



A minimalistic seasonal prediction model for Indian monsoon based on spatial patterns of rainfall anomalies

Rishi Sahastrabuddhe¹ · Subimal Ghosh^{1,2} · Anamitra Saha¹ · Raghu Murtugudde^{1,3}

Received: 28 October 2017 / Accepted: 3 July 2018 / Published online: 11 July 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Seasonal prediction of Indian Summer Monsoon Rainfall (ISMR, rainfall during June to September over India) has remained an important scientific challenge for decades, due to its complex multi-scale nature. Statistical and dynamical seasonal ISMR predictions have traditionally relied on the relatively small variability of the spatially averaged monsoon rainfall over India, known as All India Monsoon Rainfall (AIMR). While this has served to mitigate socioeconomic impacts to some extent, overall prediction skill has remained relatively low (Wang et al. in Nat Commun 6:7154. <https://doi.org/10.1038/ncomm8154>, 2015) while the spatial variability is anything but small. Here we find that the classification of deficit/ dry or surplus/ wet monsoon years based on AIMR does not add value at a regional scale due to the very high heterogeneity of monsoon rainfall, even in the extreme years. To demonstrate the need and the potential to predict this important spatial heterogeneity, we improve the classification of monsoon years by focusing on the spatial patterns of rainfall anomalies within different meteorological subdivisions of India. We apply the k -means clustering methodology and also offer cluster validation. Cluster validation reveals the existence of nine clusters of monsoon years with distinct spatial patterns of monsoon rainfall anomalies. The composite anomalies of sea surface temperature (SST) and winds during March to May (MAM) and June to September (JJAS) show distinct hydroclimatic teleconnections indicating potential predictability of regional monsoon rainfall at seasonal scale. To demonstrate the potential prediction pathways for spatial patterns of ISMR, we develop a statistical seasonal prediction model based on Classification and Regression Tree (CART) between SST over different oceanic regions as predictors and monsoon classes as predictands for the period 1901–2010. Search for the potential regressors reveals the importance of new predictors such as Atlantic Niño and SST over North Pacific region along with conventional predictors such as El Niño Southern Oscillation (ENSO), Indian Ocean Dipole/Zonal Mode (IODZM), etc. Validation of the method is performed for 2011–2015 and the model is able to predict the regional pattern of monsoon rainfall for 4 out of the 5 years. The purpose of this prediction exercise is to demonstrate the need to focus on the process and predictive understanding of these clusters and their predictability.

Keywords Indian monsoon · Seasonal prediction · Spatial patterns · Hydroclimatic teleconnections

1 Introduction

Indian Summer Monsoon Rainfall (ISMR) contributes 70% to the country's total annual rainfall which hugely impacts agricultural yield, water resources management, infrastructure and socioeconomic systems. Complex interactions between the multi-scale local and remote geophysical processes pose a major challenge to the prediction of ISMR. Past studies have revealed that the seasonal prediction of ISMR is linked to several large-scale dynamics such as Sea surface temperature (SST) over the Pacific and Indian Oceans (Rasmusson and Carpenter 1983; Rajeevan et al. 2001; Webster et al. 1998), Eurasian snow cover (Bamzai

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00382-018-4349-2>) contains supplementary material, which is available to authorized users.

✉ Subimal Ghosh
subimal@civil.iitb.ac.in

¹ Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India

² Interdisciplinary Program in Climate Studies, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India

³ Earth System Science Interdisciplinary Center (ESSIC)/DOAS, University of Maryland, College Park, MD, USA

and Shukla 1999), Land–Ocean thermal contrast (Rajeevan et al. 1998), among many others. Predictors based on hydroclimatic teleconnections were examined in the seminal work by Walker (1933) to evaluate the predictability of Indian monsoon. Many attempts have since followed for forecasting All India Monsoon Rainfall (AIMR) with statistical regression models (Gowariker et al. 1991; Rajeevan et al. 2004). Several other works regarding the selection of potential inputs for statistical prediction models of ISMR were carried out in due course (Kung and Sharif 1982; Mooley and Parthasarathy 1983; Parthasarathy et al. 1991; Rajeevan et al. 2004; Gadgil et al. 2004; Goswami et al. 2006; Maity and Nagesh Kumar 2006; Vathsala and Koolagudi 2016). Majority of these predictions largely focused on the El Niño Southern Oscillation (ENSO) teleconnections (Shukla and Paulino 1983). The relationship between equatorial Pacific SST and ISMR was found to be strong with below-normal (above-normal) monsoon rainfall during a substantial number of El Niño (La Niña) years (Sikka 1980; Rasmusson and Carpenter 1983). However, Krishna Kumar et al. (1999) showed a weakened relationship between El Niño and ISMR for recent decades. Such a weakened relationship between the Walker circulation and El Niño is argued to be due to the variability of the Indian Ocean Dipole/Zonal Mode (IODZM) (Murtugudde et al. 1998; Saji et al. 1999; Yamagata et al. 2004; Annamalai et al. 2005; Wu and Kirtman 2004). Recent global warming along with the weakened El Niño-ISMR relationship has contributed to the poor skill in forecasting ISMR (Wang et al. 2015). It must be emphasised that the AIMR offers a measure for the overall food production but is of little practical value for farmers, water managers or the overall food-energy-water nexus (Barik et al. 2016), all of which depend on local rainfall variability. There is also an indication that the spatial variability of the monsoon has been on the rise (Ghosh et al. 2012) and widespread extreme rainfall has increased threefold over the last few decades (Roxy et al. 2017). In this context, it is essential and interesting to explore whether the relatively narrow window of AIMR variability is associated with a limited number of spatial patterns. Extracting such patterns may also offer a novel way to quantify the non-stationary correlations not only between the monsoon and ENSO but also other factors such as the IODZM and the Atlantic Zonal Mode (Pottapinjara et al. 2014, 2015).

Amongst the predictors used for monsoon prediction, SST over different regions of Pacific and Indian Oceans have shown significant skill in predicting Indian summer monsoon as investigated from coupled atmospheric-oceanic models (DelSole and Shukla 2012). Atmospheric General Circulation Model (AGCM) Experiments with prescribed SSTs over the Indian Ocean show a reverse Indian monsoon–ENSO relationship (Wu and Kirtman 2004). Coupled model experiments by Lau and Nath (2000) revealed that ENSO

induces monsoon anomalies over the Indian Ocean and its feedback to the atmospheric processes opposes the remote forcing of the Indian monsoon from the Pacific (Arpe et al. 1998). This makes the hydroclimatic teleconnection processes even more complicated since land processes may also be involved such as soil moisture feedbacks and moisture source variability (Pathak and Ghosh 2014, 2017). The teleconnections between the monthly summer monsoon rainfall and the SSTs over Indian and tropical Pacific Oceans have been discussed by Maity and Nagesh Kumar (2006). Along with these mechanisms, impacts of the Equatorial Indian Ocean Oscillation (EQUINO- the fundamental output of IOD mode) on Indian monsoon are analyzed by Gadgil et al. (2004). Most recently, tropical Atlantic influences have been proposed to be modulating the interannual variability of ISMR (Kucharski et al. 2007, 2008a; Pottapinjara et al. 2014, 2015).

However, all the above-mentioned studies have focused on the relationship between SST predictors and AIMR that considers a spatial average of rainfall over entire India. The vast diversity in geographic conditions, fluctuations between periods of dry and wet spells and the large spatial extent of India lead to a large spatial variability of ISMR. Several attempts have been made to understand spatial variability of Indian monsoon using many statistical tools and mathematical conventions (Gadgil and Iyengar 1980; Iyengar and Basak 1994; Kulkarni et al. 1992; Kulkarni and Kripalani 1998) and the way they exert significant impact upon socio-economic activities. Generally, there are areas of deficit rainfall in the best monsoon years and areas with good rainfall in drought years (Parthasarathy and Pant 1984). This can cause extreme wet and dry regional conditions in the same monsoon season with a misleading definition of wet and dry monsoon years based solely on AIMR. Figure 1 shows the standardized rainfall anomalies for traditional wet and dry years, with AIMR anomalies greater than and less than 10% of the long-term mean of AIMR, respectively, within the past two decades. This shows the need of predicting seasonal monsoon at a local scale to make it useful for sowing and harvesting grains and management of food, water and energy. However, hydroclimatic teleconnections may lose their significance over smaller/regional spatiotemporal scales (Kane 1998). Regional scale monsoon prediction was performed by Kashid and Maity (2012) over five homogeneous zones with ENSO and EQUINO indices. Similarly, Saha et al. (2017) considered different predictors for the rainfall predictions in four homogeneous zones following the approach by Rajeevan et al. (2004). These homogeneous zones over India are vast and hence, they tend to have substantial spatial variability within them. To understand these spatial variability at finer scale, Kulkarni et al. (1992) used a clustering algorithm to find the relevance of recurring spatiotemporal rainfall patterns with moisture circulation.

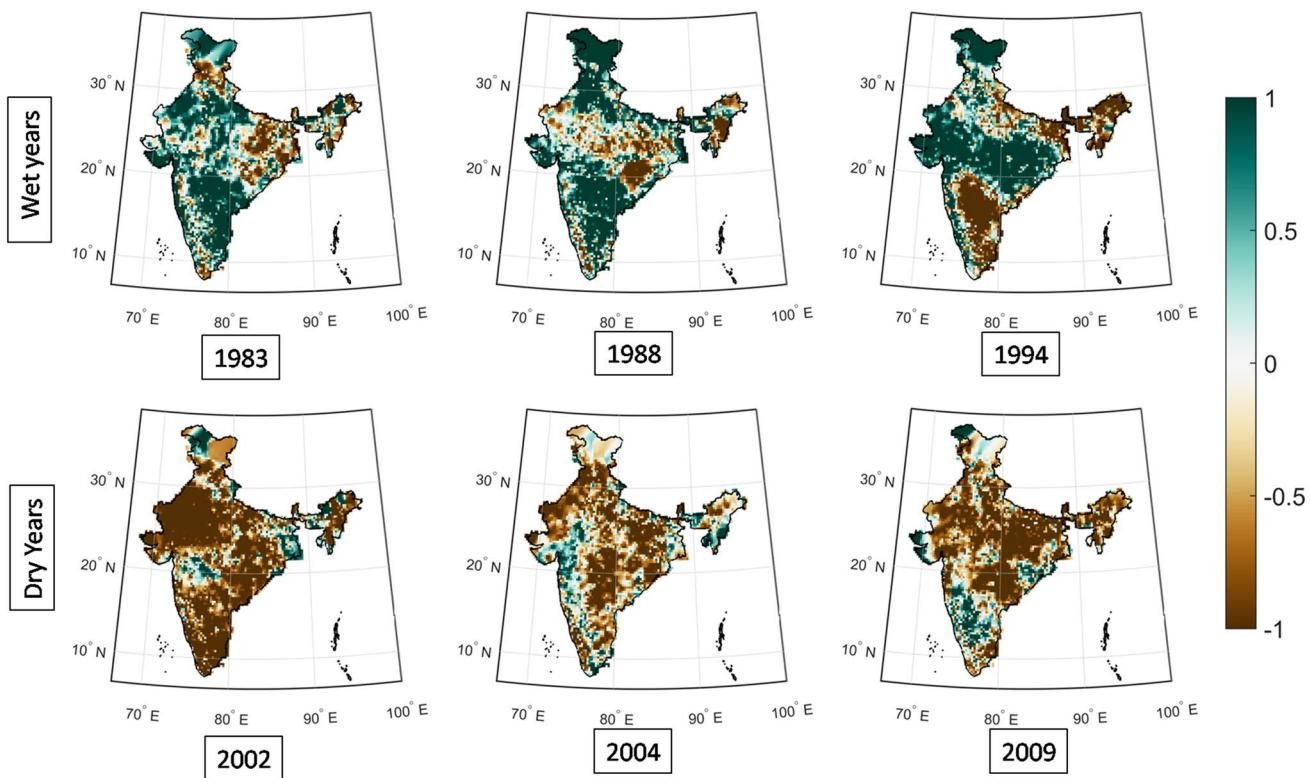


Fig. 1 Standardized rainfall anomalies for the recent dry and wet monsoon years

Attempts were also made to develop a hydroclimatic teleconnection model between streamflow at a station and SST (Kashid et al. 2010), but the results were not satisfactory due to the loss of signal from large scale to station scale. Here, we hypothesize that the hydroclimatic teleconnection may not predict rainfall at a grid or a station individually, but it can provide the spatial pattern of rainfall over a large region that gives usable information about local rainfall. Hydroclimatic teleconnections from large-scale events are accompanied by major changes in the pattern of convection (Parthasarathy and Pant 1984) which not only affect rainfall over a larger region but may also cause spatial variations over the entire subcontinent (Sardeshmukh et al. 2000). This local and regional information based on spatial patterns of rainfall may specifically be important for a prior understanding of possible wet and dry rainfall areas during a season to better plan for seasonal water transfer through the newly planned river interlinking schemes in India. Following the above-mentioned hypothesis, we develop a unique methodology where we first classify monsoon seasons based on the spatial patterns of rainfall anomalies over India and then predict these spatial patterns statistically from potential predictors. We employ a k -means clustering approach for the classification of monsoon seasons based on spatial patterns. An optimum number of clusters is selected using a cluster validation (Silhouette) index. Our study investigates

these clusters in detail and quantifies their relations with the March–May (MAM) SST using a Classification and Regression Tree (CART) model. The model is validated with the recent 5 years which are independent of the training period. Note again that the empirical prediction exercise is meant to motivate the process and predictive understanding of the dynamics of clustering and whether that is a viable approach for more skillful ISMR predictions at regional scales. The next section presents the details of data and method used in the analysis.

2 Data and methodology

2.1 Datasets

High spatial resolution ($0.25^\circ \times 0.25^\circ$) long-term (1901–2015) daily gridded rainfall data over India (Pai et al. 2014), from the India Meteorological Department (IMD), is used in this study. For preparing this gridded data, daily rainfall records from 6955 rain gauge stations over India were used. This high-resolution dataset is especially useful for better representation of certain regions such as the Western Ghats and Northeast India which show distinct regional rainfall. Seasonal average over 122 days of the Indian monsoon, i.e., from 1st June to 30th September is used for a historical

period of 1901–2015. Last 5 years of this period are used for validation and the first 110 years (1901–2010) are used for training/calibration. Observed SST monthly statistical means are extracted from the Extended Reconstructed Sea Surface Temperature version 3b (ERSST.v3b) with the resolution of $2^\circ \times 2^\circ$ based on Comprehensive Ocean–Atmosphere Data Set (COADS) (Smith et al. 2008) for the period 1901–2015. Monthly zonal and meridional winds are provided for June–September (JJAS) at 850 hPa pressure level for 1901–2010 by ERA-20C (Poli et al. 2016).

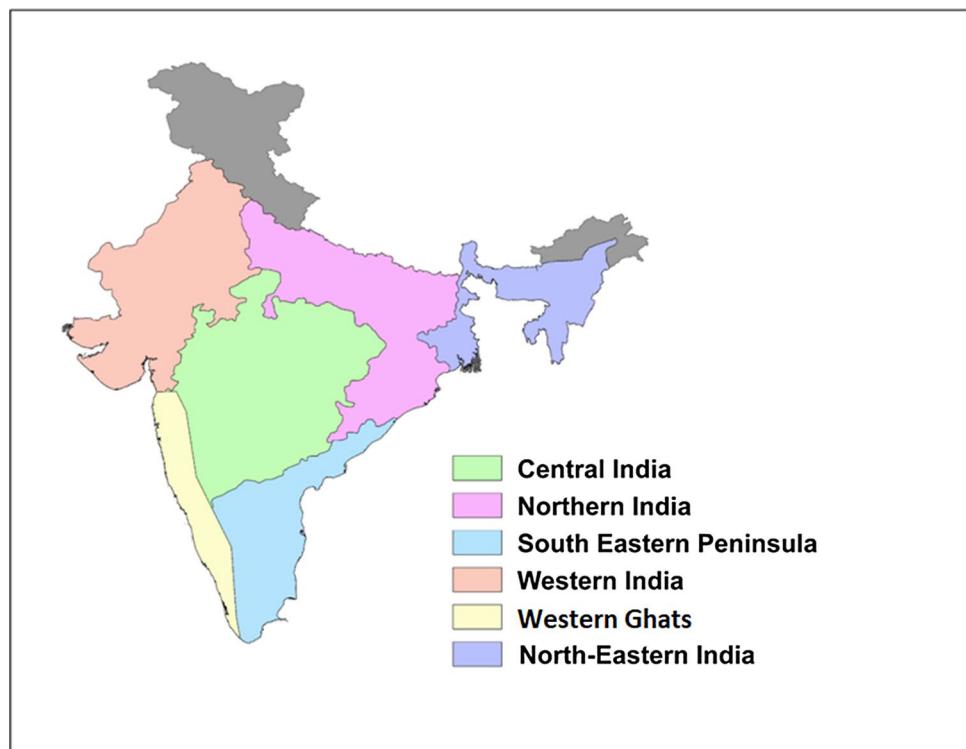
2.2 Method

The physical relationship between spatial variability of ISMR and the pre-monsoon SST (during MAM) is analyzed and used for prediction in this present study. The hypothesis of this study is that the anomalous SST conditions during March to May (MAM) drive specific spatial patterns in Indian monsoon rainfall during June to September (JJAS). To find these spatial patterns of ISMR that are generated, we employ a *k*-means clustering algorithm. The purpose is not to classify a year into a dry, wet or normal monsoon year based on spatial average, but to understand the spatial pattern of rainfall anomaly of each year that provides regional information at usable and useful timescales to facilitate the process and predictive understanding and explore potential predictability of these patterns.

The *k*-means clustering classifies multi-dimensional datasets into different classes based on Euclidean distances between the points in a multi-dimensional space. The algorithm maximizes the Euclidean distances between the points belonging to different clusters and minimizes the Euclidean distance between the points belonging to the same cluster. To classify the monsoon years into different classes, the seasonal rainfall values at different grids over India should constitute the multi-dimensional attributes that enter as input to the clustering algorithms. However, inherently high spatial variability across grids even within a meteorologically homogeneous region will not allow the key signal of spatial pattern to participate in the clustering. Hence, we consider the spatial average of rainfall over the meteorologically homogeneous regions (Fig. 2). Two zones, viz., Central India and Southern Peninsula are modified and formed into three zones (Central India, Southern Peninsula and the Western Ghats) for our comprehensive study. For training the model with the first 110 years (1901–2010) of the IMD rainfall dataset, spatial means of JJAS seasonal rainfall over all the zones are estimated. Then for each zone, the climatological mean (μ) and standard deviation (σ) are computed. The rainfall value over each zone for each year is then categorized as follows.

Category 1—Extremely low ranges of zonal mean rainfall
Value $< \mu - 2.5\sigma$

Fig. 2 India Meteorological Department (IMD) zones. Northern and extreme Northeast regions are not considered for any kind of analysis in the study



- Category 2—Much lower range of zonal mean rainfall
 $\mu - 1.0\sigma < \text{Value} < \mu - 2.5\sigma$
- Category 3—Slightly lower range of zonal mean rainfall
 $\mu - 0.5\sigma < \text{Value} < \mu - 1.0\sigma$
- Category 4—Within nearest range of zonal mean rainfall
 $\mu - 0.5\sigma < \text{Value} < \mu + 0.5\sigma$
- Category 5—Slightly higher range of zonal mean rainfall
 $\mu + 0.5\sigma < \text{Value} < \mu + 1.0\sigma$
- Category 6—Much higher range of zonal mean rainfall
 $\mu + 1.0\sigma < \text{Value} < \mu + 2.5\sigma$
- Category 7—Extremely high ranges of zonal mean rainfall
 $\mu + 2.5\sigma < \text{Value}$

These categorized rainfall values serve as the input to the clustering algorithm. Hence a newer matrix (110×6) is formed with just category values to apply the algorithm.

The number of optimum clusters is obtained with the cluster validity measure, the Silhouette index, by comparing its tightness and separation within the clusters (Rousseeuw 1987). The index compares the similarity of a point in its own cluster with the points in other clusters. In the given analysis, the measure of similarity of each year in its own cluster is compared with the years in other clusters. The silhouette value for the i th year, S_i , is defined as:

$$S_i = (q_i - p_i) / \max(p_i, q_i) \quad (1)$$

where, p_i is the average distance of the i th year from other years in the same cluster and q_i is the minimized average distance of the i th year to the years in a different cluster. The range of Silhouette index is from -1 to $+1$. The analysis is performed with multiple trials and an average value of the Silhouette index is determined for multiple clusters. A high average Silhouette value indicates that the years in each cluster are a good match to their own cluster, and poor match to the neighboring clusters.

We obtain the maximum Silhouette index value when the number of clusters is 9 as shown in Fig. 3. We also compute the index for more than 10 clusters. There are a few cases, when the number of clusters is greater than 12, the silhouette index is higher than that of 9. This is due to the heterogeneity at a very fine resolution, such as differences among neighboring grids and its interannual variability. These may act as spatial noise while deriving the classification tree. We intentionally avoid them as the meso-scale spatial features and patterns of monsoon inter-annual variations will be contaminated by the finer scale features and noise. To summarize the methodology, a flowchart of the procedure is given in Fig. 4. As shown in Fig. 3, the best optimum number is always two, as the index divides the years into dry and wet years. However, we intentionally avoid this binary split because our objective is to be more nuanced and capture the maximum number of independent spatial patterns of rainfall anomaly.

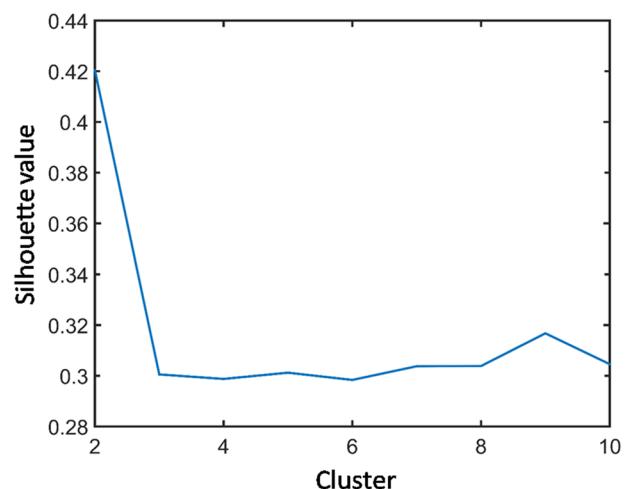


Fig. 3 Silhouette index for validation of the number of clusters

After applying the k -means clustering algorithm with 9 clusters chosen subjectively, each year is assigned to a cluster according to the centroid position of that cluster. We aim to predict the cluster of JJAS rainfall in any given year, from the SST over different oceanic regions during MAM. First, we compute the composite of standardized anomalies of SST during MAM for each cluster and explain the association or hydroclimatic teleconnection using these composites. To improve the understanding, we further use the composites of 850 hPa winds during JJAS.

After analyzing the hydroclimatic teleconnection process associated with the interannual variation of ISMR, we identified 12 oceanic regions, where the SST during MAM affects monsoon rainfall. We develop a prediction model based on the Classification and Regression Tree (CART). The SSTs during MAM over the identified oceanic regions are considered as predictors, while the cluster identities are considered as predictands. The training period is chosen to be 1901–2010. In the standard statistical analysis, it is preferable to use a smaller training period; however, we wanted a good representation of all the clusters in the training sample and hence used a longer training data set. We reemphasize that the prediction exercise is only a demonstration to motivate the dynamical analysis of the clusters and their drivers. The model is validated for the recent 5 years, 2011–2015. We present the results and discussion in the next section.

3 Results and discussions

3.1 Spatial composites

Having identified the nine clusters as above, the occurrence of years within each cluster is divided into pre-1955 and post-1955 periods to compare the dominant clusters within

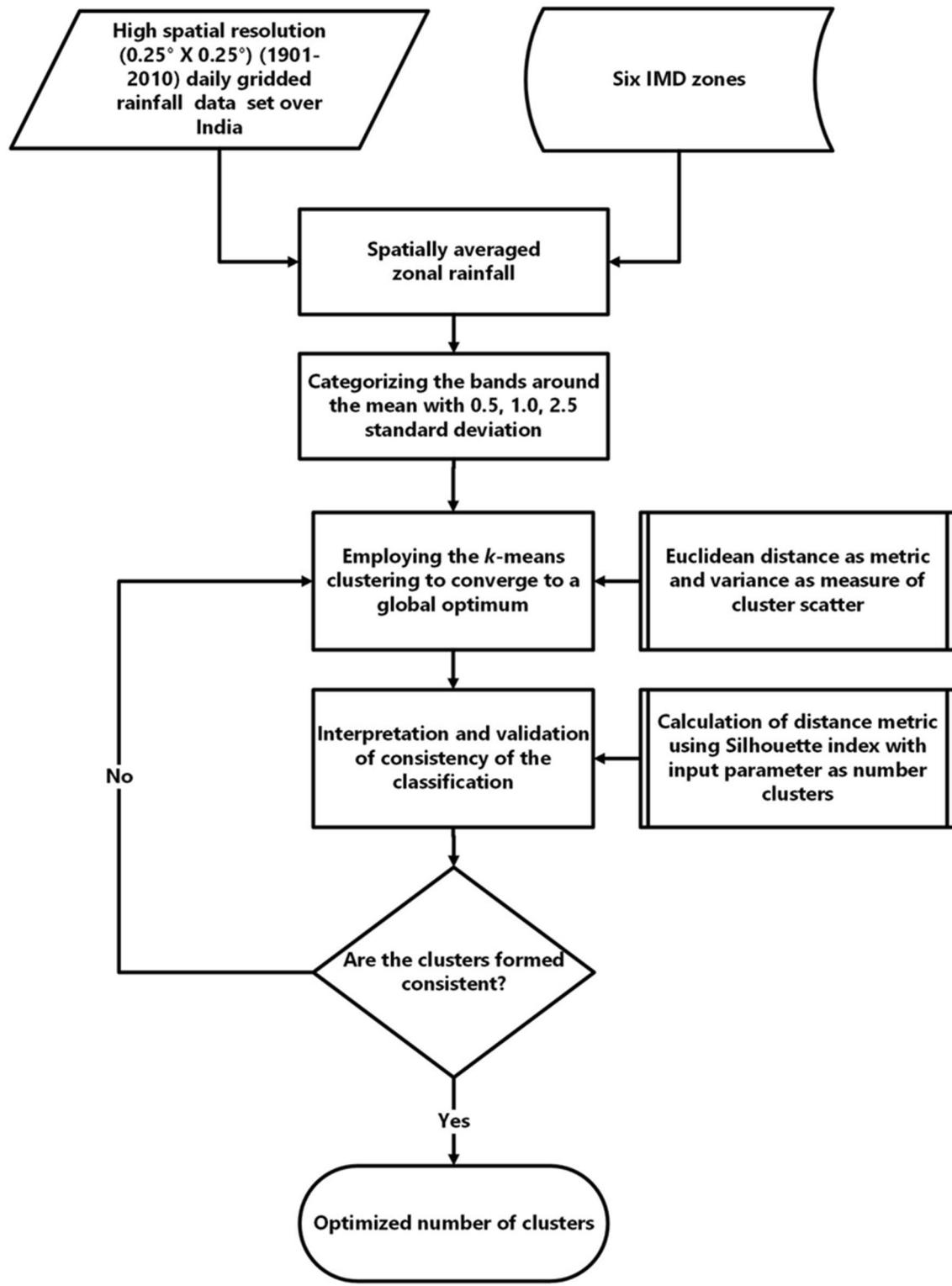


Fig. 4 Flowchart of the methodology employed for clustering spatial distribution of rainfall

each period (Table 1). Here, we discuss each of these clusters, their patterns, hydroclimatic teleconnections and their changing occurrences. We also discuss certain years in some

clusters to highlight specific details that are relevant. While this split into two periods is also subjective, it is another indicator of the trends in ISMR in the framework of our

Table 1 Number of years in each cluster for years 1901–2010, years 1901–1955, years 1955–2010 and years in each cluster

Cluster no.	Years 1901– 2010	Years 1901– 1955	Years 1956– 2010	Years in each cluster
1	5	1	4	1918, 1966, 1974, 1987, 2002
2	12	6	6	1901, 1907, 1928, 1932, 1941, 1951, 1965, 1972, 1979, 1992, 2004, 2009
3	13	8	5	1904, 1905, 1911, 1913, 1920, 1929, 1939, 1952, 1968, 1982, 1985, 1986, 1999
4	9	1	8	1902, 1971, 1984, 1989, 1991, 1993, 1995, 1997, 1998
5	19	13	6	1909, 1919, 1921, 1922, 1925, 1930, 1934, 1936, 1937, 1943, 1948, 1953, 1955, 1960, 1969, 1980, 2001, 2003, 2008
6	13	7	6	1908, 1926, 1927, 1944, 1945, 1950, 1954, 1970, 1973, 1976, 1977, 1990, 2010
7	12	6	6	1903, 1906, 1912, 1915, 1935, 1940, 1957, 1962, 1963, 1967, 1996, 2000
8	11	5	6	1914, 1923, 1931, 1933, 1946, 1958, 1959, 1961, 1994, 2005, 2006
9	16	8	8	1910, 1916, 1917, 1924, 1938, 1942, 1947, 1949, 1956, 1964, 1975, 1978, 1981, 1983, 1988, 1994

clusters and offers a clear picture of which clusters have remained nearly the same and which ones have been altered significantly in terms of the number of their occurrences in each period.

Cluster-1 During the period of 1901–2010, only 5 years belong to cluster 1, which show a dry condition (Fig. 5a)

over most of the country except the northeastern region. Interestingly, among these 5 years, 4 occurred during the recent period, viz., post-1955 (Table 1). The composites of SST anomalies during MAM for cluster 1 are presented in Fig. 5c. The figure indicates warm SSTs over the central equatorial, eastern and Northern Pacific regions. However,

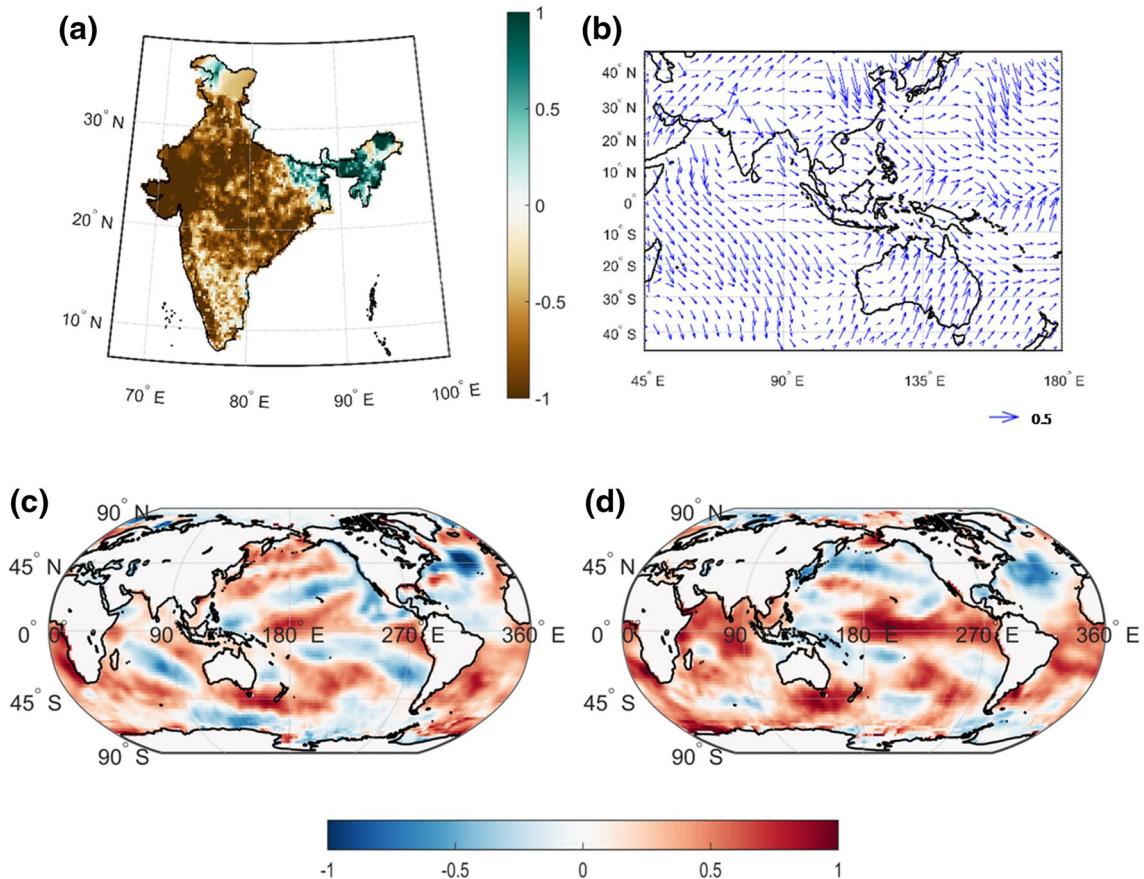


Fig. 5 Composites of standardized seasonal anomalies for Cluster 1: **a** rainfall, **b** wind speeds at 850 hPa, **c** SSTs during March–May (MAM), **d** SSTs during June–September (JJAS)

the North Atlantic shows cold SSTs. Figure 5d presents the SST composites during JJAS. Warm SSTs are observed over the equatorial Pacific, Atlantic Niño (AtlNiño), Western Indian Ocean (WIO), Southern Indian Ocean (SIO) and South China Sea (SCS) for cluster 1. Warm central and east equatorial Pacific SST is the indication of a canonical El Niño. Note that we have not delved into the details of ENSO flavors here (Chen et al. 2015) even though the ENSO-monsoon teleconnections are shown to depend on where the warming in the tropical Pacific occurs (Krishna Kumar et al. 1999). Some ENSO flavors do emerge naturally when we compute the composite SST anomalies seen in Fig. 5d. We find that 3 among the 5 years that belong to cluster 1 are El Niño as well as dry years (Table 2). The composite of 850 hPa wind anomalies for Cluster 1 is shown in Fig. 5b. The figure shows strong northwesterly wind anomalies over the Indian region implying a weakening of monsoon circulation.

Cluster-2 This cluster has 12 years, well distributed over the period of 110 years, i.e., 1901–2010. An equal number of years occur in each of the 55-year periods (Table 1). Low rainfall anomalies over the Gangetic Plain, Central India and Eastern Peninsula (Fig. 6a) are associated with cluster 2. SST anomalies in MAM and JJAS (Fig. 6c, d) are positive over the central and eastern Pacific Ocean. Warm conditions exist over the North Atlantic region in MAM but not during JJAS. However, compared to cluster 1, cluster 2 does not have very warm SST over WIO and hence, the drying of the Indian monsoon is less severe in cluster 2 (Arpe et al. 1998). We find a positive rainfall anomaly over peninsular India, despite the negative rainfall anomaly over the rest of India. The cluster consists of four El Niño and dry years as listed in Table 2. This cluster clearly captures El Niño Modoki or the Central Pacific El Niño pattern (Ashok et al. 2007). The wind anomalies (Fig. 6b) show a strong retreat towards the

Bay of Bengal suggesting strong northwesterlies. We discuss certain important years in this cluster which might help in understanding the hydroclimatic relationship for this specific spatial pattern.

1972 The year 1972 had a positive IODZM co-occurring with a strong El Niño. Generally, a positive IODZM is argued to mitigate the ENSO impact on the Indian monsoon, via the local Hadley circulation that brings moisture in JJAS (Annamalai and Slingo 2001). However, during 1972, IODZM failed to negate the rainfall deficit over the Indian subcontinent. Based on the wavelet analysis as performed by Saji and Yamagata (2003), IODZM had peak power in the 1960s and 1990s, while ENSO dominated in the 1970s and 1980s. Therefore, El Niño seemed to dominate in the 1960s and 1970s even when co-occurring with a IODZM, leading to monsoon deficits. Indian Ocean's role in the preconditioning of these decadal IODZM-ENSO phasing remains to be explored further (Annamalai et al. 2005; Wu and Kirtman 2004; Wang et al. 2016).

2009 The Indian subcontinent suffered one of the worst droughts during the boreal summer (June–September) of 2009. The sensitivity experiments by Ashok et al. (2007) using an AGCM have shown that the El Niño Modoki modifies the regional Hadley circulation over the tropical and subtropical Pacific.

Cluster 3 13 years belong to this cluster. There are 8 years prior to 1955 and 5 years post-1955, as shown in Table 1. The rainfall anomalies are negative over the western and peninsular India while a positive rainfall anomaly is observed over the eastern Ganga Basin (Fig. 7a). The composites of SST anomalies during MAM (Fig. 7c) show cold conditions over WIO and SIO. A slightly warm condition is observed over the eastern equatorial Pacific off the Mexican coast. However, during JJAS, El Niño-like condition is clearly visible (Fig. 7d). North Atlantic and AtlNiño

Table 2 Occurrences of El Niño/ La Niña and (+/−) Indian Ocean Dipole/Zonal Mode (IODZM) and consequent dry/wet year due to El Niño/ La Niña in each cluster

Cluster no.	El Niño years		La Niña years		Indian Ocean dipole/zonal mode	
	Dry years	Normal years	Wet years	Normal years	Positive IOD	Negative IOD
1	1918, 1987, 2002					1974
2	1951, 1965, 1972, 2009				1972	1928, 1992
3	1904, 1911, 1968, 1982	1913			1905, 1913, 1982	1968, 1985
4		1902, 1991, 1997			1902, 1991, 1997	1989
5		1925, 1930			1909, 1948, 1955	1909, 1930, 1980
6		1944, 1976	1970	1908, 1954, 1973, 2010	1926, 1944, 1945, 1977	1950
7		1935, 1957		2000	1935, 1957, 1963, 1967	1906
8		2006	1933		1923, 1946, 1961, 1994	1933, 1958
9			1910, 1916, 1975, 1988		1983	1910, 1916, 1917, 1942, 1975, 1981

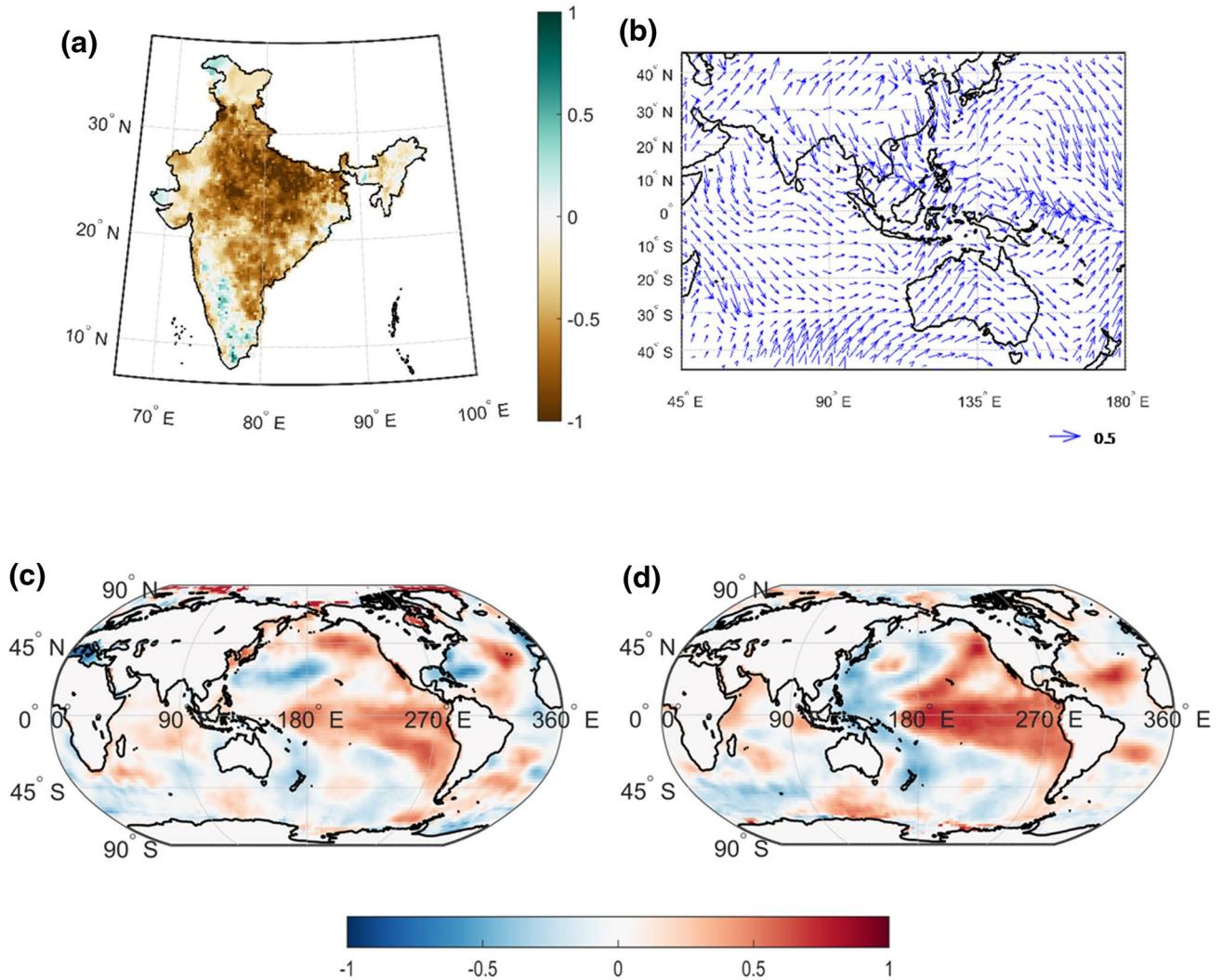


Fig. 6 Same as Fig. 5 but for Cluster 2

regions show cold SSTs in both MAM and JJAS. In clusters 1 and 2, El Niño patterns are visible even during MAM, but in cluster 3, that becomes prominent only during and after JJAS. The strength and the flavor of ENSO are likely to determine the pre-monsoon ENSO-monsoon signatures as well monsoon predictability based on ENSO indices. Table-2 depicts 4 occurrences of El Niño events in this cluster. The southwesterly wind (Fig. 7b) and moisture transport are weakened by the eastern Pacific warming in cluster 3.

1982 SSTs for 1982 display similar characteristics as seen in 1972, as discussed in cluster 2. Although a positive IODZM prevails, the strong El Niño condition suppresses the Indian monsoon (Saji and Yamagata 2003).

It is important to note that even though clusters 2 and 3 consist of dry years, the spatial patterns of JJAS rainfall are distinctly different across these clusters. This indicates that

different hydroclimatic teleconnections from the SST patterns cause different spatial variations over ISMR.

Cluster 4 This cluster contains 9 years, of which, 8 occurred during the post-1955 period. This could probably be one of the clusters which contribute to the drying of the Indian subcontinent as reported in many studies (Saha et al. 2014; Roxy et al. 2015). The composite shows negative rainfall anomalies over the eastern central India, upper Ganga basin and northwestern India (Fig. 8a). The rainfall anomalies show normal-to-excess rainfall over the northeastern and southern India. AIMR does not distinguish these years as dry years as indicated in Table 2. However, there are several dry regions over the country, for most of the years belonging to cluster 4. Warm SSTs are observed over much of the low-latitude oceanic region during both MAM and JJAS (Fig. 8c, d). The extratropical northern oceanic regions such as the North Atlantic and a small part

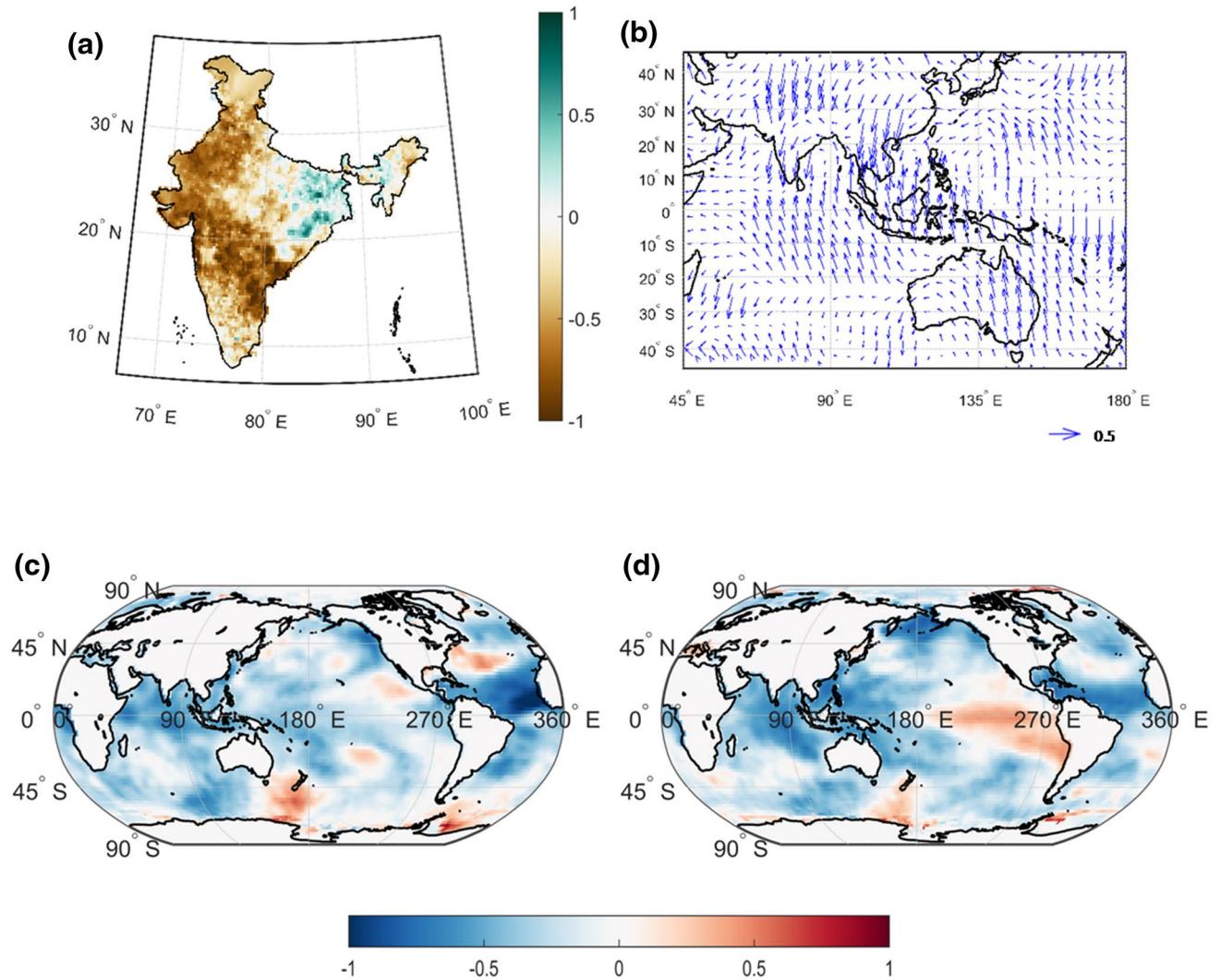


Fig. 7 Same as Fig. 5 but for Cluster 3

of North Pacific exhibit cold conditions. Cold SSTs over the North Atlantic and warm SSTs over the tropical Atlantic are observed in clusters 1 and 4. This SST pattern may be associated with a slowdown of the Atlantic meridional overturning circulation (AMOC) (Rahmstorf et al. 2015), which is a potential consequence of global warming. This slowdown in AMOC may lead to a southward shift of the Inter-Tropical Convergence Zone (ITCZ) and a weakening of ISMR (Vellinga and Wood 2002; Chiang and Bitz 2005). This cluster is also responsible for the increased spatial variability (Ghosh et al. 2012), although, more analysis is needed to quantify the details.

The 850 hPa wind anomaly composites display weak southwesterlies (Fig. 8b). There are several El Niño events in this cluster. However, the Indian subcontinent faces no dry conditions during any of these events (Table 2, years 1902, 1991, 1997). This is associated with a strong

IODZM-induced convergence that compensates the negative impacts of El Niño leading to an overall normal AIMR.

Cluster 5 Table 1 shows that the occurrences of this cluster decrease from pre-1955 to post-1955 years. There are 19 years in this cluster, out of which, 13 belong to the pre-1955 period. The rainfall spatial pattern exhibits positive anomaly over the Gangetic region but a negative anomaly in the northwest and Peninsular India (Fig. 9a). SSTs are warm over the western Pacific during MAM (Fig. 9c) and over the northern Pacific during JJAS (Fig. 9d). The composites of anomalies during both MAM and JJAS show a warm North Atlantic but the tropics show cold SST. There are a few El Niño and La Niña events in this cluster, but apparently, they produce no significant dry or wet years (Table 2). The 850hPa wind anomalies indicate stronger southwesterlies towards the landmass of India but a divergence over the deep tropics (Fig. 9b). Further analysis is needed to understand

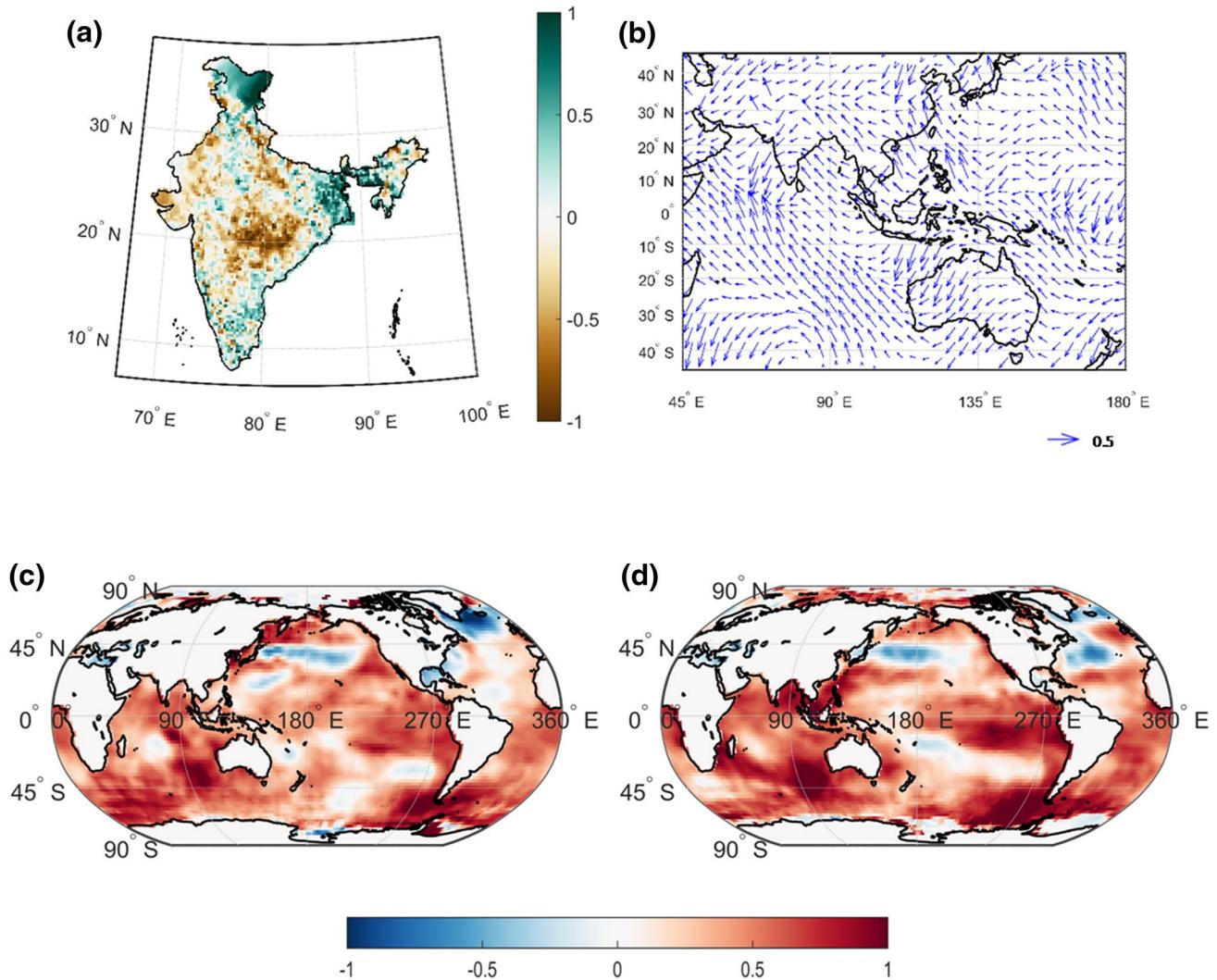


Fig. 8 Same as Fig. 5 but for Cluster 4

the circulation anomalies in each cluster in terms of the land–ocean redistribution of rainfall (Sahana et al. 2015) and the significance of the SST anomalies as drivers of rainfall anomalies over India.

Cluster 6 Table 1 shows that the occurrence of cluster 6 is well distributed over the pre-1955 and post-1955 periods. Positive rainfall anomalies over the western and central India are observed in this cluster (Fig. 10a). However, drier Gangetic and Peninsular regions are seen in the composite of rainfall anomalies. This cluster has SST anomalies with above normal temperatures in central Pacific, North Atlantic, South China Sea and AtlNiño during MAM and JJAS (Fig. 10c, d). Certain cold conditions prevail over the eastern Pacific favoring positive moisture transport to India. This cluster also has a few moderate El Niño/La Niña episodes, but they do not produce any dry/wet years (Table 2). There is a convergence of 850 hPa winds over northwest India,

which appears to cause enhanced accumulation of moisture (Fig. 10b).

Cluster 7 This cluster consists of 12 years (Table 1), with half the years in each period. Peninsular India experiences positive rainfall anomaly, while negative anomaly is observed over the Western Ghats, southeastern and western India. Rainfall all across India is observed to be near-normal as in Fig. 11a. SSTs are moderately warm over the equatorial and North Pacific Oceans in both MAM and JJAS (Fig. 11c, d). The cluster consists of a few El Niño years, but the positive IODZM seem to negate the El Niño impacts (Table 2). Wind anomalies in Fig. 11b suggest a reduced moisture flux, resulting in the deficit rainfall over central India.

Cluster 8 This cluster consists of 11 years of which 5 occur in the pre-1955 period (Table 1). This cluster indicates positive rainfall anomalies over the Western Ghats, western and central India (Fig. 12a). Negative rainfall

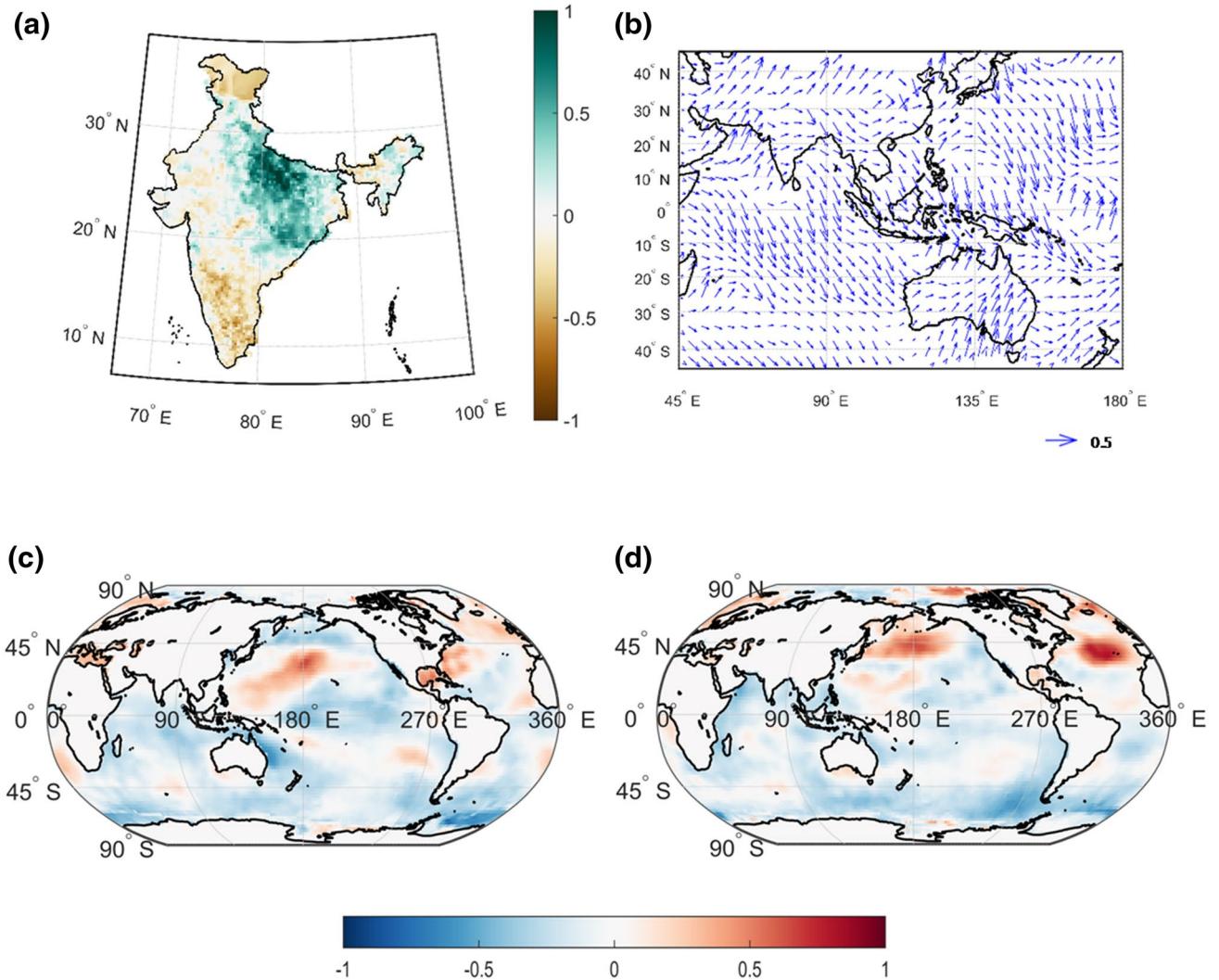


Fig. 9 Same as Fig. 5 but for Cluster 5

anomaly is observed over the Gangetic Plain, northeast India and eastern peninsular regions of the country. The MAM SSTs show positive anomalies over the equatorial region, western Indian and western Pacific Oceans (Fig. 12c). However, cold SST anomalies are observed over the eastern Pacific and western Indian Oceans during JJAS (Fig. 12d). The southwesterly wind anomalies correspond to the good rainfall over the core monsoon zone (Fig. 12b).

Cluster 9 16 years belonging to this specific cluster are equally distributed over the two periods (Table 1). Figure 13a shows positive rainfall anomaly over almost the entire Indian subcontinent except over northeast India. A cold equatorial Pacific is observed in these years, during both MAM and JJAS, suggesting consistent occurrences of La Niña (1910, 1916, 1975 and 1988; Fig. 13c, d) with a warm western Pacific Ocean. Amongst these La Niña events, a few co-occur with the negative IODZM events (1910, 1916

and 1975) (Table 2). However, negative IODZM does not seem to play an effective role in decreasing rainfall over the country. It must be pointed out that negative IODZMs do not tend to have as strong coupling as during some positive IODZM years (Annamalai et al. 2003). The high wind anomaly values extend from the Bay of Bengal suggesting more moisture flux over the country (Fig. 13b).

3.2 Relationship between sea surface temperatures and ISMR spatial variability using CART model

The discussion regarding clusters above suggests potentially direct relation of SST patterns with the spatial structure of monsoon rainfall over India. We hypothesize that spatial structures of seasonal rainfall depend on SST patterns in MAM which offers a potential to forecast seasonal rainfall with a sufficient lead time which should enhance usable

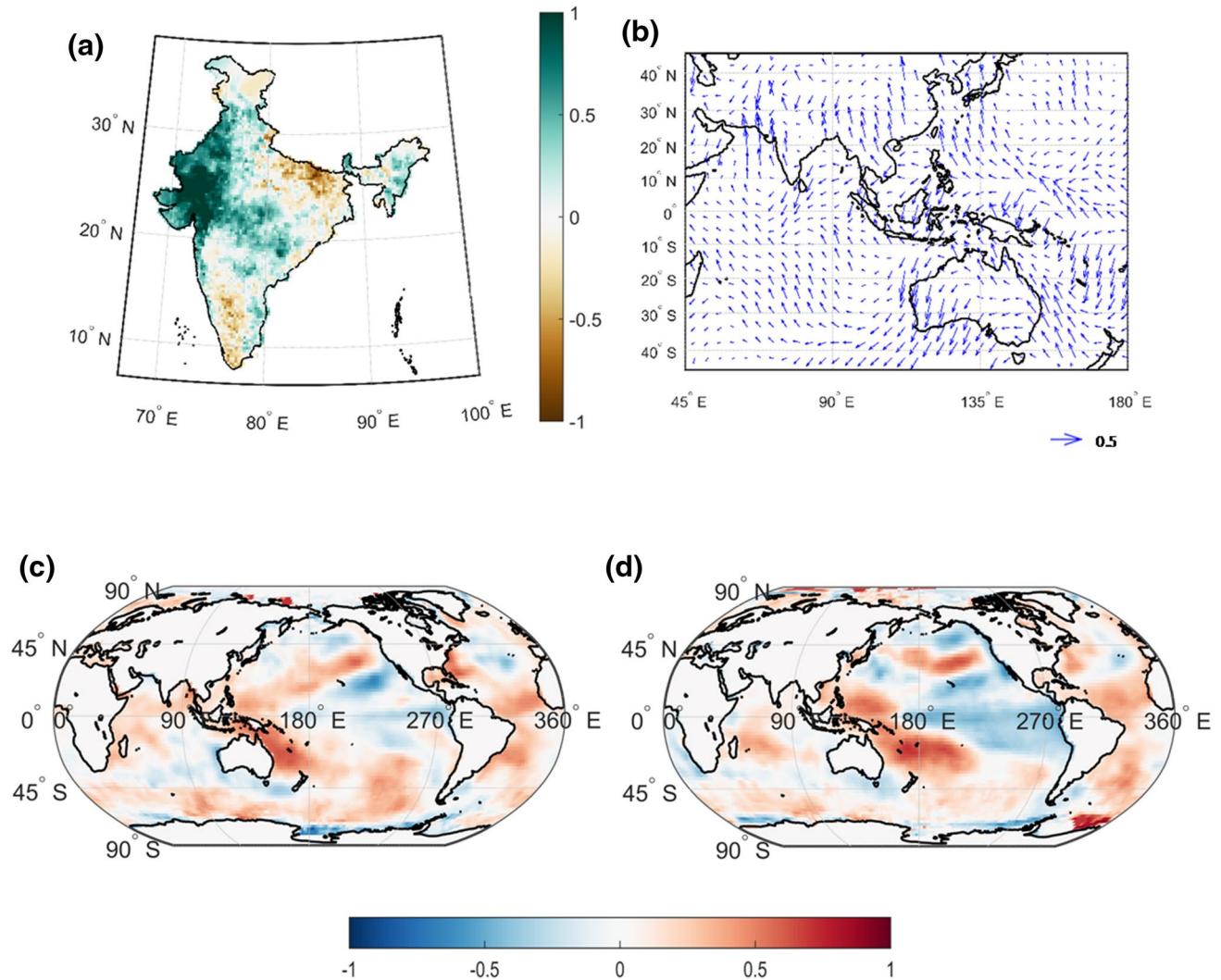


Fig. 10 Same as Fig. 5 but for Cluster 6

spatial information. We develop a simple Classification and Regression Tree (CART) to model this relationship for the training period of 1901–2010. We use SST over 12 oceanic regions as predictors, given in Table 3, and the cluster indices as predictands. Furthermore, we use this relationship to validate our model with predictions for the last 5 years (2011–2015) of the study period.

We use categorical values for individual predictors to simplify the structure of CART. The categories are based on standardized anomalies with the same logic as used for rainfall (Sect. 2.2). These categorical values are used as input in CART model. Clusters based on spatial patterns of the Indian monsoon are used as predictands. There might be different sets of input conditions which can lead to the same cluster index. The structure of CART has nodes, which signify SST of an oceanic region and the leaves as a cluster index. Figure 14a shows the tree structure derived from the

classification algorithm. Classification algorithm progresses from each node to the left (right), if the standardized value of SST over the region is less (greater or equal to) than the value at the node. The categorical value for each node is a multiple of one standard deviation of SSTs for the respective oceanic region. For example in Fig. 14a, the value assigned to the parent node (AtlNiño) is -2.5 . This indicates that cluster 3 is an outcome of the SST conditions where the standardized anomaly of AtlNiño is less than -2.5 . The composites of SST for cluster 3 at AtlNiño region shows cold condition (Fig. 7c) and this is consistent with the CART structure.

For a standardized value more than -2.5 over the AtlNiño, the tree proceeds to the Niño3 oceanic region with a standardized anomaly value of 0.5 . The left branch from Niño3 indicates standardized values lower than 0.5 which leads to SIO with standardized anomaly value of -1.0 .

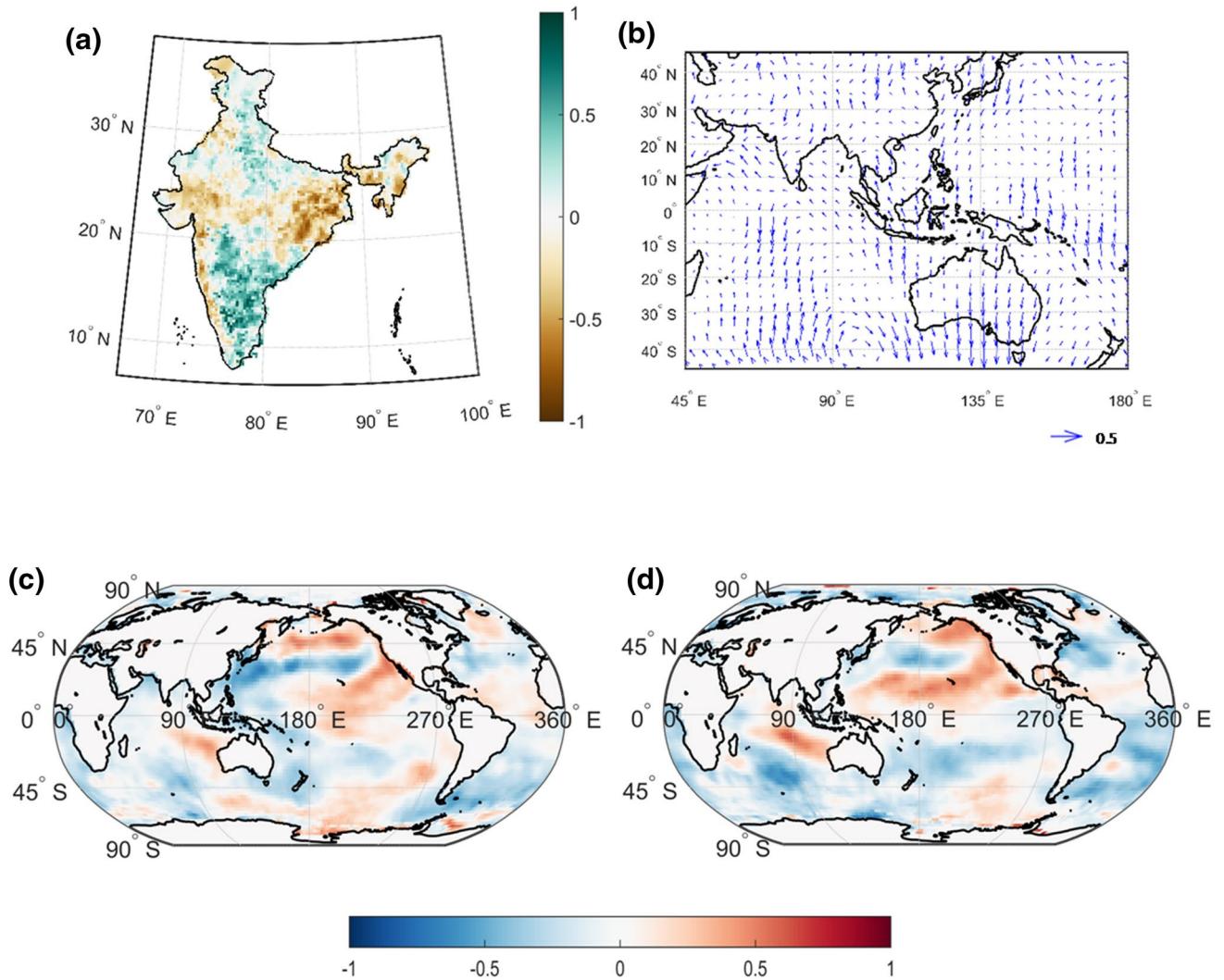


Fig. 11 Same as Fig. 5 but for Cluster 7

Traversing to the left of the SIO node leads to AtlNiño conditions with standardized anomaly value -1.0 . Now, if the SSTs over AtlNiño are cooler than the standardized anomaly of -1.0 , it will lead to cluster 2, otherwise, cluster 5. Therefore, the logical combinations of the standardized anomaly between -2.5 and -1.0 over AtlNiño, less than 0.5 over Niño3 and less than -1.0 over SIO, will lead to cluster 2. Similarly, the combinations of the standardized anomaly greater than or equal to -1.0 over AtlNiño, less than 0.5 over Niño3 and less than -1.0 over SIO, will lead to cluster 5. Comparing these logical combinations with MAM SST patterns in cluster 2 (Fig. 6c), we find that they are consistent. Similarly, cluster 5 also shows below normal SSTs over the AtlNiño, SIO and Niño3 regions, as evident in Fig. 9c. Normal SST conditions are observed over the AtlNiño region in cluster 5; whereas cluster 2 has

cooler SST. This difference of MAM SST patterns over the region is registered in both the clusters (Figs. 6c, 9c).

Furthermore, such combinations of SSTs over the 12 oceanic regions are helpful to determine the spatial pattern of rainfall from the output as cluster indices. This trained CART model is tested to predict the spatial monsoon rainfall patterns over the validation period of 2011–2015, from their individual SST anomalies during MAM.

3.3 Prediction of spatial patterns over ISMR for 2011–2015 using trained CART model

Here, we discuss the prediction of spatial patterns of seasonal monsoon rainfall for the period of 2011–2015. We employ categorical values (as used in Sect. 2.2) of the MAM SST for each year, over the 12 oceanic regions defined, as

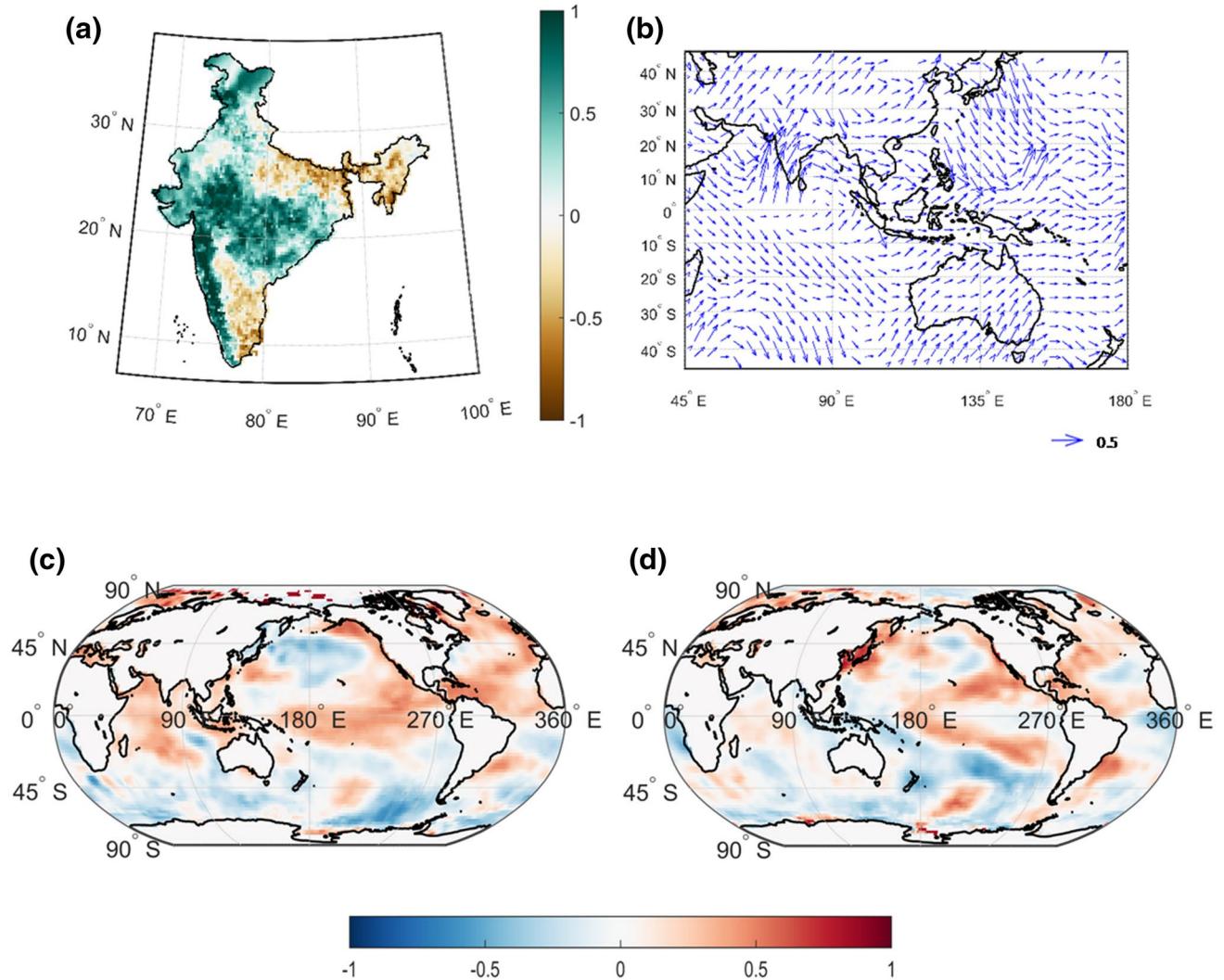


Fig. 12 Same as Fig. 5 but for Cluster 8

the predictor input for the trained CART. For evaluating the model performance, the gridwise standardized anomaly of Indian monsoon rainfall is computed for years 2011–2015, using the climatological mean and standard deviation of 1901–2010 period. These spatial patterns of yearly rainfall standardized anomalies are compared with the predicted spatial patterns as obtained from the CART. Figure 15 shows the comparison between observed standardized rainfall anomalies and CART model predictions for years 2011–2015. The rainfall anomalies during 2011 show quite a similar spatial pattern as predicted by the model (Fig. 15). A high rainfall anomaly is seen over the Western Ghats, central and northwestern India. The retrospective inspection from CART structure shows that the warming over the WIO, NP2 and AtlNiño regions consequently predicts the rainfall pattern corresponding to cluster 8. Prediction for the year 2012 produces cluster 9. However, the observed

pattern for 2012 shows a high heterogeneity in rainfall over the Indian landmass. Cluster 9 has high rainfall anomaly over peninsular India whereas the observed rainfall pattern shows drier conditions. We find that this failure is attributable to the sudden transition of Pacific SST from MAM to JJAS. Xue et al. (2013) showed a warming of tropical Pacific SSTs in 2012, where earlier, La Niña conditions existed. El Niño-like conditions started prevailing over the central-eastern Pacific Ocean during JJAS because of which rainfall subsided over India. As our model considers MAM SSTs, it fails to account the warming of SST over Pacific during JJAS. Observed monsoon rainfall during 2013 shows the similarity of dry regional conditions over northeastern and peninsular India, with CART prediction. The observed rainfall patterns for 2014 and 2015 indicate dry conditions over the northeast and central India but wet conditions over the peninsula, similar to the predicted spatial pattern from

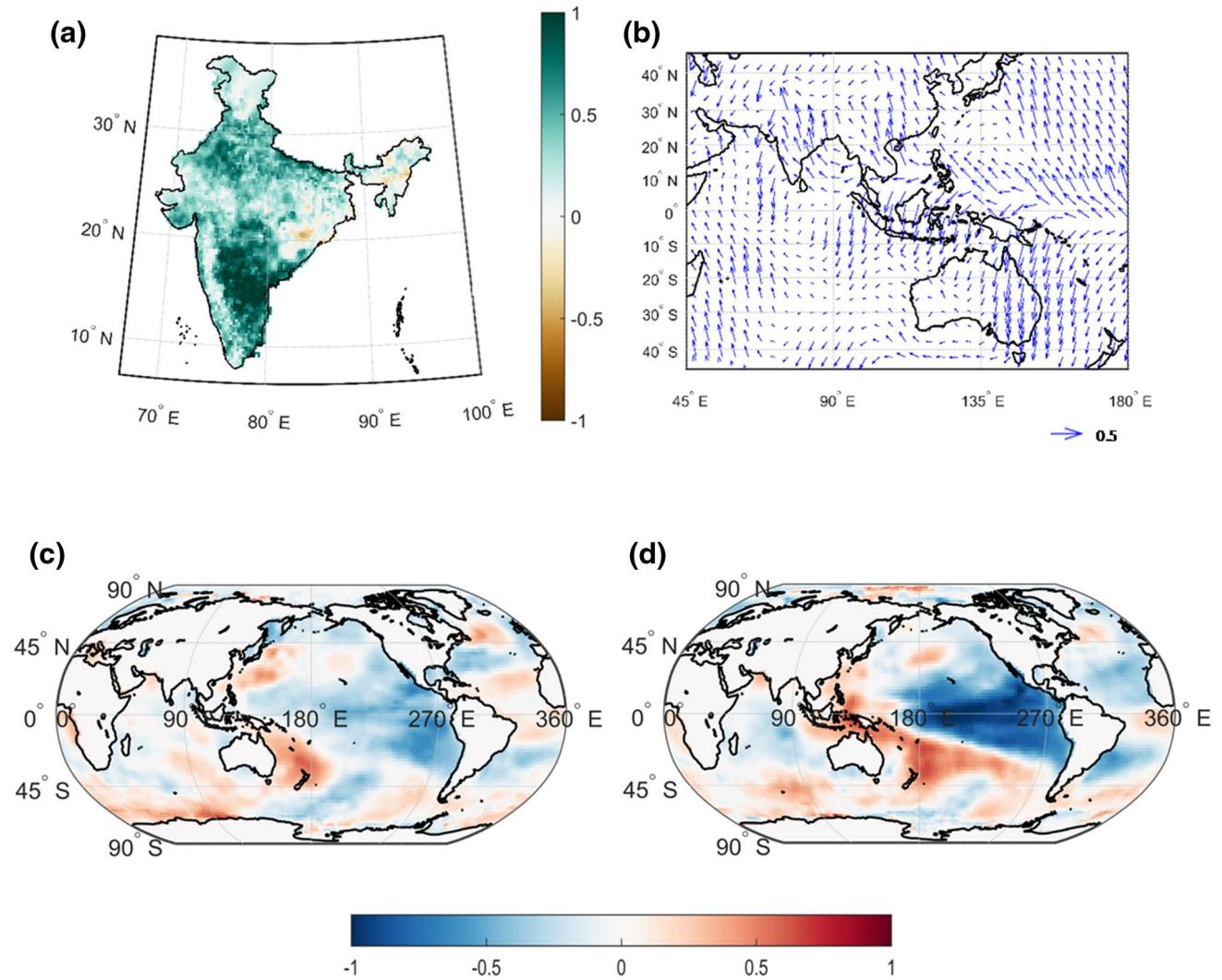


Fig. 13 Same as Fig. 5 but for Cluster 9

Table 3 SST regions for CART model

SST region	Co-ordinates of SST region	Abbreviations
Atlantic Niño	5°N–5°S, 340°E–360°E	AtlNiño
Bay of Bengal	6°N–20°N, 80°E–100°E	BoB
Indian Ocean Dipole/Zonal Mode	10°S–10°N, 50°E–70°E 10°S–0°, 90°E–110°E	IOD
Niño3	5°S–5°N, 210°E–270°E	Niño3
Niño3.4	5°S–5°N, 190°E–240°E	Niño3.4
Niño4	5°S–5°N, 160°E–210°E	Niño4
North Atlantic	30°N–45°N, 300°E–350°E	NA
North Pacific 1	30°N–65°N, 140°E–180°E	NP1
North Pacific 2	30°N–65°N, 200°E–240°E	NP2
South China Sea	5°N–15°N, 110°E–120°E	SCS
Southern Indian Ocean	5°S–15°N, 40°E–100°E	SIO
Western Indian Ocean	5°S–10°N, 50°E–65°E	WIO

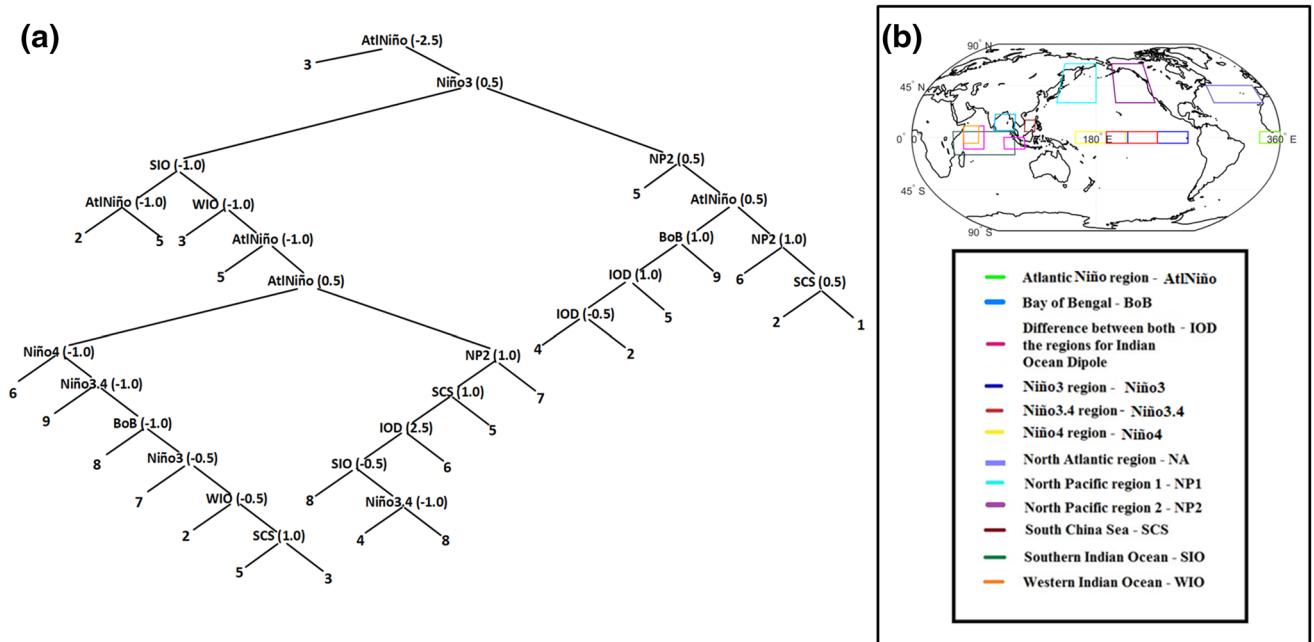


Fig. 14 **a** Classification and regression tree (CART). **b** Regions of SST over oceanic regions used as predictors in the CART model

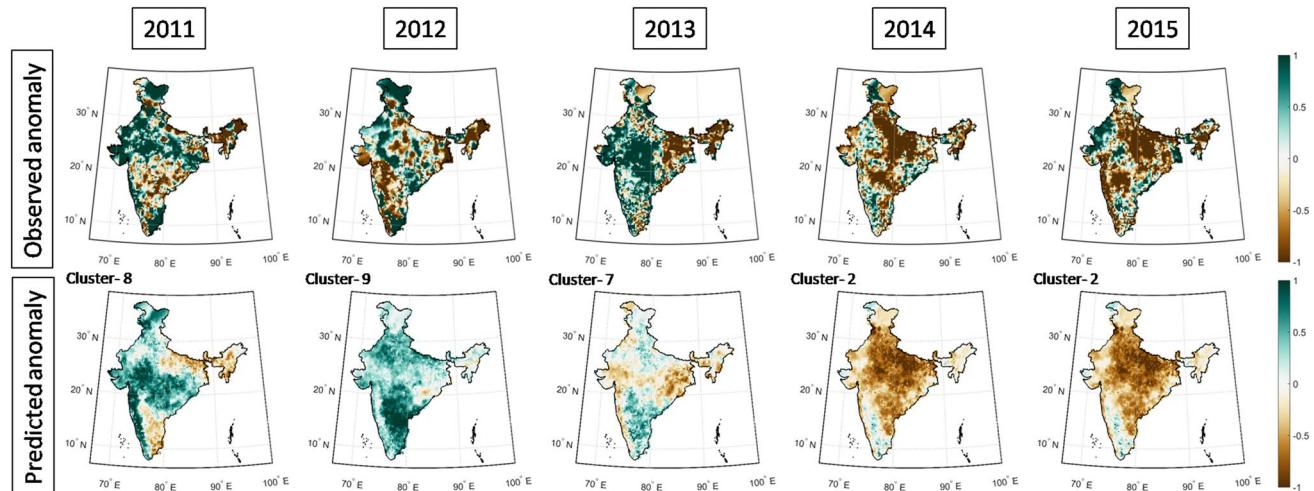


Fig. 15 Observed and predicted rainfall anomaly

CART predictions. The model captures the impact of warm SST from the central and eastern Pacific Ocean and predicts JJAS rainfall spatial pattern for years 2014 and 2015 as represented in cluster 2. From the above analysis, we find that 4 out of 5 years validate the Classification model very well. To evaluate the extent of the matches at a grid level in terms of three conventionally used classes, dry, wet or normal. The results are presented in Supplementary Fig. 1. We find more than 80% matches in majority of the grids, which demonstrates that the model can be used efficiently

for seasonal prediction at regional scale. The results of 2012 indicate that there is a need to further improve the model by considering the predicted SST of JJAS from a dynamic model as an extra input variable to CART. The evolution of El Niño-like conditions and the sudden demise during early fall of 2012 and 2014, may correspond to special cases in terms of the ocean–atmosphere interactions or the monsoon-ENSO feedbacks. This is clearly a future area of research.

To understand the sensitivity of derived CART to the selection of training period, we re-derive a classification

tree with the training period as 1901–1979 (Supplementary Fig. 2), and validate it for the period of 1980–2015. We find changes in the derived classification tree (Supplementary Fig. 2), as can be expected; however, the major nodes of the tree derived with the training data of 1901–2010, such as SIO, Niño3, AtlNiño and NP2 remain the key nodes in the newly derived tree also. This shows the robustness of the model, though it is highly sensitive to the selection of the training period. We have validated the model (newly derived tree with training period 1901–1979) for the period 1980–2015. We find more than 50% matches (Supplementary Fig. 3) in each class (dry, normal or wet based on climatology, details in Supplementary Information) even at a grid level for the validation period. Hence, the key objective of seasonal prediction at a regional scale with a depiction of spatial variability is fulfilled with this method. To achieve the best prediction with the proposed CART algorithms, it is better to consider the years as training data with significant number of occurrences of all the clusters. We consider 1901–2010 as the training period keeping that in mind and find that the model performs well for 4 years out of remaining 5, during validation. We also expect that as with all statistical forecast applications, updates of the training periods must be an integral component of the system, especially based upon regular assessment of skill and validation.

4 Summary and conclusions

Here, we analyze the spatiotemporal distributions of ISMR and find that they are directly linked with the SSTs over different oceanic regions. We acquire nine different spatial patterns and investigate the spatial heterogeneity of JJAS rainfall within each cluster. These nine clusters provide more regional information, which is not available from the operational forecasted spatially averaged AIMR. We discuss the hydroclimatic teleconnections between these spatial patterns of rainfall and standardized SST over different oceanic regions. This reveals that the SSTs are not only helpful in predicting ISMR as reported in the literature but also provides information about the spatial structure of seasonal monsoon rainfall. We posit that further corroborating these clusters and their renditions in coupled climate models is a pathway to extract the drivers of these clusters in terms of remote (e.g., SSTs) and local (e.g., soil moisture) forcing, moisture sources, and dependence on background states. This novel approach will advance the process understanding of the dominant spatial structures of excess and deficit rainfall over India, as well as spatiotemporal biases in models that affect monsoon predictions. Such dynamic and thermodynamic understanding of the process can be expected to translate to enhanced prediction skill of regional rainfall and offer more hydroclimatic information that is more useful for

agriculture and seasonal water management planning including seasonal planning of newly initiated nation-wide inter-basin water transfer.

In order to develop the seasonal prediction model for further understanding, variables and processes corresponding to each cluster are investigated. We find that different modes of large-scale events such as El Niño can induce different spatial variability over ISMR, as seen in cluster 1, 2 and 3. Moreover, cluster 4 indicates that positive IODZM conditions also impact these regional-scale rainfall distributions. However, there are no signs of negative IODZM diminishing rainfall during La Niña events as addressed in Table 2. But Atlantic Niño may have a role to play during a La Niña as reported by Pottapinjara et al. (2014, 2015). We also investigate the dominance of the clusters in pre-1955 and post-1955 years in Table 1. We find that cluster 5 is more prevalent during the pre-1955 period. However, clusters 1 and 4 are dominant during the post-1955 period, where oceanic surfaces show significant warming.

Interannual variations in SST over the chosen oceanic regions during MAM can have strong hydroclimatic teleconnections with regional patterns of ISMR. These connections are used to develop a simplistic data-driven model to predict spatial patterns of ISMR. We feature 12 oceanic regions affecting the spatial patterns of ISMR. A Classification and Regression Model is trained for the period 1901–2010 with SSTs over these 12 oceanic regions as predictors and cluster indices as predictands. Other than the conventional predictors used for Indian monsoon, we find that the Atlantic Niño region also plays an important role in the classification model (Fig. 14a). The model shows that cold SST over AtlNiño can lead to severe dry conditions with a unique spatial pattern over India. Moreover, we present other oceanic regions such as the North Pacific region 1, North Pacific region 2 and North Atlantic SSTs playing a vital role in the model. The variations of SSTs over different regions drive ISMR with a particular spatial pattern. We validate our model for the years 2011–2015. We find that the classification model is able to predict 4 out of 5 years. The failure of the model during 2012 is likely due to the sudden transition of SST over Pacific from the season MAM to JJAS.

The operational seasonal monsoon prediction systems focus on the seasonal total monsoon rainfall spatially averaged over the entire country. However, such a prediction does not serve the regional water management or food production needs due to the high spatial heterogeneity of rainfall distribution. The regional seasonal models based on teleconnection fail due to the loss of strength in the impacts of remote SSTs on regional/ local precipitation. Here, rather than focusing on individual local regions, we use a unique approach of considering the spatial patterns of precipitation over a large region as extracted by a clustering process that can still provide valuable local information while also

retaining large-scale dynamical links which are potentially easier to capture in prediction models. We find that this minimalistic technique works well for the majority of the monsoon seasons considered here. At present this method considers only the past SSTs for demonstration purposes but may be improved further with consideration of evolving SSTs as simulated by a dynamic prediction model. SST anomalies over the identified forcing regions can be used in forcing AGCMs to explore initial driving mechanisms and existing hindcasts can be explored for model skill in capturing the clusters and their teleconnections to the various oceanic regions. We deem the clusters identified here as a novel way forward for analyzing monsoon simulations and potentially enhancing monsoon prediction skills based on the process understanding extracted from this approach. Verifying model renditions of these clusters should also offer an additional metric for evaluating model performances in terms of local and remote impacts on ISMR especially when combined with moisture source analysis depicted in the approach of Pathak et al. (2014, 2017). In addition, the clusters offer a framework to bring together seemingly disparate studies that report shifts in onset and withdrawal (Sahana et al. 2015; Sabeerali et al. 2013), trends in spatial and temporal extremes (Ghosh et al. 2012; Roxy et al. 2017), and impacts of distinct modes of variability such as ENSO, IODZM and Atlantic Niño as well as extratropical modes.

Acknowledgements The work presented in financially supported by Ministry of Earth Sciences, Government of India through Project No. MoES/PAMC/H&C/35/2013-PC-II. The rainfall data is obtained from India Meteorological Department. The reanalysis data is obtained from ERA-20C. The SST data is obtained from ER-SST. The authors sincerely thank the editor and the reviewer for providing very useful comments.

References

- Annamalai H, Slingo JM (2001) Active/break cycles: diagnosis of the intraseasonal variability of the Asian summer monsoon. *ClimDyn* 18:85–102. <https://doi.org/10.1007/s003820100161>
- Annamalai H, Murtugudde R, Potemra J, Xie SP, Liu P, Wang B (2003) Coupled dynamics in the Indian Ocean: spring initiation of the zonal mode. *Deep-Sea Res II* 50:2305–2330. [https://doi.org/10.1016/S0967-0645\(03\)00058-4](https://doi.org/10.1016/S0967-0645(03)00058-4)
- Annamalai H, Potemra J, Murtugudde R, McCreary JP (2005) Effect of preconditioning on the extreme climate events in the tropical Indian Ocean. *Am Metereol Soc.* <https://doi.org/10.1175/JCLI3494.1>
- Arpe K, Dumenil L, Giorgetta MA (1998) Variability of the Indian monsoon in the ECHAM3 model: sensitivity to sea surface temperature, soil moisture, and the stratospheric quasi-biennial oscillation. *J Clim* 11:1837–1858. <https://doi.org/10.1175/1520-0442-11.8.1837>
- Ashok K, Behera S, Rao AS, Weng H, Yamagata T (2007) El Niño Modoki and its teleconnection. *J Geophys Res* 112:C11007. <https://doi.org/10.1029/2006JC003798>
- Bamzai AS, Shukla J (1999) Relation between Eurasian snow cover, snow depth and the Indian summer monsoon: an observational Study. *J Clim* 12:3117–3132. [https://doi.org/10.1175/1520-0442\(1999\)012<3117:RBESCS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<3117:RBESCS>2.0.CO;2)
- Barik B, Ghosh S, Sahana AS, Pathak A, Sekhar M (2016) Water-food-energy nexus with changing agricultural scenarios in India during recent decades. *Hydrol Earth SystSci* 21:3041–3060. <https://doi.org/10.5194/hess-21-3041-2017>
- Chen D, Lian T, Fu C, Cane MA, Tang Y, Murtugudde R, Song X, Wu Q, Zhou L (2015) Strong influence of westerly wind bursts on El Niño diversity. *Nature Geosci* 8:339–345. <https://doi.org/10.1038/ngeo2399>
- Chiang JCH, Bitz CM (2005) Influence of high latitude ice cover on the marine Intertropical convergence zone. *Clim Dyn* 25:5:477–496. <https://doi.org/10.1007/s00382-005-0040-5>
- DeSole T, Shukla J (2012) Climate models produce skillful predictions of Indian summer monsoon rainfall. *Geophys Res Lett* 39:L09703. <https://doi.org/10.1029/2012GL051279>
- Gadgil S, Iyengar RN (1980) Cluster analysis of rainfall stations of the Indian peninsula. *Q J R Meteorol Soc* 106:873–886. <https://doi.org/10.1002/qj.49710645016>
- Gadgil S, Vinayachandran PN, Francis PA, Gadgil S (2004) Extremes of the Indian summer monsoon rainfall, ENSO and equatorial Indian Ocean oscillation. *Geophys Res Lett* 31:L12213. <https://doi.org/10.1029/2004GL019733>
- Ghosh S, Das D, Kao SC, Ganguly AR (2012) Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. *Nat Clim Change* 2:86–91. <https://doi.org/10.1038/nclimate1327>
- Goswami BN, Venugopal V, Sengupta D, Madhusoodanan MS, Xavier PK (2006) Increasing trend of extreme rain events over India in a warming environment. *Science* 314(5804):1442–1445. <https://doi.org/10.1126/science.1132027>
- Gowariker V, Thapliyal V, Kulshrestha SM, Mandal GS, Sen Roy N, Sikka DR (1991) A power regression models for long range forecast of southwest monsoon rainfall over India. *Masaum* 42:125–130
- Iyengar RN, Basak P (1994) Regionalization of Indian monsoon rainfall and long term variability signals. *Int J Clim* 14:1095–1114. <https://doi.org/10.1002/joc.3370141003>
- Kane RP (1998) Extremes of the ENSO phenomenon and Indian summer monsoon rainfall. *Int J Clim.* [https://doi.org/10.1002/\(SICI\)1097-0088\(19980615\)18:7<775::AID-JOC254>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1097-0088(19980615)18:7<775::AID-JOC254>3.0.CO;2-D)
- Kashid SS, Maity R (2012) Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using genetic programming. *J Hyd* 45:42–41. <https://doi.org/10.1016/j.jhydrol.2012.05.033>
- Kashid SS, Ghosh S, Maity R (2010) Streamflow prediction using multi-scale rainfall obtained from hydroclimatic teleconnection. *J Hyd* 395:23–38. <https://doi.org/10.1016/j.jhydrol.2010.10.004>
- Krishna Kumar K, Rajgopalan B, Cane MK (1999) On the weakening relationship between the Indian monsoon and ENSO. *Science* 284:2156–2159. <https://doi.org/10.1126/science.284.5423.2156>
- Kucharski F, Bracco A, Yoo JH, Molteni F (2007) Low-frequency variability of the Indian monsoon—ENSO relationship and the tropical Atlantic: the ‘weakening’ of the 1980s and 1990s. *J Clim* 20:4255–4266. <https://doi.org/10.1175/JCLI4254.1>
- Kucharski F, Bracco A, Yoo JH, Molteni F (2008a) Atlantic forced component of the Indian monsoon interannual variability. *Geophys Res Lett* 33:L04706. <https://doi.org/10.1029/2007GL033037>
- Kulkarni A, Kripalani RH (1998) Rainfall patterns over India: classification with fuzzy c-means method. *Theor App Climtol* 59:137–146. <https://doi.org/10.1007/s007040050019>

- Kulkarni A, Kripalani RH, Singh SV (1992) Classification of summer monsoon rainfall patterns over India. *Int J Clim* 12:269–280. <https://doi.org/10.1002/joc.3370120304>
- Kung EC, Sharif TA (1982) Long-range forecasting of the Indian summer monsoon onset and rainfall with upper air conditions. *J Meteorol Soc Jpn* 60:672–681
- Lau NC, Nath MJ (2000) Impact of ENSO on the variability of the Asian–Australian monsoons as simulated in GCM experiments. *J Clim* 13:4287–4307. [https://doi.org/10.1175/1520-0442\(2000\)013<4287:IOEOTV>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<4287:IOEOTV>2.0.CO;2)
- Maity R, Nagesh Kumar D (2006) Hydroclimatic association of monthly summer monsoon rainfall over India with large-scale atmospheric circulation from tropical Pacific Ocean and Indian Ocean region. *Atmos Sci Lett R Meteorol Soc Wiley InterSci* 7:101–107. <https://doi.org/10.1002/asl.141>
- Mooley DA, Parthasarathy B (1983) Indian summer monsoon and El Niño. *Pageoph* 121 2:339–352
- Murtugudde R, Goswami BN, Busalacchi AJ (1998) Air–sea interaction in the southern tropical Indian Ocean and its relations to interannual variability of the monsoon over India. In: Proceedings of the international conference on monsoon and hydrologic cycle. Korean Meteorological Agency, pp 184–188
- Pai DS, Sridhar L, Rajeevan M, Sreejith OP, Satbhai NS, Mukhopadhyay B (2014) Development of a new high spatial resolution ($0.25^\circ \times 0.25^\circ$) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *MAUSAM* 65 1:1–18
- Parthasarathy B, Pant GB (1984) The spatial and temporal relationships between the Indian summer monsoon rainfall and the Southern Oscillation. *Int J Clim* A36:269–277. [https://doi.org/10.1002/\(SICI\)1097-0088\(199612\)16:12%3C1391::AID-JOC94%3E3.0.CO;2-X](https://doi.org/10.1002/(SICI)1097-0088(199612)16:12%3C1391::AID-JOC94%3E3.0.CO;2-X)
- Parthasarathy B, Rupa Kumar K, Deshpande VR (1991) Indian summer monsoon rainfall and 200 mb meridional wind index: application for long range prediction. *Int J Clim* 11:165–176
- Pathak A, Ghosh S (2014) Precipitation recycling in the Indian subcontinent during summer monsoon. *J Hydromet* 15(5):2050–2066. <https://doi.org/10.1175/JHM-D-13-0172.1>
- Pathak A, Ghosh S, Martinez JA, Dominguez F, Kumar P (2017) Role of oceanic and land moisture and transport in the seasonal and interannual variability of summer monsoon in India. *J Clim* 30:1839–1859. <https://doi.org/10.1175/JCLI-D-16-0156.1>
- Poli P, Hersbach H, Dee DP et al (2016) ERA-20C: an atmospheric reanalysis of the twentieth century. *J Clim*. <https://doi.org/10.1175/JCLI-D-15-0556.1>
- Pottapinjara V, Girishkumar MS, Ravichandran M, Murtugudde R (2014) Influence of the Atlantic zonal mode on monsoon depressions in the Bay of Bengal. *J Geophys Res Atmos* 119:11. <https://doi.org/10.1002/2014JD021494>
- Pottapinjara V, Girishkumar MS, Sivareddy S, Ravichandran M, Murtugudde R (2015) Relation between the upper ocean heat content in the equatorial Atlantic during boreal spring and the Indian monsoon rainfall during June–September. *R Meteorol Soc*. <https://doi.org/10.1002/joc.4506>
- Rahmstorf S, Box JE, Feulner G, Mann ME, Robinson A, Rutherford S, Schaffernicht EJ (2015) Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nat Clim Change* 5:475–480. <https://doi.org/10.1038/nclimate2554>
- Rajeevan M (2001) Prediction of Indian summer monsoon: status, problems and prospects. *Curr Sci* 81:1451–1457
- Rajeevan M, Pai DS, Thapliyal V (1998) Spatial and temporal relationships between global and surface air temperature anomalies and Indian summer monsoon. *Meteorol Atmos Phys* 66:157–171. <https://doi.org/10.1007/BF01026631>
- Rajeevan M, Pai DS, Dikshit SK, Kelkar RR (2004) IMD's new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003. *Curr Sci* 86(3):22–431
- Rasmusson EM, Carpenter TH (1983) The relationship between eastern equatorial Pacific Sea temperature and rainfall over India and Sri Lanka. *Mon Weather Rev* 111:517–528. [https://doi.org/10.1175/1520-0493\(1983\)111%3C0517:TRBEEP%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1983)111%3C0517:TRBEEP%3E2.0.CO;2)
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comp Appl Math* 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Roxy MK, Kapoor R, Terray P, Murtugudde R, Ashok K, Goswami BN (2015) Drying of Indian subcontinent by rapid Indian Ocean warming and a weakening land-sea thermal gradient. *Nat Comm*. <https://doi.org/10.1038/ncomms8423>
- Roxy MK, Ghosh S, Pathak A, Athulya R, Murtugudde R, Terray P, Rajeevan M (2017) A threefold rise in widespread extreme rain events over central India. *Nat Comm* 8:708. <https://doi.org/10.1038/s41467-017-00744-9>
- Sabeerali CT, Dandi AR, Dhakate A, Salunke K, Mahapatra S, Rao SA (2013) Simulation of boreal summer intraseasonal oscillations in the latest CMIP5 coupled GCMs. *J Geophys Res Atmos* 118:4401–4420. <https://doi.org/10.1002/jgrd.50403>
- Saha A, Ghosh S, Sahana AS, Rao EP (2014) Failure of CMIP5 climate models in simulating post-1950 decreasing trend of Indian monsoon. *Geophys Res Lett* 41(20):7323–7330. <https://doi.org/10.1002/2014GL061573>
- Saha M, Mitra P, Nanjundiah RS (2017) Deep learning for predicting the monsoon over the homogeneous regions of India. *J Earth Syst Sci* 126:54. <https://doi.org/10.1007/s12040-017-0838-7>
- Sahana AS, Ghosh S, Ganguly A, Murtugudde R (2015) Shift in Indian summer monsoon onset during 1976/1977. *Environ Res Lett* 10:054006. <https://doi.org/10.1088/1748-9326/10/5/054006>
- Saji NH, Yamagata T (2003) Possible impacts of Indian Ocean dipole mode events on global climate. *Clim Res* 25:151–169. <https://doi.org/10.3354/cr025151>
- Saji NH, Goswami BN, Vinayachandran PN, Yamagata T (1999) A dipole mode in the tropical Indian Ocean. *Nature* 401:360–363
- Sardeshmukh PD, Compo GP, Penland C (2000) Changes of probability associated with El Niño. *J Clim* 13 (24):4268–4286. [https://doi.org/10.1175/1520-0442\(2000\)013<4268:COPAWE>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<4268:COPAWE>2.0.CO;2)
- Shukla J, Paolino DA (1983) The Southern Oscillation and long range forecasting of the summer monsoon over India. *Mon Weather Rev* 111:1830–1837
- Sikka DR (1980) Some aspects of large scale fluctuations of summer monsoon rainfall over Indian in relation to fluctuations in the planetary and regional scale circulation parameters. *Proc Ind Acad Sci Earth Planet Sci* 89:179–195. <https://doi.org/10.1007/BF02913749>
- Smith TM, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements NOAAs Historical Merged Land-Ocean Temp Analysis (1880–2006). *J Clim* 21:2283–2296
- Vathsala H, Koolagudi SG (2016) Long-Range prediction of Indian summer monsoon rainfall using data mining and statistical approaches. *Theor Appl Clim* 130:19–33. <https://doi.org/10.1007/s00704-016-1862-2>
- Walker GT (1933) Seasonal weather and its prediction. *Nature* 132:805–808
- Wang B, Xiang B, Li J, Webster PJ, Rajeevan MN, Liu J, Ha K (2015) Rethinking Indian monsoon rainfall prediction in the context of recent global warming. *Nat Comm* 6:7154. <https://doi.org/10.1038/ncomms8154>
- Wang H, Murtugudde R, Kumar A (2016) Evolution of Indian Ocean dipole and its forcing mechanisms in the absence of ENSO. *Clim Dyn* 47:2481–2500
- Webster PJ, Magaña VO, Palmer TN, Shukla J, Tomas RA, Yanai M, Yasunari T (1998) Monsoons: processes, predictability,

- and the prospects for prediction. *J Geophys Res.* <https://doi.org/10.1029/97JC02719>
- Wu R, Kirtman BP (2004) Impacts of the Indian Ocean on the summer monsoon-ENSO relationship. *J Clim* 17:3037–3054. [https://doi.org/10.1175/15200442\(2004\)017%3C3037%3AIOTIOO%3E2.0.CO;2](https://doi.org/10.1175/15200442(2004)017%3C3037%3AIOTIOO%3E2.0.CO;2)
- Xue Y, Hu Z, Kumar A, Banzon V, Smith TM, Rayner NA (2013) [Global oceans] sea surface temperatures [in “state of the climate in 2012”]. *Bull Am Meteorol Soc* 94(8):S47–S50
- Yamagata T, Behera SK, Luo J, Masson S, Jury MR, Rao AS (2004) Coupled ocean–atmosphere variability in the tropical Indian Ocean. *Earth Clim Ocean Atmos Interact Geophys Monogr.* <https://doi.org/10.1029/147GM12>