Originally published as:

Agarwal, A., Marwan, N., Maheswaran, R., Merz, B., Kurths, J. (2018): Quantifying the roles of single stations within homogeneous regions using complex network analysis. - *Journal of Hydrology*, *563*, pp. 802—810.

DOI: http://doi.org/10.1016/j.jhydrol.2018.06.050

Quantifying the roles of single stations within homogeneous regions using complex network analysis

A. Agarwal[1, 2, 3], N. Marwan[2], R. Maheswaran[2], B. Merz[1, 3], J Kurths[1, 2]

[1]Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany

[2]RDIV, Potsdam Institute for Climate Impact Research, Telegrafenberg, Potsdam, Germany

[3]GFZ German Research Centre for Geosciences, Section 5.4: Hydrology, Telegrafenberg, Potsdam, Germany

Correspondence to: A. Agarwal (aagarwal@uni-potsdam.de)

Abstract

Regionalization and pooling stations to form homogeneous regions or communities are essential for reliable parameter transfer, prediction in ungauged basins, and estimation of missing information. Over the years, several clustering methods have been proposed for regional analysis. Most of these methods are able to quantify the study region in terms of homogeneity but fail to provide microscopic information about the interaction between communities, as well as about each station within the communities. We propose a complex network-based approach to extract this valuable information and demonstrate the potential of our approach using a rainfall network constructed from the Indian gridded daily precipitation data. The communities were identified using the network-theoretical community detection algorithm for maximizing the modularity. Further, the grid points (nodes) were classified into universal roles according to their pattern of within- and between-community connections. The method thus yields zoomed-in details of individual rainfall grids within each community.

Keywords: Complex network, event synchronization, rainfall network, Z-P approach

1. Introduction

Reliable and accurate information about precipitation is essential for most hydrological studies. For example, precipitation observations are required for the design of hydraulic structures, flood estimation and forecasting, assessment of water availability, or climate impact studies. However, in most situations, raingauges are scarce, requiring knowledge about how precipitation characteristics at neighboring stations are related. These interrelationships can be viewed in a statistical sense (e.g. by applying correlation analysis), in a physical sense (as in dynamical meteorology), or in a topological sense (as in complex network analysis). Knowledge of these interrelationships will be crucial for various purposes, including (1) applying interpolation/extrapolation techniques to generate rainfall at locations where raingauge measurements are not available (Yang et al., 2015), (2) filling gaps in historical rainfall records using available rainfall observations at neighboring stations (Jha et al., 2015), (3) determining the optimal density and locations for the installation of new raingauges (Mishra and Coulibaly, 2009; Pardo-Igúzquiza, 1998), and (4) analysing regional flood frequency (Hassan and Ping, 2012; Smith et al., 2015; Zrinji and Burn, 1994, 1996).

Even though there is a plethora of methods available for identifying homogeneous regions, such as clustering algorithms (Agarwal et al., 2016; Hsu and Li, 2010), principal component analysis (Darand and Mansouri Daneshvar, 2014), region-of-influence approach (Zrinji and Burn, 1994, 1996), or multiple regression (Sivakumar et al., 2015), there are some important challenges which need to be addressed.

(i) A common assumption in studies (Razavi and Coulibaly, 2013; Salinas et al., 2013) dealing with interpolation/extrapolation, missing values and prediction in ungauged basins (PUB) is that the variables of interest, such as precipitation characteristics, at nearby points are more closely related than those at distant points, as described by

(Tobler, 1970) in his ‘First Law of Geography’. This assumption is also the foundation of geostatistics, which in turn is fundamental to many classical approaches to spatial data analysis and interpolation throughout hydrology and other geoscientific disciplines. While this assumption is often reasonable, it may not hold in every situation, especially in regions with complex topography (Jha et al., 2015). In such areas, statistics of rainfall recorded at neighboring stations can significantly vary due to the high topographic gradients and, hence, changes in rainfall patterns between them (Berndtsson, 1988; Li et al., 2014; Niu, 2013; Özger et al., 2010).

(ii) A significant disadvantage of these methods is that the selection of factors for identifying the similarity in rainfall patterns is highly subjective. They rely on the preconceived notion of the existence of linear relationship between the factors that influence the precipitation in a region. For instance, in PCA method the subjectivity is introduced in terms of extraction method, rotation method, number of compoeenents to be retined etc. For more details refer to Saxena et al., 2017.

(iii) More importantly, the traditional methods for pooling stations within homogeneous regions are not capable of unraveling the role of each raingauge station within the community. This includes the interactions within the community, the role of the stations, and the strength and number of inter- and intra-community connections.

The main aim of this paper is to address this last point by proposing a network-based approach for unravelling the role of each node in a community. This microscopic analysis is essential to understand the role of each of the member stations of the community and is very useful in many applications. For example, by knowing the connections and their strength, it is possible to reduce the uncertainty of predictions at ungauged locations by including only those stations that have strong connections in that community. Similarly, the reliability of filling gaps in observational time series can be improved by identifying the stations that share strong connections with that particular station. The relative importance of the stations in the community will also help in understanding the connection between the communities and is particularly useful for selecting stations that share characteristics with more than one community.

In the context of connections within rainfall systems, recent developments in network theory, especially regarding complex networks, have been found useful for identifying the spatial connections in rainfall (Malik et al., 2012). Steinhaeuser et al. (2010) explored the utility of complex networks to analyze climate data, i.e., air temperature, pressure, relative humidity and perceptible water. They used the WalkTrap community detection algorithm to identify communities. They concluded that these communities have a climatological interpretation and that alterations in community structure can be an indicator of climatic events. Tsonis et al. (2011) applied complex networks and modularity based community detection to observed and simulated model data and concluded that the complexity of the system condenses into small interacting components called communities. This approach provided information about the nature of different climate subsystems. Jha et al. (2015) demonstrated the use of the clustering coefficient, a complex network based measure (Stolbova et al., 2014), on two rainfall networks in Australia. They attempted to relate the strength of spatial connections in rainfall to topographic and rainfall properties, towards identifying dominant factors governing spatial connections and for offering a better physical interpretation on spatial rainfall variability. Eustace et al. (2015) identified community structures by proposing local community neighborhoods ratio algorithm and showed that the algorithm detects well-defined communities in networks by a wide margin. Conticello et al. (2017) applied the Louvain community detection algorithm to identify clusters of rainfall stations using the concept of event synchronization and Self Organizing Maps. Even though the study of Halverson and Fleming (2015) on streamflow regionalization is not directly relevant for rainfall, it showed that the choice of the community detection algorithm does not strongly impact the community structure.

All above-mentioned studies have used complex network based community detection algorithm to identify homogenous regions but little attention has been paid to the different characteristics or roles of each of the member stations of a community. Although Halverson and Fleming, 2015 have identified the high priority stations, based on high betweenness centrality values, but have not discussed the role of other stations. This study shows that the microscopic analysis of homogeneous regions provides additional insights into the behavior and dynamics of single stations within the homogeneous region, which can be vital for many engineering and water management purposes.

This study builds on emerging ideas in the very fast-evolving field of complex network theory and contributes to work in hydro-monitoring system design. Although studies in different fields, such as physics (Quian Quiroga et al., 2002b; Quiroga et al., 2000) or neurology (Rubinov and Sporns, 2010; Zhou et al., 2007), have seen immense use of complex network theory, event synchronization, and Z-P space, our study is the first combined application of these methods in hydrology to date. It clearly demonstrates the large potential of these methods in hydrology.

As advancement to the research in the application of complex networks in rainfall network analysis, we use a network-based measure to provide a comprehensive analysis of the stations in a community and their roles. For this, we apply the concept of cartographic representation of networks by Guimerà and Amaral (2005). The proposed approach is demonstrated using the synthetic network and then applied to the Indian Precipitation gridded precipitation dataset. The paper is organized in the following manner. Section 2 describes the basic aspects of network construction, and network measurement and Section 3 briefly discusses the methods used in the study. The application of the methodology and the subsequent results obtained are discussed in detail in Section 4. The conclusions are reported in Section 5.

2. Methods

2.1 Network definition

A network or a graph is a collection of entities (nodes, vertices) interconnected by lines (links, edges) as shown in Fig. 1. These entities could be anything from humans defining social networks (Arenas et al., 2008), computers in web networks (Zlatić et al., 2006), neurons of the brain (Pfurtscheller and Lopes da Silva, 1999; Zhou et al., 2007), streamflow stations defining hydrological networks (Halverson and Fleming, 2015; Sivakumar and Woldemeskel, 2014) to raingauge stations defining climate networks (Stolbova et al., 2014; Malik et al., 2012; Rheinwalt et al., 2016).

Formally, a network or graph is defined as an ordered pair $G = (N, E)$, containing a set of nodes $N = \{N_1, N_2, \dots . N_N\}$ together with a set E of edges $\{i, j\}$ which are 2-element subsets of N. In this work, we consider undirected and unweighted graph (G), where only one edge can exist between a pair of nodes and self-loops of the type $\{i, i\}$ are not allowed. Hence, edges simply show connections between nodes, and each edge can be traversed in either direction. This type of graph can be represented by the symmetrical adjacency matrix (Stolbova et al., 2014)

$$A_{i,j} = \begin{cases} 0 & \{i,j\} \quad \notin \quad E \\ 1 & \{i,j\} \quad \in \quad E \end{cases} \tag{1}$$

Figure 1 is a simple example of an undirected and unweighted network. In general, large graphs with non-trivial topological characteristics, used to represent real systems, are called complex networks. To define whether a link between two nodes exists, any similarity measure can be used, such as correlation (Donges et al., 2009; Jha et al., 2015), synchronization (Conticello et al., 2017; Malik et al., 2012; Stolbova et al., 2016) or mutual information (Paluš, 2018). Depending on the topological structure of the network, groups of nodes can be pooled together forming communities (Jha et al., 2015).

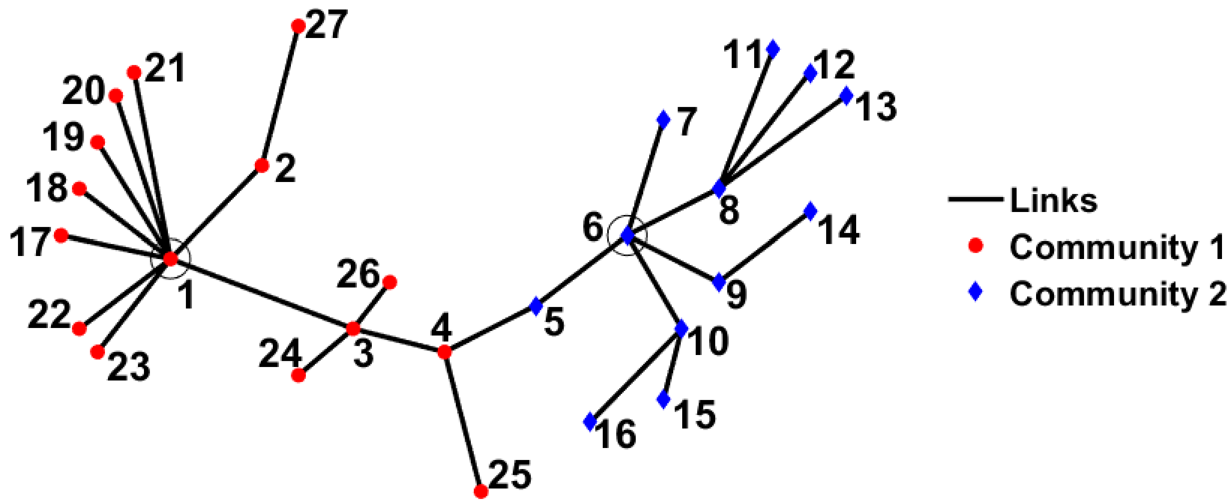

Figure 1. The topology of the sample network used to explain the network construction and universal role of a node. Different colors represent different communities, i.e., community 1 (red) and community 2 (blue). Nodes 4 and5 are the hybrid nodes connecting their community to the other community. Nodes 1 and 6 are the hubs of their respective community.

2.2 Event synchronization

We use event synchronization (Stolbova et al., 2014) to define whether a link between two nodes exists. Event synchronization (ES) has been specifically designed to calculate nonlinear relations between timeseries with events defined on them. A simple algorithm proposed by (Quian Quiroga et al., 2002a) can be used for any time series for which we can define events, such as single-neuron recordings, epileptiform spikes in electroencephalograms (EEG), heartbeats, stock market crashes, or rainfall events. When dealing with signals of a different character, the events could be defined differently in each time series, since their common cause might manifest itself differently in different time series. ES has advantages over other time-delayed correlation techniques (e.g., Pearson lag correlation), as it uses a dynamic (not fixed) time delay (Agarwal et al., 2018, 2017). The latter refers to a time delay that is adjusted according to the two time series being compared, which allows its application to different situations. Another advantage of ES is that it can be applied to non-Gaussian data (Stolbova et al., 2014; Tass et al., 1998). Having its roots in neuroscience, ES only considers events beyond a threshold and ignores the absolute magnitude of events, which could be a challenge to incorporate in future, work.

A number of modifications have been proposed to the basic algorithm, considering various issues such as boundary effects or bias toward the number of events (Agarwal et al., 2017; Rheinwalt et al., 2016). The modified algorithm proposed by (Rheinwalt et al., 2016) can be explained as follows: An event above a threshold α percentile occurs in the signals $x(t)$and $y(t)$ at timest_l^x and t_m^y where $l = 1,2,3,4 \ldots S_x$, $m = 1,2,3,4 \ldots S_y$ and within a time lag $\pm\tau_{lm}^{xy}$ which is defined as (Stolbova et al., 2014)

$$\tau_{lm}^{xy} = min\{t_{l+1}^x - t_l^x, t_l^x - t_{l-1}^x, t_{m+1}^y - t_m^y, t_m^y - t_{m-1}^y\}/2 \qquad \textbf{(2)}$$

where S_x and S_y are the total number of events (greater than the threshold α) in the signals $x(t)$ and $y(t)$, respectively. This definition of the time lag helps to separate independent events, which in turn allows to take into account the fact that different processes are responsible for the generation of events. To count the number of times an event occurs in $x(t)$ after it appears in $y(t)$ and vice versa, $C(x|y)$and $C(y|x)$ are defined as follows:

$$C(x|y) = \sum_{l=1}^{S_x} \sum_{m=1}^{S_y} J_{xy} \qquad \textbf{(3)}$$

And

$$J_{xy} = \begin{cases} 1 & if \;\; 0 < t_l^x - t_m^y < \tau_{lm}^{xy} \\ \frac{1}{2} & if \qquad\qquad t_l^x = t_m^y \\ 0 & else, \end{cases} \tag{4}$$

$C(y|x)$ is defined accordingly, and from these quantities we obtain:

$$Q_{xy} = \frac{C(x|y) + C(y|x)}{\sqrt{(S_x - 2)(S_y - 2)}} \tag{5}$$

Q_{xy} is a measure of the strength of the event synchronization between $x(t)$ and $y(t)$. It is normalized to $0 \leq Q_{xy} \leq 1$. This implies that$Q_{xy} = 1$ for perfect synchronization between $x(t)$ and $y(t)$.

2.3 Network construction

To construct a rainfall network, each grid cell is considered as a node and links between each pair of nodes are setup based on the statistical relationship between them. The similarity measure used is the ES which gives a Q matrix (Eq.5). Applying a certain threshold (θ) on the Q matrix (Eq.5), we yield an adjacency matrix (rewriting Eq. 1)

$$A_{i,j} = \begin{cases} 1, & if \; Q_{i,j} \; \geq \; \theta_{i,j}^{Q} \\ 0, & \qquad else, \end{cases} \tag{6}$$

Here, $\theta_{i,j}^{Q} = 95^{th}\; percentile$ is a chosen threshold, and $A_{i,j} = 1$ denotes a link between the i^{th} and j^{th}nodes and 0 denotes otherwise. The adjacency matrix represents the connections in the rainfall network. In this study, we use an undirected network, meaning we do not consider which of the two synchronized events happened first, in order to avoid the possibility of misleading directionalities of event occurrences between nodes that are topographically close to one another.

2.4 Network measures

To analyze and quantify the topological features of complex networks, a large number of network measures have been introduced (Blondel et al., 2008; Malik et al., 2016). We use the *within-module degree Z-score* (Z) and the *participation coefficient* (P) (Guimerà and Amaral, 2005) to investigate the role of individual nodes within a community. Z identifies hubs and non-hubs within the community. Hubs are nodes with a significantly larger number of links compared to the other nodes in the network. P is a measure of the diversity of the connections between individual nodes and identifies to which extent a node has intra-community or inter-community links.

The within-module degree ($\boldsymbol{Z_i}$ or Z-score) is a within-community version of degree centrality (total number of link of any node) and shows how well a node is connected to other nodes in the same community. It is estimated as (Guimer and Amaral, 2005)

$$Z_i = \frac{K_i - \overline{K_{s_i}}}{\sigma_{k_{s_i}}} \tag{7}$$

where $\boldsymbol{K_i}$ is the total number of links (degree) of node $\boldsymbol{i}$ in the community$\boldsymbol{s_i}$, $\overline{\boldsymbol{K_{s_i}}}$ is the average degree of all nodes in the community $\boldsymbol{s_i}$, and $\boldsymbol{\sigma_{k_{s_i}}}$ is the standard deviation of $\boldsymbol{K}$ in $\boldsymbol{s_i}$. Since two nodes having the same *Z*-score may play different roles within the community, this measure is often combined with the participation coefficient $\boldsymbol{P_i}$.

The participation coefficient ($\boldsymbol{P_i}$) compares the number of links of node i to nodes in all communities with the number of links within its own community. We define the $\boldsymbol{P_i}$ of node $\boldsymbol{i}$ as (Guimer and Amaral, 2005)

$$P_i = 1 - \sum_{s_j=1}^{N_M} \left(\frac{k_{is_j}}{k_i}\right)^2 \qquad (8)$$

where k_{is_j} is the number of links of node i to nodes in community s_j, and k_i is the total number of links (degree) of node i. N_M represent the number of communities in the network. The participation coefficient of a node is therefore close to one if its links are uniformly distributed among all the communities, and zero if its entire links are within its own community because in later case $\mathrm{K}_{\mathrm{is_j}} = \mathrm{K_i}$ hence $\mathrm{P_i} = 0$.

2.5 Community detection

Complex networks often show subsets of nodes that are densely interconnected. These subsets are called communities. The community structure of a complex network provides insight into the network (Girvan and Newman, 2002). For instance, different communities within a network may have very different properties compared to the averaged properties of the complete network.

There exist several community detection approaches aiming at stratifying the nodes into communities in an optimal way (see (Fortunato, 2010) for an extensive review). The question which community detection algorithm should be used is difficult to answer. However, it has been found that the choice of the community detection algorithm has a small impact on the resultant communities in geophysical data science studies (Halverson and Fleming, 2015). In this study, we use the Louvain method which maximizes the modularity to find the optimal community structure in the network. The optimal community structure is a subdivision of the network into non-overlapping groups of nodes, which maximizes the number of within-group edges and minimizes the number of between-group edges (Blondel et al., 2008; Rubinov and Sporns, 2011).

Modularity is defined, besides a multiplicative constant, as the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random. Positive modularity values suggest the presence of communities. Thus, one can search for community structures by looking for the network divisions that have positive, and preferably large, modularity values (Newman, 2004). Modularity (*M*) is calculated as:

$$M = \frac{1}{2m}\sum_{i,j}\left\lfloor A_{ij} - \frac{k_i k_j}{2m}\right\rfloor \delta(C_i C_j) \qquad (9)$$

where A_{ij} represents the number of edges between i and $j, k_i = \sum_j A_{ij}$ is the sum of the number of the edges (degree) attached to vertex i, C_i is the community to which vertex i is assigned, the $\delta -$ function $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise, and $m = {}^{1}/_{2} \sum_{ij} A_{ij}$.

Equation (9) is solved using the two-step iterative algorithm proposed by Blondel et al. (2008), also known as the Louvain method. The first step consists in optimizing the modularity by permitting only a local modification of communities; in the second step, the communities identified are pooled to assemble a new network of communities. High modularity networks are densely linked within communities but sparsely linked between communities. The algorithm stops when the highest modularity is achieved. The algorithm was implemented using the Brain

Connectivity Toolbox (BCT), provided by (Rubinov and Sporns, 2010), and is available at https://sites.google.com/site/bctnet/.

2.6 Z-P space approach

Following the approach proposed by Guimerà and Amaral (2005), we calculate for each node the participation coefficient P_i and the within-module degree Z_i, and plot all nodes onto the Z-P space. Both measures are calculated once the network communities have been determined (Guimerà and Amaral, 2005; Guimera et al., 2007). Guimerà et al. (2007) propose to divide the $Z - P$ space into seven classes (R1 – R7) which express the different roles of the nodes (Table 1). In the first step, the nodes are broadly categorized as hubs or non-hubs using the within-module degree (Z). Nodes with $Z \geq 2.5$ are classified as community hubs and nodes with $Z < 2.5$ as non-hubs. At the second level, the hub and non-hub nodes are further characterized using the participation coefficient. Hence, each node is assigned to one of these seven classes.

Table 1.Definition and interpretation of R classes according to Guimerà et al. (2007), defining the role of each node.

R-Class	Z	P	Remarks	Characteristics of R class
R1	<2.5	≈ 0	ultra-peripheral nodes, i.e.,nodes with almost all their links within their community ($P \approx 0$).	representative nodes (almost all intramodular links)
R2	<2.5	$0 < P \leq 0.625$	peripheral nodes, i.e.,node has at least 60% its links within the community.	node has more intramodular links than intermodular links
R3	<2.5	$0.625 < P \leq 0.80$	satellite connectors, i.e.,nodes have half of its connection outside the community.	node has more intermodular links than intramodular links
R4	<2.5	$P > 0.80$	kinless nodes, i.e., nodes with a maximum of links (>70%)outside the community.	wrongly assigned nodes

R5	>2.5	P ≤ 0.30	provincial hubs,i.e.,hubs with the vast majority of links within their community.	local centers, representative nodes if $P \approx 0$
R6	>2.5	0.30 < P ≤ 0.75	connector hubs,i.e., hubs with atleast half of its links toother community.	hybrid nodes (connecting two different communities)
R7	>2.5	P > 0.75	global hubs,i.e.,hubs with links homogeneously distributed among all community.	global connector nodes hence cannot be assigned to the single community.

Nodes in the classes R1 and R5 with $P \approx 0$ have almost all links within the own community. Since class R5 have provincial hubs (Table 1) which contain both intracommunity and intercommunity links, the limit on the participation coefficient ($P \approx 0$) helps to identify nodes that have almost all intracommunity links. These nodes with almost all intracommunity links ($P \approx 0$) are local centers in the region and can only be selected as a representative node of the community (Halverson and Fleming, 2015).

Nodes in the classes R2 and R3 are peripheral and satellite connectors respectively (Table 1). Both the class contains hybrid non-hub nodes which generally connect two different communities. The only difference between R2 and R3 is that R3 nodes have more intercommunity links (outside of its own community).

Similarly, R6 nodes represent the nodes that have many intercommunity links but are hubs. In the given community we interpret them as hybrid hubs which have a maximum connection outside of its own community. Kinless nodes (R4) have the greatest proportion of links outside the community and are interpreted as wrongly assigned nodes in the community. If there exist many R4 nodes in the community a reformation of the communities or reallocation of such nodes to appropriate community is suggested. The nodes in class R7 maintain homogeneous links with all the communities. We surmise that such nodes may not be clearly associated with a single community hence termed as the global hubs or global connectors (nodes connecting many different climate sub-systems).

The above characterization of nodes is important as it helps in understanding their specific roles in terms of non-hubs, hubs, local centers, hybrid nodes, global hubs. In the context of climate systems, local centers correspond to nodes which are important for local climate phenomena, while bridges correspond to nodes which connect different subsystem of climatology, leading to non-local interaction (teleconnections).

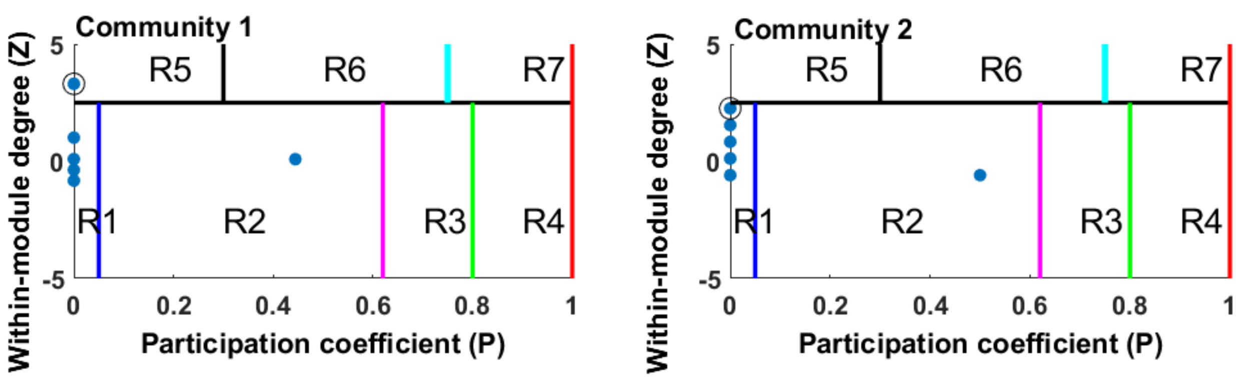

Figure 2. Nodes of the sample network of Figure 1 plotted onto the Z-P-space. Nodes 1 and 6 (both encircled) are the representative stations for community 1 and 2, respectively. Nodes 4 and 5 in community 1 and 2, respectively, are the only hybrid nodes and are thus in the R2 class. All other nodes have only intracommunity links and are assigned to the R1 class. Many stations have the same values for Z and P and are thus on top of each other in the R1 class. Nodes 1 and 6 are the local center ($P \approx 0$)and are thus in the R5 and R1 class respectively.

Using the classification of Table 1, Figure 2 shows the $Z - P$ space for the sample network of Figure 1 and the assigned R classes. Node 1 is a hub in community 1, having all of its nodes within the community, and hence can be considered as a representative station. Node 4 of community 1 (non-hub) has intercommunity links and thus falls in the R2 class. For community

2, station 6 is a representative node with all links within the community, and the non-hub node 5 has intercommunity links falling in class R2. There is no kinless node (R4 and R7) in both communities.

If there exists a node fully unsynchronized to the other nodes in the network, i.e. there are no links to other nodes, the proposed $Z - P$ approached will detect this station given its unique characteristcis. This unsynchronized station will lie at the origin of $Z - P$ space and will fall in a community on its own. As an extreme example, one might imagine that in a meteorological sub-region, characterized by fine-scale convective thunderstorms with sparse raingauge coverage, precipitation event synchronization across all raingauges in that sub-region would be poor and each stations would form a separate community.

3. Model application

The method was tested on a gridded rainfall dataset for two reasons: i) the availability and the access to raingauge data is limted, and ii) gridded datasets provide an effective platform to understand the precipitation dynamics. Owing to the assumptions underlying the spatial interpolation, the gridding process used to build the dataset might affect the relationships between nodes. However, these effects can be neglected considering the extent of the study area. The high-resolution (0.25° × 0.25°) daily gridded rainfall data (Pai et al., 2015) was developed by the Indian Meteorological Department (IMD) for a spatial domain of 66.5°E to 100°E and 6.5°N to 38.5°N covering the mainland region of India. The gridded data was generated from the observed data of 6995 gauging stations across India using spatial interpolation for the period 1901-2013. Several studies in the past using the same dataset have reported such as downscaling (Lakhanpal et al., 2017; Sehgal et al., 2016) and rainfall variability (Krishnamurthy and Shukla, 2000). This shows that the data are highly accurate and capable of capturing the spatial distribution of rainfall over the country. In this study, out of total 6995 grid stations, 4631 stations were identified for which continuous rainfall data for 63 years (Jan 1951 to Dec 2013) was available without any missing values.

The rainfall network is constructed (as explained in section 2.3) by extracting an event series from 4631 raingauges (Fig. 3), i.e., by applying a threshold we identify extreme rainfall events in the given time series (Agarwal et al., 2017; Rheinwalt et al., 2015). We define extreme events as precipitation that is greater than the 95th percentile at that station. The 95th percentile is a good compromise between having a sufficient number of events at each location and a rather high threshold to study heavy precipitation. Next, we compute the Q (Eq. 5) between each pair of 4631 rainfall grid points. Applying a threshold ($\theta_{i,j}^{Q} = 95^{th}$ percentile) on the Q matrix (Eq. 5) yields an adjacency matrix (Eq.6), representing the connections in the rainfall network. In this study, we use an undirected network, meaning we do not consider which of the two synchronized events happened first, in order to avoid the possibility of misleading directionalities of event occurrences between rain gauges that are topographically close to one another. After formation of the rainfall network, we aimed to obtain a small set of communities representing relevant sub-processes of the rainfall network. In this study, we apply Louvain algorithm (section 2.5) on the constructed network to unravel the community structure.

The resultant community structure is the rainfall network mapped in Figure 3.

(a) (b)

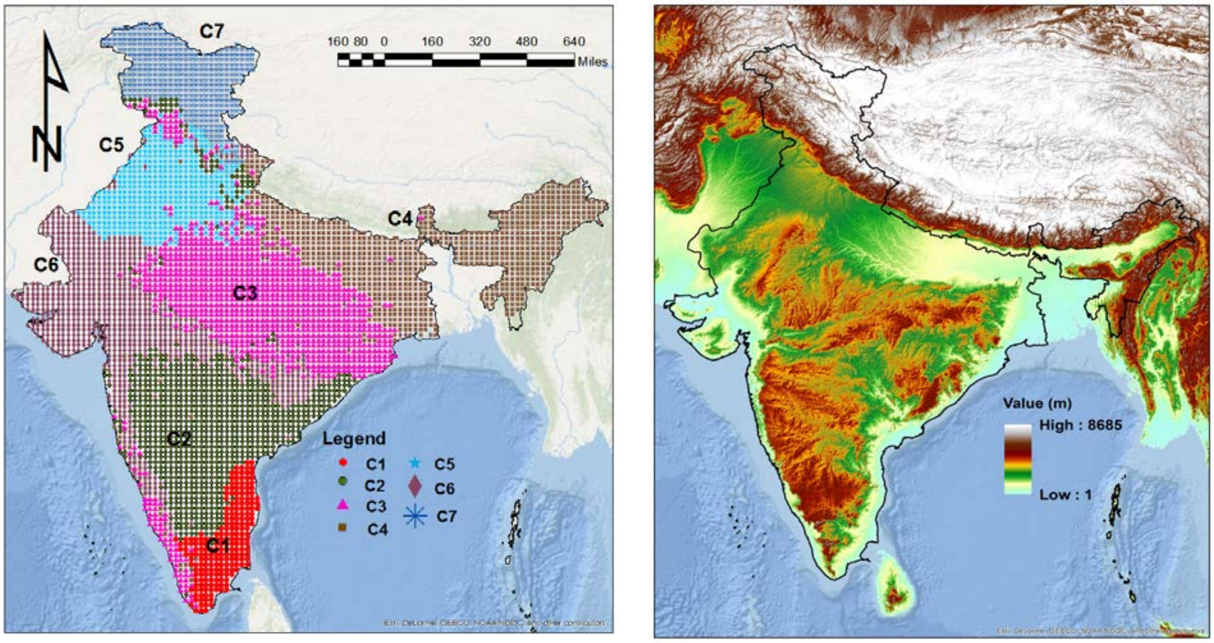

Figure 3. (a) Community structure of precipitation data in the rainfall network resulting from the Louvain algorithm. (b) Elevation map of the Indian continent.

The obtained community structure (Fig.3) shows some similar patterns to those provided by the Indian Institute of Tropical Meteorology (Vinnarasi and Dhanya, 2016) and (Malik et al., 2016). It is also important to emphasize that the formation of the regions using complex networks is based on a cluster of actual connections, rather than on our traditional criteria of geographic proximity, nearest neighbors, regional patterns, and linear correlations.

Table 2 shows the geographical and statistical interpretation of the resultant community which includes the mean, standard deviation, and coefficient of skewness of the precipitation distribution for each community. Higher mean precipitation shows a greater total amount of precipitation, a larger standard deviation shows a stronger variation of data for the collecting period, and a larger coefficient of skewness indicates more extreme (monthly) precipitation events (Hsu and Li, 2010).

Table 2. Summary of geographical and statistical analysis for each individual community. Communities formed by maximizing the modularity using Louvain algorithm. Elevation map for India is presented in the Fig.3b.

C. No.	Number of stations	Monthly mean (mm)	Stand. Deviation (mm)	Skewness	Remarks
1	214	79.70	98.29	2.04	smallest community, eastern coastline, low elevation region, warm, humid climate regime
2	876	76.30	104.45	2.16	mild elevation, semi-arid climate regime (south)
3	1028	105.01	154.69	1.91	moderate elevation, equatorial grassland (south) semi-arid climate regime
4	865	150.89	178.92	1.60	high elevation, subtropical humid climate regime (Himalayan foothills and northeast)
5	433	48.26	79.39	2.71	moderate elevation, semi-arid climate regime (Central India)
6	843	75.50	127.89	2.79	low elevation, northwest and western coastline, arid and warm, humid climate regime (northwest)
7	372	66.26	85.41	2.48	very high elevation, alpine climate regime

Considering statistical properties, community 4 (Fig.3), which covers almost all of the greenest and most mountainous regions of India (northeastern India), has the highest monthly mean (150.89 mm), the largest variation (178.92 mm) and low skewness (1.6) of precipitation in the region (Table 2). Meanwhile, community 5 (Fig.3), covering dry and lowland areas (northwestern India), shows the lowest monthly mean (48.26mm) with lower variation. Community 6 (western coastline) shows the greatest skewness along with high variability. One possible reason for the high variability and skewness could be that these regions are near to both coastlines and are low-lying areas with two different climate regimes (arid and humid). Community 3 (southeastern India) shows a high coefficient of skewness (1.91) and second high monthly rainfall (105.01mm) and variability (154.69mm). All the communities show the positive coefficient of skewness, which indicates precipitation with a long tail toward high values.

Community 7 (mountainous region) shows low monthly precipitation mean, moderate variability and high skewness. In South India, both communities 1 and 2 (Fig. 3) almost have similar rainfall characteristics but are differentiated by topological (elevation, land, coastline) features.

Further, using a node-to-node connection approach (Guimerà et al., 2007; Guimerà and Amaral, 2005) we explore the microscopic details of each individual station within the community. We fit all raingauges of the rainfall network in the ZP space (Fig.4) according to the estimated network measures (Section 2.4) of the within-module degree (Z) and participation coefficient (P).

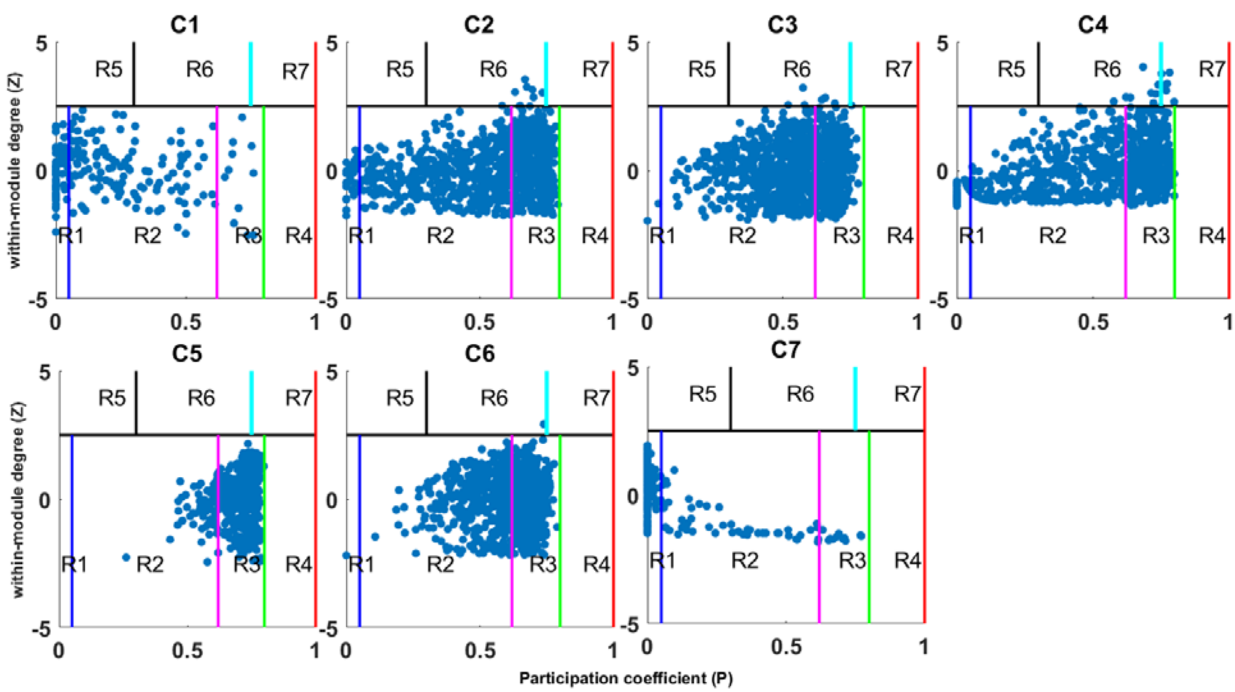

Figure 4 Role-specific representation of node behavior in the $Z - P$ space (Section 2.6) plotted for each community (C1 to C7). Within-module degree (Z) differentiates between hubs and non-hubs and the participation coefficient (P) quantifies the percentage of intra-/inter-community links. Blue colored dots in Z-P space in a particular community represent the raingauge station (node) of that particular community. The significance of each R class is explained in Section 2.6. Many stations have the same values for Z and P and are thus on top of each other in the different R class.

Figure 4 shows the Z-P space plot for each community (C1 to C7) separately. Table 3 shows the percentage of each class of stations in each community. From Fig. 4 and Table 3, we find that none of the communities has a kinless node (R4 class node),i.e., no wrongly assigned node. This explains the robustness of the method (edge-betweenness) used for clustering.

It can be seen that all the communities (C1 to C6) have a dominance of hybrid nodes in their respective community except for community 7, which shows the dominance of nodes with intra-community links. This observation falls along the expected lines, as the Indian sub-continent's precipitation shows the vast variability in topography, climate diversity,etc. The results are quite different from those shown by Agarwal et al., 2017, for German regions. In Germany, the raingauge stations were mostly connected by intra-community links, indicating more homogeneity in the precipitation compared to Indian precipitation.

As explained in Section 2.6, stations with the almost all $(P \approx 0)$ intra-community links can be considered a spatially representative station of the community. We argue that such stations have climatological properties (rainfall time series) that are representative of the other members of their respective communities (Halverson and Fleming 2015).This information has significant importance in big data analysis and uncertainty analysis, as the information from the entire community is available in the form of the representative station.

Further analyzing the Z-P space, we see that the eastern coastline region (C1) to some extent shows good interconnectedness (high number of R1 and R2) and also does not show any hubs (R5 to R7) in the region. This suggests that rainfall in this region is more localized and does not show any long-range connections. This is in congruence with the general understanding that the eastern coastline region is dominated by the northeastern (NE) monsoonal rainfall while the rest of the country receives rainfall from southwestern (SW) monsoons (Jain et al., 2013).

The mild and moderate-elevation inland regions of India (C2, C3, C5,and C6) show negligible intracommunity links (R1) compared to other high-elevation regions (C4 and C7) and low-elevation regions (C1).These mild and moderate-elevation regions (C2, C3, C5, and C6) are strongly dominated by hybrid stations (R3 and R6) sharing some common dynamics with other regions. For instance, C2 (Southeast) and C3 (Central-East) have very few nodes in the R1 class; the majority of nodes fall in R2 and a significant amount in R3 class stations. This shows that the southeastern and central-eastern regions of the country have short-range and long-range connections. A significant number of R6 class stations reveal that the long-range connections are prevalent over these regions. The ability to detect both short-range and long-range connections is one of the advantages of the complex network approach used in this study, compared to commonly used geostatistical methods which are based on the assumption of a semi-variogram having a decreasing correlation with increasing distance.

Similarly, the western coastline (C6) of India is also dominated equally by R2 and R3 class stations representing short- and long-range connection dynamics in the region. On the contrary, the central-western region (C5) of India is strongly dominated by only R3 class-type stations having a maximum number of links outside the community. This suggests that central-western (C5) regions have no intra-community links to stations. The above observations fall along the expected lines since westerlies enter in India from the West and travel to an entirely different part. Because of a lack of sufficient orographic barriers, we do not see any localized rainfall in this region.

The northeastern region of India (C4) shows a unique kind of pattern, with a significant number of intra-community links, inter-community links, connector hubs and global hubs. This region has a sufficient number of orographic barriers,which helps to accumulate more localized rainfall, represented by short-range connections. Hence, some of the rainfall features in this area are regionally bound and short-range. This region also shows a significant number of inter-community links owing to its long-range connections with the easterlies moisture movement from the C5 regions.

The Himalayan region (C7) shows dominance of R1 class stations, representing a very high degree of interconnectedness in the region. In other words, it suggests that this region receives localized rainfall, having short-range connections. Also, it can be said from the results that this region features a different climatology characterized by seasonal snow and a colder climate than the rest of the regions. Furthermore, it is entirely possible that this region may have connections to regions beyond what is considered in the present study.

From the above analysis, we infer that Z-P space is a useful tool to provide more insight into the qualitative and quantitative connections between the nodes within and outside a community. It also shows the strength of the connections between the communities and is useful in understanding how extreme events in one community affect the other regions. The physical reasoning for the classification of the nodes into seven classes is inline with the general understanding of the precipitation dynamics in India.

Table 3. Summary of the total number of each type of R class stations in the induvial community. The significance of each R class is described in Section 2.6.

C. No.	Explanation of R class	P(%)= Percentage of stations in particular R class in each community						
		R1	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
1	Eastern coastline, low-land region having no hubs, mostly dominated by intracommunity links and short-range connections	33.2	61.7	5.1	0	0	0	0
2	Mild-elevation inland region with connector hubs shows the dominance of both intra-community and inter-community links.	4.3	51.6	44.1	0	0	.9	0

3	Eastern-central region with moderate elevation shows a lower number of intra-community links to stations.	0.9	59.8	39.3	0	0	.7	0
4	Northeastern region of India shows all kinds of connections. Intra-community, inter-community, hubs, non-hubs, global hubs, etc.	13.0	44.7	40.7	0	0	1.3	.5
5	No intra-community links, highly dominated by hybrid stations; community shows short-range connections.	0	14.5	85.5	0	0	0	0
6	Negligible intra-community links, dominated by inter-community links and hybrid stations	.1	50.2	49.5	0	0	.1	0
7	No hubs, the community has all (ultra-) peripheral nodes that have links within the community, hence well isolated.	78.5	18.3	3.2	0	0	0	0

$\mathbf{P}(\%) = \left(\boldsymbol{total\ stations\ in\ any\ R\ class\ of\ community\ C} \middle/ \boldsymbol{total\ stations\ in\ community\ C}\right) \times \mathbf{100}$

4. Conclusion

This study proposed a novel, complex, network-based approach for quantifying the role of a single (rainfall) station within homogeneous regions, which is of great interest in regionalization studies, estimating missing information, etc. The study used a network information-theoretical approach known as Z-P space for understanding the qualitative and quantitative aspects of the members of a community. The Z-P approach categorizes the members into different classes based on the relative roles they play in the community and their strength of connections within and outside the community. The utility of the method was demonstrated using a synthetic case and then applied to the real-world case of the Indian rainfall network. The entire Indian rainfall network was divided into seven communities, and each community was analyzed using the Z-P approach. The results from the Z-P space approach provided important information such as how the communities are connected within themselves and with others. It was observed that the high-elevation, northern part of India was disconnected from other regions (communities). On the other hand, the southern peninsular region had strong intra-community links as well as inter-community links. It was also observed that the central and eastern parts of the country had many connecter hubs, indicating that these regions have long-range connections with other communities. The stations from the northeastern regions of the country, interestingly, have strong connections with other communities. The results of the study have significant implication in identifying key node locations in climate systems which play a major role in affecting the climate in the given community.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG) (GRK 2043/1) within the graduate research training group "Natural risk in a changing world (NatRiskChange)" at the University of Potsdam (http://www.uni-potsdam.de/natriskchange). The authors gratefully thank the Dr. Stephanie Natho and Roopam Shukla for helpful suggestion and reading the papers.

References

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

Agarwal, A., Marwan, N., Rathinasamy, M., Merz, B., Kurths, J., 2017a. Multi-scale event synchronization analysis for unravelling climate processes: a wavelet-based approach. Nonlinear Process. Geophys. 24, 599–611. https://doi.org/10.5194/npg-24-599-2017

Agarwal, A., Marwan, N., Rathinasamy, M., Ozturk, U., Merz, B., Kurths, J., 2018. Optimal Design of Hydrometric Station Networks Based on Complex Network Analysis. Hydrol. Earth Syst. Sci. Discuss. 1–21. https://doi.org/10.5194/hess-2018-113

Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y., Zhou, C., 2008. Synchronization in complex networks. Phys. Rep. 469, 93–153. https://doi.org/10.1016/j.physrep.2008.09.002

Berndtsson, R., 1988. Temporal variability in spatial correlation of daily rainfall. Water Resour. Res. 24, 1511–1517. https://doi.org/10.1029/WR024i009p01511

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 2008, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Conticello, F., Cioffi, F., Merz, B., Lall, U., 2017. An event synchronization method to link heavy rainfall events and large-scale atmospheric circulation features:. Int. J. Climatol. https://doi.org/10.1002/joc.5255

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

Donges, J.F., Zou, Y., Marwan, N., Kurths, J., 2009. Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods. Eur. Phys. J. Spec. Top. 174, 157–179. https://doi.org/10.1140/epjst/e2009-01098-2

Eustace, J., Wang, X., Cui, Y., 2015. Community detection using local neighborhood in complex networks. Phys. Stat. Mech. Its Appl. 436, 665–677. https://doi.org/10.1016/j.physa.2015.05.044

Fortunato, S., 2010. Community detection in graphs. Phys. Rep. 486, 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Guimera, R., Amaral, L.A.N., 2005. Cartography of complex networks: modules and universal roles. J. Stat. Mech. Theory Exp. 2005, P02001. https://doi.org/10.1088/1742-5468/2005/02/P02001

Guimera, R., Sales-Pardo, M., Amaral, L.A.N., 2007. Classes of complex networks defined by role-to-role connectivity profiles. Nat. Phys. 3, 63–69. https://doi.org/10.1038/nphys489

Halverson, M.J., Fleming, S.W., 2015. Complex network theory, streamflow, and hydrometric monitoring system design. Hydrol. Earth Syst. Sci. 19, 3301–3318. https://doi.org/10.5194/hess-19-3301-2015

Hassan, B.G.H., Ping, F., 2012. Regional Rainfall Frequency Analysis for the Luanhe Basin – by Using L-moments and Cluster Techniques. APCBEE Procedia 1, 126–135. https://doi.org/10.1016/j.apcbee.2012.03.021

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

Jain, S.K., Kumar, V., Saharia, M., 2013. Analysis of rainfall and temperature trends in northeast India. Int. J. Climatol. 33, 968–978. https://doi.org/10.1002/joc.3483

Jha, S.K., Zhao, H., Woldemeskel, F.M., Sivakumar, B., 2015. Network theory and spatial rainfall connections: An interpretation. J. Hydrol. 527, 13–19. https://doi.org/10.1016/j.jhydrol.2015.04.035

Krishnamurthy, V., Shukla, J., 2000. Intraseasonal and Interannual Variability of Rainfall over India. J. Clim. 13, 4366–4377. https://doi.org/10.1175/1520-0442(2000)013<0001:IAIVOR>2.0.CO;2

Lakhanpal, A., Sehgal, V., Maheswaran, R., Khosa, R., Sridhar, V., 2017. A non-linear and non-stationary perspective for downscaling mean monthly temperature: a wavelet coupled second order Volterra model. Stoch. Environ. Res. Risk Assess. 31, 2159–2181. https://doi.org/10.1007/s00477-017-1444-6

Li, Z., Yang, D., Hong, Y., Zhang, J., Qi, Y., 2014. Characterizing Spatiotemporal Variations of Hourly Rainfall by Gauge and Radar in the Mountainous Three Gorges Region. J. Appl. Meteorol. Climatol. 53, 873–889. https://doi.org/10.1175/JAMC-D-13-0277.1

Malik, N., Bookhagen, B., Marwan, N., Kurths, J., 2012. Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks. Clim. Dyn. 39, 971–987. https://doi.org/10.1007/s00382-011-1156-4

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

Mishra, A.K., Coulibaly, P., 2009. Developments in hydrometric network design: A review. Rev. Geophys. 47. https://doi.org/10.1029/2007RG000243

Newman, M.E.J., 2004. Detecting community structure in networks. Eur. Phys. J. B - Condens. Matter 38, 321–330. https://doi.org/10.1140/epjb/e2004-00124-y

Niu, J., 2013. Precipitation in the Pearl River basin, South China: scaling, regional patterns, and influence of large-scale climate anomalies. Stoch. Environ. Res. Risk Assess. 27, 1253–1268. https://doi.org/10.1007/s00477-012-0661-2

Özger, M., Mishra, A.K., Singh, V.P., 2010. Scaling characteristics of precipitation data in conjunction with wavelet analysis. J. Hydrol. 395, 279–288. https://doi.org/10.1016/j.jhydrol.2010.10.039

Pai, D.S., Sridhar, L., Badwaik, M.R., Rajeevan, M., 2015. Analysis of the daily rainfall events over India using a new long period (1901–2010) high resolution (0.25° × 0.25°) gridded rainfall data set. Clim. Dyn. 45, 755–776. https://doi.org/10.1007/s00382-014-2307-1

Paluš, M., 2018. Linked by Dynamics: Wavelet-Based Mutual Information Rate as a Connectivity Measure and Scale-Specific Networks, in: Tsonis, A.A. (Ed.), Advances in Nonlinear Geosciences. Springer International Publishing, Cham, pp. 427–463. https://doi.org/10.1007/978-3-319-58895-7_21

Pardo-Igúzquiza, E., 1998. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. J. Hydrol. 210, 206–220. https://doi.org/10.1016/S0022-1694(98)00188-7

Pfurtscheller, G., Lopes da Silva, F.H., 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. 110, 1842–1857. https://doi.org/10.1016/S1388-2457(99)00141-8

Quian Quiroga, R., Kraskov, A., Kreuz, T., Grassberger, P., 2002a. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. Phys. Rev. E 65. https://doi.org/10.1103/PhysRevE.65.041903

Quian Quiroga, R., Kreuz, T., Grassberger, P., 2002b. Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. Phys. Rev. E 66. https://doi.org/10.1103/PhysRevE.66.041904

Quiroga, R.Q., Arnhold, J., Grassberger, P., 2000. Learning driver-response relationships from synchronization patterns. Phys. Rev. E 61, 5142–5148. https://doi.org/10.1103/PhysRevE.61.5142

Razavi, T., Coulibaly, P., 2013. Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods. J. Hydrol. Eng. 18, 958–975. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690

Rheinwalt, A., Boers, N., Marwan, N., Kurths, J., Hoffmann, P., Gerstengarbe, F.-W., Werner, P., 2016. Non-linear time series analysis of precipitation events using regional climate networks for Germany. Clim. Dyn. 46, 1065–1074. https://doi.org/10.1007/s00382-015-2632-z

Rheinwalt, A., Goswami, B., Boers, N., Heitzig, J., Marwan, N., Krishnan, R., Kurths, J., 2015. Teleconnections in Climate Networks: A Network-of-Networks Approach to Investigate the Influence of Sea Surface Temperature Variability on Monsoon Systems, in: Lakshmanan, V., Gilleland, E., McGovern, A., Tingley, M. (Eds.), Machine Learning and Data Mining Approaches to Climate Science. Springer International Publishing, Cham, pp. 23–33. https://doi.org/10.1007/978-3-319-17220-0_3

Rubinov, M., Sporns, O., 2011. Weight-conserving characterization of complex functional brain networks. NeuroImage 56, 2068–2079. https://doi.org/10.1016/j.neuroimage.2011.03.069

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: Uses and interpretations. NeuroImage 52, 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Salinas, J.L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies. Hydrol. Earth Syst. Sci. 17, 2637–2652. https://doi.org/10.5194/hess-17-2637-2013

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W., Lin, C.-T., 2017. A review of clustering techniques and developments. Neurocomputing 267, 664–681. https://doi.org/10.1016/j.neucom.2017.06.053

Sehgal, V., Lakhanpal, A., Maheswaran, R., Khosa, R., Sridhar, V., 2016. Application of multi-scale wavelet entropy and multi-resolution Volterra models for climatic downscaling. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2016.10.048

Sivakumar, B., Singh, V.P., Berndtsson, R., Khan, S.K., 2015. Catchment Classification Framework in Hydrology: Challenges and Directions. J. Hydrol. Eng. 20, A4014002. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000837

Sivakumar, B., Woldemeskel, F.M., 2014. Complex networks for streamflow dynamics. Hydrol. Earth Syst. Sci. 18, 4565–4578. https://doi.org/10.5194/hess-18-4565-2014

Smith, A., Sampson, C., Bates, P., 2015. Regional flood frequency analysis at the global scale. Water Resour. Res. 51, 539–553. https://doi.org/10.1002/2014WR015814

Steinhaeuser, K., Chawla, N.V., Ganguly, A.R., 2010. An exploration of climate data using complex networks. ACM SIGKDD Explor. Newsl. 12, 25. https://doi.org/10.1145/1882471.1882476

Stolbova, V., Martin, P., Bookhagen, B., Marwan, N., Kurths, J., 2014. Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka. Nonlinear Process. Geophys. 21, 901–917. https://doi.org/10.5194/npg-21-901-2014

Stolbova, V., Surovyatkina, E., Bookhagen, B., Kurths, J., 2016. Tipping elements of the Indian monsoon: Prediction of onset and withdrawal: TIPPING ELEMENTS OF MONSOON. Geophys. Res. Lett. 43, 3982–3990. https://doi.org/10.1002/2016GL068392

Tass, P., Rosenblum, M.G., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., Schnitzler, A., Freund, H.-J., 1998. Detection of n : m Phase Locking from Noisy Data: Application to Magnetoencephalography. Phys. Rev. Lett. 81, 3291–3294. https://doi.org/10.1103/PhysRevLett.81.3291

Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Econ. Geogr. 46, 234. https://doi.org/10.2307/143141

Tsonis, A.A., Wang, G., Swanson, K.L., Rodrigues, F.A., Costa, L. da F., 2011. Community structure and dynamics in climate networks. Clim. Dyn. 37, 933–940. https://doi.org/10.1007/s00382-010-0874-3

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

Yang, X., Xie, X., Liu, D.L., Ji, F., Wang, L., 2015. Spatial Interpolation of Daily Rainfall Data for Local Climate Impact Assessment over Greater Sydney Region. Adv. Meteorol. 2015, 1–12. https://doi.org/10.1155/2015/563629

Zhou, C., Zemanová, L., Zamora-López, G., Hilgetag, C.C., Kurths, J., 2007. Structure–function relationship in complex brain networks expressed by hierarchical synchronization. New J. Phys. 9, 178–178. https://doi.org/10.1088/1367-2630/9/6/178

Zlatić, V., Božičević, M., Štefančić, H., Domazet, M., 2006. Wikipedias: Collaborative web-based encyclopedias as complex networks. Phys. Rev. E 74. https://doi.org/10.1103/PhysRevE.74.016115

Zrinji, Z., Burn, D.H., 1996. Regional Flood Frequency with Hierarchical Region of Influence. J. Water Resour. Plan. Manag. 122, 245–252. https://doi.org/10.1061/(ASCE)0733-9496(1996)122:4(245)

Zrinji, Z., Burn, D.H., 1994. Flood frequency analysis for ungauged sites using a region of influence approach. J. Hydrol. 153, 1–21. https://doi.org/10.1016/0022-1694(94)90184-8