

# Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks

Nishant Malik · Bodo Bookhagen · Norbert Marwan · Jürgen Kurths

Received: 29 April 2011 / Accepted: 23 July 2011  
© Springer-Verlag 2011

**Abstract** We present a detailed analysis of summer monsoon rainfall over the Indian peninsular using nonlinear spatial correlations. This analysis is carried out employing the tools of complex networks and a measure of nonlinear correlation for point processes such as rainfall, called event synchronization. This study provides valuable insights into the spatial organization, scales, and structure of the 90th and 94th percentile rainfall events during the Indian summer monsoon (June–September). We furthermore analyse the influence of different critical synoptic atmospheric systems and the impact of the steep Himalayan topography on rainfall patterns. The presented method not only helps us in visualising the structure of the extreme-event rainfall fields, but also identifies the water vapor pathways and decadal-scale moisture sinks over the region. Furthermore a simple scheme based on complex networks is presented to decipher the spatial intricacies and temporal evolution of monsoonal rainfall patterns over the last 6 decades.

**Keywords** Indian summer monsoon · Event synchronization · Complex networks · Rainfall patterns

## 1 Introduction

Indian Summer monsoon (ISM) rainfall over South Asia is the result of the interaction of several complex atmospheric processes evolving at many different spatial and temporal scales (e.g., Webster 1987). Apart from the influences of the interplay of synoptic scale weather phenomena, the ISM rainfall patterns are also modulated by the steep topography of the Himalayas (e.g., Bookhagen and Burbank 2010). Hence monsoonal rainfall has highly intricate spatiotemporal patterns. Here, we will analyse these spatiotemporal ISM rainfall patterns over South Asia employing nonlinear correlation measures called event synchronization and complex networks.

The methodology of complex networks has emerged as an important mathematical tool in the analysis of complex systems in the last decade and has been applied to a wide variety of disciplines within the natural and social sciences (e.g., Watts and Strogatz 1998; Newman 2003; Albert and Barabási 2002). The spatiotemporal structure of complex networks, dynamics on them, and phenomena such as network synchronization has been of large interest to the nonlinear-dynamics but also to the climate communities (e.g., Boccaletti et al. 2006; Arenas et al. 2008). In recent years, the tools of complex network theory has also found an application in the data driven analysis of the global climate system (e.g., Donges et al. 2009; Tsonis et al. 2006; Tsonis and Swanson 2008; Yamasaki et al. 2008; Donges et al. 2011). The correlation structure of climate variables and their teleconnections have been studied

---

N. Malik (✉) · N. Marwan · J. Kurths  
Potsdam Institute for Climate Impact Research,  
P.O. Box 60 12 03, 14412 Potsdam, Germany  
e-mail: malik@pik-potsdam.de

N. Malik  
Institute of Physics, University of Potsdam,  
Karl-Liebknecht-Str. 24/25, 14476 Potsdam-Golm, Germany

B. Bookhagen  
Department of Geography, University of California Santa  
Barbara, 1832 Ellison Hall, Santa Barbara,  
CA 93106-4060, USA

J. Kurths  
Department of Physics, Humboldt University,  
Newtonstr. 15, 12489 Berlin, Germany

employing this mathematical tool. Other applications of this approach in climate data analysis include the identification of community structures in the climate system (Tsonis et al. 2010) and the linkages of different regional climate phenomena (Steinhaeuser et al. 2010), leading to the discovery of a new dynamical mechanism for major climate shifts (Swanson and Tsonis 2009). A similar approach based on a shared-nearest neighbor method has been used to discover climate indices from sea surface temperature data (Steinbach et al. 2003). Here, we apply a complex networks approach to a specific regional climatic phenomena of the ISM and study the spatiotemporal pattern of rainfall.

Although rainfall has a significant impact on society, agriculture and fresh-water generation in this region (Bookhagen and Burbank 2010), it is not easy to decipher its dynamics due to its spatiotemporal complexity and involved small-scale processes. Furthermore, rainfall is a point process with large spatial and temporal discontinuities ranging from very weak to strong events within small temporal and spatial scales (Wulf et al. 2010). Nonlinear correlations called event synchronization can overcome these difficulties (e.g., Quiroga et al. 2002). The methodologies of complex networks and event synchronization take into account the nonlinearities existing in the correlation structure of the rainfall field. For the analysis presented here, we have considered the rainfall events at the 90th and 94th percentiles and thus focus on the extreme rainfall events and the associated atmospheric processes. By further developing the methodology and building upon previous investigations (Malik et al. 2010), we also attempt to study the evolution of monsoonal rainfall pattern over the last decades. The key advantage of applying complex network theory is that it does not require details of a several of climatic variables and indices one may have to analyse to study spatiotemporal rainfall patterns.

This study provides several critical insights into the underlying atmospheric processes responsible for the evolution of the ISM and related extreme rainfall events. Before introducing the methodology and data, we provide a climatic overview of the region.

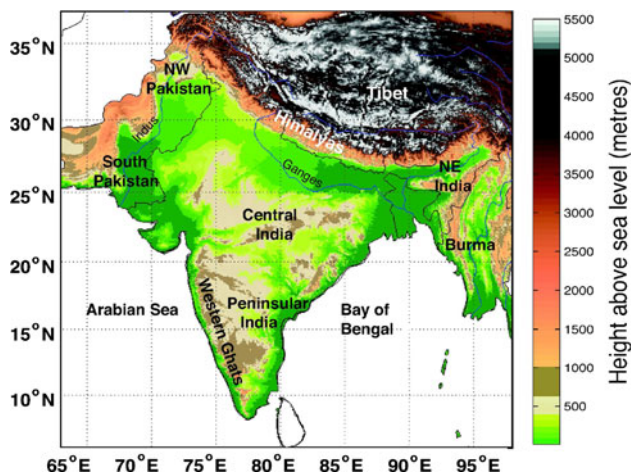
## 2 Climatic setting of the Indian summer monsoon (ISM)

We do not intend to give a full account of the ISM, its driving factors and dynamics and instead refer to summaries and references in Webster (1987), Webster et al. (1998), Gadgil (2003), and Wang (2006). In this section, we aim to synthesize characteristics that are pertinent to this study.

The ISM system is one of the active components of the global climate system in the tropics. It has been argued that the basic origin of the monsoon lies in the differential heating of the land and the sea during the summer season, which results in a positive moisture advection feedback leading to widespread rainfall over the Indian subcontinent (Webster 1987; Zickfeld et al. 2005; Levermann et al. 2009). The ISM accounts for a large part of the annual rainfall budget over much of the Indian subcontinent. For instance, the Ganges Plain and central parts of the Himalaya receive  $\sim 80\%$  of their annual rainfall budget during the ISM (Bookhagen and Burbank 2010). This, in turn leads to high societal and economic significance of this climatic phenomena (Webster et al. 1998; Gadgil 2003).

The onset of the ISM occurs on average at June 1st at the southern tip of the Indian peninsular (Webster 1987). While the onset of the monsoon is relatively stable through time, the intra-seasonal variation is high (Fasullo and Webster 2003). ISM rainfall is result of inherently large spatial scale processes in the atmosphere. An important feature of the ISM dynamics is the existence of two well pronounced rainfall regimes, the active and break phases (e.g., Webster et al. 1998; Waliser 2006). During the active phase, the rainfall is spatially widespread over the Indian subcontinent. In the break phase, rainfall ceases over the vast majority of land mass, except the foothills of Himalaya where it is observed to increase (Krishnamurthy and Shukla 2000). During the break phases, other meteorological features are observed that lead to a shifting of the monsoon trough to the Himalayan foothills and lead to an absence of low level easterly winds over the north of the subcontinent (Webster et al. 1998; Waliser 2006). ISM tends to abruptly fluctuate between these two regimes several times during a season (e.g., Webster et al. 1998; Waliser 2006). In addition to these rainfall-modulating factors there exists synoptic-scale weather systems such as monsoon depressions and lows. Previous studies also indicate that rainfall has large spatial extents during monsoonal depressions (Stephenson et al. 1999; Sikka 1977) or mid-tropospheric cyclones (Keshavamurthy 1973).

Large topographic barriers, such as the Himalaya and western Ghats (Fig. 1), have major influences on rainfall-generating processes (Roe 2005). In the case of the Himalaya and adjacent Tibetan Plateau, the topography also alters the pathways of water-vapour transport. It has been previously documented that the steep topography of the Himalaya significantly influences spatiotemporal rainfall distribution (Bookhagen and Burbank 2006, 2010; Bookhagen 2010). The large elevated Tibetan plateau region is thought to heavily influence the energy budget of this region, also leading to teleconnections with different synoptic-scale weather phenomena (Yanai and Wu 2006).



**Fig. 1** Topographic map (based on ETOPO1 data provided by NOAA) of the Indian peninsular and the Himalaya. The Himalaya form a high topographic barrier in the north resulting in orographic rainfall. Along the west coast of India are the western Ghats forming an additional orographic barrier. The blue lines show the two important river system draining the Himalaya, the Indus in the west and the Tsangpo–Brahmaputra–Ganges system in the central and eastern Himalaya

### 3 Data

In this study, we used daily, gridded rainfall data from 1951 to 2007 developed as part of the project—Asian Rainfall Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources (APHRODITE) (Yatagai et al. 2009). It is available from the website—<http://www.chikyu.ac.jp/precip>. We have extracted the data for the South Asian region (Fig. 2) with a horizontal resolution of 0.5 degree (~55km) (APHRO–V1003R1). We refer to this data set as APHRO–V01003R1. We have also employed the zonal ( $u$ ) and meridional ( $v$ ) wind components and surface rainfall data from the NCEP/NCAR reanalysis data set from 1951 to 2004 with 2.5 degree resolution, provided by the NOAA and available at <http://www.esrl.noaa.gov/psd/>.

### 4 Methodology

In this section we will first introduce the nonlinear correlation measure of event synchronization; second we will provide a description of network construction methodology and third, we list the details of network measures and terminology used in this work.

#### 4.1 Event synchronization (ES)

We have employed event synchronization (ES) as a nonlinear correlation to measure the strength of synchronization of rain events between two different grid points and

their delay behaviour. ES serves also as the basis for constructing the complex networks. ES has been previously introduced in Quiroga et al. (2002) and modified by Malik et al. (2010). Only those rain events and their time indices are considered, which are above a certain  $\alpha$  percentile of all the wet days during the 4 summer monsoon months of JJAS (June, July, August and September). An  $\alpha$  percentile threshold for each grid point gives a unique net amount of rainfall per day as the threshold for each grid point. For  $\alpha = 94\%$  and  $\alpha = 90\%$  the thresholds on net rainfall amount per day is shown in Fig. 2a, b respectively. Also, the annual standard deviation of the number of events per year is plotted in the Fig. 2c, d for both thresholds. Thresholds used in this study are considered to be extreme rainfall events (Groisman et al. 1999; Kripalani and Kulkarni 1999; Goswami et al. 2006). We have chosen these thresholds, because rainfall at or above this threshold only occurs during the active phase of the ISM. Hence these thresholds are extremely useful in studying the spatial structures of underlying atmospheric processes which are responsible for the active phase of monsoon. Events with the above mentioned thresholds are less effected by the sampling uncertainty, as compared to very high thresholds and hence provide more robustness to calculation of ES. This is one of the approaches for characterizing extreme rainfall events, an alternative approach is to describe extreme daily rainfall events by suitable statistical models of extreme value theory (Coles 2001). A detailed analysis between two approaches in context of the ISM could be found in for e.g., May (2004a, b).

Let us say an event above  $\alpha$  occurs at time  $t_l^i$  at grid point  $i$  and  $t_m^j$  at grid point  $j$   $l = 1, 2, \dots, s_i, m = 1, 2, \dots, s_j$  and within a time lag  $\pm \tau_{lm}^{ij}$ , which is defined as following

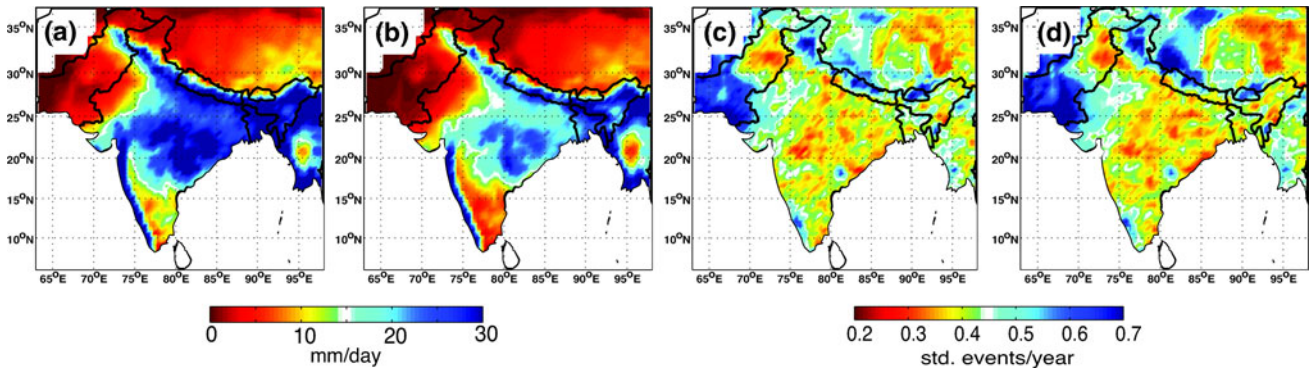
$$\tau_{lm}^{ij} = \min\{t_{l+1}^i - t_l^i, t_l^i - t_{l-1}^i, t_{m+1}^j - t_m^j, t_m^j - t_{m-1}^j\} / 2 \quad (1)$$

where  $s_i$  and  $s_j$  are the total number of such events that occurred at the grid point  $i$  and  $j$  respectively. The above definition of time lag  $\pm \tau_{lm}^{ij}$  helps to separate of independent events. Which in turn allows to take into account the fact that different atmospheric processes responsible for generation of rain events evolve at different time scales. We need to count the number of times an event occurs at  $i$  after it appears at  $j$  and vice versa and, this is achieved by defining quantities  $c(ilj)$  and  $c(jli)$ . Where,

$$c(i|j) = \sum_{l=1}^{s_i} \sum_{m=1}^{s_j} J_{ij} \quad (2)$$

and

$$J_{ij} = \begin{cases} 1 & \text{if } 0 < t_l^i - t_m^j < \tau_{lm}^{ij} \\ 1/2 & \text{if } t_l^i = t_m^j \\ 0 & \text{else,} \end{cases} \quad (3)$$



**Fig. 2** Rainfall thresholds and their annual variability calculated from 57 years of data (1951–2007). We show the annual daily rainfall amount for thresholds of  $\alpha = 94\%$  (a) and  $\alpha = 90\%$  (b). Note the generally high daily rainfall amounts in the Ganges plain and at the orographic barriers of the eastern Swats and the southern Himalaya.

Similarly, we can define  $c(j|i)$  and from these quantities we can obtain

$$Q_{ij} = \frac{c(i|j) + c(j|i)}{\sqrt{s_i s_j}} \tag{4}$$

$$q_{ij} = \frac{c(i|j) - c(j|i)}{\sqrt{s_i s_j}} \tag{5}$$

$Q_{ij}$  is the measure of the *strength of event synchronization* between grid points  $i$  and  $j$ . It is normalized to  $0 \leq Q \leq 1$ . This implies  $Q = 1$  for complete synchronization. The  $q_{ij}$  measures the *delay behaviour* and  $-1 \leq q \leq 1$ . And  $q_{ij} = 1$  implies that an event at  $i$  always precedes an event at  $j$ . The matrix  $q_{ij}$  open up several unique possibilities to analyse delay directions.  $Q_{ij}$  is a square symmetric matrix and  $q_{ij}$  is square anti-symmetric matrix. ES has been very specifically designed to calculate nonlinear correlations among time series with events. And better suited than other correlation measure such as cross-correlation (linear) and mutual information (nonlinear) for measuring correlations among bivariate time series with events defined on them. There exist several other measures for estimating dependencies among events too. A distinctively different but similarly analytical rigorous method to measure dependencies between extremes based on extreme value theory is provided in Coles et al. (1999). In this approach, it is not required to pre define events like in ES but it has no option of analysing delays and their directions.

#### 4.2 Constructing adjacency matrices

To construct the network out of the ES matrices described above, we treat a grid point over land as the vertex of the network and any edge between them will be referred to as a link. In other words, vertex and grid point have the same

We show the standard deviation from the mean number of rainfall events per year for  $\alpha = 94\%$  (c) and  $\alpha = 90\%$  (d). Areas in *blue* indicate high inter-annual variability and are generally spatially disconnected from the mean annual rainfall pattern

meaning in the following discussion. The links between different grid points exist if the strength of the ES is above a certain predetermined threshold. Directionality to the links is introduced using the delay direction from the sign of  $q$ . Let us say a grid point is connected to  $k$  other different grid points. Then we call  $k$  the *degree* of the grid point where  $k$  is an integer between 0 and  $N - 1$ , where  $N$  is the total number of grid points. For the purpose of constructing the network we have used a fixed global link density  $K$ . In the case of undirected networks it is related to the probability  $P(k)$ —the number of grid points having  $k$  connections as  $\frac{1}{N-1} \sum_{k=1}^{k=k_{max}} P(k)k = K$ , where  $k_{max}$  is the maximum of degree. We will only analyse the minimalist correlation structure of the ISM rainfall. In order to obtain it, we have taken the value of  $K = 0.05$ , i.e., we assume that only 5% of the total grid points are connected. These links represent the 5% strongest correlations. The underlying assumption is that the extracted minimalist correlation structure contains the statistically most significant and essential features of correlations in the ISM rainfall field. Thereby, we analyse only the statistically most significant correlations and, in turn, we will be able to remove much of the redundant information. From a meteorological perspective, we thus analyse the most persistent atmospheric features during an ISM season responsible for generation of extreme rainfall events. We calculate a threshold  $\theta_{ij}^Q$  on  $Q_{ij}$  by setting  $K = 0.05$ . Next we, then convert  $Q_{ij}$  into a binary matrix called the *adjacency matrix*  $A$  where,

$$A_{ij} = \begin{cases} 1 & \text{if } Q_{ij} > \theta_{ij}^Q \\ 0 & \text{else,} \end{cases} \tag{6}$$

and  $A$  is a symmetric matrix such as  $Q$ . Similarly,  $q_{ij}$  can also be constructed into an adjacency matrix  $A^q$  with a difference that now

$$A_{ij}^q = \begin{cases} 1 & \text{if } q_{ij} > |\theta_{ij}^q| \\ 0 & \text{else,} \end{cases} \quad (7)$$

We now have sense of direction in  $A^q$ , i.e.,  $A_{ij}^q = 1$  means that the link is from  $i$  to  $j$  and not vice-versa. Therefore  $A^q$  is not a symmetric matrix. A further adjacency matrix  $A^{Qq}$  can be constructed as,

$$A_{ij}^{Qq} = A_{ij}^q A_{ij} \quad (8)$$

and it has the characteristic of having both, the information of direction and strength of the link. Again  $A^{Qq}$  will not be a symmetric matrix. In Fig. 3 we provide a schematic representation of steps involved in the construction of a network, starting from rainfall time series at each grid point.

### 4.3 Complex network measures

We use several basic measures from complex network theory to characterize the rainfall network constructed from the ES matrices (Boccaletti et al. 2006) (c.f. Fig. 3) The simplest measure is the *degree centrality* of grid point  $j$ ,  $C_{Dj}$ , which is given as

$$C_{Dj} = \frac{\sum_{i=1}^N A_{ij}}{N - 1} \quad (9)$$

where  $N$  as above is the total number of grid points.  $C_{Dj}$  measures the number of grid points linked to a particular grid point  $j$ . A grid point having higher degree centrality is expected to have higher influence on the functioning of the network. In the presented study grid points where critical atmospheric processes responsible for the development of ISM rainfall take place should show higher degree

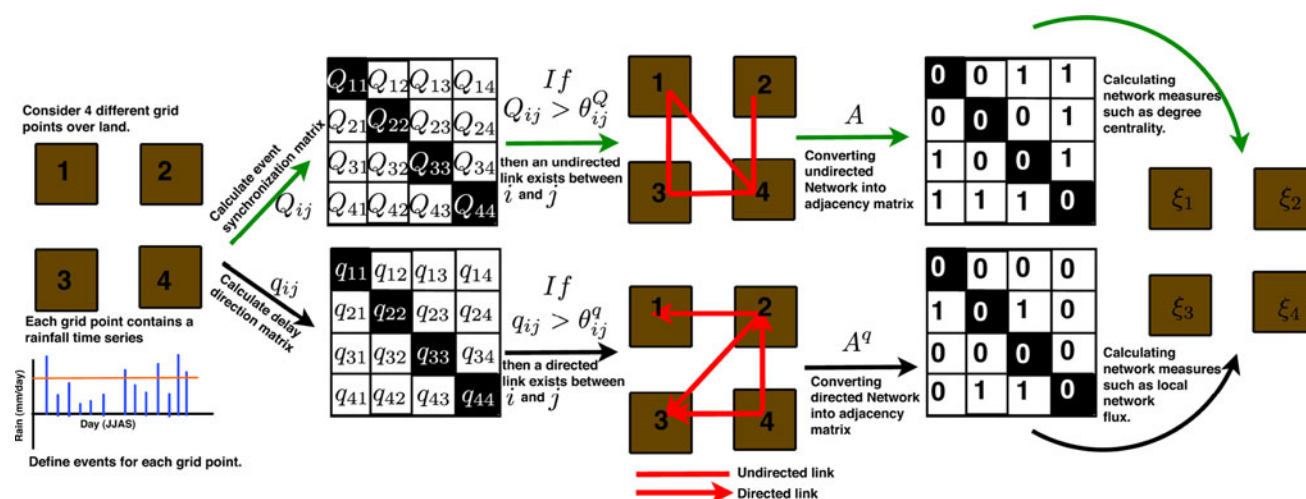
centrality. The *local clustering coefficient*  $\mathcal{C}_j$  for the grid  $j$  tells us the probability if two different connected grid points that are also connected to the same third grid point. In the language of graph theory this relates to how close the neighbours of a vertex are to a complete graph (also called clique). Let us say that grid point  $j$  has  $k_j$  links, i.e., its is connected it with  $k_j$  other nodes. If each of these grid points are also connected with each other then we need  $\frac{k_j(k_j-1)}{2}$  links. If the actual number of links existing is  $\mathcal{E}_j$  then

$$\mathcal{C}_j = \frac{2\mathcal{E}_j}{k_j(k_j - 1)} \quad (10)$$

In our approach, we will use  $\mathcal{C}_j$  to estimate the spatial continuity of rainfall fields. An additional sophisticated centrality measure is the *closeness centrality*  $C_{Cj}$ . The mathematical definition used here measures the network vulnerability. Grid points with a high value  $C_{Cj}$  are very critical for the functioning of the network (Dangalchev 2006). Let us say that the mean *geodesic distance*, i.e., the shortest path between a grid point  $j$  to all the other grid points  $i$  connected to it is  $d(j, i)$  then closeness centrality is

$$C_{Cj} = \sum_{i \in \{V \setminus j\}} 2^{-d(j,i)} \quad (11)$$

where  $\{V \setminus j\}$  is the set of all the vertices or grid points excluding  $j$ . One of the physical interpretation of this measure is that it gives the speed of information propagation. For example, any perturbation in the system travels fastest to the vertices with highest vales of  $C_{Cj}$ . We have normalized  $C_{Cj}$  between 0 and 1 by dividing it with the maximum of  $C_{Cj}$ .



**Fig. 3** Schematic flow diagram showing the steps involved in the construction of the network (read from left to right) and in calculating some network measure  $\xi$ . First, we define events for a rainfall time series at the each grid point and then calculate event synchronization

matrix  $Q_{ij}$  and delay direction matrix  $q_{ij}$ . Depending on matrix characteristics, we establish undirected or directed links and convert the network to an adjacency matrix. And from the adjacency matrix we estimate the network measure  $\xi$

Betweenness centrality  $C_{Bj}$  is the sum of the ratio of the number of shortest paths between two vertices passing through a particular grid point to the total number of shortest paths between those two vertices. Mathematically it is given by

$$C_{Bj} = \sum_{j \neq s \neq t \in V} \frac{\sigma_j(l, m)}{\sigma(l, m)} \tag{12}$$

where  $\sigma_j(l, m)$  is the number of shortest paths between  $l$  and  $m$  passing through  $j$ . Physically  $C_{Bj}$  indicate the information pathways if we assume that the information travels using shortest path. In our case, we hypothesise that rain events (or in the wider sense water vapour) are the quantity traveling through the network. This hypothesis is subject to further research in order to provide a robust and solid physical interpretations.

We can also measure the length of the links in physical units by defining the *geographical distance*. If two grid points  $i$  and  $j$  are connected then the length of this link  $L_{ij}$  can be calculated using the formula for spherical earth projected on to plane, i.e.,

$$L_{ij} = R \sqrt{(\delta\phi_{ij})^2 + (\cos(\phi_m)\delta\lambda_{ij})^2} \tag{13}$$

where  $\delta\phi_{ij}$  and  $\delta\lambda_{ij}$  are differences in latitude and longitude in radians between grid point  $i$  and  $j$ ,  $\phi_m$  is the mean of the latitudes of  $i$  and  $j$  and  $R$  is the radius of the earth.

#### 4.4 Directed networks: local network flux

Directed networks are the networks where every link has a sense of direction, i.e., either it is outgoing or incoming. The indegree  $k^{in}$  is the number of links incident on a grid point; the outdegree  $k^{out}$  is the number of links leaving a grid point. We define the local network flux as the difference of the two, i.e.,  $\Delta k = k^{in} - k^{out}$ . A strong positive value of  $\Delta k$  will indicate accumulation of moisture at the grid point, i.e., it will be a moisture sink. To calculate the local network flux, we make use of the matrix  $A^q$  evaluated as defined in Sect. 4.2.

##### 4.4.1 Identifying anomalous monsoon years

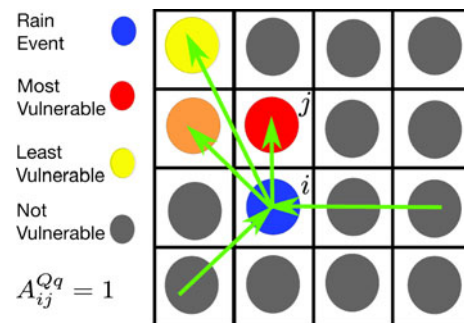
The above described measures can help us in studying the spatial structures of rainfall fields and their properties but not its temporal evolution. In the following section, we develop and present a new scheme, which uncovers some details of the temporal evolution of the ISM rainfall patterns. We apply this to the same data set over last 6 decades and provide new insights into the spatiotemporal complexity of monsoonal rainfall. Our underlying assumption is that the above constructed network has the minimum essential correlation structure of the rainfall field. We thus

suggest that extreme events occur within this structure and will deviate from this synoptic structure only if the ISM is abnormal. We can use this observation to find anomalous behaviours of the ISM and to identify regions where the ISM rainfall has the most intricate spatial structure. Usually, the method employed for discovering anomalous monsoon behaviour is based on a standard deviation of a rainfall index, which may be biased by inhomogeneous spatial rainfall distributions. The method described in the following paragraphs is not impacted by the large spatio-temporal discrepancies in rainfall data.

Let us assume that a rain event occurs at some grid point  $i$ . This will make any other grid point  $j$  vulnerable to such an event too, if there exists a link from  $i$  to  $j$ . This type of vulnerability must be inversely depended on the distance from  $i$  to  $j$ . Apart from the link strength its directionality is also important, as only those grid points will be vulnerable which have incoming links from  $i$ . All this information can be obtained from the adjacency matrix  $A^{Qq}$ . A schematic explanation of the above is presented in the diagram in Fig. 4. We can now write the *vulnerability* of grid point  $j$  receiving rainfall to be

$$\rho_j = \frac{\sum_{i \in V(t)} \frac{1}{L_{ij}} A_{ij}^{Qq}}{\sum_{i=1}^N \frac{1}{L_{ij}} A_{ij}^{Qq}} \tag{14}$$

where  $V(t)$  the set of grid points where rain event of the type  $\alpha$  happened at time  $t$  and  $N$  is the total number of grid points.  $L_{ij}$  is the geographical distance between grid points  $i$  and  $j$ .  $\rho_j$  is calculated from the first half of the data set and predicted for the second half. We set a pre-assigned value for  $\rho_j$  to find the prediction accuracy and assess the evolution of monsoonal rainfall patterns and also their spatial complexity.



**Fig. 4** A simple schematic representation of the calculation of  $\rho_j$  (Eq. 14). The matrix depicts the square grid points over land.  $j$  is most vulnerable as it is geographically closest to  $i$ . Only incident links contribute to the vulnerability of a grid point and may have an extreme rain event

## 5 Results and discussion

First, we discuss the degree centrality and its distribution obtained from the adjacency matrix  $A$ . Then we present the analysis of spatial scales and clustering coefficients. Further, we introduce our results on centrality measures, followed by an attempt to visualise the links within the network. At the end of the discussion, we analyse the spatiotemporal evolution of ISM rainfall patterns using the new scheme described in Sect. 4.5.

### 5.1 Degree centrality and degree distribution

The spatial patterns of degree centrality  $C_{Dj}$  are very similar for both thresholds of  $\alpha = 94\%$  and  $\alpha = 90\%$ . Higher degrees are observed in northwest Pakistan and lowest values occur in southeast India (Fig. 5). We suggest that higher degree emerges mainly due to longer spatial connections in these regions. This potentially can be related to the large spatial scale of ISM rainfall over these regions. To distinguish between regions where monsoonal rainfall is also due to localised or synoptic (large scale) activity of ISM, we show the distribution of the degree in Fig. 6a for the case of  $\alpha = 90\%$ . We obtain a bimodal distribution and a model fit of the type

$$P(k) = \frac{n_r^{n_k} \exp(-n_r)}{n_k!} + \frac{n_d^{n_k} \exp(-n_d)}{n_k!} \quad (15)$$

where we found  $n_r = 17.4$ ,  $n_d = 8.0$  and  $n_k = k/40$  (red curve in Fig. 6a). Hence,  $P(k)$  is sum of two poissonian distributions with different means. This implies that there must exist two different kinds of regions with their own characteristic number of links or distinct spatial scales of rainfall. Previous work has documented that the ISM has two modes, the active phase and the break phase. Therefore, it is possible that certain regions continue to receive rainfall during the break phases but certain other regions receive rainfall only during the active phases of ISM especially areas bounded by the monsoonal trough. To construct such a division of regions, we divide grid points into different zones (Fig. 6b). This figure depicts the spatial regions associated with the degree distribution. Rainfall in central Pakistan northwestern India, and partly in the western Tibetan Plateau (Zada Basin) is expected to be the result of large spatial monsoonal activity. Heavy rainfall processes in the northwestern Indian subcontinent are associated with atmospheric interactions between western disturbances and the Indian monsoon system. Where the trough existing in mid latitude westerlies penetrates southward and interacts with monsoonal trough over the Indo-Pakistan region causing recurvature of depressions and lows in  $75^\circ\text{E}$ – $78^\circ\text{E}$  belt (Ding and Sikka 2006). This complex interaction is fundamentally composed of cold

westerly winds interacting with warm, moisture-laden monsoonal winds over a very dry and hot region during the peak of summer with low pressure fields. Hence such interaction should lead to volatile convective instabilities in the atmosphere and which should not only produce extreme convective rainfall but also has far reaching effects on the internal dynamics of ISM. It is understood that this interaction first enhances monsoonal rainfall over the northwest and may ultimately lead to withdrawal of monsoonal trough to foothills of Himalayas i.e a break phase of ISM (Ding and Sikka 2006 and references therein). A recent study detailing the causes of floods in North West Pakistan during late July–August 2010 also hints that a similar mechanism was responsible for these floods (Hong et al. 2011). We will go into further details of this interaction and its possible influences in the next sections.

### 5.2 Median length of links

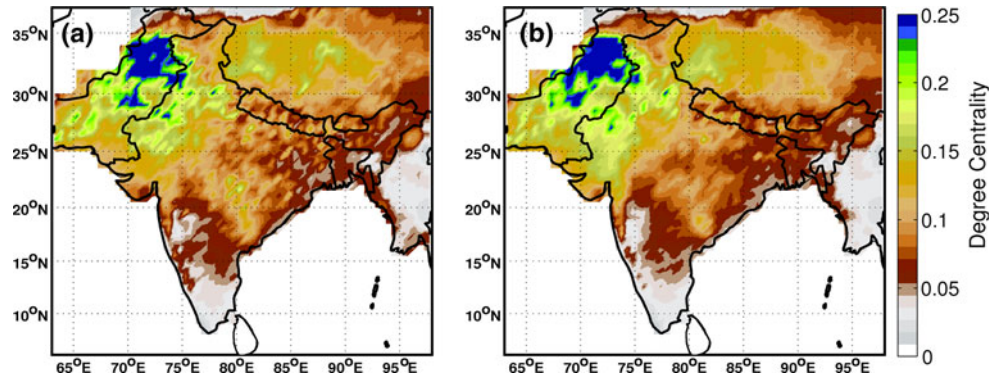
We provide further insight into the spatial scales involved in these regions by analysing the geographical length of these links. We will use the formula introduced in Eq. 13. This distance metric gives us the advantage of expressing spatial scales in length units. The median of  $L_{ij}$  for each grid point is shown in Fig. 7a. Clearly the characteristic scale of monsoonal rainfall for the 90th percentile seems to be below 250 km for most of the region. Also, some larger spatial scales above 500 km exist, such as in northwest Pakistan and the southwestern coast of peninsular India. We observe that most of the larger spatial scales exist in the region with medium to high degree distributions (green and blue colors in Fig. 6b). This supports our statement that these regions receive rainfall from large spatial monsoon activity stretching over distances of 500 km, a characteristic for the active monsoon phase. Figure 7a also provides some supplementary information to Figs. 5 and 6, about the south west coast of peninsular India along. Due to the significant topographic barrier of the western Ghats (c.f. Fig. 1) this region has a generally smaller degree centrality (see Fig. 5). However, we observe high spatial scales for this region (Fig. 7a) and associated these with rainfall coeval with rainfall south of the Himalaya during the active monsoon phases.

Next, we analyze if the spatial scales follow an analytical form. For this purpose, we plot the distribution of distance versus the number of links (Fig. 7). The fitted bold lines are gamma distribution of the form

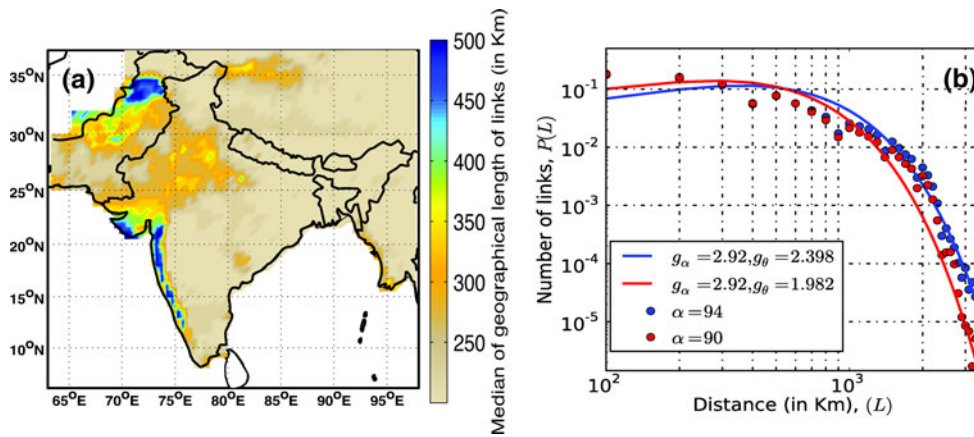
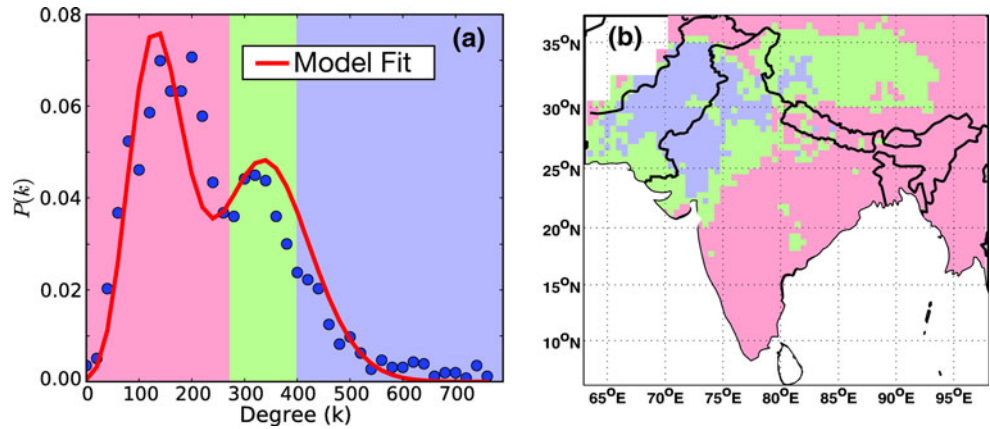
$$P(L) = n_L^{(g_\alpha-1)} \frac{\exp(-n_L/\theta)}{\Gamma(g_\alpha)\theta^{g_\alpha}} \quad (16)$$

where  $n_L = L/100.0$  and the values of other parameter are given in the legend of the Fig. 7b for the fitted curves.

**Fig. 5** Degree centrality  $C_{Dj}$  **a**  $\alpha = 94\%$ , **b**  $\alpha = 90\%$ . Obtained from the matrix  $A$  it gives the number of links to a grid point and it is normalized between 0 and 1 by dividing it by  $N - 1$  i.e., the total possible number of links to grid point. Note higher  $C_{Dj}$  in northwest Pakistan and lower values on the southern Indian peninsular



**Fig. 6** Spatial degree distribution. **a** Measured degree distributions are shown in blue dots and red line indicates the model fit. **b** Depicts the spatial distribution of colored regions shown in **a**. The region in blue and green receive rainfall only during the large-scale spatial activity of monsoon, i.e., during its active phases, whereas the regions in pink also receive rainfall during break phases



**Fig. 7** **a** Median of geographical length (km) of links calculated using formula for spherical Earth projected onto a plane. **b** The distribution of length scales for  $\alpha = 94\%$  (blue) and  $\alpha = 90\%$  (red) and their corresponding fits. Note that more extreme rainfall events

( $\alpha = 94\%$ ) have longer characteristic spatial scales as the blue curve results in higher numbers of links at longer distances. However, the longest recorded scales are associated with  $\alpha = 90\%$ , because the red curve has a longer tail

Figure 7b indicates that we observe longer spatial scales in the smaller events. Although characteristic scales must be larger for stronger rainfall events at  $\alpha = 94\%$ , because these values are higher for most part of the distribution (blue points in Fig. 7b). Our analytical form implies that more extreme rain events are more spatially localised.

### 5.3 Clustering coefficient

The local clustering coefficient  $\mathcal{C}_j$  shows the spatial organization of rainfall with respect to a reference grid point. The field may have a large spatial extent but whether it is highly fragmented or spatially continuous cannot be



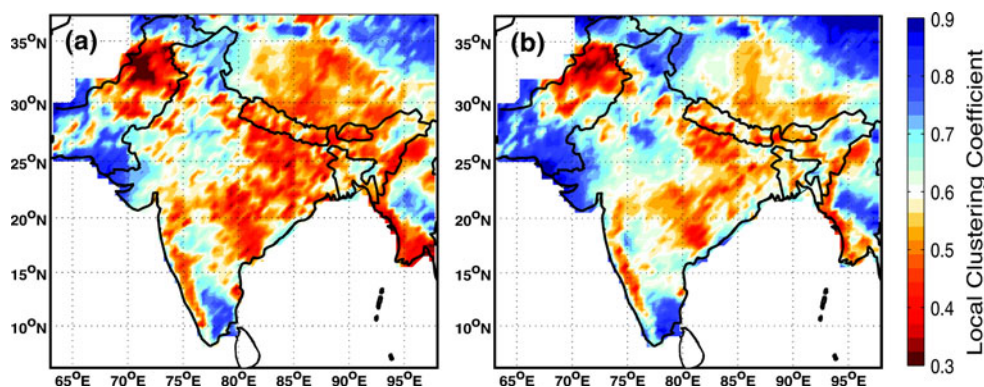
inferred from the above measures and analysis. However, we can derive this additional information from the values of  $\mathcal{C}_j$ . We associate lower values of the clustering coefficient with more fragmented or spatially discontinuous rainfall fields, whereas larger values represent clustered activity. Importantly, this measure is independent of the involved spatial scales. The spatial pattern of  $\mathcal{C}_j$  for the two rainfall threshold we are using is very similar to each other (Fig. 8a, b). Large areas of low clustering values  $\mathcal{C}_j$  related to fragmented rainfall are located in northwest Pakistan and in central and eastern India. We observe higher clustering coefficients in south Pakistan, parts of the Tibetan plateau, and in the northwestern and southeastern parts of India (c.f. Figs. 1, 8) This suggests that stronger rainfall events are more spatially clustered in these regions. Northwest Pakistan has lower values of clustering coefficients, which indicates that this region receives rainfall due to the large spatial activity of the ISM, but with a spatially fragmented rainfall field. Also, comparing the two panels in Fig. 8a, b indicates that stronger rainfall events are more spatially fragmented than smaller ones.

#### 5.4 Centrality measures

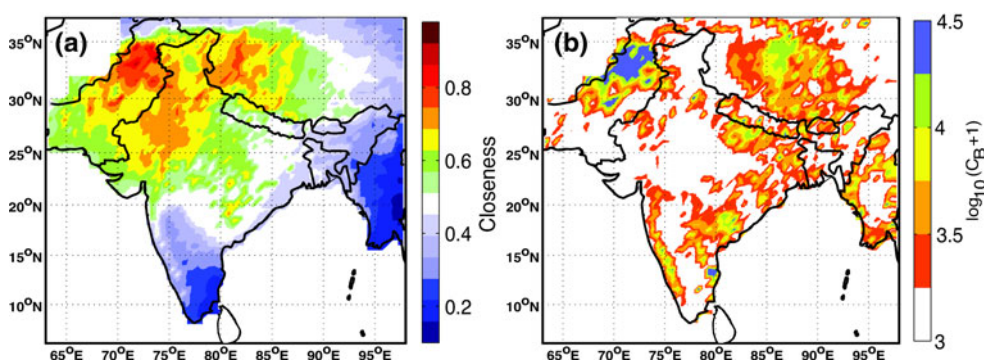
Closeness centrality  $C_{Cj}$  has been introduced in Sect. 4.3 and it can be employed in the task of identifying the grid points which perform a critical role in the functioning of

this network structure.  $C_{Cj}$  is plotted in Fig. 9a and shows that regions of highest closeness centrality lies in the northwestern subcontinent with a focus on northwest Pakistan. As described in Sect. 4.3 this indicates that the information travels fastest to and from these points. Any perturbation occurring in this region will effect the monsoonal rainfall patterns at rapid temporal scales. The existence of any atmospheric instability over this region will have an immediate and widespread effect over ISM rainfall. It has also been observed that this region is near the boundary between two synoptic systems: the westerly and the monsoonal trough. Spatial fluctuations in the westerly trough and its southward penetration could lead to an interaction resulting in a complex modulation of the ISM activity over rest of the land mass (Ding and Sikka 2006). To provide further evidence that this may be an important mechanism during the most active phase of the ISM, we make the following calculation. First, we identify a set of 50 grid points with 50 top most values of  $C_{Cj}$ . We found that all the grid points in this set have  $C_{Cj} > 0.7585$ , (see Fig. 9a). Next, we calculate the linear cross correlation  $r$  between the total number of events above the threshold  $\alpha = 90\%$  that occurred within a set of 50 grid points and the remainder of the land mass. The value  $r$  was found to be 0.535 (refer Table 1) and  $C_{Cj} > 0.7585$ . Most of these 50 grid points are located in northwest Pakistan (see Fig. 9a). This indicates that interaction of western

**Fig. 8** Local clustering coefficient  $\mathcal{C}_j$  **a**  $\alpha = 94\%$ , **b**  $\alpha = 90\%$ . Red colours indicate that the rainfall field is less spatially continuous, i.e., it is fragmented. In contrast, blue colours outline more spatially continuous rainfall fields. We observe that  $\mathcal{C}_j$  is independent of the spatial scales involved in the rainfall (compare with Fig. 7a)



**Fig. 9 a** Closeness centrality  $C_{Cj}$ . We observe high  $C_{Cj}$  in the northwestern parts of the subcontinent suggesting the importance of atmospheric processes in modulating the ISM activity. **b** Betweenness centrality  $C_{Bj}$ . Higher values of  $C_{Bj}$  represent the moisture transport pathways over the land during the active phase of the ISM. For both **a** and **b** we chose  $\alpha = 90\%$



**Table 1** Linear correlation  $r$  between extreme rain events in a region and to the remainder of the subcontinent for  $\alpha = 90\%$

Region	No. of grid points	Ref. Fig.	$r$
Tibet	50	10b	0.4495
NW Pakistan	50	10a	0.4675
Central India	50	10d	0.3182
West coast of India	50	10c	0.2992
Closeness $> 0.7585$	50	9a	0.5351
$\log(\text{Betweenness} + 1) > 4.358$	50	9b	0.5325
Degree centrality $> 0.26$	50	9c	0.4893

disturbances with the ISM plays an important role in generation of the extreme rainfall events over large parts of the Indian subcontinent. Our results indicate that an influence of such interactions must be far more spatially extensive than limited to the northwestern parts of the subcontinent. We will further analyse the underlying atmospheric mechanism in the next section.

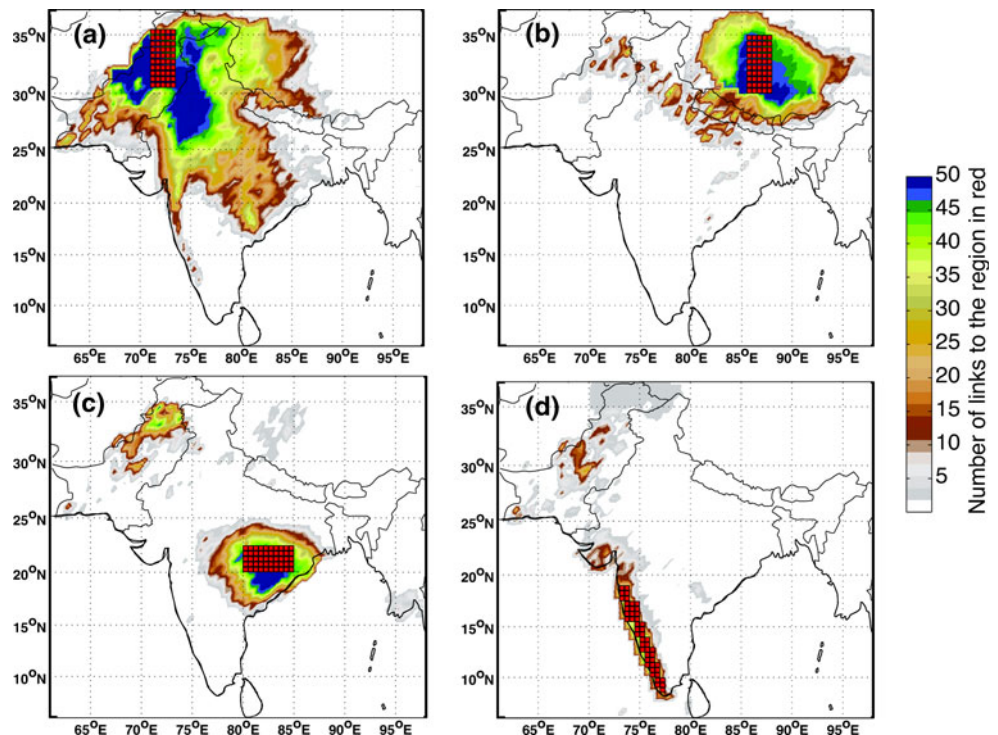
We have plotted the betweenness centrality  $C_{Bj}$  to analyze and visualize its spatial structure (Fig. 9b). Higher values of  $C_{Bj}$  are observed over large parts of Tibet, the east coast of peninsular India, parts of central India, the central Gangetic plains, northwest Pakistan and along the western Ghats. From a mathematical point of view, the higher values of  $C_{Bj}$  highlight the main pathways of information travel in a network. In the above case, moisture is the quantity assumed to be traveling through the network

and therefore our analysis highlights the main pathways of moisture transport during the ISM. Pathways of moisture transport are modulated and facilitated by the existence of deep convection and the underlying topography (Roe 2005; Webster et al. 1998; Bookhagen and Burbank 2006, 2010; Bookhagen 2010). Hence, higher values of  $C_{Bj}$  in Fig. 9b represents the region where deep convection ceases to exist during the active phase of monsoon.

### 5.5 Visualising links of complex networks for the study region

These links that are obtained from the matrix  $A$  and are non directional. In Fig. 10 we have plotted the number of links in relation to 50 grid points selected for a particular region (highlighted by a gridded, red colour matrix in Fig. 10). The number of links to other regions are shown by the colour scale ranging from 0 to 50. In Fig. 10a we observe that links over northwest Pakistan extend deep into central India and also connect to parts of Tibet and to almost the entire western and central Himalayan region. As stated previously, this indicates that stronger rainfall events over northwest Pakistan are likely the result of large spatial scale monsoonal activity. In Fig. 10a we are able to visualise this spatial extent and spatial structure of monsoonal activity. Clearly the spatial structure in Fig. 10a reemphasises the significance of the mechanism mentioned in Sect. 5.3. It indicates the extensive influence on the extreme rain events that occur in other parts of the

**Fig. 10** Links between a set of 50 reference grid points (gridded red matrix) to other grid points (colour bar) at  $\alpha = 90\%$ . Note the spatially extensive links for a reference area in northwestern Pakistan (a). In contrast, extreme rainfall linkages in the western Ghats (d) have a limited spatial extent



subcontinent. The number of links on the Tibetan plateau are high and they appear to be much more localised (see Fig. 10b), as high topographic barriers exist to the west and south. In Fig. 10c we observe an interesting feature for the central Indian region: there are two major geographically disconnected regions with high number of links. This may happen due to formation of deep convective cells over the region of NW Pakistan and adjoining regions and is dominating during the active phase of the ISM. We also observe some localisation in the southwest India Fig. 10d. This localization is caused by the topographic barriers formed by the western Ghats. We also observe a few long-range connections from this region to almost the entire west coast of India and western parts of the subcontinent. These links may emerge due to the existence of an offshore trough along the west coast of India and embedded mesoscale vortices during active phase of monsoon (Ding and Sikka 2006). This synoptic configuration facilitates convection along the west coast and the orographic forcing of western Ghats acts as boundary to this convection.

In Table 1 we provide the linear cross-correlation  $r$  between rain events at  $\alpha = 90\%$  occurring in the marked regions during the ISM season to similar events that occur in the rest of the region (Fig. 10). We observe that the highest linear cross-correlation are the ones in northwest Pakistan. This strengthens the argument that extreme rainfall events in this region are due to the large spatial activity of the ISM. The interaction of western disturbances and the ISM could be an intensifying monsoonal activity over the Indian subcontinent (Dimri 2004). However, no detailed study, at least to our knowledge, exists about this interaction and the penetration depth across the Himalaya and Tibet.

An immediate practical application of our methodology can be derived from Fig. 10 and Table 1: These regions outline areas that are best suited for paleoclimatic proxies that reconstruct the Holocene and historical monsoonal activity. Our results suggest that the best region lies in the northwestern Indian subcontinent centered in northwest Pakistan and within the northwestern Himalaya because only very strong and significant active phases of the ISM transport rainfall to this region.

### 5.6 Directed networks

To obtain the sinks of moisture over land, we plot the local network flux  $\Delta k$  as defined in Sect. 4.4 in Fig. 11. Higher positive values exhibit moisture sinks over land and are highlighted in red colours in Fig. 11. One of the pronounced moisture sinks spreads from the central to lower Gangetic plains along the foothills of the central Himalaya. This region is known for the formation of a high number of low-pressure systems (monsoonal lows and depressions)

(Mooley and Shukla 1989). The large spread of sinks is related to the movement of low-pressure systems over land, as they are not constant spatial feature (Sikka 1977). The extent of the red areas in the map indicates the region in which most of these systems form. A second major moisture sink shown in Fig. 11 is located in Pakistan near the western boundary of the average monsoonal-trough location (Ding and Sikka 2006). Some other minor sinks are observed on the Tibetan plateau and along the western Ghats. These sinks are rainfall accumulation from different directions, i.e., moisture convergence zones and do not necessarily indicate higher rainfall amounts. This type of moisture convergence is likely generated by the underlying heat balance and orographic effects over these regions (Bhide et al. 1997). The regions with high negative values of local network flux are the regions, which are closer to the moisture sources.

In this study we have considered extreme rain events (90 and 94% percentile), which are usually a result of convective rainfall processes (Schumacher and Houze 2003). The mesoscale convective systems (MCS) play an important role for generating these rainfall amounts (Johnson 2006; Houze et al. 2007). Therefore, the identified sinks and sources can be associated to spatial patterns of MCS within the ISM region. Apart from the underlying heat balance over the land, the vorticity and divergence of wind vectors can also provide information about the spatial structures of the MCS and corresponding locations of sources and sinks of moisture (moisture convergence zones). To better understand the origin of moisture sinks over land, we present a comparison of Fig. 11 with wind vorticity and divergence (NCEP/NCAR data) (Fig. 11). For obtaining the time indices of the wind vector, we have used the same thresholding procedure as used for APHRO-V1003R1 and thus we maintain a statistical consistency.

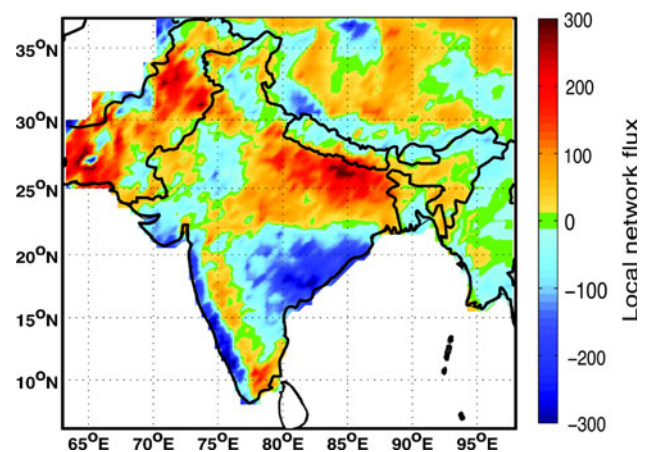


Fig. 11 Local network flux obtained from directed networks i.e., matrix  $A^q$ ,  $\alpha = 90\%$

The relative vorticity ( $\zeta$ ) and divergence ( $\delta$ ) were computed using the formulas

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \quad \delta = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$$

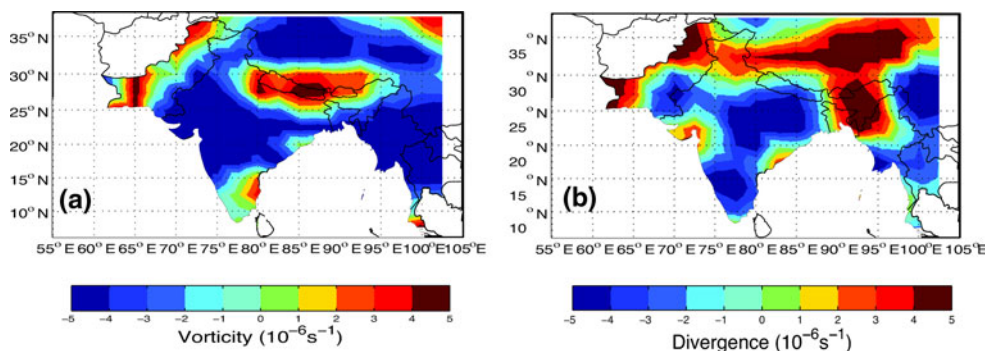
where  $u$  and  $v$  are zonal and meridional wind components at the height of 850 hPa. We observe positive vorticity over the central Gangetic plains extending over the central Himalaya and Tibet (Fig. 12a). Also, positive vorticity is observed over Pakistan and southeast India. These are also the two regions where we observe higher local network flux. Positive vorticity is associated with cyclonic rotation and can often be related to low pressure areas, which drive the moisture inward into the land. Divergence shows convergent winds over the central Gangetic plains and in central India (Fig. 12b). The shape of the divergence zone in Pakistan appears to show the boundary of monsoonal trough during these events. The similarity in spatiotemporal extent of divergence and vorticity during the extreme rainfall events suggests that these structures are required for causing extreme rainfall events.

### 5.7 Identifying anomalous monsoon years

In this section, we will present an application and description of the mathematical scheme developed in Sect. 4.5 to (i) identify regions in which ISM rainfall has the most intricate and temporally unstable structure, to (ii) identify anomalous monsoon years and their deviation from the normal spatial structure of the ISM rainfall field, and to (iii) decipher the temporal evolution of spatial structure of the ISM rainfall.

The first requirement for using the scheme described in Sect. 4.5 is to determine a threshold on  $\rho_j$  (c.f. Eq. 14). Then we will use this threshold to ascertain the accuracy of our prediction. Our prediction based on  $\rho_j$  will be in binary space, i.e., if  $\rho_j$  is greater than a predetermined threshold,

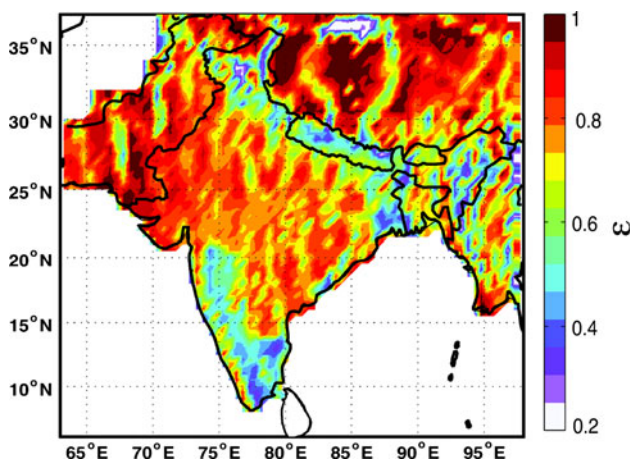
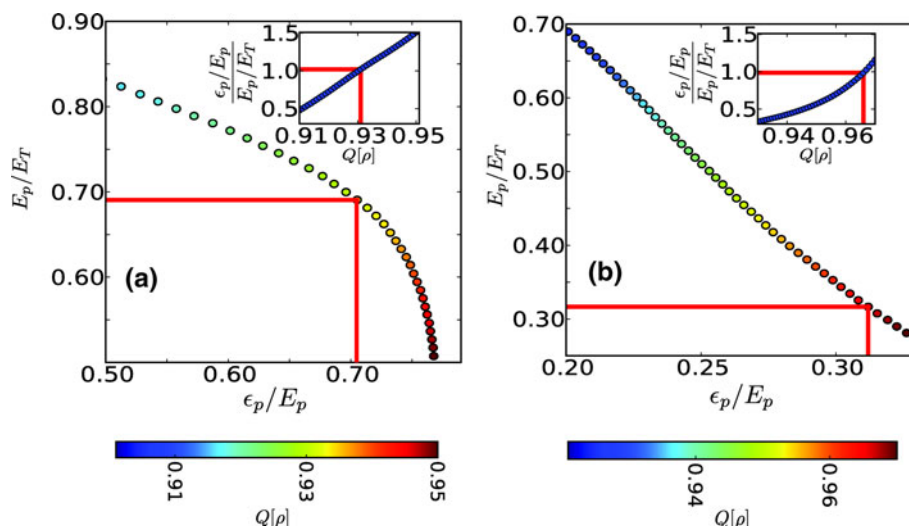
an event (1) occurs at grid point  $j$ , else no event (0) occurs at  $j$  on this day. We thus need two different kinds of ratios to examine the quality of the prediction. First, we have  $\frac{\varepsilon_p}{E_p}$ , which is the ratio of the correctly predicted events ( $\varepsilon_p$ ) to the total number of events predicted ( $E_p$ ). Second, we will use a ratio  $\frac{E_p}{E_T}$ , i.e., total number of predicted events to the total number of events that occurred ( $E_T$ ). We will use a quantile value of  $\rho_j$ , i.e.,  $Q[\rho]$  as a threshold on  $\rho_j$ . The value of  $Q[\rho]$  is chosen such that  $\frac{\varepsilon_p}{E_p} \sim \frac{E_p}{E_T}$  has the best accuracy in the first half of the data. The curves between the above two ratios are shown in Fig. 13a, b without delay and with 1-day delay, respectively. The red lines indicate the points where  $\frac{\varepsilon_p}{E_p} \sim \frac{E_p}{E_T}$  is satisfied. The delay of 1 day was introduced by changing the set  $V(t)$  to  $V(t - 1)$  in Eq. 14. The value of  $Q[\rho]$  has been estimated from the first half of the data set and the second half is predicted. We will use the symbol  $\varepsilon$  for the accuracy of prediction and it is given by  $\varepsilon = \frac{\varepsilon_p}{E_p}$  where  $\frac{\varepsilon_p}{E_p} \sim \frac{E_p}{E_T}$ . In Fig. 13a we observe  $\varepsilon = \frac{\varepsilon_p}{E_p} \sim \frac{E_p}{E_T} \sim 0.7$ , i.e., 70% of the events were correctly predicted when we did not use a time delay. The curve in Fig. 13b suggests that the scheme described above does not have practical usefulness for prediction of extreme rainfall events. Introducing a delay of 1 day, we observe that the accuracy drops to 30%. We will only be using the  $\varepsilon$  (also referred as accuracy in this text) without delay as a measure of spatiotemporal intricacy of the ISM rainfall field (Fig. 14). The accuracy ( $\varepsilon$ ) will be higher if the correlation structure of extreme rainfall events obtained using the above methodology is least intricate and more temporally stable. The accuracy metric  $\varepsilon$  is known as *probability of detection* (POD). Using the same threshold on  $Q[\rho]$  we compare the quality of prediction with other skill scores in Fig. 15. We have used the *threat score* (TH) which is defined as  $TH = \frac{\varepsilon_p}{\varepsilon_p + \varepsilon_f + \varepsilon_m}$ . Where  $\varepsilon_f$  the number of false



**Fig. 12** Relative vorticity (a) and divergence (b) obtained from the NCEP/NCAR reanalysis data set for wind at the height of 850 hPa. We observe high vorticity over the central Gangetic plains and adjacent regions in Nepal, Tibet, and parts of Pakistan. When

comparing a with Fig. 11, most high vorticity regions correspond to moisture sinks. The divergence zone over northern Pakistan shows the western boundary of the monsoonal trough during extreme events considered in the study

**Fig. 13** Estimation of  $\rho_j$  from the first half of the data set. The value of the  $\rho_j$  used is one where the condition  $\frac{\epsilon_p}{E_p} \sim \frac{E_p}{E_T}$  is satisfied. **a** No delay, **b** 1-day delay



**Fig. 14** The accuracy of prediction  $\epsilon$  on a scale from 0 to 1 for the last half of the data set over the map, calculated without delay in the prediction scheme. The lower values of  $\epsilon$  indicate higher intricacies in rainfall patterns. Observe lower values of  $\epsilon$  along the Himalaya and the western Ghats in southern peninsular India

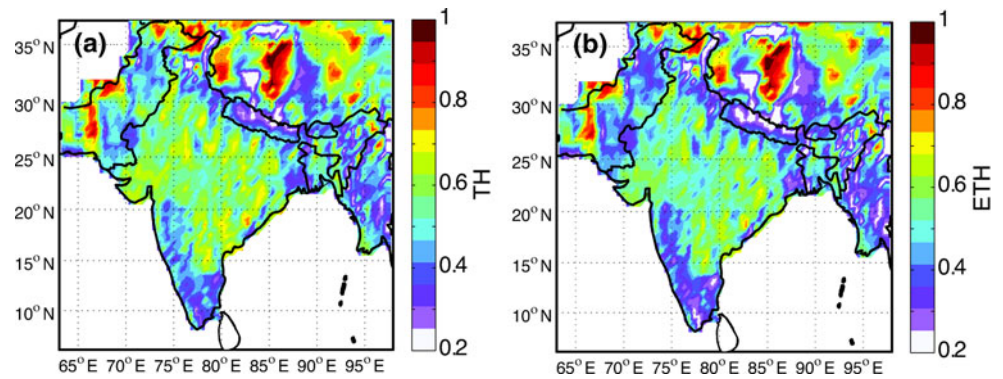
alarms, i.e., predicted events that did not occur and  $\epsilon_m$  is the number of misses, i.e., no predicted event but an actual event occurred. Also,  $E_P = \epsilon_p + \epsilon_f$  and  $E_T = \epsilon_p + \epsilon_m$ . TH is more balanced score and it ranges between 0 and 1 (Fig. 15a). An additional similar skill score is the *equitable threat score* (ETH), which is defined as  $ETH = \frac{\epsilon_p - \epsilon_r}{\epsilon_p + \epsilon_f + \epsilon_m - \epsilon_r}$ . Where  $\epsilon_r = \frac{E_P E_T}{N_d}$ ,  $N_d$  is the number of predicted days.  $\epsilon_r$  gives the number of prediction correct by chance (Fig. 15b). We observe that the basic pattern of prediction quality remains almost the same for all measures. However, values of prediction are slightly lower for most regions based on TH and ETH as compared to POD.

Following the procedure described above, we calculated  $\epsilon$  for each grid point. We have plotted  $\epsilon$  in map view and we observe that for most of the subcontinent the accuracy

( $\epsilon$ ) was above 70% and even reaches up to 100% in places (Fig. 14). We also clearly show that regions with high and complex topography such as the Himalaya and the western Ghats are characterized by the lowest  $\epsilon$  values. From this spatial pattern we infer that these regions are characterized by the most intricate rainfall patterns and high temporal fluctuations.

Next, we employ the above scheme to analyse the anomalous behaviour of the ISM. Our basic assumption for this task is that the complex-network construct inherits the most essential structures and patterns of extreme rainfall events. Hence, any considerable deviation from this essential structure should suggest anomalous behaviour of the ISM—an exceptionally weak or strong, or an abnormal monsoon. We will consider our prediction accuracy  $\epsilon$  per year for the entire region to be the measure of this deviation. In Fig. 16a we plot a scatter plot of our prediction accuracy  $\epsilon$  per year for the entire region versus the z-score of AIMRI (All-India Monsoon Rainfall Index) (Parthasarathy et al. 1995). The significance band of  $\epsilon$  was obtained by bootstrapping the spatial sum of  $\epsilon$  for each day (Davison and Hinkley 2006). That is, we randomly draw 122 days with replacement from the spatial sum over  $\epsilon$  for all the grid points. We observe that the prediction accuracy  $\epsilon$  is lower for weak monsoon years and is high for strong monsoon years (Fig. 16a). All the weak monsoon years are below the upper limit of the significance band and similarly all the strong monsoon years are above the upper limit of the significance band. This indicates that the prediction accuracy  $\epsilon$  can distinguish between normal and abnormal monsoon. We emphasise that  $\epsilon$  tends to distinguish a monsoon year in terms of the spatial organization and structure of the extreme rain events (90th percentile) in a particular year. Years lying within the significance band are characterized by no significant change in spatial organization and structure of the extreme rain events.

**Fig. 15** Comparison of the prediction quality using several skill scores (see text for explanation). **a** The threat score (*TH*). **b** The equitable threat score (*ETH*). The basic spatial prediction pattern remains the same regardless of skill score (c.f. Fig. 13). However, we note the values derived from *TH* and *ETH* are slightly lower than  $\varepsilon$  for almost all the regions



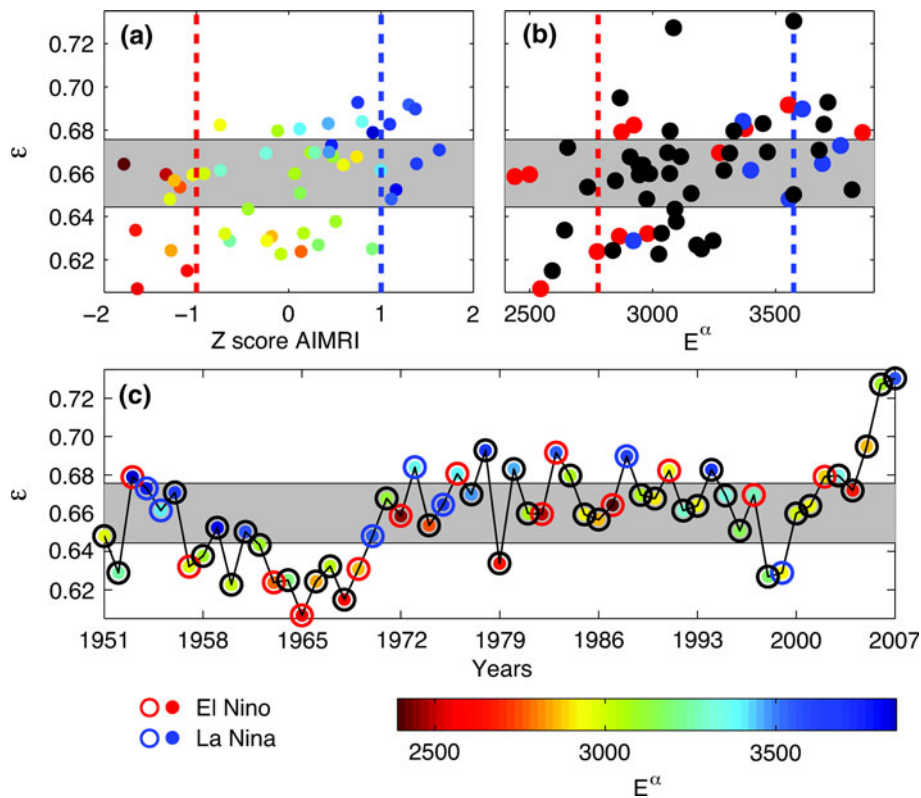
$E^z$  is the total number of events above the threshold  $\alpha = 90\%$  in a year over the whole subcontinent. The cross correlation between  $E^z$  and AIMRI was found to be 0.83. We use  $E^z$  instead of AIMRI for further analysis as AIMRI is limited to the year 2000. In Fig. 16b we plot a scatter plot of  $E^z$  and  $\varepsilon$ . It is thought that there exists a dynamical coupling between El Niño Southern Oscillation (ENSO) and ISM. This coupling has been of wide interest in the scientific literature (e.g., Kumar et al. 1999, 2006, Mokhov et al. 2011; Maraun and Kurths 2005). Thus, we attempt to identify possible influences of ENSO on the spatial structure of the ISM rainfall field. We have indicated the El Niño and La Niña years in Fig. 16b and we observe many events where the ISM has been weak during El Niño years (left to the vertical red lines in Fig. 16b) and the accuracy has been low. When a strong monsoon has followed La Niña (right to the vertical blue line in Fig. 16b) there is a generally higher accuracy. We also observe several exceptions to this rule: some (El Niño years) are not only within the significance band but also above the significance band. In summary, our above results suggest that El Niño and La Niña are not resulting in a complete breakdown of the spatial structure of monsoonal rainfall or in its complete re-organization. ENSO's influence on ISM rainfall pattern appears to be highly complex. These exceptions could be caused by the temporal evolution of the suggested weakening of the coupling between ENSO and ISM (Kumar et al. 1999). The dynamical nature of this coupling could also be another reason, as it has been hypothesised that ENSO-ISM relationship is a bi-directional phenomena rather than ENSO directly influencing ISM only (Mokhov et al. 2011).

We can also take this analysis a step further to understand how the ISM has evolved during the time span of the data set. This will give additional information about the changes in the organization and structure of the ISM extreme rainfall field during times of global warming (Goswami et al. 2006; Ramanathan et al. 2005; Levermann et al. 2009; Zickfeld et al. 2005). For this purpose we plot the temporal evolution of  $\varepsilon$  in Fig. 16c. Between 1951 and

2000, we do not observe any drastic trend in the values of  $\varepsilon$ . It is merely fluctuating between the higher values for strong monsoon and the opposite for weak monsoon years. However, a small change in the mean level of fluctuation is observed around the late 1960s. Interestingly, we observe a characteristically distinct evolution of  $\varepsilon$  since the year 2000 with  $\varepsilon$  not following the previously determined rule with high values for strong monsoon and the opposite for weak monsoon. It is continually and strongly increasing, for example the years 2005, 2006, and 2007 show the highest values of  $\varepsilon$ . This indicates that some basic structural change may have occurred in the rainfall patterns over the Indian subcontinent during this period. Increase in  $\varepsilon$  can only be due to considerable increase in spatial correlations during this period and it is known that increasing spatial correlations are an early warning of approaching tipping points or abrupt dynamical transitions in dynamical systems (Lenton 2011; Lenton et al. 2008). We speculate that we are reading a tipping point in this system, which has some particular precursor in the dynamics (Lenton et al. 2008; Zickfeld et al. 2005; Levermann et al. 2009). A plausible reason for this is a shift in the distribution of magnitude and frequency of extreme rainfall events, caused either by rapid changes in surface heat fluxes due to the increase in aerosols content (Ramanathan et al. 2005) or that some inherent monsoonal dynamics has changed during this time (Levermann et al. 2009). A few recent studies suggested a significant increasing trend in the frequency and magnitude of extreme rain events and a significant decreasing trend in the frequency of moderate events over central India during the monsoon seasons from 1951 to 2000 (Goswami et al. 2006) and 1901–2004 (Rajeevan et al. 2008). Our above arguments are only speculative and subject to further research.

## 6 Summary of key findings

We have analyzed the spatial structure and organization of the monsoonal rainfall field. We have obtained the



**Fig. 16** a Scatter plot between AIMRI (All-India Monsoon Rainfall Index) and  $\epsilon$  (accuracy of prediction). The red dashed vertical line is the  $-1$  sigma standard deviation of AIMRI, i.e., to the left of it are the weakest monsoons years and the blue dashed vertical line is the  $+1$  standard deviation of AIMRI, i.e., to the right of it are the strongest monsoon years. The colour code gives  $E^\alpha$ , which is the total number of events above the used threshold  $\alpha$  in a year over the entire region. We observe that years with large number of events accumulate at the top of the significance band indicating that  $\epsilon$  is higher for stronger monsoon years and lower for weak monsoon years, i.e., years with

less events accumulating at the bottom of the significance band. b Scatter plot between  $E^\alpha$  and  $\epsilon$  and different colours representing the El Niño years (red), La Niña years (blue) and Non El Niño/La Niña years (black). Left to the red vertical line are the weakest 15% monsoon years in terms of  $E^\alpha$  and right to the blue line are top 15% of the monsoon years in terms of  $E^\alpha$ . c The temporal evolution of  $\epsilon$ . El Niño years (red circles), La Niña years (blue circles) and Non El Niño/La Niña years (black circles). Colours within the circles are  $E^\alpha$  values. We note a strong increase in  $\epsilon$  since the year 2000 with highest values of  $\epsilon$  occurring for the last 3 years

correlation structure of extreme rainfall events ( $>90$ th and  $>94$ th percentiles) employing nonlinear correlation of event synchronization. Furthermore, we have carried out a comprehensive spatiotemporal analysis using a complex network approach. Some of our findings provide new insights into the interaction of atmospheric processes responsible for the generation of extreme rainfall events during the ISM. Our methodological approach is strengthened by reiterating previous findings and observation of the climatic features of the ISM. Because of the somewhat extensive and new methodology, we synthesize and list our important findings:

1. It has been previously shown that there exist two distinctive phases of activity within during the ISM (Webster et al. 1998): (1) the active phase, during which large areas of the Indian subcontinent receive extensive rainfall and (2) the break phase during which most regions receive no to very little rainfall. With the

- presented approach, we have been able to document the spatial manifestation of these phases and are thus able to support and validate our approach. In addition, our approach provides new insights and we were able to identify regions, which receive rainfall only during the most active phase of the ISM. In a second step, we were able to provide a quantitative measure of the median length scale involved in rainfall during the active phase of monsoon. This analysis of geographical-length scales shows that spatial scales above the 90th and 94th percentile rain events follow a gamma distribution. We determined that median length scale in these events are up to 250 km for most of the region.
2. We were able to identify the structure and organization of the rain field in terms of its spatial discontinuity. Using clustering coefficients, we have determined that in northwest and southeast India, south Pakistan and in parts of the Tibetan plateau rainfall activity of the ISM occurs in more defragmented forms as compared to

other parts of the subcontinent. This may be related to the fact that extreme rainfall events (at the 90 and 94% percentiles) in these regions are often associated with localized convective rainfall cells.

3. Our approach using centrality measures (degree, closeness, betweenness) suggests that atmospheric processes in northwest Pakistan play crucial role in the generation of large rainfall events over other parts of the Indian subcontinent during the ISM. In northwest Pakistan, mid-latitude westerlies interact with the monsoonal trough in a very dry and hot region enabling formation of convection instabilities. This particular interaction is also said to be responsible for generation of the ISM break phase. Hence we are able to establish with our methodology the importance of this particular mechanism on the internal dynamics of monsoon.
4. We have identified that the central Gangetic plains and parts of Pakistan are the major moisture sinks characterized by high amounts of moisture accumulation during the ISM. The location of these moisture sinks was also found to be consistent with vorticity and divergence of wind vector during the ISM. The central Gangetic plains are known for the formation high number of monsoonal depressions during the ISM season.
5. We have developed a methodology based on causalities of rain events and the complex network approach to identify anomalous monsoon years. We find that regions with high topography and relief have the most intricate and unstable ISM rainfall patterns. Using a similar approach, we studied the temporal evolution of the ISM and its linkages to El Niño Southern Oscillation (ENSO). This analysis reveals that ENSO is not always resulting in a complete breakdown of the spatial ISM rainfall structure nor in its complete re-organization. This findings supports the previously established understanding that the coupling between ENSO and ISM is of complex dynamical nature.
6. Our analysis reveals that since the year 2000 the ISM exhibits characteristically different rainfall patterns as compared to the time period from 1951 to 2000. This is a new insight into the evolving complexity of monsoonal precipitation in a warming environment. As it has been stated in previous studies, there exists an increase in magnitude and frequency in extreme rainfall events during the ISM, while moderate rainfall events are decreasing. This characteristic feature in ISM rainfall distribution could be responsible for the observed change in rainfall patterns since 2000.

## 7 Conclusion

In this study, we have presented an analysis of the spatiotemporal Indian Summer Monsoon (ISM) rainfall

distribution using nonlinear methods and complex networks. Our study improves previous analysis because it specifically takes into account the temporal disparities and spatial complexity of ISM rainfall. The analysis provides new insights into the interaction of different atmospheric processes responsible for generation of extreme rainfall events (at the 90 and 94% percentile) during the ISM. In summary, this study not only opens up new opportunities for meteorologist to look at regional climate using tools from complex networks, but also provides new and valuable insights into the phenomena of ISM rainfall.

**Acknowledgments** This work is supported by a doctoral scholarship from the DFG Graduate School 1364. The authors thank Dr. Yong Zou, Jonathan F. Donges and Arghya Mondal for helpful discussion.

## References

- Albert R, Barabási AL (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74:47
- Arenas A, Díaz-Guilera A, Kurths J, Moreno Y, Zhou CS (2008) Synchronization in complex networks. *Phys Rep* 469(3):93–153
- Bhide UV, Muzamdar R, Ghanekar SP, Paul DK, Chen TC, Rao GV (1997) A diagnostic study on heat sources and moisture sinks in the monsoon trough area during active break phase of the Indian monsoon of 1979. *Tellus Ser A Dyn Meteorol Oceanol* 49:455–473
- Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU (2006) Complex networks: structure and dynamics. *Phys Rep* 424(4–5): 175–308
- Bookhagen B (2010) Appearance of extreme monsoonal rainfall events and their impact on erosion in the himalaya. *Geomat Nat Hazards Risk* 1:37–50
- Bookhagen B, Burbank DW (2006) Topography, relief, and tmm-derived rainfall variation along the himalaya. *Geophys Res Lett* 33:L08405
- Bookhagen B, Burbank DW (2010) Toward a complete himalayan hydrological budget: spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *J Geophys Res* 115:F03019
- Coles S, Heffernan J, Tawn J (1999) Dependence measures for extreme value analyses. *Extremes* 2(4):339–365
- Coles S (2001) An introduction to statistical modeling of extreme values. Springer, Berlin
- Dangalchev C (2006) Residual closeness in networks. *Physica A* 365(2):556–564
- Davison AC, Hinkley D (2006) Bootstrap methods and their application, 8th edn. Cambridge University Press, Cambridge
- Dimri AP (2004) Impact of horizontal model resolution and orography on the simulation of a western disturbance and its associated precipitation. *Meteorol Appl* 11:115–127
- Ding Y, Sikka DR (2006) Synoptic systems and weather. In: Wang B (ed) *The Asian monsoon*. Springer, Berlin
- Donges JF, Zou Y, Marwan N, Kurths J (2009) The backbone of the climate network. *EPL* 87(4):48007
- Donges JF, Schultz HCH, Marwan N, Zou Y, Kurths J (2011) Investigating the topology of interacting networks—theory and application to coupled climate subnetworks. *Eur Phys J B*. doi: 10.1140/epjb/e2011-10899-1



- Fasullo J, Webster PJ (2003) A hydrological definition of Indian monsoon onset and withdrawal. *J Clim* 16:3200–3211
- Gadgil S (2003) The Indian monsoon and its variability. *Annu Rev Earth Planet Sci* 31:429–467
- Goswami BN, Venugopal V, Sengupta D, Madhusoodanan MS, Xavier Prince K (2006) Increasing trend of extreme rain events over India in a warming environment. *Science* 314:1442–1445
- Groisman PY, Karl TR, Easterling DR, Knight RW, Jamason PF, Hennessy KJ, Suppiah R, Page CM, Wibig J, Fortuniak K et al (1999) Changes in the probability of heavy precipitation: important indicators of climatic change. *Clim Change* 42(1):243–283
- Hong C-C, Hsu H-H, Lin N-H, Chi H (2011) Roles of European blocking and tropical? extratropical interaction in the 2010 Pakistan flooding. *Geophys Res Lett* 38:L13806
- Houze RA, Wilton DC, Smull BF (2007) Monsoon convection in the Himalayan region as seen by the TRMM Precipitation Radar. *Q J R Meteorol Soc* 133:1389–1411
- Johnson RH (2006) Mesoscale processes. In: Wang B (ed) *The Asian monsoon*. Springer, Berlin
- Keshavamurthy RN (1973) Power-spectra of large scale disturbances of the Indian south-west monsoon. *Indian J Meteorol Geophys* 24:117–124
- Kripalani RH, Kulkarni A (1999) Rainfall variability over south-east Asia—connections with Indian monsoon and enso extremes: new perspectives. *Int J Climatol* 17(11):1155–1168
- Krishnamurthy V, Shukla J (2000) Interseasonal and inter annual variability of rainfall over India. *J Clim* 13:4366–4377
- Kumar KK, Rajagopalan B, Cane M (1999) On the weakening relationship between the Indian monsoon and enso. *Science* 284(5423):2156–2159
- Kumar KK, Rajagopalan B, Hoerling M, Bates G, Cane M (2006) Unraveling the mystery of Indian monsoon failure during el niño. *Science* 314(5796):115–119
- Lenton TM, Held H, Kriegler E, Hall JW, Lucht W, Rahmstorf S, Schellnhuber HJ (2008) Tipping elements in the earth's climate system. *PNAS* 105(6):1786–1793
- Lenton TM (2011) Early warning of climate tipping points. *Nat Clim Change* 1:201–209
- Levermann A, Schewe J, Petoukhov V, Held H (2009) Basic mechanism for abrupt monsoon transitions. *PNAS* 106(49):20572–20577
- Malik N, Marwan N, Kurths J (2010) Spatial structures and directionalities in monsoonal precipitation over south Asia. *Nonlinear Process Geophys* 17:371–381
- Maraun D, Kurths J (2005) Epochs of phase coherence between el nino/southern oscillation and Indian monsoon. *Geophys Res Lett* 32:L15709
- May W (2004) Simulation of the variability and extremes of daily rainfall during the Indian summer monsoon for present and future times in a global time-slice experiment. *Clim Dyn* 22:183–204
- May W (2004) Variability and extremes of daily rainfall during the Indian summer monsoon in the period 1901–1989. *Glob Planet Change* 44:83–105
- Mokhov II, Smirnov DA, Nakonechny PI, Kozlenko SS, Seleznev EP, Kurths J (2011) Alternating mutual influence of el-nino southern oscillation and Indian monsoon. *Geophys Res Lett* 38:L00F04
- Mooley DA, Shukla J (1989) Main features of the westward moving low pressure systems which form over the Indian region which form over the Indian region during the monsoon season and their relationship with monsoonal rainfall. *Mausam* 40:137–152
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256
- Parthasarathy B, Munot AA, Kothawale DR (1995) Monthly and seasonal rainfall series for all-India homogeneous regions and meteorological subdivisions: 1871–1994. Contributions from Indian Institute of Tropical Meteorology, Research Report RR-065, Aug. 1995, Pune 411 008 India
- Quián Quiroga R, Kreuz T, Grassberger P (2002) Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Phys Rev E* 66:041904-1
- Ramanathan V, Chung C, Kim D, Bettge T, Buja L, Kiehl JT, Washington WM, Fu Q, Sikka DR, Wild M (2005) Atmospheric brown clouds: impact on south Asian climate and hydrologic cycle. *PNAS* 102:5326–5333
- Rajeevan M, Bhatte J, Jaswal AK (2008) Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophys Res Lett* 35:L18707
- Roe GH (2005) Orographic precipitation. *Annu Rev Earth Planet Sci* 33:64571
- Schumacher C, Houze RA (2003) Stratiform rain in the tropics as seen by the trmm precipitation radar. *J Clim* 16:1739–1756
- Sikka DR (1977) Some aspects of the life history, structure and movement of monsoon depressions. *Pure Appl Geophys* 115:1501–1529
- Stephenson DB, Rupa Kumar K, Doblas-Reyes FJ, Royer JF, Chavin F (1999) Extreme rainfall events and their impact on ensemble forecasts of Indian monsoon. *Mon Weather Rev* 127:1954–1966
- Steinbach M, Tan PN, Kumar V et al (2003) Discovery of climate indices using clustering. In: Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining, pp 446–455
- Steinhaeuser K, Chawla NV, Ganguly AR (2010) An exploration of climate data using complex networks. *ACM SIGKDD Explor* 12:25–32
- Swanson KL, Tsonis AA (2009) Has the climate recently shifted? *Geophys Res Lett* 36:L06711
- Tsonis AA, Swanson KL (2008) Topology and predictability of el nino and la niña networks. *Phys Rev Lett* 100(22):228502
- Tsonis AA, Swanson KL, Roebber PJ (2006) What do networks have to do with climate? *Bull Am Meteorol Soc* 87(5):585
- Tsonis AA, Wang G, Swanson KL, Rodrigues F, Costa L (2010) Community structure and dynamics in climate networks. *Clim Dyn*. doi:10.1007/s00382-010-0874-3
- Waliser DE (2006) Interseasonal variability. In: Wang B (ed) *The Asian monsoon*. Springer, Berlin
- Wang B (ed) (2006) *The Asian monsoon*. Springer, Berlin
- Watts DJ, Strogatz SH (1998) Collective dynamics of small-world networks. *Nature* 393(6684):440–442
- Webster PJ (1987) The elementary monsoon. In: Fein JS, Stephens PL (eds) *Monsoons*. Wiley, New York
- Webster PJ, Magana VO, Palmer T, Shukla J, Tomas RA, Yanai M, Yasunari T (1998) Monsoons: Processes, predictability, and the prospects for prediction. *J Geophys Res* 103:14,451–14,510
- Wulf H, Bookhagen B, Scherler D (2010) Seasonal precipitation gradients and their impact on fluvial sediment flux in the northwest himalaya. *Geomorphology* 118:13–21
- Yamasaki K, Gozolchiani A, Havlin S (2008) Climate networks around the globe are significantly affected by el niño. *Phys Rev Lett* 100(22):22850
- Yanai M, Wu G (2006) Effects of the tibetan plateau. In: Wang B (ed) *The Asian monsoon*. Springer, Berlin
- Yatagai A, Arakawa O, Kamiguchi K, Kawamoto H (2009) A 44 year daily gridded precipitation dataset for Asia based on dense network of rain gauges. *SOLA* 5:137–140
- Zickfeld K, Knopf B, Petoukhov V, Schellnhuber HJ (2005) Is the Indian summer monsoon stable against global change? *Geophys Res Lett* 32:L15707