https://doi.org/10.3390/rs11060637

Satyanarayana, P., Srinivas, V. V., 2011. Regionalization of precipitation in data sparse areas using large scale atmospheric variables - A fuzzy clustering approach. J. Hydrol. 405, 462–473. https://doi.org/10.1016/j.jhydrol.2011.05.044

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

Sehgal, V., Lakhanpal, A., Maheswaran, R., Khosa, R., Sridhar, V., 2018. Application of multi-scale wavelet entropy and multi-resolution Volterra models for climatic downscaling. J. Hydrol. 556, 1078–1095. https://doi.org/10.1016/j.jhydrol.2016.10.048

Seneviratne, S.I., Nicholls, N., Easterling, D., Goodess, C.M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., others, 2012. Changes in climate extremes and their impacts on the natural physical environment.

Shanmugasundaram, J., Lee, E., 2018. Oceanic and atmospheric conditions associated with the pentad rainfall over the southeastern peninsular India during the North-East Indian Monsoon season. Dyn. Atmos. Ocean. 81, 1–14. https://doi.org/10.1016/j.dynatmoce.2017.11.001

Shannon, C.E., 1948. A Mathematical Theory of Communication. Bell Syst. Tech. J. 27, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data, in: Proceedings of the 1968 23rd ACM National Conference On -. ACM Press, New York, New York, USA, pp. 517–524. https://doi.org/10.1145/800186.810616

Shukla, R., Agarwal, A., Sachdeva, K., Kurths, J., Joshi, P.K., 2019. Climate change perception: an analysis of climate change and risk perceptions among farmer types of Indian Western Himalayas. Clim. Change 152, 103–119. https://doi.org/10.1007/s10584-018-2314-z

Singh, V.P., 2015. Entropy Theory in Hydrologic Science and Engineering. McGraw-Hill Education, New York.

Singh, V.P., 2013. Entropy Theory and its Application in Environmental and Water Engineering, Entropy Theory and its Application in Environmental and Water Engineering. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/9781118428306

Singh, V.P., 1997. the Use of Entropy in Hydrology and Water Resources. Hydrol. Process. 11, 587–626.

Sreekala, P.P., Rao, S.V.B., Rajeevan, M., 2012. Northeast monsoon rainfall variability over south peninsular India and its teleconnections. Theor. Appl. Climatol. 108, 73–83. https://doi.org/10.1007/s00704-011-0513-x

Sugahara, S., da Rocha, R.P., Silveira, R., 2009. Non-stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil. Int. J. Climatol. 29, 1339–1349. https://doi.org/10.1002/joc.1760

Türkeş, M., Tatlı, H., 2011. Use of the spectral clustering to determine coherent precipitation regions in Turkey for the period 1929-2007. Int. J. Climatol. 31, 2055–2067. https://doi.org/10.1002/joc.2212

Vesanto, J., Alhoniemi, E., 2000. Clustering of the self-organizing map. IEEE Trans. Neural Networks 11, 586–600. https://doi.org/10.1109/72.846731

Vinnarasi, R., Dhanya, C.T., 2016. Changing characteristics of extreme wet and dry spells of Indian monsoon rainfall. J. Geophys. Res. Atmos. 121, 2146–2160. https://doi.org/10.1002/2015JD024310

Vul, T.M., Mishra, A.K., 2019. Nonstationary frequency analysis of the recent extreme precipitation events in the United States. J. Hydrol. 575, 999–1010. https://doi.org/10.1016/j.jhydrol.2019.05.090

Weijs, S. V., Van De Giesen, N., Parlange, M.B., 2013. Data compression to define information content of hydrological time series. Hydrol. Earth Syst. Sci. 17, 3171–3187. https://doi.org/10.5194/hess-17-3171-2013

Wolski, P., Jack, C., Tadross, M., van Aardenne, L., Lennard, C., 2018. Interannual rainfall variability and SOM-based circulation classification. Clim. Dyn. 50, 479–492. https://doi.org/10.1007/s00382-017-3621-1

Yadav, R.K., Roxy, M.K., 2019. On the relationship between north India summer monsoon rainfall and east equatorial Indian Ocean warming. Glob. Planet. Change 179, 23–32. https://doi.org/10.1016/j.gloplacha.2019.05.001

Yadav, R.K., Rupa Kumar, K., Rajeevan, M., 2007. Role of Indian Ocean sea surface temperatures in modulating northwest Indian winter precipitation variability. Theor. Appl. Climatol. 87, 73–83. https://doi.org/10.1007/s00704-005-0221-5

Yang, T., Shao, Q., Hao, Z.-C., Chen, X., Zhang, Z., Xu, C.-Y., Sun, L., 2010. Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin, China. J. Hydrol. 380, 386–405. https://doi.org/10.1016/j.jhydrol.2009.11.013

Zhang, L., Singh, V.P., 2006. Bivariate Flood Frequency Analysis Using the Copula Method. J. Hydrol. Eng. 11, 150–164. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:2(150)

Zhang, Q., Zheng, Y., Singh, V.P., Xiao, M., Liu, L., 2016. Entropy-based spatiotemporal patterns of precipitation regimes in the Huai River basin, China. Int. J. Climatol. 36, 2335–2344. https://doi.org/10.1002/joc.4498

Zhao, C., Ding, Y., Ye, B., Yao, S., Zhao, Q., Wang, Z., Wang, Y., 2011. An analyses of long-term precipitation variability based on entropy over Xinjiang, northwestern China. Hydrol. Earth Syst. Sci. Discuss. 8, 2975–2999. https://doi.org/10.5194/hessd-8-2975-2011

Zhao, Y., Duan, A., Wu, G., Sun, R., 2019. Response of the Indian Ocean to the Tibetan Plateau thermal forcing in late spring. J. Clim. 32, 6917–6938. https://doi.org/10.1175/JCLI-D-18-0880.1