Unfolding Community Structure in Rainfall Network of Germany Using Complex Network-Based Approach

A. Agarwal, N. Marwan, U. Ozturk and R. Maheswaran

Abstract Many natural systems can be represented as networks of dynamical units with a modular structure in the form of communities of densely interconnected nodes. Unfolding structure of such densely interconnected nodes in hydro-climatology is essential for reliable parameter transfer, model inter-comparison, prediction in ungauged basins, and estimating missing information. This study presents the application of complex network-based approach for regionalization of rainfall patterns in Germany. As a test case study, daily rainfall records observed at 1,229 rain gauges were selected throughout Germany. The rainfall data, when represented as a complex network using event synchronization, exhibits small-world and scale-free network topology which are a class of stable and efficient networks common in nature. In total, eight communities were identified using Louvain community detection algorithm. Each of the identified communities has a sufficient number of rain gauges which show distinct statistical and physical rainfall characteristics. The method used has wide application in most of the real systems which can be represented by network enabling to understand modular patterns through time series analysis.

Keywords Complex network · Event synchronization · Rainfall network

1 Introduction

A complex network is a collection of nodes, interconnected with links in a non-trivial manner. In a functional network, links are set up between each pair of nodes based on how the nodes interact with each other. For example, in a family network, each person is considered as a node and the relationship between them is a link; in a computer network, each computer is a node and links are the connections between computers;

A. Agarwal (✉) · U. Ozturk
Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany
e-mail: aagarwal@uni-potsdam.de

A. Agarwal · N. Marwan · U. Ozturk · R. Maheswaran
Potsdam Institute for Climate Impact Research (PIK), 601203, 14412 Potsdam, Germany

M. Rathinasamy et al. (eds.), *Water Resources and Environmental Engineering II*,
https://doi.org/10.1007/978-981-13-2038-5_17

in brain networks, neurons are nodes and links represent the pairwise neurons' interaction. In the last decade, climatologists and hydrologists have successfully applied the same network concept to analyze different research questions of hydro-climatic science. Each node represents a geographical location of climatological data (rainfall, stream flow, temperature, air pressure, etc.) and links between nodes are set up based on their interaction or similar variability (correlation, synchronization).

Hydro-climatic systems often show the topology of interacting nodes embedded in space. Such spatial networks are usually organized in modules (communities) of densely interconnected nodes. The spatial embedding of the network can hide the underlying community structure, rendering the identification of communities a challenging task. The investigation of the community structure in such networks helps in better understanding the functional mechanism of the highly complex systems. For instance, the identification of communities in rainfall network is essential to obtain reliable information about rainfall in case of missing values or no observation. Reliable records of rainfall are vital for many hydraulic (flood studies, dam, dikes, diversion structures, power plants) and environmental studies (land use planning and management, stream habitat assessment, extreme events, and climate impact studies). The traditional approach for estimating missing values in hydrology is to pool the information from other hydrologically homogeneous watersheds; by performing clustering analysis, [1–3] which have been identified to lead to erroneous estimations.

In the past, there have been several attempts [4–10] to develop a general unified framework for identifying rainfall communities (rainfall coherent sub-systems) using different approaches. The methods, in general, range from similarities in the rainfall signature, [11–13], via geographical location and catchment characteristics [14, 15] to rainfall network complexity, model parameters, and uncertainty [16–19]. Razavi and Coulibaly provides a detailed review of several methods for clustering in hydro-climatology [19]. Even though there are a plethora of methods available for community detection, most of these methods are subjective and consider the spatial proximity of the region, which in turn is not always true [8]. Also, traditional clustering methods are not capable of unraveling the numerous connections of each rain gauge station within and outside the community. Information on node connections/interconnections is essential to understand the role of each station in the rainfall network. For instance, dead ends (stations having few connections) which might be influenced by sampling size whereas stations connecting two communities are hybrid nodes hence play an important role in the rainfall network. Also, stations having a high number of links within the community can be termed as a local center. This kind of analysis by community detection is essential to understand the relative roles of each of the member stations of the community and is critical information in uncertainty analysis for predictions in ungauged basins, regionalization, missing values, and hydro-monitoring [20].

To add such insights into community detection, we use a more data-centric approach based on the complex network which extracts a spatial pattern from the network. The fundamental advantage of using network approach is that it considers the entire network structure in contrast to the only node attributes used by traditional approaches [21]. Also, complex network uses network distance for partition compared to pairwise distance used by various clustering algorithm which is suitable for portioning such a spatiotemporal dynamical system [22, 23]. This study aims to identify a homogeneous region in rainfall network as defined by similarity in the long-term rainfall variability. Such regions (communities) are of interest both to reveal inherent structure with the rainfall coherent sub-system and to use as potential climate indicators [24]. It is vital to note that spatial proximity is not taken into consideration while forming a network and identifying communities.

In this study, to construct a rainfall network, we use event synchronization (ES) similarity measure. ES has advantages over other time-delayed correlation techniques (e.g., Pearson lag correlation), as it allows us to use dynamics time delay (not fixed) which is suitable to study interrelations between series of non-Gaussian data, data with heavy tails [25, 26]. After calculating synchronization between all possible pair of stations, we apply a suitable threshold to construct rainfall network. Louvain community detection algorithm employed to identify homogeneous regions by maximizing modularity. As a test case study, we use 1,229 available observed rain gauges across Germany to form a rainfall network. Eight communities were identified, and each of the identified communities has a sufficient number of rain gauges which show distinct statistical and physical rainfall characteristics. In this paper, the only preliminary result has been shown on German rainfall network which has immense potential to be extended in the future for hydro-monitoring purposes.

The paper is organized in the following manner. Section 2 describes the methods used in this study such as event synchronization, community detection, and various network measures. Section 3 discusses the application of network on observed daily rainfall data of Germany, and subsequent results obtained are discussed in detail in Sect. 4. The summary of the study is briefed in Sect. 5.

2 Methods

2.1 Event Synchronization

Adapting the state-of-the-art method, event synchronization, we measure nonlinear synchronization between all possible pair of rain gauges [27]. The modified algorithm proposed works as follows: An event occurs in the signals $x(t)$ and $y(t)$ at time t_l^x and t_m^y, where $l = 1, 2, 3, 4 \ldots S_x$, $m = 1, 2, 3, 4 \ldots S_y$, and S_x, and S_y are the total number of events, respectively. In our study, we derive events from a more or less continuous time series by selecting all time steps with values above a threshold (α = 95th percentile). These events in $x(t)$ and $y(t)$ are considered as synchronized when they occur within a time lag $\pm\tau_{lm}^{xy}$ which is defined as follows:

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

where S_x and S_y are the total number of such events (greater then threshold α) that occurred in the signal $x(t)$ and $y(t)$, respectively. The above definition of the time lag helps to separation of independent events which in turn allows to take into account the fact that different processes are responsible for the generation of events. We need to count the number of times an event occurs in the signal $x(t)$ after it appears in the signal $y(t)$ and vice versa, and this is achieved by defining quantities $C(x|y)$ and $C(y|x)$. Where

$$C(x/y) = \sum_{l=1}^{S_x} \sum_{m=1}^{S_y} J_{xy} \tag{2}$$

and

$$J_{xy} = \begin{cases} 1 \ \text{if}\, 0 < t_l^x - t_m^y < \tau_{lm}^{xy} \\ \frac{1}{2} \ \text{if}\, t_l^x = t_m^y \\ 0 \ \text{else}, \end{cases} \tag{3}$$

Similarly, we can define $C(y|x)$, and from these quantities, we can obtain

$$Q_{xy} = \frac{C(x|y) + C(y|x)}{\sqrt{(S_x - 2)(S_y - 2)}} \tag{4}$$

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

2.2 Community Detection

Complex networks often show subsets of nodes that are densely interconnected. These subsets are often known as communities [7]. The understanding and visualization of community structure provide insight into the network [28]. For instance, different communities within a network may have very different properties compared to the averaged properties of the complete network [29].

There exist several community detection approaches aiming at stratifying the nodes into communities in an optimal way (see [30] for an extensive review), but very few of those are applicable for a hugely complex network having more than thousand nodes [24]. In this study, we adopted the Louvain algorithm proposed by [31, 32]). The Louvain algorithm works to optimize modularity (Q), a network measure, an indicator of "community's partition correctness" in a way that the number of edges falling within the community should be maximum and minimum in between communities [33]. Modularity (Q) is calculated as:

$$Q = \frac{1}{2m} \sum_{i,j} \lfloor A_{ij} - P_{i,j} \rfloor \delta(C_i C_j) \tag{5}$$

where A_{ij} represents the number of edges between i and j and $P_{i,j} = \frac{k_i k_j}{2m}$ represent the expected number of edges between node i and j. k_i and k_j are the total number of links of nodes i and j, respectively. m is the total number of edges in a network calculated as $m = 1/2 \sum_{ij} A_{ij}$. C_i and C_j are the communities to which node i and j are assigned, and the $\delta -$ function $\delta(C_i C_j)$ is 1 if nodes i and j are in the same community and 0 otherwise.

2.3 Network Measures

Various network measures exist to characterize the network dynamics, but in this study, we use only three prime and widely used properties: the degree (k) and degree distribution, the clustering coefficient (CC), and the average path length (L).

The degree [26] of a node in the network emphasizes the number of connections linked to the node directly. It can explain the type of nodes to some extent such as hubs having the highest degree and non-hubs having a low degree. The degree distribution $p(k)$ of a network is then defined to be the fraction of nodes in the network with degree k. Thus, if there are N nodes in total in a network and N_k of them have degree k, we have $P(k) = N_K/N$.

In general, the clustering coefficient (CC) [7] is used to identify the modular organization of the network by quantifying the tendency of a node to share same neighbors of directly connected nodes (tendency to form a triangle). High values of CC represent well-interconnected nodes and suggest redundancy of information in the network. Also, high values of CC are interpreted to exhibit significant spatial coherence [7, 34, 35].

The average path length (L) is the average number of steps taken along the shortest paths between all possible pairs of network nodes [26]. It is a measure of the efficiency of information or mass transport in a network. Efficiency in the network is inversely related to path length. A network with small average path length is highly efficient because two nodes are likely to be separated by few links (Table 1).

2.4 Random and Scale-Free Network

We generate an equivalent random and scale-free network for the same number of nodes ($N = 1229$) and links (96,384) as rainfall network (RN). We have used MATLAB toolbox for network analysis provided by MIT [36].

A random network is constructed by starting with a set of N isolated nodes and adding successive edges between them at random. There exist several random

Table 1 Network measures

Degree	Clustering coefficient	Average path length
$D_i = \frac{\sum_{j=1}^{N} A_{i,j}}{N-1}$	$C = \frac{1}{N}\sum_{i\in N} C_i = \frac{1}{N}\sum_{i\in N}\frac{2E}{k_i(k_i-1)}$	$a = \sum_{v_i,v_j\in N}\frac{d(v_i,v_j)}{N(N-1)}$

N the total number of nodes in a network. D_i degree of node i.; the clustering coefficient for the ith node is represented as C_i where E is the number of links that are actually observed to exist between the k_i neighbors of node i. We use the average of all the local clustering coefficients over the network as a bulk measure of the clustering tendency or cliquishness of the network as a whole. a is average path length, V is the set of nodes in the network, and $d(v_i, v_j)$ is the shortest path from v_i to v_j

network models, and each of them produces different probability distributions on graphs. Most commonly studied is **Erdős–Rényi model** which is denoted as $G(N, p)$. In the $G(N, p)$ model, a network is formed by linking nodes randomly with probability p $(0 < p < 1)$ independently from every other edge [37, 38].

$$\text{The expected number of edges } (L) \text{ in } G(n, p)\, is = C(N, 2)$$

For given L and N, we can estimate the probability (p) for the random network.

For the scale-free network, we use the **Barabási–Albert (BA) model** which uses a preferential attachment algorithm. Many existing natural and human-made systems such as the Web network, social network seem to be approximately scale-free and certainly contain few supernodes (called hubs) with an exceptionally high degree as compared to the other nodes of the network. The BA model tries to explain the existence of such nodes in real networks [39, 40].

The network begins with an initial connected $m_o = 2$ node, the links between which are chosen arbitrarily, as long as each node has at least one link. At each time step, we add a new node with m $(\leq m_0)$ links that connect the new node to m nodes already in the network. Repeating the procedure, total number of edges (L) in the network will be mN, where N is the number of nodes. For a known total number of edges (L) and nodes (N), we can estimate the value of m.

3 Rainfall Network Construction

Precipitation data from Germany is studied to explore the utility of network theory in identifying the communities in rainfall network. Daily data from an extensive network of 1,440 rain gauge stations (Fig. 1a) in contiguous Germany is available. Out of which 1,229 rain gauge stations lie inside Germany (red dots in Fig. 1a). Two hundred and eleven stations outside Germany (green dots in Fig. 1a) are included in the analysis to minimize the spatial boundary effect in the network formation (these stations are finally excluded from the discussion). One hundred and ten years of daily data, from 1 January 1901 to 31 December 2010, are available from various stations

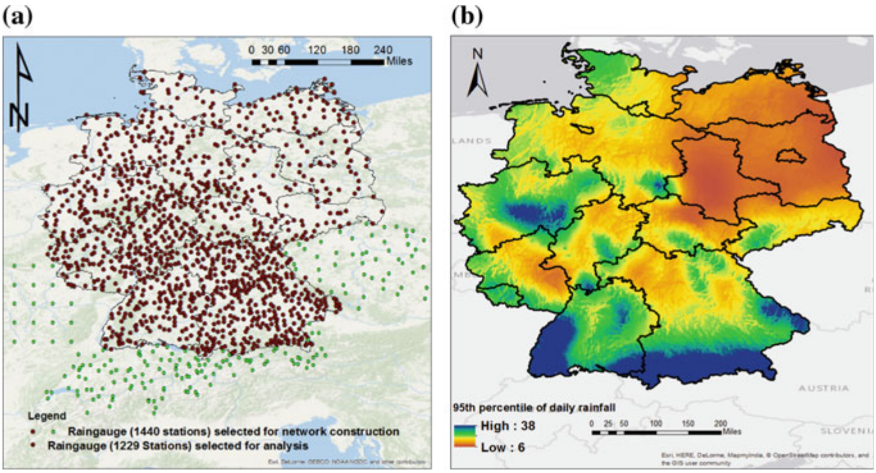

Fig. 1 (Left) Geographical location of gauging stations selected from Germany. (Right) 95th percentile of daily rainfall amounts for the observed data

operated by the German Weather Service. Data processing and quality control were performed according to Österle et al. [41].

We begin the network construction [29] by extracting event series from the 1,440 rain gauges representing extreme rainfall events, i.e., precipitation exceeding the 95th percentile at that station. By applying a threshold (95th percentile), we draw out extreme events from the given time series [24]. 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. The strength of the connection between any two stations is established using the concept of event synchronization as discussed in [42]. If the value of ES is close to 1, it implies the two stations are highly synchronized, and if the value is close to 0, it indicates no synchronization [43].

In this study, we use an undirected network; i.e., we do not consider which of the two synchronized events happened first, to avoid the possibility of misleading directionalities of event occurrences in between rain gauges that are topographically close. Although the reconstructed network is based on all 1,440 nodes (to minimize the boundary effect), the subsequent topological analysis is performed only for the 1,229 stations lying inside Germany (red dots in Fig. 1a).

4 Result and Discussion

The *rainfall network* (*RN*) obtained using the 1,229 nodes contain total 96,384 pair-wise links between them. The average degree for the entire RN is 95, the minimum is 0, and the maximum is 413. Several characteristics of the rainfall network are

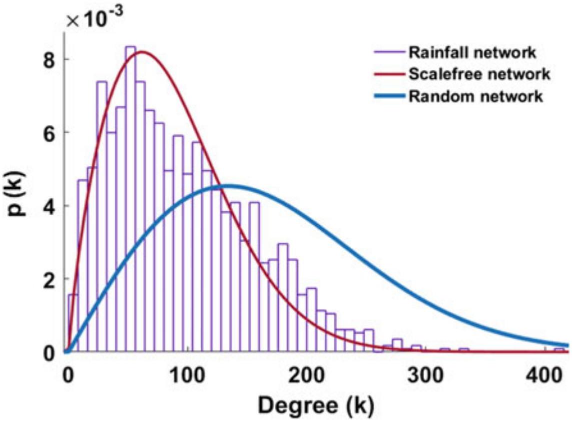

Fig. 2 Degree distribution of the German rainfall network (RN), random network, and scale-free network generated for an equal number of nodes and links

immediately clear. For instance, obtained RN is not a regular network since each node of the regular network exhibits equal degree. Also, the existence of supernodes (hubs) in the network, i.e., some nodes has a very high degree.

To understand the network topology, we compute the network properties: the degree distribution ($p(k)$), the clustering coefficients (CC),, and the average path length (L) [7] and then place the rainfall network (RN) into the context of known topologies. The degree distribution of the RN (Fig. 2) shows some resemblance with the degree distribution of a scale-free network [44] because scale-free networks have an asymmetric degree distribution which asymptotes to $P(k) \propto k^{-\gamma}$ at sufficiently large values of k, where γ ranges from 2.1 to 4 for a wide array of the observed network [7].

The small-world property of the RN is checked by clustering coefficient and average path length. A network has a small-world property if $C \gg C_{\text{random}}$ and $L \gtrsim L_{\text{random}}$ [7, 45]. For the *RN,* we find a global clustering coefficient of $C=0.64$ and an average path length of $L=3.92$, whereas the equivalent random graph has a clustering coefficient of $C_{\text{random}} = 0.12$, and an average path length of $L_{\text{random}} = 1.87$; thus, the rainfall network behaves as small-world network that exhibits scale-free behavior.

The scale-free network represents a network having supernodes (as already interpreted) also termed as "hubs" which have many more connections as compared to the rest of the nodes in the network as a whole following the power law/Pareto distribution. In general, a small-world network is characterized regarding stability and efficiency [7, 46]. Stability signifies that the network holds its integrity if some nodes of the network are randomly removed; i.e., the removal of the node will likely not fragment the network structure. In the context of rainfall network, this implies that if a randomly selected station is removed, then it is possible to recover most of the information. The efficiency of the network qualitatively characterizes the ease of information propagation in the network. A network with small average path length is highly efficient because two nodes are likely to be separated by few links.

The rainfall network constructed behaves as small-world network which, in turn, indicates the possible modular organization (communities) in the network. Hence, community mining has been performed in the study.

Identification of Modular Organization of the Network

While analyzing the network, it always remains a concern to explore the densely interconnected small subgroups, the communities. Here, it can be stated that with the enhanced knowledge of the community, their pattern of formation, the structure can give a deep insight into the network which is particularly true for small-world network. Also, the identification of the modular organization is helpful for coarse-graining the network [47, 48] which is very crucial while dealing with billions of nodes such as Web networks, brain networks, climate networks.

As already mentioned, a vast range of community detection algorithms exists (see [30] for an extensive review). However, considering the finding that the particular choice of the community detection algorithms has only a small impact [7], we only use one community detection algorithm, i.e., the Louvain algorithm which works for optimizing modularity on each step of the algorithm. In general, high modularity networks are densely linked within communities but sparsely linked between communities; i.e., the algorithm stops when the highest modularity is achieved.

The Louvain (maximizing modularity) community detection algorithm detects eight communities within the RN of Germany (Fig. 3).

Community structure, identified by modularity-based network analysis, shares specific elements which can be tied back to two general categories such as climate characteristics and physical characteristics. Therefore, the number of communities reflects the climatological diversity of Germany, and the number of stations per community sets the extent to which each distinct climatology "family" is sampled.

The spatial extent of the stations in the communities prominently shows the ability of the methods to capture the underlying driving forces. For instance, we do not impose any spatial constraints on the network, and there is no guarantee that community will be geographically cohesive, but as shown in Fig. 3 this is often the case. In fact, we observe a relationship between clustering coefficient and the number of stations in the community; namely, the higher the value of C, a large number of stations exist in the community and more geographically coherent they are. Although geographical proximity plays a vital role apart from that there is another governing mechanism, viz. physiographic features, climatic patterns, and statistical similarity of rainfall regimes which might influence the climate pattern. It is also important to emphasize, however, that such a modular structure is identified based on a cluster of actual connections, rather than based on our traditional way of geographic proximity, nearest neighbors, regional patterns, and linear correlations.

Table 2 shows the statistical and geographical 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 higher total amount of precipitation, larger standard deviation shows a stronger variation of data for the collecting period, and a larger coefficient of skewness indicates more extreme precipitation events [49].

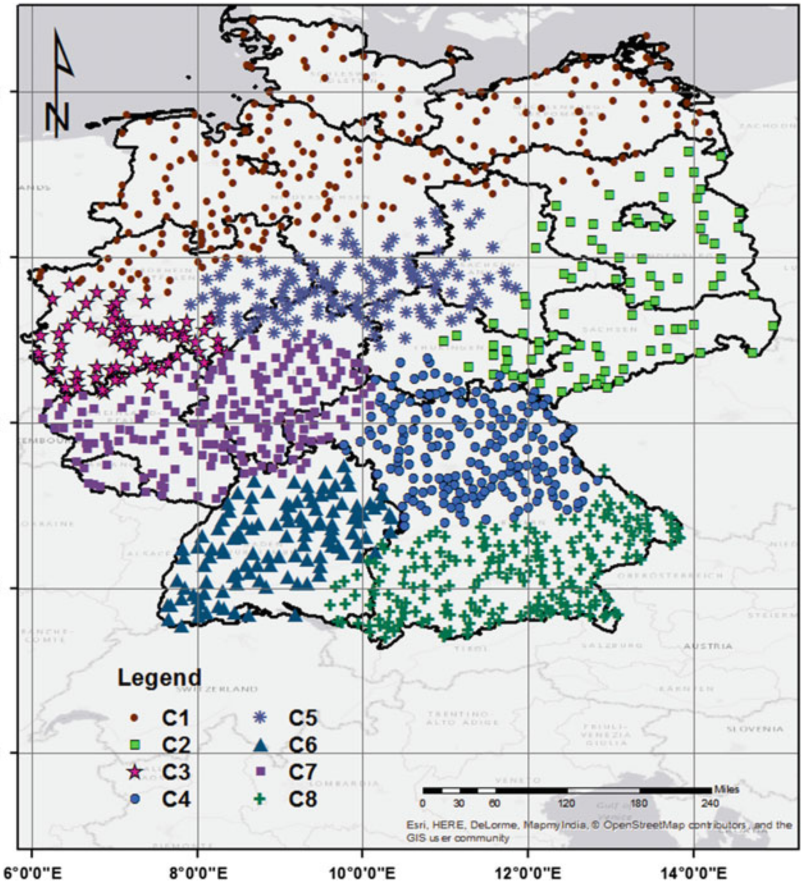

Fig. 3 Rain gauge stations map colored according to community membership. The communities were identified with maximizing modularity using Louvain algorithm

Intriguingly, resultant communities show numerous known relationships of precipitation along with new insights that are not obvious and hence may be of interest to a climate scientist. For example, community 8 (Fig. 3), which covers an almost mountainous region (Bavarian Alps), with daily mean of 3.0 mm and standard deviation of 6.34 mm (Table 2), represents the area with the highest daily mean and largest variation in the precipitation while community 2 (Fig. 3), which cover slow land areas (Mecklenburg lowlands), represents the region with the lowest mean and uniform precipitation. Community 2 (Mecklenburg lowlands) has a large coefficient of skewness, whereas communities 3 (Rhenish Massif region) and 6 (Black Forest region) show the smallest coefficient of skewness. All the communities show a positive coefficient of skewness, which indicates precipitation with a long tail toward high values.

Table 2 Summary of geographical and statistical analysis for each individual community. Communities formed by maximizing the modularity using Louvain algorithm. Elevation map for Germany is presented in the Fig. 1b

C. No.	Number of stations	Daily mean	Standard deviation	Skewness	Remarks
1	240 (20%)	1.92	4.02	4.46	Low elevation, wide geographic range
2	91 (7%)	1.69	4.01	6.04	Small community, mid elevation, Mecklenburg lowlands
3	68 (5%)	2.56	4.99	3.8	Smallest community, mid elevation, Rhenish Massif
4	193 (16%)	1.97	4.24	4.33	High elevation, Bavarian Forest
5	131 (11%)	2.03	4.31	4.65	Mid elevation, Rhenish Massif
6	118 (10%)	2.63	5.47	3.98	High elevation, Black Forest region
7	190 (15%)	2.06	4.38	4.23	Mid elevation, Rhineland-Palatinate
8	198 (16%)	3.0	6.34	4.13	Very high elevation, Bavarian Alps

Community 1 also shows low daily mean and uniform precipitation in a wide longitudinal geographic range consisting of low elevation stations. Communities 2 and 3 are the smallest communities which may have characteristics that are rare because they show either undersampled or uncommon hydro-climatological regimes, which make them vital if the aim of a rainfall network is to sample the inherent hydro-climatological diversity of that area [7]. In South Germany, both communities 4 and 6 (Fig. 3) represent the forest-dominated (Bavarian Forest and Black Forest region, respectively) high elevation regions having nearly same statistical and geographical properties (Table 2); then what makes them different? The answer must lie in the orographic barrier which lies along the 10°E longitude which causes an abrupt change of topography of the area resulting in different climate regimes for communities 4 and 6. These results also follow the study by Rheinwalt [50], which shows an abrupt change in the precipitation isochrones along the orographic barrier.

It is also interesting to understand how these network communities present different climate regime properties. For instance, both day-to-day precipitation dynamics and overall seasonal regimes exhibited by data from a particular rain gauge station are determined to a significant extent by the elevation whether the station receives daily precipitation as rain, snow, or some mixture of the two. Thus, it might be possible to understand the community structure, at least in part, by their stations' elevation. Comparing the communities with the corresponding stations' elevation reveals, to some degree, that they are stratified by elevation (Fig. 4). Overall, it can be stated that the singularities of sharp transition hidden in precipitation data are more significant

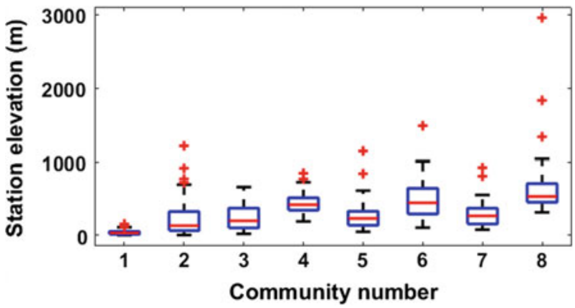

Fig. 4 Boxplots of mean basin elevation grouped by the communities. The number of stations in each community is mentioned in Table 2

than changes in the periodicity or data structure of the time series for community detection.

5 Summary

This study has shown the application of complex network-based method to unfold modular structure in complex natural systems using the dynamics of observed time series. We applied the method on observed rainfall time series and uncovered eight communities that are consistent with statistical and physical characteristics of rainfall. The results lead to the following concluding remarks:

- Preliminary investigation on the rainfall network of Germany shows that the network exhibits small world and scale-free behaviour which is a common class in many disciplines such as brain network, airport network. A small-world network implies stability, and the network is resilient to the loss of nodes whereas scale-free behaviour suggested that network consist of supernodes in the network which are vital to consider for many hydrologic applications such as missing information, prediction in ungauged basin etc. Hence, the advantage of using complex network-based approach is that network topology gives a beforehand idea of the behavior of communities and rain gauges in the network.
- The rainfall network based on event synchronization seems to be a formidable statistic in capturing the rainfall system dynamics. The 1,229 stations considered are categorized into eight communities each exhibiting distinct rainfall characteristics. We then show that these eight communities appear to be defined by their geographical proximity, which in turn corresponds to shared or different meteorological forcing. The number of communities reflects the diversity of such rainfall dynamical classes, and the number of stations per community sets the extent to which each regime is sampled.

In future, it would be interesting to extend the current work to understand the strength of connection among rain gauges within the community. In other words, microscopic investigation of the (intra-) intercommunity connection is essential to have better hydro-monitoring design of rain gauges. Also, it would be vital to extend the work to identify the universal role of each raingauge (node) such as hubs, non-hubs, dead-ends in communities and whole network.

Acknowledgements The author acknowledges Deutsche Forschungsgemeinschaft (DFG) (GRK 2043/1) for funding this research within the graduate research training group Natural Risk in a Changing World (NatRiskChange) at the University of Potsdam (http://www.uni-potsdam.de/natriskchange).

References

1. Agarwal, A., Maheswaran, R., Kurths, J., Khosa, R.: Wavelet spectrum and self-organizing maps-based approach for hydrologic regionalization—a case study in the Western United States. Water Resour. Manag. **30**(12), 4399–4413 (2016)
2. Blöschl, G., Sivapalan, M.: Scale issues in hydrological modelling: a review. Hydrol. Process. **9**(3–4), 251–290 (1995)
3. Saf, B.: Regional flood frequency analysis using L-moments for the West Mediterranean region of Turkey. Water Resour. Manag. **23**(3), 531–551 (2009)
4. Agarwal, A., Maheswaran, R., Sehgal, V., Khosa, R., Sivakumar, B., Bernhofer, C.: Hydrologic regionalization using wavelet-based multiscale entropy method. J. Hydrol. **538**, 22–32 (2016)
5. Carlón Allende, T., Mendoza, M.E., López Granados, E.M., Morales Manilla, L.M.: Hydrogeographical regionalisation: an approach for evaluating the effects of land cover change in watersheds. a case study in the Cuitzeo Lake Watershed, Central Mexico. Water Resour. Manag. **23**(12), 2587–2603 (2009)
6. Franchini, M., Suppo, M.: Regional analysis of flow duration curves for a limestone region. Water Resour. Manag. **10**(3), 199–218 (1996)
7. Halverson, M.J., Fleming, S.W.: Complex network theory, streamflow, and hydrometric monitoring system design. Hydrol. Earth Syst. Sci. **19**(7), 3301–3318 (2015). https://www.doi.org/10.5194/hess-19-3301-2015
8. Jha, S.K., Zhao, H., Woldemeskel, F.M., Sivakumar, B.: Network theory and spatial rainfall connections: an interpretation. J. Hydrol. **527**, 13–19 (2015)
9. Lakhanpal, A., Sehgal, V., Maheswaran, R., Khosa, R., Sridhar, V.: 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 (2017)
10. Sehgal, V., Lakhanpal, A., Maheswaran, R., Khosa, R., Sridhar, V.: Application of multi-scale wavelet entropy and multi-resolution Volterra models for climatic downscaling. J. Hydrol. (2016)
11. Atiem, I.A., Harmanciolu, N.B.: Assessment of regional floods using L-moments approach: the case of the River Nile. Water Resour. Manag. **20**(5), 723–747 (2006)
12. de Vos, N.J., Rientjes, T.H.M., Gupta, H.V.: Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering. Hydrol. Process. **24**(20), 2840–2850 (2010)
13. Yang, L., Tian, F., Smith, J.A., Hu, H.: Urban signatures in the spatial clustering of summer heavy rainfall events over the Beijing metropolitan region: urban modification of heavy rainfall. J. Geophys. Res. Atmos. **119**(3), 1203–1217 (2014)
14. Ramachandra Rao, A., Srinivas, V.V.: Regionalization of watersheds, vol. 58. Springer, Netherlands, Dordrecht (2008)

15. Burn, D.H., Boorman, D.B.: Estimation of hydrological parameters at ungauged catchments. J. Hydrol. **143**(3–4), 429–454 (1993)
16. Agarwal, A., Rathinasamy, M., Kurths, J., Khosa, R.: SOM technique. Unpublished (2016)
17. Bock, A.R., Hay, L.E., McCabe, G.J., Markstrom, S.L., Atkinson, R.D.: Parameter regionalization of a monthly water balance model for the conterminous United States. Hydrol. Earth Syst. Sci. **20**(7), 2861–2876 (2016)
18. Oudin, L. Andréassian, V. Perrin, C. Michel, C., Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: a comparison of regionalization approaches based on 913 French catchments. Water Resour. Res. **44**(3) (2008)
19. Razavi, T., Coulibaly, P.: Streamflow prediction in ungauged basins: review of regionalization methods. J. Hydrol. Eng. **18**(8), 958–975 (2013)
20. Agarwal, A. Marwan, N., Rathinasamy, M., Ozturk, U., Merz, B., Kurths, J.: Optimal design of hydrometric station networks based on complex network analysis. Hydrol. Earth Syst. Sci. Discuss. 1–21 (2018)
21. Jia, C. Li, Y., Carson, M.B., Wang, X., Yu, J.: Node attribute-enhanced community detection in complex networks. Sci. Rep. **7**(1) (2017)
22. Steinhaeuser, K., Chawla, N.V., Ganguly, A.R.: An exploration of climate data using complex networks. ACM SIGKDD Explor. Newsl. **12**(1), 25 (2010)
23. Steinhaeuser, K., Chawla, N.V., Ganguly, A.R.: Complex networks as a unified framework for descriptive analysis and predictive modeling in climate science. Stat. Anal. Data Min. **4**(5), 497–511 (2011)
24. Sun, A.Y., Xia, Y., Caldwell, T.G., Hao, Z.: Patterns of precipitation and soil moisture extremes in Texas, US: a complex network analysis. Adv. Water Resour. **112**, 203–213 (2018)
25. Tass, P., et al.: Detection of n: m phase locking from noisy data: application to magnetoencephalography. Phys. Rev. Lett. **81**(15), 3291–3294 (1998)
26. Stolbova, V., Martin, P., Bookhagen, B., Marwan, N., Kurths, J.: Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka. Nonlinear Process. Geophys. **21**(4), 901–917 (2014)
27. Quian Quiroga, R., Kraskov, A., Kreuz, T. Grassberger, P.: Performance of different synchronization measures in real data: a case study on electroencephalographic signals. Phys. Rev. E. **65**(4) (2002)
28. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Nat. Acad. Sci. **99**(12), 7821–7826 (2002) https://www.doi.org/10.1073/pnas.122653799
29. Agarwal, A., Marwan, N., Maheswaran, R., Merz, B., Kurths, J.: Quantifying the roles of single stations within homogeneous regions using complex network analysis. Hydrology (2018)
30. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010)
31. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech Theory Exp. **2008**(10), P10008 (2008)
32. Rubinov, M., Sporns, O.: Weight-conserving characterization of complex functional brain networks. NeuroImage **56**(4), 2068–2079 (2011)
33. Conticello, F., Cioffi, F., Merz, B., Lall, U.: An event synchronization method to link heavy rainfall events and large-scale atmospheric circulation features. Int. J. Climatol. (2017)
34. Boers, N., Bookhagen, B., Barbosa, H.M.J., Marwan, N., Kurths, J., Marengo, J.A.: Prediction of extreme floods in the eastern Central Andes based on a complex networks approach. Nat. Commun. **5**, 5199 (2014)
35. Sivakumar, B., Woldemeskel, F.M.: Complex networks for streamflow dynamics. Hydrol. Earth Syst. Sci. **18**(11), 4565–4578 (2014)
36. Bounova, G., de Weck, O.: Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles. Phys. Rev. E **85**(1) (2012)
37. Gilbert, E.N.: Random Graphs. Ann. Math. Stat. **30**(4), 1141–1144 (1959)
38. Newman, M.E.J., Strogatz, S.H., Watts, D.J.: Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E **64**(2) (2001)
39. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74**(1), 47–97 (2002)

40. Cohen, R., Havlin, S.: Scale-free networks are ultrasmall. Phys. Rev. Lett. **90**(5) (2003)
41. Österle, H., Werner, P., and Gerstengarbe, F.: Qualitätsprüfung, Ergänzung und Homogenisierung der täglichen Datenreihen in Deutschland, 1951–2003: ein neuer Datensatz, 7. Deutsche Klimatagung, Klimatrends: Vergangenheit und Zukunft, 9–11. Oktober 2006, München, (2006)
42. Agarwal, A., Marwan, N., Rathinasamy, M., Merz, B., Kurths, J.: Multi-scale event synchronization analysis for unravelling climate processes: a wavelet-based approach. Nonlinear Process. Geophys. **24**(4), 599–611 (2017)
43. Ozturk, U., Wendi, D., Crisologo, I., Riemer, A., Agarwal, A., Vogel, K., López-Tarazón, J. A. and Korup, O.: Rare flash floods and debris flows in southern Germany. Sci. Total Environ. **626**, 941–952 (2018) https://doi.org/10.1016/j.scitotenv.2018.01.172
44. Barabási, A.-L. and Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999) https://doi.org/10.1126/science.286.5439.509
45. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440–442 (1998)
46. Newman, M.E.J.: Detecting community structure in networks. Eur. Phys. J. B Condens. Matter **38**(2), 321–330 (2004) https://www.doi.org/10.1140/epjb/e2004-00124-y
47. Chen, J., Lu, J., Lu, X., Wu, X., Chen, G.: Spectral coarse graining of complex clustered networks. Commun. Nonlinear Sci. Numer. Simul. **18**(11), 3036–3045 (2013)
48. Gfeller, D., De Los Rios, P.: Spectral coarse graining of complex networks. Phys. Rev. Lett. **99**(3) (2007)
49. Hsu, K.-C., Li, S.-T.: Clustering spatial–temporal precipitation data using wavelet transform and self-organizing map neural network. Adv. Water Resour. **33**(2), 190–200 (2010)
50. Rheinwalt, A., Boers, N., Marwan, N., Kurths, J., Hoffmann, P., Gerstengarbe, F.-W. and Werner, P.: Non-linear time series analysis of precipitation events using regional climate networks for Germany. Climate Dynamics **46**(3–4), 1065–1074 (2016) https://www.doi.org/10.1007/s00382-015-2632-z