A complex network approach to study the extreme precipitation patterns in a river basin

Cite as: Chaos **32**, 013113 (2022); doi: 10.1063/5.0072520
Submitted: 23 September 2021 · Accepted: 20 December 2021 ·
Published Online: 11 January 2022

Ankit Agarwal,[1,a)] Ravi Kumar Guntu,[1] Abhirup Banerjee,[2,3,4] Mayuri Ashokrao Gadhawe,[1] and Norbert Marwan[2,5]

AFFILIATIONS

[1]Department of Hydrology, Indian Institute of Technology Roorkee, Roorkee 247667, India
[2]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, 14412 Potsdam, Germany
[3]Institute of Environmental Science and Geography, University of Potsdam, 14476 Potsdam, Germany
[4]Institute of Physics and Astronomy, University of Potsdam, 14476 Potsdam, Germany
[5]Institute of Geoscience, University of Potsdam, 14476 Potsdam, Germany

Note: This article is part of the Focus Issue, Theory-informed and Data-driven Approaches to Advance Climate Sciences.
[a)]**Author to whom correspondence should be addressed:** ankit.agarwal@hy.iitr.ac.in

ABSTRACT

The quantification of spatial propagation of extreme precipitation events is vital in water resources planning and disaster mitigation. However, quantifying these extreme events has always been challenging as many traditional methods are insufficient to capture the nonlinear interrelationships between extreme event time series. Therefore, it is crucial to develop suitable methods for analyzing the dynamics of extreme events over a river basin with a diverse climate and complicated topography. Over the last decade, complex network analysis emerged as a powerful tool to study the intricate spatiotemporal relationship between many variables in a compact way. In this study, we employ two nonlinear concepts of event synchronization and edit distance to investigate the extreme precipitation pattern in the Ganga river basin. We use the network degree to understand the spatial synchronization pattern of extreme rainfall and identify essential sites in the river basin with respect to potential prediction skills. The study also attempts to quantify the influence of precipitation seasonality and topography on extreme events. The findings of the study reveal that (1) the network degree is decreased in the southwest to northwest direction, (2) the timing of 50th percentile precipitation within a year influences the spatial distribution of degree, (3) the timing is inversely related to elevation, and (4) the lower elevation greatly influences connectivity of the sites. The study highlights that edit distance could be a promising alternative to analyze event-like data by incorporating event time and amplitude and constructing complex networks of climate extremes.

Extreme precipitation networks are constructed over a river basin using event synchronization (ES) and edit distance (ED). Edit distance is an alternative to event synchronization accounting for the sequences and the magnitude of events. Network-based measure degree is employed to understand the spatial synchronization pattern of the extreme precipitation in the Ganga river basin (GRB). The influence of the topography and rainfall characteristics on the extreme precipitation networks is also quantified. The proposed study can estimate the impact of artificial boundaries, thereby better understanding the extreme precipitation network's topology and spatial risk quantification of extreme events.

I. INTRODUCTION

Extreme events are of great societal interest as they may lead to many meteorological hazards that majorly affect lives and economic assets.[1,2] The evolution of such events in hydrology is nonuniformly distributed with time resulting in irregularly spaced data series. Traditionally, linear measures like Pearson and Spearman correlation are recognized as effective and robust estimators to quantify linear dependency between any two variables in irregular time series.[3] The Pearson correlation coefficient computes the linear dependency by looking into entire distribution of the data. For an extreme event-like data (tail part of the distribution), its application becomes suboptimal.[4] Likewise, other linear approaches

such as wavelet analysis and Fourier transform are insufficient to capture nonlinear interrelationships.[5] As a result, defining a principled nonlinear approach is required to analyze extreme event time series, particularly when the dependency between event time series across locations is explored.

Numerous methods are available to study nonlinear dynamical behavior, such as recurrence,[6,7] phase difference,[8,9] Kolmogorov entropy,[10,11] or mutual information (MI).[12,13] However, by definition, extreme events are those that occur rarely and are separated by long periods. So, defining a suitable distance measure that can aid in analyzing the dynamics of extreme event-like time series becomes important in such instances. In the last decade, event synchronization (ES) has gained popularity, particularly in climate science studies. The strengths of ES include capturing dynamic time-delays between various spatially distributed processes and thereby determining typical spatiotemporal patterns in climate systems and synchronization strength. Malik *et al.* used ES to analyze the extreme precipitation events (EPEs) in the Indian Monsoon System.[14] Boers *et al.* conducted a similar study to characterize the extreme precipitation synchronicity in the South American Monsoon System.[15] Later, Agarwal *et al.* proposed a multi-scale event synchronization method by combining the wavelet transform and ES.[16] Furthermore, Kurths *et al.* used wavelet-based ES to unravel the spatial diversity of Indian rainfall teleconnections.[17] An overview by Moreno and Perc complements nicely the hugely important role that network science has played in the analysis of various extreme events.[18]

Event-like time series, a discrete event series generated by a point process, can be examined in their unaltered state. As a measure of similarity, Victor and Purpura proposed a distance metric for calculating the distance between two spike trains (binary event sequences).[19] Later, Hirata and Aihara developed a method for transforming a spike train into a time series known as edit distance.[20] Then, this method was used by Suzuki *et al.* to construct a recurrence plot by measuring a distance between marked point processes to analyze financial data.[21] Later, several studies employed similar distance measures for analyzing irregularly sampled paleoclimate data.[22,23] Recently, we proposed a modification of the edit distance (ED) to investigate extreme event-like time series dynamics by integrating a nonlinear cost function.[6] The robustness of the ED measure is demonstrated by several prototypical examples and real flood event series constructed from discharge data of the Mississippi River in USA. However, to the best of the authors' knowledge, there is no study comparing the ES and ED-based extreme precipitation networks over a river basin with a diverse climate[24] and with complicated topography like the Ganga river basin (GRB) in India.

The GRB is highly vulnerable to EPEs due to its complex topography and altitude-dependent climate. Spatiotemporal changes in the EPEs at a basin-scale are crucial for preventing and mitigating water-related disasters and providing critical information for successful water resource management. EPEs over the Himalayan region are considered to be caused by a combination of thermodynamics and orographic uplifting.[25] However, some studies argue that topography has no direct influence on precipitation.[26] Conversely, Houze identified topographic variables as fundamental contributors to precipitation fallout.[27] Topography significantly impacts precipitation, but the relationship between topographical characteristics and precipitation is poorly understood.[28,29] Bharti *et al.* made an effort to study the EPEs in association with the elevation and found that the frequency and intensity of EPE have exhibited an inverse relationship with elevation.[30] Furthermore, basin-scale studies investigating the influence of topography on EPEs are rare. Therefore, there is a need to study the impact of topographic measures on the EPEs over the river basin having complex terrain toward quantifying the variability of EPEs.

In the present study, event synchronization and edit distance are compared to characterize the EPEs in the Ganga river basin; this involves constructing extreme precipitation networks over the basin. The topology of the resulting networks is quantified using graph theory-based measures such as network degree. Finally, a statistical comparison is made between the network-based measures and topographic measures to quantify the impact of topography on EPEs.

II. STUDY AREA AND DATASET

A. Study area

The Ganga river is the longest in India and is located in the northern part of the country, covering the western and central Himalayas. The GRB is a transboundary basin and lies in India, China, Nepal, and Bangladesh. The present study is carried out over the region falling in India only due to gauge-based observations. The spatial extent of GRB falling in India covers 21°6′ N to 31°21′ N and 73°2′ E to 89°5′ E in latitude and longitude directions, respectively. An undulating terrain pattern characterizes the topography of GRB. The elevation ranges from about 7000 m north to less than 100 m above mean sea level in the south [Fig. 1(a)]. This complex topography leads to high spatiotemporal variability in precipitation and diverse climate classifications ranging from semi-tropical to semi-arctic.[30] The GRB's typical average annual rainfall ranges from 300 mm at the western side to 3700 mm at the eastern end [Fig. 1(b)]. The GRB experiences intense cyclone activity during the monsoon season resulting in loss of life and infrastructure. Flow in the basin is majorly driven by the precipitation coming from the Indian summer monsoon (June to September) and snowmelt from the Himalayas during the Spring season (March to May).

B. Dataset

In the present study, we use the gridded daily precipitation (mm) data with a high spatial resolution of $0.25° \times 0.25°$ covering the entire GRB. The gridded dataset was generated from a network of 6995 gauging stations across India using the inverse distance weighted interpolation technique proposed by Shepard.[31] The dataset was developed by the Indian Meteorological Department (IMD)[32] and can be retrieved from IMD's website.[33] We use daily data for 22 years (1998–2019) to examine the EPEs over the GRB. In the past, several studies used the same dataset for various hydrometeorological applications. These include extreme precipitation analysis,[34] precipitation regionalization,[35] intrinsic predictability of Indian summer monsoon,[36] spatiotemporal variability of precipitation over India,[37] and spatial diversity of Indian precipitation teleconnections,[17] among others. Various applications confirm that the data are highly accurate and reliable in capturing the spatial distribution of precipitation over the GRB.

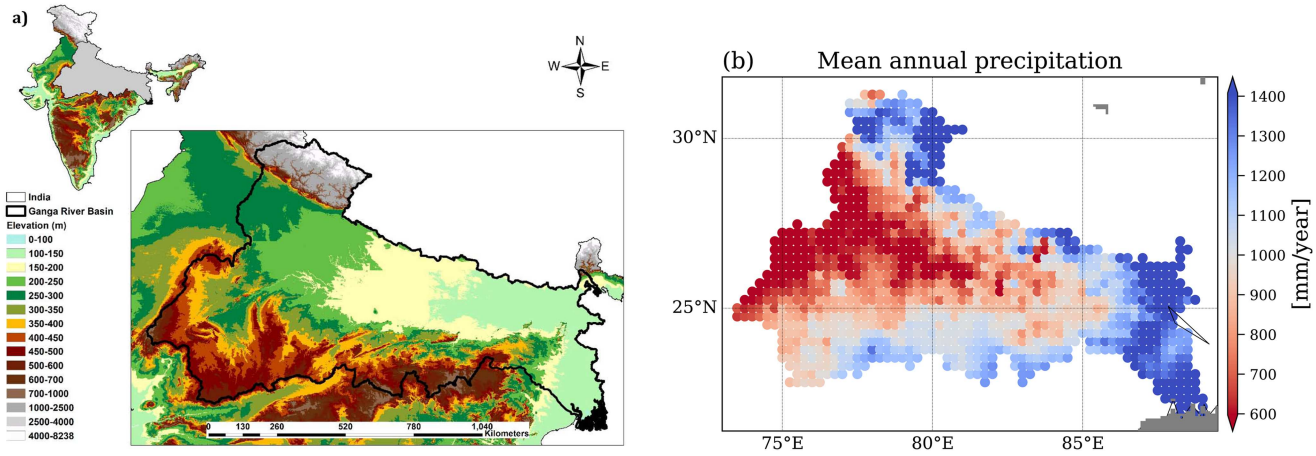

FIG. 1. (a) Geographical location of the Ganga river basin in India showing elevation of the study area in meters above mean sea level. The elevation map is prepared by using the shuttle radar topography mission digital elevation model. This figure is generated using ArcGIS 10.6 (https://www.esri.com/en-us/arcgis/products/arcgis-pro). (b) Spatial distribution of mean annual precipitation (in mm) over the period 1998–2019 based on IMD observations.

III. METHODOLOGY

In this work, we use two event-based nonlinear similarity measures, event synchronization (ES)[38] and edit distance (ED),[19] to study the extreme precipitation pattern across the GRB. In ES and ED, we consider events or sequences of events at one grid point and measure the strength of synchronicity with the other grid points. In this way, we construct the similarity matrix for all the grid points. To understand the underlying pattern of extreme events in more detail, elevation and precipitation seasonality need to be studied. Additionally, to the timing of events, ED also considers the similarity of amplitude variations.

A. Edit distance

Victor and Purpura developed a specific distance metric to measure the similarity between the two spike trains based on the cost to transform one spike sequence into the other one.[19] Hirata and Aihara extended this idea for converting a spike train into a real-valued time series and named it edit distance.[20] Edit distance was used for marked point processes to study the recurrence properties of exchange tick data of foreign currencies,[21] recurrence analysis of irregularly sampled data,[39] and extreme event-like data.[6]

We consider the event series at every grid point as a spike train. Let us consider two grid points i and j to find the similarity between the events; we transform one spike train into another by using the following operations: shifting in time, change in amplitude, and deletion/insertion [Fig. 2(a)]. Each of these operations is assigned with a cost. We use a combination of all three elementary operations and find the minimum path to convert one segment to another. The minimum transformation cost is defined as[40]

$$D(S_i, S_j) = \sum_{(l,m)\in C} \{p_0||t_l^i - t_m^j|| + p_1||L_l^i - L_m^j||\} + p_2(|I| + |J| - 2C), \tag{1}$$

where C is a set of pairs involved in the shifting process; the first term on the right-hand side of Eq. (1) corresponds to the shifting of an event l by time $t_j - t_m$ and change in amplitude by $L_l - L_m$, and the second term involves the deletion/insertion of events. t_l^i and t_m^j denote the time of events and L_l^i and L_m^j denote the amplitude of events.

In edit distance analysis, we have three cost parameters, namely, p_0, p_1, and p_2, which are associated with the cost of shifting in time, change in amplitude, and cost of deletion/insertion, respectively. In the present study, we adopted the definition of the cost parameters proposed by Ozken *et al.*[39,41] The first cost parameter p_0 provides the weightage for shifting in time for two event series and is calculated as $p_0 = \frac{M}{\text{total time}}$, where M is the number of events. So, when we calculate the transformation cost between two events series, we count the total number of events in two grid points divided by the total time, and it could be considered an event rate. The dimension of this parameter is $\frac{1}{\text{time}}$. The second cost parameter is responsible for the change in amplitude, given by $p_1 = \frac{M-1}{\sum_t^{M-1}||x_i - x_{i+1}||}$, where x_i is the amplitude of ith event in the data. The dimension of this parameter is $\frac{1}{\text{Length}}$. The cost of deletion/insertion p_2 is set to be 1,

$$D(S_i, S_j) = \sum_{(l,m)\in C} \{p_0||t_l^i - t_m^j|| + p_1||L_l^i - L_m^j||\} + (|I| + |J| - 2C). \tag{2}$$

We apply the distance measure Eq. (2) for all the pairs of grid points and store the value as a similarity matrix R_{ij}.

In the present study, the cost of shifting in time p_0 depends on the number of events M. As we fixed the number of events in each grid point, the relative contribution in every pair of event series would be the same. Hence, the cost for shifting operation (in time + difference in amplitude) is mainly influenced by the parameter p_1. Unlike p_0, the value of P_p will change depending on

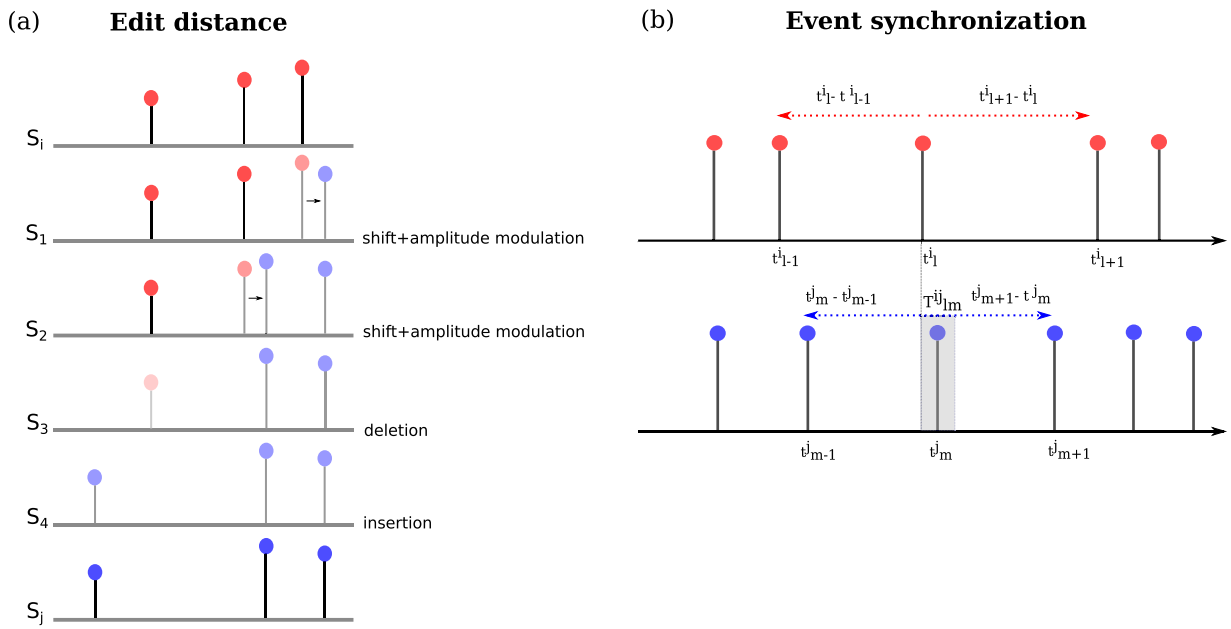

FIG. 2. Schematic illustration of (a) the edit distance method—segment S_i transformed to S_j by following the path from S_1 to S_4 operations; the final cost is obtained by minimizing all the possible combination. (b) The event synchronization—t_l^i and t_m^j are occurrence of events in the event series at grid point i and j, and T_{lm}^{ij} is the adaptive time lag.

the difference in amount of rainfall. So, the final cost is mainly influenced by parameter p_1. ED works recursively, i.e., the shifting operation is associated with the events of more or less similar time and amplitude. So, in this way, ED can distinguish between the events with small variations in time and amplitude and with the events with a large difference in time and amplitude. This work chooses ∼36 events (top 10 percentile in a year) for each grid point and constructs the event series. The threshold value may differ from grid to grid. In this way, we manage to find a way to reduce the bias due to different event numbers.

B. Event synchronization

Event synchronization is a method to measure the synchronicity between climate extremes like extreme precipitation[42] or heat-wave pattern.[43] In precipitation network analysis, we consider each precipitation grid point as a node. To measure the synchronicity of events between two grid points i and j, let us consider an extreme event l occurring at a timet_l^i supposed to be synchronized with another event m at grid point j at time t_m^jwithin a time lag $\pm\ T_{lm}^{ij}$ [Fig. 2(b)], where T_{lm}^{ij} is a flexible time lag that adopts the density of events (sparse events = larger lags, dense events = smaller lags) and is expressed as follows:

$$T_{lm}^{ij} = \frac{\min\left(t_{l+1}^i - t_l^i, t_l^i - t_{l-1}^i, t_{m+1}^j - t_m^j, t_m^j - t_{m-1}^j\right)}{2}. \quad (3)$$

Here, $l = 1, 2, 3, \ldots, S_i$ and $m = 1, 2, 3, \ldots, S_j$ are the number of events at ith and jth grid points, respectively. We count the coincidence of events at grid points i and j with the following scheme:[39]

$$J_{ij} = \begin{cases} 1 & \text{if } 0 < t_l^i - t_m^j < T_{lm}^{ij}, \\ 1/2 & \text{if } t_l^i = t_m^j, \\ 0 & \text{else.} \end{cases} \quad (4)$$

The synchronization is measured by counting the weights of the synchronized events in both the time series $c(i|j) = \sum_{l=1}^{s_i} \sum_{m=1}^{s_j} J_{ij}$ and vice versa: $c(j|i)$ at the grid points i and j. Finally, the strength of synchronization is defined as[14]

$$Q_{ij} = \frac{c(i|j) + c(j|i)}{\sqrt{(s_i - 2)(s_j - 2)}}, \quad (5)$$

where Q_{ij} lies in the range of 0–1, where 1 implies the complete synchronization and 0 the absence of synchronization. We use this method for all pairs of grid points ($i \neq j$) and obtain the similarity matrix, which stores the strength of synchronization between extreme precipitation at all grid points.

C. Study design to test the efficacy of ED and ES

Finding a suitable similarity measure is important to investigate the interactions between different time series. However, when we deal with event-like data, the available options of similarity measures become very specific and have some limitations.

Kreuz *et al.* proposed a framework to compare the synchronization measurements from different approaches.[44] We adopted a similar prototype system to study the interaction between two time

series and their strength of synchronization for different coupling values and is described as follows.

Consider two unidirectionally coupled Hénon maps with the following equations:[9]

$$x_1(i+1) = 1.4 - x_1^2(i) + bx_2(i)x_2(i+1) = x_1(i), \tag{6}$$

$$y_1(i+1) = 1.4 - [\mu x_1(i)y_1(i) + (1-\mu)y_1^2(i)] + by_2y_2(i+1) = y_1(i), \tag{7}$$

where x_1 and x_2 are the components of the drive system, y_1 and y_2 are the components of the response system, and μ is the coupling coefficient that varies in range of 0–0.8. A similar model system has been adopted in previous studies[45–47] to investigate the direction of coupling. In the first step, the event series is constructed. To do so, scan through x_1 and y_1 component and find the local maxima ($x(t_i) > x(t_{i\pm1})$) and store the times t_i^x, t_j^y, and amplitudes L_i^x, L_j^x ($i = 1, 2, \ldots, n_x; j = 1, 2, \ldots, n_y$), where n_x and n_y denote the number of events. The local maxima procedure has been successfully used to study event-like data.[44,46,48]

When the coupling is minimal [see x axis in Fig. 3(c)], the similarity between two time series is less; hence the Q value is low and simultaneously D value is high as the cost of transformation is very high. As we increase the coupling parameter (μ), the value of D decreases [Fig. 3(a)], and the value of Q increases [Fig. 3(b)]. We detected the first four transition points using change-point detection,[49] where abrupt changes in Q and D values are detected. In the previous studies,[45–47] the onset of identical synchronization was found to start for $\mu \geq 0.6$. To confirm the direction of interdependency, a mutual information (MI) estimator is used between two time series x_1 and y_1. The original time series is considered in the analysis but not the local maxima from the driver and response system. Kraskov *et al.* proposed a method to compute mutual information based on the k nearest neighbor.[50] MI measure has less bias and is robust against short time series.[51] For readers who are unfamiliar with MI, refer to the study by Kraskov *et al.* for detailed methodology.[50]

Interestingly, from Fig. 3, we observed that all transition points lie early for relatively low coupling in ES [Fig. 3(b)] when compared with the mutual information [Fig. 3(c)]. However, in the case of ED [Fig. 3(a)], the transition points (vertical colored lines) are placed when the coupling is relatively strong, i.e., the similarity between two time series is high. The prototypical system shows that both ED and ES are suitable enough to study the interdependence between two event-like time series. Event series with almost coinciding events but strong differences in amplitudes will be considered very similar by ES but more different by ED; less synchronous event series with very similar amplitude variation can, in contrast, be considered different by ES but more similar by ED. With the additional information of amplitude in ED, it helps in capturing a more realistic picture.

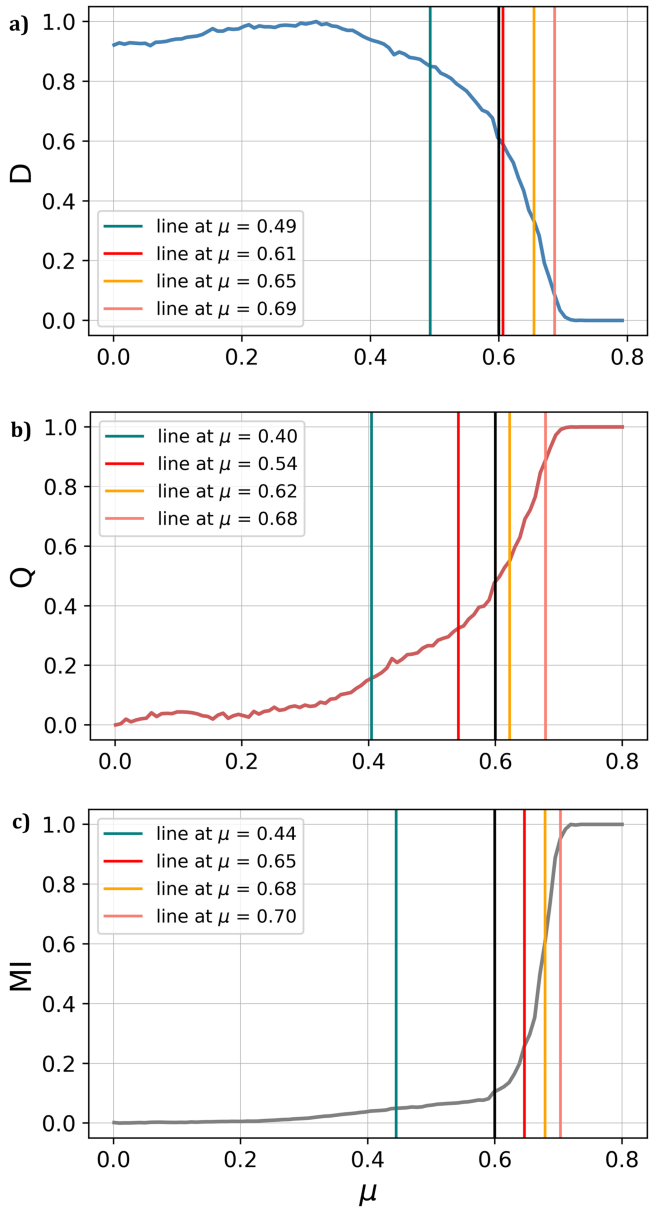

FIG. 3. Detection of changes in the direction of coupling with coupling parameter value (μ) ranging from 0 to 0.8 for (a) event synchronization (ES) and (b) edit distance (ED) with (c) mutual information (MI). Vertical color lines represent the transition in the direction of coupling. The black vertical line signifies the onset of identical synchronization.

D. Network construction and network measure

We obtain the adjacency matrix by applying a certain threshold upon the similarity matrix. In this study, the threshold is defined based on the link density [Eq. (8)],

$$\rho = \frac{2E}{(N(N-1))}, \tag{8}$$

where E is the number of edges and N is the number of nodes, and the corresponding threshold $\tau = \tau(\rho)$ is chosen. For ES, the

adjacency matrix defined as

$$A_{ij}^{ES} = \begin{cases} 1 & \text{if } Q_{ij} > \tau, \\ 0 & \text{else.} \end{cases} \tag{9}$$

For the adjacency matrix, it defined as

$$A_{ij}^{ED} = \begin{cases} 1 & \text{if } R_{ij} < \tau, \\ 0 & \text{else.} \end{cases} \tag{10}$$

For both ED and ES, we choose the threshold τ for an edge density of 0.05. In the previous climate network applications,[14,52,53] a link density of 0.05 satisfies a high similarity and retains the significant links.

In the study of complex networks, there are multiple network measures available based on the research question. Our research is particularly interested in finding the grid points that show a similar pattern of extreme events. Accordingly, the node degree (k) of a node i is calculated, indicating the number of connections with other nodes in a network,

$$k_i = \sum_{j=1}^{N} A_{ij}, \tag{11}$$

where N is the number of nodes and A_{ij} is the adjacency matrix.

E. Precipitation seasonality

In the following, we assess the difference in node degree (k) obtained from event synchronization and edit distance, which can be applied in understanding the difference in the topology of the extreme precipitation network patterns. Feng *et al.* calculated the first moment of area for monthly precipitation to derive the timing of 50th percentile precipitation within a year [Fig. 4(a)].[54] Later, several studies[35,55] used the same centroid concept to derive the timing (T_i) of 50th percentile precipitation over different time scales.

The calculation of the long-term average timing is a two-step procedure. We considered 22 years of data, and the median value of precipitation is estimated for all the months [Fig. 4(a)]. Next, the timing (T_i) was computed using Eq. (12) and is indicated in Fig. 4(b),

$$T_i = \frac{\sum_{m=1}^{12} m p_m}{\sum_{m=1}^{12} p_m}, \tag{12}$$

where m represents the month number from 1 to 12 and p_m represents the corresponding precipitation amount during that particular month.

IV. RESULTS AND DISCUSSION

We present the results in four categories: first, investigating network characteristics of extreme precipitation; second, linking network characteristics with precipitation seasonality; third, with topographic variability; and finally, quantifying node importance over GRB.

A. Network characteristics of extreme precipitation over the Ganga river basin

Network characteristics are obtained by thresholding the similarity matrix Q_{ij} for ES and R_{ij} for ED. Network characteristics depend on the choice of threshold τ. In the present study, we constrain the link density $\rho = \frac{2E}{N(N-1)}$, where E gives the total number of edges in the current network and the denominator is the number of all possible edges in the network (if fully connected), with the number of nodes N and selecting the corresponding threshold $\tau = \tau(\rho)$.[12] In both cases, we choose the threshold so that it corresponds to 10% link density. The precipitation network comprises nodes representing the time series of each grid point and the edges showing statistically significant interdependencies between the nodes.[56]

We compute the node degree for all the grid points in the GRB, following the procedure described in Sec. III C. Figures 5(a) and 5(b) present the spatial distribution of the degree of each node using edit distance and event synchronization. We observe significant variations in the spatial distribution of node degree values across the GRB. The grid points with high degree (more links) are more likely

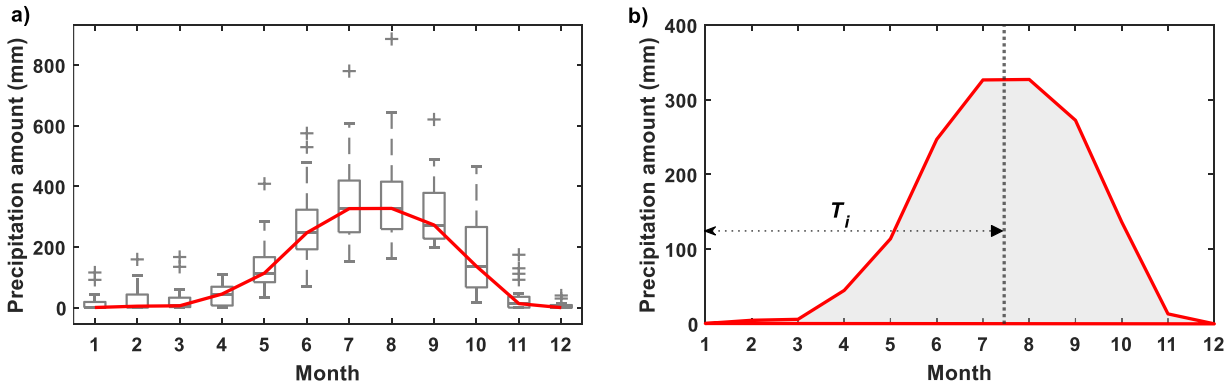

FIG. 4. (a) Long-term average monthly precipitation time series for one grid point (red color). The box-plot statistics of each month are represented in gray color. (b) Computation of the timing of 50th percentile precipitation (T_i) linked with the area of monthly precipitation. The vertical dotted line represents the timing of the peak precipitation.

to be found in the southwest part of the basin, while degree values decrease toward the northeast part and periphery of the basin in both ED- and ES-based networks. In general, a high degree of a node (higher connectivity) indicates that the extreme precipitation data at that grid point shares a piece of similar information with many other grid points over the basin and is likely to have greater impact on functioning of the network. Since similar information is shared, it is not considered a unique one. On the other side, the grid point with degree zero does not share similar information with any other grid point.

The artificially imposed boundary (in this case, the Ganga river basin border) may impact the estimated network measures because spatial boundaries cut links that would connect the considered region with outside regions. We use the boundary correction procedure suggested by Rheinwalt *et al.* to avoid spurious bias on node degree arising from this boundary effect.[57] We use the procedure as follows; we construct 100 spatially embedded random networks (SERNs) that preserve the position of the nodes in space and link length probability depending on the spatial link length of the original network. We compute the network measure for all the SERN surrogates. The boundary effect is measured by taking the average of the network measure for all the surrogates. The corrected network measure is obtained by dividing the original measure with the estimated boundary error for all the nodes. The corrected network measure is dimensionless as it provides the network measure relative to the expected value of a SERN.

Figures 5(c) and 5(d) present the spatial distribution of the corrected degree value computed by edit distance and event synchronization of each grid cell in the basin. We obtain the boundary corrected degree for ED and ES [Figs. 5(a) and 5(b)]. In the case of ED, we find high degree concentrated at a particular (southwest) part of the basin. In contrast, the corrected degree value of the ES-based network is varying all over the basin, but a high number of nodes having high degree values are at the southwest part of the basin.

By comparing the results of the ES and the ED networks, we find a slight difference in the spatial distribution of degree values. The southwest part shows a high degree in the ED network [Fig. 5(c)] and spreads out to the bay's periphery. In the ES network [Fig. 5(d)], the degree value follows a similar pattern as ED networks but with lower degree values in the stronger connected region and with a larger extension of this stronger connected region.

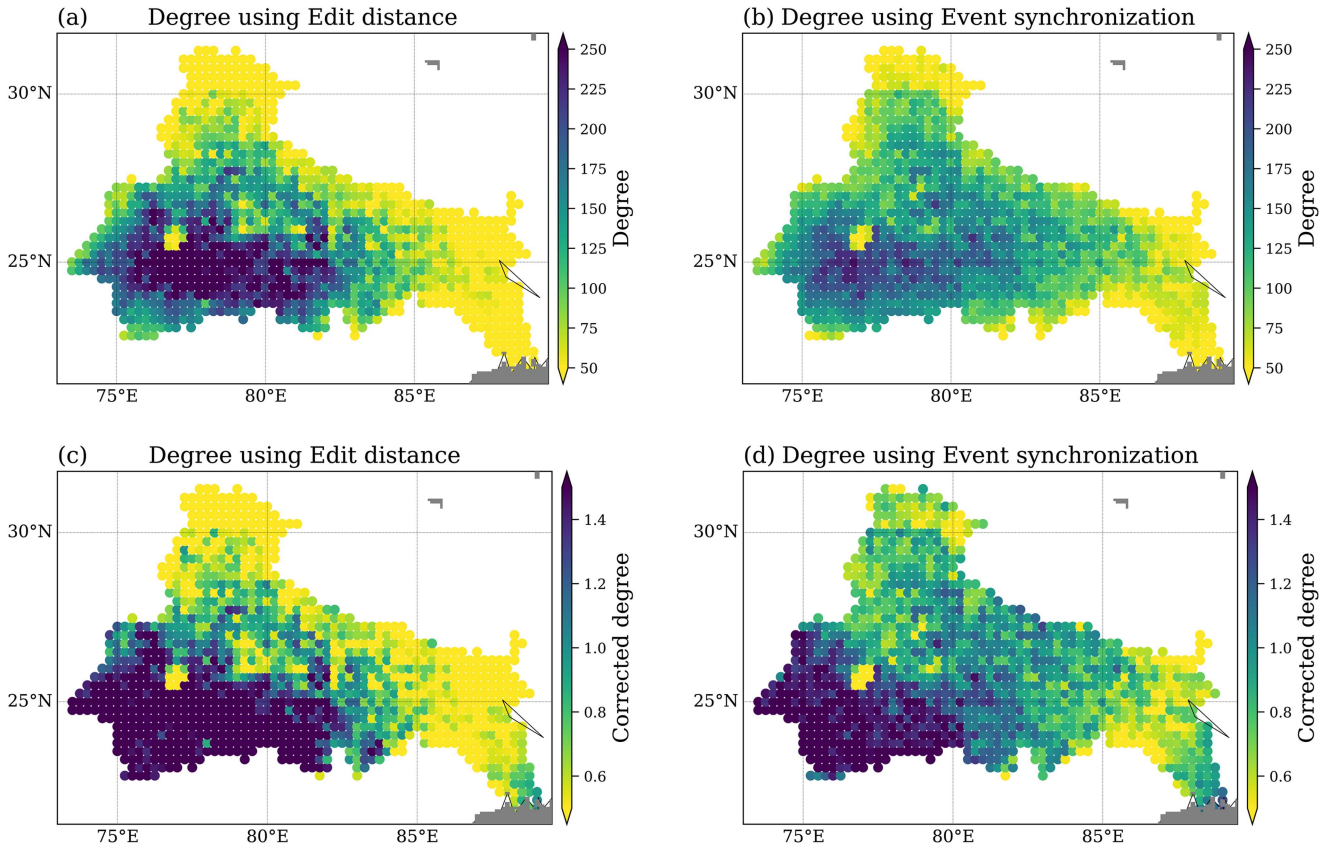

FIG. 5. Spatial distribution of node's degree for (a) the edit distance and (b) event synchronization, and corrected degree for (c) the edit distance and (d) event synchronization.

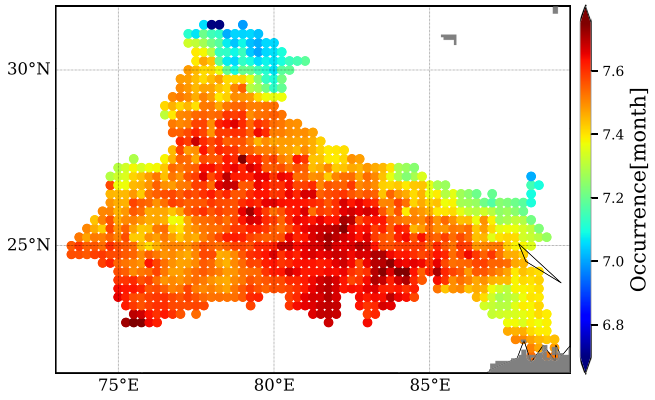

FIG. 6. Spatial pattern of the timing of 50th percentile precipitation for the temporal window 1998–2019.

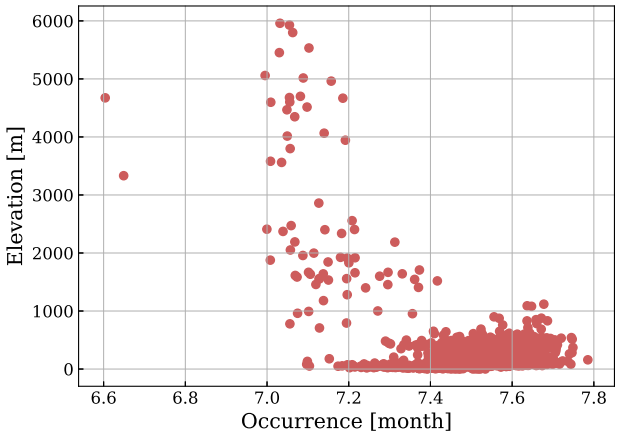

FIG. 7. Relationship between the timing of 50th percentile precipitation and elevation.

Therefore, by comparing Figs. 5(a)–5(d), we see that these nonlinear methods efficiently capture the general network topology without being affected by the artificial boundary.

B. Relationship between network characteristics and precipitation seasonality

We quantify the centroid of the monthly precipitation distribution over the GRB to derive the timing of 50th percentile precipitation (see Sec. III D). The timing is mainly during June to August over the GRB (Fig. 6). The Himalayan region (north side) receives its 50th percentile precipitation at the beginning of June, and in contrast, in the middle of July over the plain region (south region) and at the end of July over the northeast and deltaic region (southwest region) of the GRB. To know if there is any statistical relationship between 50th percentile precipitation and elevation, we consider the occurrence of rainfall and elevation (Fig. 7).

Next, we consider the relationship between the timing of 50th percentile precipitation and the corrected node degree values for ED and ES (Fig. 8). In both cases, grid stations with early timing (before July) and intermediate timing (from July to mid-July) have lower degree values, i.e., fewer connections, while stations with relatively later timing have greater degrees varying from 0 to 2.7. For ED, the relationship between the timing and degree is more apparent than for ES due to the inclusion of the amplitude variations. Thus, by comparing Figs. 8(a) and 8(b), ED could be a promising alternative to analyze event-like data by incorporating event time and amplitude.

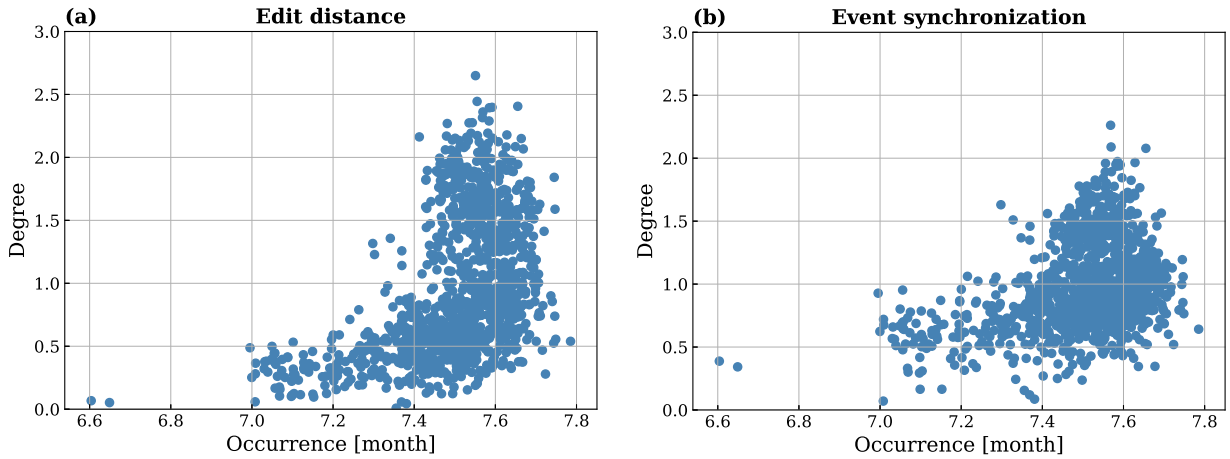

FIG. 8. Relationship between occurrences of peak precipitation and the connectedness between the grid points (degree) for the network constructed using (a) edit distance and (b) event synchronization.

C. Relationship between network characteristics and topographic measure

We further consider the relationship between node degree and elevation of each grid cell to quantify the influence of elevation on the network topology (Fig. 9). The degree varies for the grid cells at different elevations. For example, sites with low elevation (<1000 m) have degree values largely ranging from 0 to 2.5 and from 0 to 2.1 in ED and ES, respectively, while the stations at intermediate to high elevation tend to have fewer connections, i.e., low degree values (below 0.5 for ED and between 0.5 and 1 for ES). The degree in ED-based networks [Fig. 9(a)] is relatively lower than for ES-based networks [Fig. 9(b)] for stations located at elevations between 1000 and 6000 m.

The southwest of the GRB has moderate elevation and grid points over that region receive precipitation during the Indian summer monsoon only [point A in Figs. 10(a) and 10(b)]. The north of the GRB has the highest elevation compared to the rest of the basin [Fig. 1(a)] and is mainly dominated by western disturbances[58] during December to May and receives its highest precipitation during the Indian summer monsoon [point B in Figs. 10(a) and 10(b)]. As a result, the timing of 50th percentile precipitation of the southwest region is more compared to the northern region. The degree and timing are low and the connectedness is limited to a few grid points only in the north of the GRB [Figs. 10(c) and 10(d)]. In particular, connections made by ED [Fig. 10(c)] provide additional information as it includes amplitude as well. We find that grid points in the southwest have a higher degree but not to the central part of the GRB. The western and southwestern region receives less precipitation and has lesser spatiotemporal variability.[37] As most of the grid points in the southwest region receive similar information, it has higher connectedness [Figs. 10(c) and 10(d)]. Interestingly, the timing is the same across the central part and southwest of the GRB (Fig. 6). The primary factor in getting a higher degree across the southwest region is the elevation. As a result, ED connections mostly lie in the same topographical region [Fig. 10(c)], whereas ES links are going to another topographical region as well [Fig. 10(d)]. Therefore, by comparing Figs. 10(c) and 10(d), ED as a similarity measure provides additional information in segregating the extreme events originating during western disturbances and Indian summer monsoon.

D. Quantifying node importance over GRB

In several studies of complex network dynamics and structure, node importance is becoming a hot topic.[3,59] Node importance is evaluated and primarily researched regarding the network's structural properties and application requirements. It is critical to identify significant nodes to create effective networks. The essential nodes in the network are a small number of particular nodes that have a considerable impact on the network's structure and function.[60] Therefore, we argue that those nodes of large importance might indicate locations that are suitable for representing the overall precipitation state/regime or for predicting the extreme events. Degree, closeness, betweenness, clustering coefficient, weighted degree betweenness, and other centrality measures have been developed to evaluate node importance in complex networks.[59] For a comparative understanding of node importance derived from the network constructed using ED and ES, corrected degree values obtained over GRB are used.

Figure 11 shows the spatial distribution of location and network connections of the top 5% of the highest degree grid cells in the ED [Fig. 11(a)] and ES-based network [Fig. 11(b)]. The 5% top degree nodes infer that the time series at the grid stations in the south and southwest region of the basin have more robust connectivity than the rest of the stations in the GRB. Compared to ED, in ES-based networks, important grid points are concentrated in the outer southwest region of the basin [Fig. 11(b)]. Furthermore, we plotted the connections corresponding to this high degree grid station to reduce the complexity in visualizing the interactions. Theoretically and practically, identifying node importance in complex networks is vital for enhancing network robustness and invulnerability.

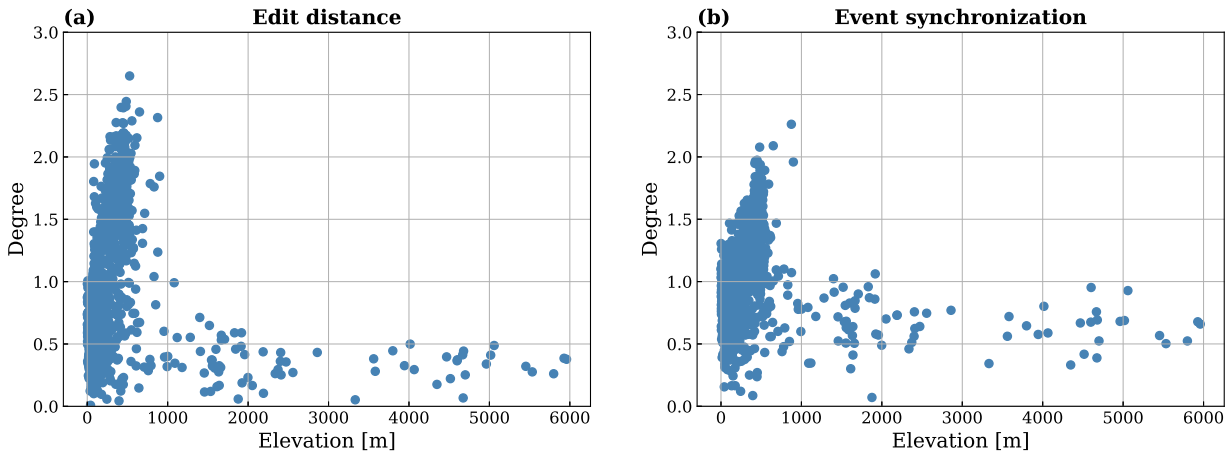

FIG. 9. Relationship between degree and elevation for the precipitation network constructed using (a) ED and (b) ES.

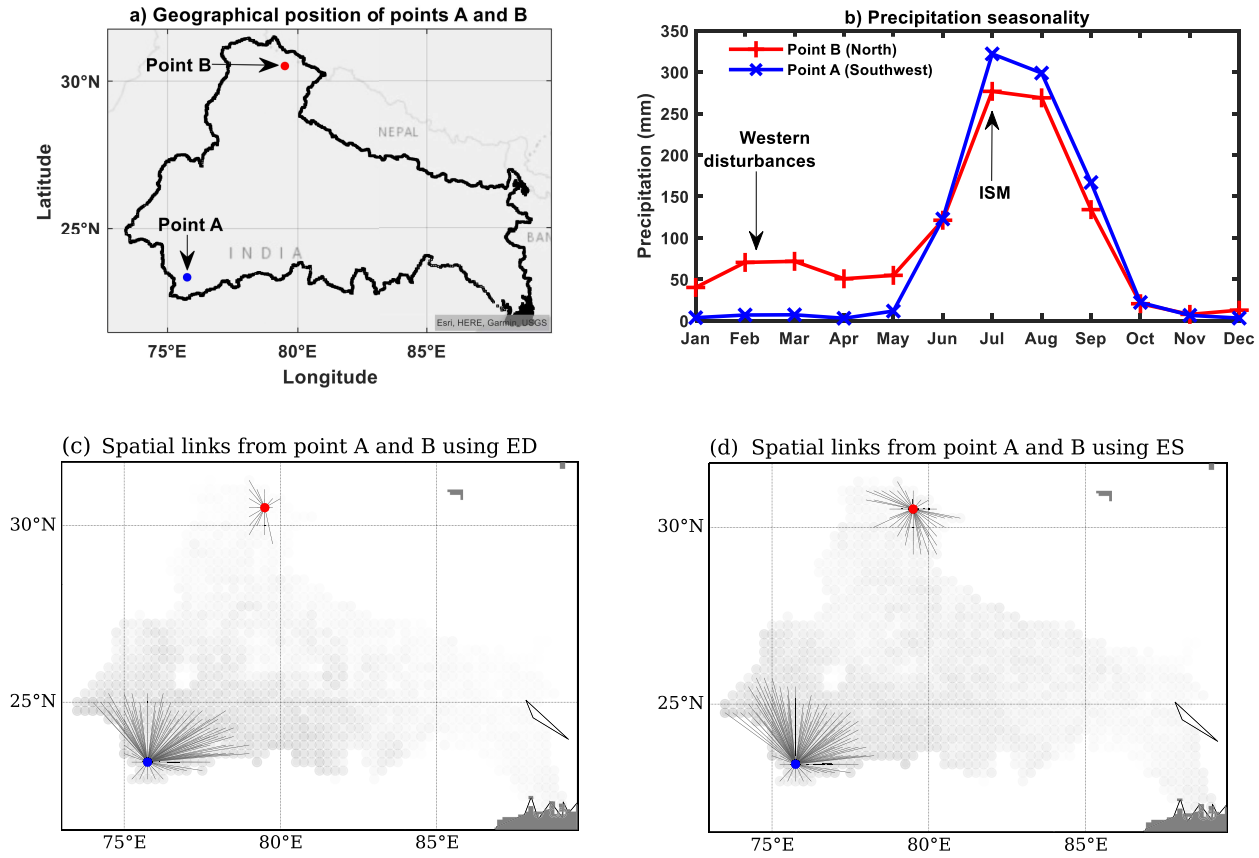

FIG. 10. (a) Map of GRB showing points in the southwest and north and (b) line plot illustrating precipitation seasonality and their connections found using ED (c) and ES (d).

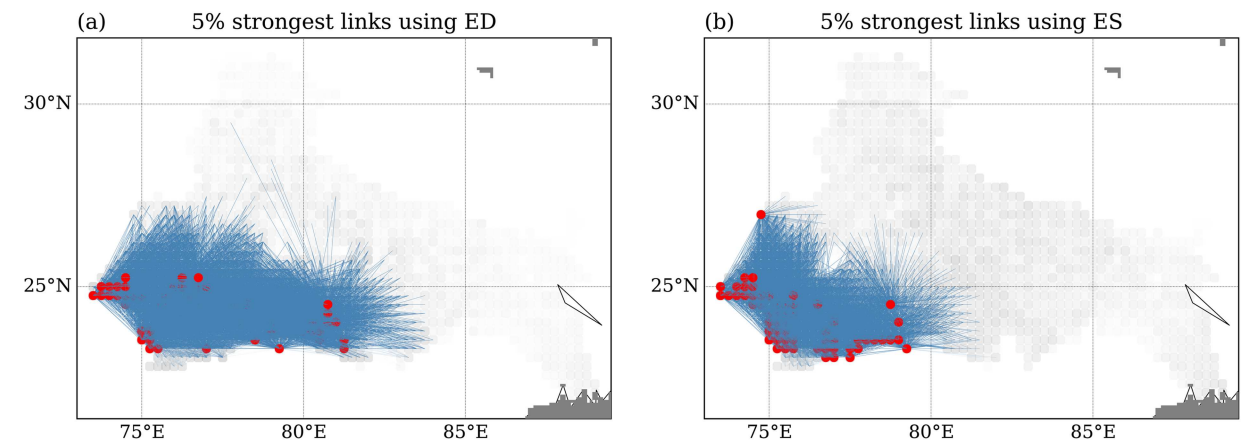

FIG. 11. Spatial distribution of the strongest 5% of nodes obtained over the GRB using (a) ED and (b) ES. All 1174 grid stations are plotted in the background in light gray color. Red colored nodes indicate the top 5% degree nodes.

V. CONCLUSION

In this study, 1174 daily precipitation records during 1998–2019 in the Ganga river basin are studied by their EPE network characteristics. Two nonlinear methods, event synchronization and edit distance, are used to compare the precipitation events. We juxtapose the network topology resulting from the edit distance-based network with the event synchronization-based network. The network measure node degree found that the degree is decreasing from the southwest to the northwest direction. As it is known that the extreme precipitation patterns are influenced by complicated topography, we quantify the variability of EPEs by investigating the impact of topography on the EPEs across the basin. We find that the connectivity of events is greatly affected by elevation. Our analysis revealed an inverse relationship between elevation and the timing of 50th percentile precipitation. The network-based degree suggests that the event-like time series at the grid stations located at the southwest region of the basin are more connected compared to the rest of the basin. When an extreme event occurs across sub-basins under this area, it contributes to a higher flooding risk.

Furthermore, as a similarity measure, edit distance includes amplitude variations in the pairwise tests. In our work, the degree pattern using ED indicates the node of high similarity in terms of their amplitude and occurrence. Precipitation extreme events with almost coinciding events but significant differences in amplitudes will be considered very similar by event synchronization but more different by edit distance. More minor synchronous extreme events with very similar amplitude variation can, in contrast, be considered different by event synchronization but more similar by edit distance. One of the primary purposes of this research is to understand the spatial distribution of EPEs in a specific area. Although we illustrated the preliminary results about the spatial connectivity of EPEs, there is still a need to understand the propagation of EPEs at different spatial and temporal scales. Also, a statistical significance testing is required for more credible network interpretation. We recognize that the application of network theory in hydrological systems is still in its infancy. We believe that current research on applying complex networks to analyze the extreme precipitation pattern will improve our understanding of the spread of extreme events in the GRB.

ACKNOWLEDGMENTS

A.A., R.K.G., and M.A.G. acknowledge the joint funding support from the University Grant Commission (UGC) and DAAD under the framework of the Indo-German Partnership in Higher Education (No. IGP2020-24/GREKO) at the IIT Roorkee. R.K.G. acknowledges the financial support received from the Prime Minister's Research Fellowship (No. PM-31-22-695-414) and DAAD Research Grant (ID:57552338/Ref. No. 91800544). The research of A.B. and N.M. was supported by the Deutsche Forschungsgemeinschaft (DFG) within graduate research training group 2043/1 "Natural risk in a changing world (NatRiskChange)" at the University of Potsdam.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The dataset was developed by the Indian Meteorological Department (IMD) and can be retrieved from the IMD website (last accessed 1 September 2021), https://imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html.

REFERENCES

[1] A. Kalyan, D. K. Ghose, R. Thalagapu, R. K. Guntu, A. Agarwal, J. Kurths, and M. Rathinasamy, Atmosphere **12**, 480 (2021).
[2] R. K. Guntu and A. Agarwal, Sci. Rep. **11**, 16447 (2021).
[3] M. A. Gadhawe, R. K. Guntu, and A. Agarwal, Eur. Phys. J. Spec. Top. **230**, 3343–3357 (2021).
[4] J. Hlinka, D. Hartman, M. Vejmelka, D. Novotná, and M. Paluš, Clim. Dyn. **42**, 1873 (2014).
[5] N. Marwan, M. Carmenromano, M. Thiel, and J. Kurths, Phys. Rep. **438**, 237 (2007).
[6] A. Banerjee, B. Goswami, Y. Hirata, D. Eroglu, B. Merz, J. Kurths, and N. Marwan, Nonlin. Processes Geophys. **28**, 213–229 (2021).
[7] R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, New J. Phys. **12**, 033025 (2010).
[8] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, Phys. Rev. Lett. **78**, 4193 (1997).
[9] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, Phys. Rev. E **54**, 6708 (1996).
[10] S. Pincus, Chaos **5**, 110 (1995).
[11] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, Physica D **134**, 419 (1999).
[12] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, Eur. Phys. J. Spec. Top. **174**, 157 (2009).
[13] J. Runge, J. Heitzig, N. Marwan, and J. Kurths, Phys. Rev. E **86**, 061121 (2012).
[14] N. Malik, B. Bookhagen, N. Marwan, and J. Kurths, Clim. Dyn. **39**, 971 (2012).
[15] N. Boers, B. Bookhagen, N. Marwan, J. Kurths, and J. Marengo, Geophys. Res. Lett. **40**, 4386, https://doi.org/10.1002/grl.50681 (2013).
[16] A. Agarwal, N. Marwan, M. Rathinasamy, B. Merz, and J. Kurths, Nonlinear Processes Geophys. **24**, 599 (2017).
[17] J. Kurths, A. Agarwal, R. Shukla, N. Marwan, M. Rathinasamy, L. Caesar, R. Krishnan, and B. Merz, Nonlin. Processes Geophys. **26**, 251–266 (2019).
[18] Y. Moreno and M. Perc, New J. Phys. **22**, 010201 (2019).
[19] J. D. Victor and K. P. Purpura, Netw. Comput. Neural Syst. **8**, 127 (1997).
[20] Y. Hirata and K. Aihara, J. Neurosci. Methods **183**, 277 (2009).
[21] S. Suzuki, Y. Hirata, and K. Aihara, Int. J. Bifurcation Chaos **20**, 3699 (2010).
[22] D. Eroglu, F. H. McRobie, I. Ozken, T. Stemler, K.-H. Wyrwoll, S. F. M. Breitenbach, N. Marwan, and J. Kurths, Nat. Commun. **7**, 12929 (2016).
[23] Y. Hirata, and K. Aihara, Chaos **25**, 123117 (2015).
[24] B. C. Yadav, R. J. Thayyen, and K. Jain, Int. J. Climatol. **40**, 6002 (2020).
[25] R. J. Thayyen, A. P. Dimri, P. Kumar, and G. Agnihotri, Nat. Hazards **65**, 2175 (2013).
[26] A. P. Barros, G. Kim, E. Williams, and S. W. Nesbitt, Nat. Hazards Earth Syst. Sci. **4**, 29 (2004).
[27] R. A. Houze, Rev. Geophys. **50**, RG1001, https://doi.org/10.1029/2011RG000365 (2012).
[28] E. Palazzi, J. von Hardenberg, and A. Provenzale, J. Geophys. Res. Atmos. **118**, 85, https://doi.org/10.1029/2012JD018697 (2013).
[29] H. Wulf, B. Bookhagen, and D. Scherler, Geomorphology **118**, 13 (2010).
[30] V. Bharti, C. Singh, J. Ettema, and T. A. R. Turkington, Int. J. Climatol. **36**, 3949 (2016).
[31] D. Shepard, in *Proceedings of the 1968 23rd ACM National Conference* (ACM Press, 1968), pp. 517–524.
[32] D. Pai, M. Rajeevan, O. Sreejith, B. Mukhopadhyay, and N. Satbha, Mausam **65**, 1 (2014).
[33] See https://imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html for retrieving gridded rainfall dataset over India.
[34] M. Rathinasamy, A. Agarwal, B. Sivakumar, N. Marwan, and J. Kurths, Stochastic Environ. Res. Risk Assess. **33**, 2053 (2019).

[35]R. K. Guntu, R. Maheswaran, A. Agarwal, and V. P. Singh, J. Hydrol. **590**, 125236 (2020).
[36]R. K. Guntu, P. K. Yeditha, M. Rathinasamy, M. Perc, N. Marwan, J. Kurths, and A. Agarwal, Chaos **30**, 033117 (2020).
[37]R. K. Guntu, M. Rathinasamy, A. Agarwal, and B. Sivakumar, J. Hydrol. **587**, 124916 (2020).
[38]R. Quian Quiroga, T. Kreuz, and P. Grassberger, Phys. Rev. E **66**, 041904 (2002).
[39]I. Ozken, D. Eroglu, T. Stemler, N. Marwan, G. B. Bagci, and J. Kurths, Phys. Rev. E **91**, 062911 (2015).
[40]Y. Shimada, K. Yamamoto, and T. Ikeguchi, Chaos **31**, 013122 (2021).
[41]I. Ozken, D. Eroglu, S. F. M. Breitenbach, N. Marwan, L. Tan, U. Tirnakli, and J. Kurths, Phys. Rev. E **98**, 052215 (2018).
[42]N. Boers, B. Goswami, A. Rheinwalt, B. Bookhagen, B. Hoskins, and J. Kurths, Nature **566**, 373 (2019).
[43]S. Mondal, A. K. Mishra, and L. R. Leung, Geophys. Res. Lett. **47**, e2020GL088185, (2020).
[44]T. Kreuz, F. Mormann, R. G. Andrzejak, A. Kraskov, K. Lehnertz, and P. Grassberger, Physica D **225**, 29 (2007).
[45]M. Paluš, V. Komárek, Z. Hrnčíř, and K. Štěrbová, Phys. Rev. E **63**, 046211 (2001).
[46]R. Q. Quiroga, J. Arnhold, and P. Grassberger, Phys. Rev. E **61**, 5142 (2000).
[47]M. C. Romano, M. Thiel, J. Kurths, and C. Grebogi, Phys. Rev. E **76**, 036211 (2007).
[48]A. Odenweller and R. V. Donner, Phys. Rev. E **101**, 052213 (2020).
[49]R. Killick, P. Fearnhead, and I. A. Eckley, J. Am. Stat. Assoc. **107**, 1590 (2012).
[50]A. Kraskov, H. Stögbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).
[51]M. Vejmelka and M. Paluš, Phys. Rev. E **77**, 026214 (2008).
[52]F. Wolf, A. Voigt, and R. V. Donner, Earth Syst. Dyn. **12**, 353 (2021).
[53]V. Stolbova, P. Martin, B. Bookhagen, N. Marwan, and J. Kurths, Nonlinear Processes Geophys. **21**, 901 (2014).
[54]X. Feng, A. Porporato, and I. Rodriguez-Iturbe, Nat. Clim. Change **3**, 811 (2013).
[55]S. Pascale, V. Lucarini, X. Feng, A. Porporato, and S. ul. Hasson, Clim. Dyn. **44**, 3281 (2015).
[56]A. A. Tsonis, K. L. Swanson, and P. J. Roebber, Bull. Am. Meteorol. Soc. **87**, 585 (2006).
[57]A. Rheinwalt, N. Marwan, J. Kurths, P. Werner, and F.-W. Gerstengarbe, Europhys. Lett. **100**, 28002 (2012).
[58]S. Raj, R. Shukla, R. M. Trigo, B. Merz, M. Rathinasamy, A. M. Ramos, and A. Agarwal, Int. J. Climatol. **41**, 6602 (2021).
[59]A. Agarwal, N. Marwan, R. Maheswaran, U. Ozturk, J. Kurths, and B. Merz, Hydrol. Earth Syst. Sci. **24**, 2235 (2020).
[60]A. Agarwal, N. Marwan, R. Maheswaran, B. Merz, and J. Kurths, J. Hydrol. **563**, 802 (2018).