Optimal Design of Hydrometric Station Networks Based on Complex Network Analysis

Agarwal Ankit[1,2,3,4,*], Marwan Norbert[3], Maheswaran Rathinasamy[5], Ozturk Ugur[2], Kurths Jürgen[2,3], and Merz Bruno[1,2]

[1]GFZ German Research Centre for Geosciences, Section 4.4: Hydrology, Telegrafenberg, Potsdam, 14473 Germany
[2]Institute for Environmental Sciences and Geography, University of Potsdam, Potsdam, 14476 Germany
[3]Research division Complexity Science, Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Telegrafenberg, Potsdam, 14473 Germany
[4]Department of Hydrology, Indian Institute of Technology Roorkee, 247667, India
[5]Department of Civil Engineering, MVGR College of Engineering, Vizianagaram, 535005, India

[*]Corresponding author: ankit.agarwal@hy.iitr.ac.in

Abstract

Hydrometric networks play a vital role in providing information for decision-making in water resources management. They should be set up optimally to provide as much and as accurate information as possible, and at the same time, be cost-effective. Although the design of hydrometric networks is a well-identified problem in hydrometeorology and has received considerable attention, yet there is scope for further advancement. In this study, we use complex network analysis, defined as a collection of nodes interconnected by links, to propose a new measure that identifies critical nodes of station networks. The approach can support the design and redesign of hydrometric station networks. The science of complex networks is a relatively young field and has gained significant momentum in the last years in different areas such as brain networks, social networks, technological networks, or climate networks. The identification of influential nodes in complex networks is an important field of research. We propose a new node ranking measure, the weighted degree-betweenness (WDB), to evaluate the importance of nodes in a network. It is compared to previously proposed measures on synthetic sample networks and then applied to a real-world rain gauge network comprising 1229 stations across Germany to demonstrate its applicability. The proposed measure is evaluated using the decline rate of network efficiency and the kriging error. The results suggest that WDB effectively quantifies the importance of rain gauges, although the benefit of the method needs to be investigated in more detail.

Keywords: Rainfall network, complex networks, event synchronization, kriging error.

1 Introduction

Hydrometric observation networks monitor a wide range of water quantity and water quality parameters such as precipitation, streamflow, groundwater, or surface water temperature (Keum et al., 2017). Designing adequate hydrometric monitoring is key in water resources management; e.g., flood estimation, water budget analysis,
5 hydraulic design, and monitoring climate change. Even after the advent of remote sensing based information, such as satellite precipitation estimates, in-situ observations are considered as an essential source of information in hydrometeorology (Rossi et al., 2017).

The basic characteristics of hydrometric networks comprise the number of stations, their locations, observation periods, and sampling frequency (Keum et al., 2017). The general understanding is that the higher the number of
10 monitoring stations, the more reliable the quantification of areal average estimates and point estimates at any ungauged location. However, a higher station number elevates the cost of installation, operation, and maintenance, but may provide redundant information and, therefore, not increase the information content obtained from the observation network. Scarcity of funds for hydrometric monitoring has led to a slow but steady teardown of hydrometric stations in the last decades globally, raising the need for cost-effective design (Mishra and Coulibaly,
15 2009). For example, Putthividhya and Tanaka (2012) made an effort to design an optimal rain gauge network based on the station redundancy and the homogeneity of the rainfall distribution. Adhikary et al. (2015) proposed a kriging based geostatistical approach for optimizing rainfall networks, and Chacon-Hurtado et al. (2017) provided a generalized procedure for optimal rainfall and streamflow monitoring in the context of rainfall-runoff modelling. Yeh et al. (2017) optimized a rain gauge network applying the entropy method on radar datasets. Most of the
20 aforementioned studies inherently assume that expanding the gauge network with supplementary stations aids more information that ultimately leads to less uncertainty (Wadoux et al., 2017). However, increasing the number of stations does not necessarily decrease uncertainty (Stosic et al., 2017). There may be expendable (little significant) stations contributing little to no information though having the same maintenance cost as influential (highly significant) stations (Mishra and Coulibaly, 2009).

25 This study aims to discriminate influential and expendable stations in hydrometric station networks based on their relative information content. We propose complex networks as a suitable tool for this optimization problem. A complex network is defined as a collection of nodes, such as rain gauge stations, interconnected with links, where a link represents statistical similarity of the connected rain gauge stations. Complex networks are powerful tools in extracting information from large high-dimensional datasets (Donges et al., 2009; Kurths et al., 2019). This non-
30 parametric method allows investigating the topology of local and non-local statistical interrelationships. An example

for non-local connections in a climate network, i.e., a complex network using climate variables, are the global influence of El Niño Southern Oscillation (ENSO) on regional rainfall (Agarwal, 2019; Ferster et al., 2018), and of the Atlantic Meridional Overturning Circulation (AMOC) on air surface temperature (Agarwal et al., 2019) via teleconnections and ocean circulation, respectively. Once the spatial network of stations has been constructed, statistical network measures (e.g., degree, betweenness centrality) are used to quantify the behaviour of the network and its components for a range of applications. Examples are the identification of the community structure of stations or homogeneous regions to unravel dominant climate modes (Agarwal et al., 2018a; Halverson and Fleming, 2015), catchment classification indicating hydrologic similarity (Fang et al., 2017), short and long-range spatial connections in rainfall (Agarwal et al., 2018a; Boers et al., 2014; Jha et al., 2015) and spatio-temporal hydrologic patterns (Halverson and Fleming, 2015; Konapala and Mishra, 2017). Complex network analysis complements classical Eigen techniques, such as empirical orthogonal functions (EOFs) or coupled patterns (CP) maximum covariance analysis (Donges et al., 2015). EOFs, CPs, and related methods rely on dimensionality reduction, whereas complex network approach allows studying the full complexity and different aspects of the statistical interdependence structure and are not limited to linear and spatial-proximity connections. Also, higher-order complex network measures (betweenness centrality, closeness centrality, participation coefficient) provide additional information on the hidden structure of statistical interrelationships in climatological data (Donges et al., 2015).

In this study, we propose a complex network-based method to identify the influential and expendable stations in a rainfall network. Several methods in the field of complex networks have been proposed to evaluate the importance of nodes (Chen et al., 2012; Hou et al., 2012; Jensen et al., 2016a; Kitsak et al., 2010; Zhang et al., 2013), however, the application and interpretation of complex network in hydrology (or meteorological observations) is in its infancy. Degree (k), betweenness centrality (B), and closeness centrality (CC) are measures commonly used in complex networks (Gao et al., 2013). Studies in different disciplines have shown that degree and betweenness centrality often outperform other node-ranking measures (Gao et al., 2013; Liu et al., 2016). We propose a novel measure, the weighted degree-betweenness (WDB), which combines k and B, to identify the stations providing the largest information to the network. Our main objective is to develop a node ranking method using complex network theory that can be used to identify not only influential but also the expendable stations in large hydrometric station networks. Our study is a first effort to explore the benefit of complex networks in hydrology, and we acknowledge that further studies are necessary before the methodology can be considered a trustworthy optimization tool for measurement networks. Our aim is not to question the credibility of operating stations, but to propose an alternative evaluation procedure towards optimal design and redesign of observational hydrometric monitoring networks based on complex networks.

2 Basics of Complex Networks

2.1 Network Construction

A network or a graph is a collection of entities (nodes, vertices) interconnected with lines (links, edges) as shown in Fig. 1. These entities could be anything, such as humans defining a social network (Arenas et al., 2008), computers constructing a web network (Zlatić et al., 2006), neurons forming brain networks (Bullmore and Sporns, 2012), streamflow stations creating a hydrological network (Halverson and Fleming, 2015) or climate stations describing a climate network (Agarwal et al., 2018b). Formally, a network or graph is defined as an ordered pair $Z = \{N, E\}$; containing a set $N = \{N_1, N_2, \ldots N_N\}$, of nodes together with a set E of link $\{i, j\}$, which are 2-element subsets of N. In this work, we consider undirected and unweighted simple networks, where only one link can exist between a pair of vertices and self-loops of the type $\{i, i\}$ are not allowed. This type of network can be represented by the symmetric adjacency matrix (Eq..1).

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

(1)

$A_{i,j} = 1$ denotes a link between the ith and j^{th} station, and 0 denotes otherwise. The adjacency matrix represents the connections in the network. Fig. 1 is a simple representation of such a network, i.e., one with a set of identical nodes ($N_i, where\ i = 1\ to\ 4$) connected by identical links. In general, (large) networks of real-world entities with irregular topology are called complex networks. The links represent similar evolution or variability at different nodes and can be identified from data using a similarity measure such as Pearson correlation (Ekhtiari et al., 2019), synchronization (Agarwal et al., 2017; Boers et al., 2019; Conticello et al., 2018) or mutual information (Paluš, 2018).

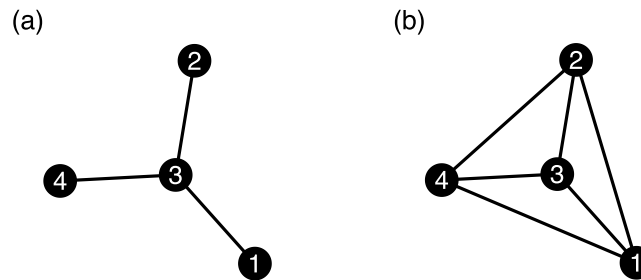



Figure 1: Topology of two sample networks to explain network structures and measures. (a) Network N1 with four nodes and three links; (b) network N2 with four nodes and six links.

2.2 Event synchronization

Event synchronization (ES) has been specifically designed to calculate nonlinear correlations among bivariate time series with events defined on them (Quiroga et al., 2002). This method has advantages over other time-delayed

correlation techniques (e.g., Pearson lag correlation), as it allows us to investigate extreme event series (such as non-Gaussian and event-like data sets) and uses a dynamic time delay (Ozturk et al., 2018). The latter refers to a time delay that is adjusted according to the two time series being compared, which allows for better adaptability to the variable and region of interest. Various extensions for ES have been proposed, addressing, for instance, boundary
5   effects (Rheinwalt et al., 2016) and bias by varying event rates.

In the following, we define events by applying an α percentile threshold at the signals $x(t)$ and $y(t)$. The α percentile threshold is selected to trade-off between a sufficient number of rainfall events at each location and a rather high threshold to study heavy precipitation. Events occur then at times t_l^x and t_m^y where $l = 1,2,3,4 \dots S_x$, $m = 1,2,3,4 \dots \dots S_y$. Events in $x(t)$ and $y(t)$ are considered to coincide if they occur within a time lag $\pm\tau_{lm}^{xy}$, which is
10   defined as follows

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

where S_x and S_y are the total number of such events (greater than threshold α) that occurred in the signal $x(t)$ and $y(t)$, respectively. The above definition of the time lag helps to separate independent events, which in turn allows to take into account the fact that different processes may be 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
15   achieved by defining quantities $C(x|y)$ and $C(y|x)$ where

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

and

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

This definition of J_{xy} prevents counting a synchronized event twice. When two synchronized events match exactly $(t_l^x = t_m^y)$, we use a factor $1/2$ since they double count in $C(x|y)$ and $C(y|x)$. Similarly, we can define $C(y|x)$ 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 normalized measure of the strength of event synchronization between signal $x(t)$ and $y(t)$. This implies
20    $Q_{xy} = 1$ for perfect synchronization and $Q_{xy} = 0$ if no events are synchronized. After repeating this procedure for

all pairs ($x \neq y$) of stations, we obtain a similarity matrix. In this case, the similarity matrix for precipitation data is a square, symmetric matrix, which represents the strength of synchronization of the extreme rainfall events between each pair of stations.

5 **2.3 Node Ranking Measures**

A large number of measures have been defined to characterize the behaviour of complex networks. We focus here on those traditional and contemporary network measures, which have been proposed to quantify the importance of nodes in a network: degree k, betweenness centrality B (Agarwal et al., 2018a), bridgeness Bri (Jensen et al., 2016), and degree and influence of line DIL (Liu et al., 2016).

10 *Traditional network measures*

The degree k of a node in a network counts the number of connections linked to the node directly. The degree of any i node is calculated as

$$k_i = \sum_{j=1}^{N} A_{i,j} \tag{6}$$

where N is the total number of nodes in a network. For example, the degree of nodes 1, 2, and 4 in network N1 (Fig. 1a) is 1 and for node 3 is 3. In the network N2 (Fig. 1b), all nodes have degree 3. The degree can explain the importance of nodes to some extent, but nodes that own the same degree may not play the same role in a network. For instance, a bridging node connecting two important nodes might be very relevant though its degree could be much lower than the value of less important nodes.

The betweenness centrality B is a measure of control that a particular node exerts over the interaction between the remaining nodes. In simple words, B describes the ability of nodes to control the information flow in networks. To calculate betweenness centrality, we consider every pair of nodes and count how many times a third node can interrupt the shortest paths between the selected node pair. Mathematically, betweenness centrality B of any i node is

$$B_i = \sum_{i \neq j \neq v \in \{V\}}^{N} \frac{\sigma_i(j,k)}{\sigma(j,k)} \tag{7}$$

where $\sigma(j,k)$ represents the number of links along the shortest path between node j and k; while $\sigma_i(j,k)$ is the number of links of the shortest path running through node i. In network N1 (Fig. 1a), B of node 3 is 3, i.e., node 3 can disturb the information transfer between all of the three pairs 1-2, 1-4, 2-4, and for other nodes $B = 0$. In the network

N2 (Fig. 1b), all nodes have $B = 0$ because no node can interrupt the information flow. Thus, node 3 is a critical node in the network N1 but not in the network N2.

Contemporary network measures

Jensen et al. (2016) developed the Bridgeness measure Bri to distinguish local centres, i.e. nodes that are highly

5   connected to a part of the network (e.g. highly correlated stations in a homogeneous region), from global bridges, i.e., nodes that connect different parts of a network (Fig. 2, e.g., teleconnection between Indian rainfall and climate indices).

Bri is a decomposition of betweenness centrality B into a local and a global contribution. Therefore, the Bri value of node i is always smaller or equal to the corresponding B value, and they only differ by the local contribution of the

10   first direct neighbours. To calculate Bri we consider the shortest path between nodes outside the neighbourhood of node i, $N_G(i)$. Mathematically, it is represented as

$$Bri_i = \sum_{j \notin N_G(i) \, \vee \, k \notin N_G(i)}^{N} \frac{\sigma_i(j,k)}{\sigma(j,k)} \tag{8}$$

The neighbourhood of node i ($N_G(i)$) consists of all direct neighbours of node i. For example, in the networks N1 and N2, all nodes (except node 3 in N1) have $B = 0$ hence $Bri = 0$. However, node 3 in the network N1 has all the nodes in direct neighbourhood hence, it also has $Bri = 0$.

15   The degree and influence of line (DIL), introduced by Liu et al. (2016), considers the node degree k and importance of line I to rank the nodes in a network:

$$DIL_i = k_i + \sum_{j = N_G(i))} I_{e_{ij}} \cdot \frac{k_i - 1}{k_i + k_j - 2} \tag{9}$$

where the line between node i and j is e_{ij} and its importance is defined as $I_{e_{ij}} = \frac{U}{\lambda}$ where $U = (k_i - p - 1) \cdot (k_j - p - 1)$ reflects the connectivity ability of a line (link), p is the number of triangles having one edge e_{ij} and $\lambda = \frac{p}{2} + 1$ is defined as an alternative index of line e_{ij}. $N_G(i))$ is the set of neighbours of node i (for detailed explanation see Liu et

20   al., 2016). The equation for DIL suggests that all the nodes having $k_i = 1$ will have $DIL_i = 1$, since the second term of the equation will be zero. Hence, in the network N1 all nodes, except node 3, have $DIL = 1$. Node 3 has $DIL = 3$ equal to its degree, since the second term is zero (all the connected nodes 1, 2 and 4 have $k_j = 1$, hence $I_{e_{ij}} = 0$). All the nodes in the network N2 have $DIL = 3$.

3 Methodology

We will first propose a new node ranking measure that we call weighted degree-betweenness (WDB). We will then compare the efficacy of this measure with the existing traditional and contemporary node ranking methods using two synthetic networks.

3.1 Weighted Degree-Betweenness

WDB is a combination of two network measures, degree and betweenness centrality. We define WDB of a particular node i as the sum of the betweenness centrality of node i and all directly connected nodes $j, j = 1,2,3 \dots k_i$ in proportion to their contribution to node i. The WDB of a node i is given by

$$WDB_i = B_i + I_i \tag{10}$$

where B_i is the betweenness centrality of node i, and I_i stands for the cumulative effect of the influence or contribution of the directly connected nodes of i, which are $j = 1,2,3, \dots, k_i$, as

$$I_i = \sum_{j=1}^{k_i} \frac{B_j * (k_j - 1)}{(k_i + k_j - 2)} \tag{11}$$

where k_i is the degree of node i, k_j is the degree of the nodes j which are directly connected to node i.

3.2 Comparison with Existing Node Ranking Measures Using Synthetic Networks

In this section, we motivate the development of the new node ranking measure WDB by comparing it to existing measures. Identifying nodes that occupy interesting positions in a real-world network using node ranking helps to extract meaningful information from large datasets with little cost. Usually, the measures degree (k_i) and betweenness centrality (B_i) are common node ranking metrics (Gao et al., 2013; Okamoto et al., 2008; Saxena et al., 2016). The network measures k_i, B_i and WDB_i of each node are given for an undirected and unweighted network $Z = (N, E)$ with 8 nodes and 11 edges shown in Fig. 2 along with the node number.

In general, high degree nodes represent most connected (highly correlated) nodes in a network. Rheinwalt et al. (2015) considered these highly correlated nodes of homogeneous precipitation community as local centres representing homogenous precipitation patterns for that particular community. Agarwal et al., (2018) defined local centres as the nodes having maximum intra-community links and minimum inter-community links based on the Z-P space approach. However, degree alone cannot distinguish the roles of nodes in the sample network as seen for nodes 5, 7, and 8, which have the same degree (k_i=2), though node 5 serves as a bridge node linking the two parts of the

network. In a larger complex network, such bridge nodes have strategic relevance as most of the information can be accessed quickly just by capturing those nodes. For example, Kurths et al. (2019) quantified the spatial diversity of Indian rainfall teleconnections at different timescales by identifying linkages between climatic indices (e.g., El Niño/Southern Oscillation, Indian Ocean Dipole, North Atlantic Oscillation, Pacific Decadal Oscillation, and Atlantic
5   Multidecadal Oscillation) and seven Indian rainfall stations (bridge nodes).

Betweenness centrality has a higher power in significantly discriminating different roles compared to k_i. For example, nodes 4 and 5 have the highest B_i ($B_4 = B_5 = 24$) followed by node 6 ($B_6 = 20$). On the other hand, B_i gives equal scores to local centers (node 4), i.e., nodes of high k_i to a single region, and to global bridges (node 5), which connect detached regions. As mentioned, global bridges connect different parts of a network (e.g., teleconnection
10   between Indian rainfall and ENSO). Measuring and interpretation of large spatial variability, process identification, interpolation of measurements and transferability of precipitation measurements across locations, would be limited in the absence of high B_i nodes.

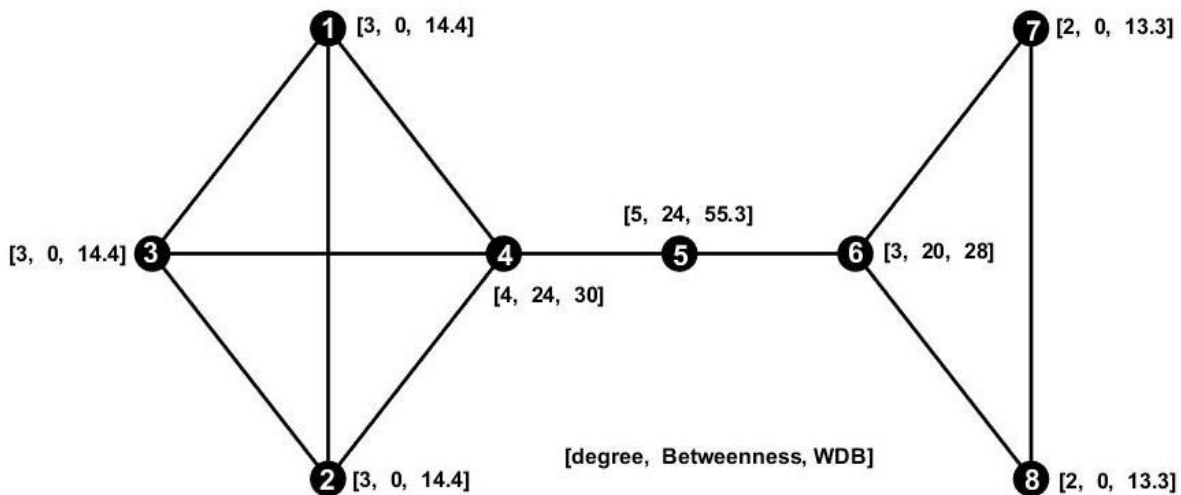



**Figure 2: Synthetic network to explain the degree *(k)*, betweenness centrality (B) and weighted degree-
15   betweenness (WDB) measures, with node number (1 to 8) followed by the degree, betweenness centrality value and WDB value in brackets *[k, B, WDB]*. Degree and betweenness are limited in distinguishing the role of different nodes in the network and centers from bridges, respectively.**

The proposed measure *WDB* has higher discrimination power compared to betweenness centrality. Node 5 has the highest *WDB* score and is ranked as the most influential node, which reflects its role as a global bridge node. *WDB*

distinguishes between nodes 1, 2, 3 (*WDB* = 14.4), and nodes 7, 8 (*WDB* = 13.3), which is important in case we need to sequentially rank nodes.

We further evaluate *WDB* with the network measures *Bri*. For this comparison, we use the same synthetic network as Jensen et al. (2016) shown in Fig. 3. Betweenness centrality once again assigns a smaller value to the global bridge (node 6) than to the local centers (nodes 4, 7). Bridgeness expresses the higher importance of node 6 compared to nodes 4, 7, however, it does not distinguish between all other nodes in the network (nodes 1, 2, 3…have *Bri* = 0). Similarly, DIL misses representing the bridge nodes by assigning higher values to local centres. *WDB* ranks the nodes, preferably following their role in the network as global bridges, local centers, and end nodes. For example, WDB is also able to differentiate between nodes 4 and 7 for which the bridgeness measure provides equal scores.

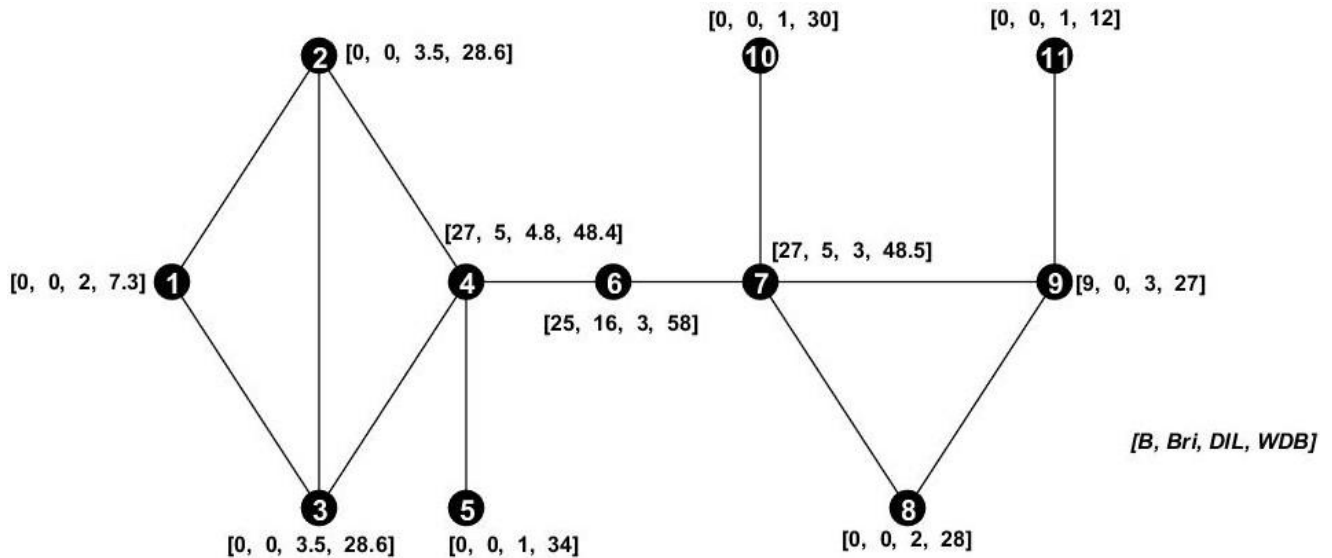



Figure 3: Synthetic network used to compare the network measures betweenness centrality, bridgeness, and DIL with the proposed measure WDB. Numbers 1 to 11 are node counts, and values in brackets represent the network measure values in order of $[B, Bri, DIL, and\ WDB]$. Node 6 is a global bridge node that connects two sub-networks. Node 4 and 7 are hubs that are connected to most of the nodes in the sub-networks. Node 5, 10, and 11 are the dead-end nodes.

3.3 Evaluation of the Proposed Measure for a Rain Gauge Network

In the context of hydrometric station networks, we hypothesize that higher ranking nodes are more influential stations in the complex network and also in the observation network. Losing such stations could reduce the network stability and efficiency given their role as bridging different communities (processes), capturing detailed process information compared to lower ranking stations. Stations with the lowest ranks in the network are the least influential and are seen as expendable stations. For example, a bridging node would be located between two regions of different variability and plays, therefore, an important role in estimating the spatial border between these regions. A low ranked node would be located within a (more or less) homogenous region and would not provide additional knowledge about the spatial variability. To test this hypothesis, we apply the proposed node ranking measure to a hydrometric station network, consisting of more than 1000 stations in Germany. The benefit of WDB is to capture the bridge nodes in the hydrometric station network that are adequate to quantify the local and non-local rainfall variability for process identification, for interpolation of measurements and for transferability of precipitation measurements across locations. In contrast, expandable stations correspond to sites of spatially extended coherent rainfall, surrounding a local centre which represents the variability of such regions. Stations within such regions of coherent rainfall provide redundant information and can be removed (except the local centre) without loss of information. The information loss caused by removing stations is quantified by two measures: (a) decline rate of network efficiency, and (b) relative kriging error.

Decline Rate of Network Efficiency

The decline rate of network efficiency quantifies the decrease in information flows within a network when nodes are removed as,

$$\eta = \frac{1}{N(N-1)} \sum_{n_i \neq n_j} \eta_{ij} \tag{12}$$

where N is the total number of nodes in a network. η_{ij} is the efficiency between nodes n_i and n_j. η_{ij} is inversely related to the shortest path length: $\eta_{ij} = 1/d_{ij}$, where d_{ij} is the shortest path between nodes n_i and n_j. The average path length L measures the average number of links along the shortest paths between all possible pairs of network nodes. A network with small L is highly efficient, because two nodes are likely to be separated by a few links only. The decline rate of network efficiency μ is defined as,

$$\mu = 1 - \frac{\eta_{new}}{\eta_{old}} \tag{13}$$

where η_{new} is the efficiency of the network after removing nodes, and η_{old} is the efficiency of the complete network.

We hypothesise that the network efficiency decreases more strongly when higher ranking stations are removed, i.e., bridge nodes.

Relative Kriging Error

As second measure to evaluate the information loss, when stations are removed from the network, we use a kriging based geostatistical approach (Adhikary et al., 2015; Keum et al., 2017). Kriging is an optimal surface interpolation technique assuming that the variance in a sample of observations depends on their distance (Adhikary et al., 2015). The algorithm estimates unknown variable values at unsampled locations in space, where no measurements are available, based on the known sampling values from the surrounding areas (Hohn, 1991; Webster and Oliver, 2007). Ordinary Kriging is used in this study for interpolating rainfall data and estimating the kriging error. The kriging estimator is expressed as

$$Z^*(x_o) = \sum_{i=1}^{n} w_i Z(x_i) \tag{14}$$

where $Z^*(x_0)$ refers to the estimated value of Z at the desired location x_0; w_i represents weights associated with the observation at the location x_i with respect to x_o, and n indicates the number of observations within the domain of the search neighbourhood of x_0 for performing the estimation of $Z^*(x_0)$. Ordinary Kriging is implemented through ArcGISv10.4.1 (Redlands, CA, USA) and its geostatistical analyst extension (Johnston et al., 2001).

The kriging variance $\sigma_z^2(x_o)$ in the Ordinary Kriging can be computed as (Adhikary et al., 2015; Xu et al., 2018)

$$\sigma_z^2 = \mu_z + \sum_{i=1}^{n} w_i \gamma(h_{oi}) \quad with \sum_{i=1}^{n} w_i = 1$$

where $\gamma(h)$ is the variogram value for the distance h; h_{0i} is the distance between observed data points x_i and x_j; μ_z is the Lagrangian multiplier in the Z scale; h_{0j} is the distance between the unsampled location x_0 (where the estimation is desired) and sample locations x_i; and n is the number of sample locations.

The square root of the kriging variance, also named as kriging standard error (KSE), is used as a gauge network evaluation factor. We estimate the increase in the kriging standard error across the study area when stations are removed to evaluate the performance of the WDB measure in identifying influential and expendable stations in a large network.

The relative kriging error before and after removing the stations is denoted as

$$\Re(\%) = \frac{KSE_{new} - KSE_{old}}{KSE_{old}} \times 100 \tag{15}$$

where KSE_{new} denotes the standard kriging error after removing stations, and KSE_{old} is the error for the original network. We hypothesise that the increase in the relative kriging error is higher when removing high ranking stations. To cover a broad range of rainfall characteristics, the error is calculated for different statistics, i.e., the mean, 90th, 95[th,] and 99th percentile rainfall and the number of wet days (precipitation > 2.5mm).

5   **4 Application to an Extensive Rain Gauge Network**

4.1 Rainfall Data

To evaluate the proposed measure in the context of the optimal design of hydrometric networks, we apply it to an extensive network of rain stations in Germany and adjacent areas (Fig. 4). The data covers 110 years at daily resolution (1 January 1901 to 31 December 2010). The 1229 rain stations in Germany (blue dots in Fig. 4) are

10   operated by the German Weather Service. Data processing and quality control were performed according to Österle et al. (2006), and in this study, we assume that data is free from measurement errors. 211 stations from different sources outside Germany (red dots in Fig. 4) were included in the analysis to minimize spatial boundary effects in the network construction; however, these stations were excluded from the node ranking analysis. For parts of France, precipitation data on a 0.22° x 0.22° rotated pole grid from E-OBS is used (Haylock 217 et al., 2008).

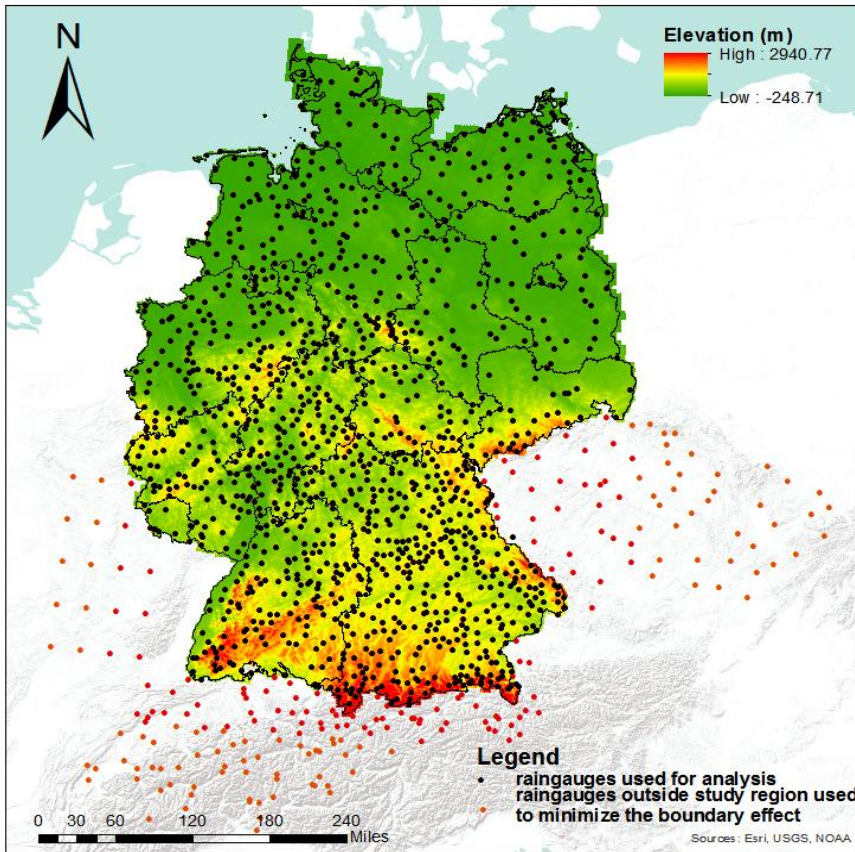



Figure 4: Location of rain stations in Germany and adjacent areas. Black dots indicate stations lying inside Germany that are used in the analysis. Red dots indicate stations outside of Germany that are used for network construction only to minimize the boundary effect. © Esri, USGS, NOAA

4.2 Network Construction

We begin the network construction by extracting event time series from the 1229 daily rainfall time series. The event series represent heavy rainfall events, i.e., precipitation exceeding the $\alpha = 95th$ percentile at that station (Rheinwalt et al., 2016). The 95th percentile is a trade-off between having a sufficient number of rainfall events at each location

10 and a rather high threshold to study heavy precipitation. All rainfall event series are compared with each other using event synchronization (section 2.2), which is the base for deriving a complex network. This results in the similarity

matrix Q, where the entry at index pair (*i,j*) defines synchronization in the occurrence of heavy rainfall events at station *i* and station *j* (Eq. 5).

Applying a certain threshold (θ) to the Q matrix yields the adjacency matrix (Eq. 1). Here, θ_{xy}^{Q} is a chosen threshold, and $A_{ij} = 1$ denotes a link between the *ith* and *jth* sites, and $A_{ij} = 0$ denotes otherwise. The adjacency matrix

5 represents a rain gauge network, and complex network theory can subsequently be employed to reveal properties of the given network.

Two criteria have been proposed to generate an adjacency matrix from a similarity matrix, such as the fixed amount of link density (Agarwal et al., 2018b, 2019) or global fixed thresholds (Jha et al., 2015; Sivakumar and Woldemeskel, 2014). However, both criteria are subjective and may lead to the presence of weak and non-significant links in the

10 complex network. These non-significant links might obscure the topology of strong and significant connections. To minimize these threshold effects, we choose the threshold $\theta_{i,j}^{Q}$ objectively by considering all links in the network that are significant. A link is significant (i.e., two stations are significantly synchronized) if the synchronization value exceeds the $\theta_{i,j}^{Q}$ =95th percentile (corresponding to a 5% significance level) of the synchronization obtained by two synthetic variables that have the same number of events but distributed randomly in the time series (i.e., both event

15 series are independent). We calculate ES for 100 pairs of such random time series and derive the 95th percentile of the resulting ES distribution. Using this 5% significance level, we assume that synchronization cannot be explained by chance if the ES value between two stations is larger than the 95th percentile of the test distribution. Here, we select the 5% significance level since it is a well-accepted criterion in general in statistics. To validate the results, we repeated the analysis for the 90-99th percentile threshold range and observed that the node ranking is robust against

20 the threshold selection. For the sake of brevity, detailed results are presented for the 95th percentile threshold only.

4.3 Decline Rate of Network Efficiency

In this section, we evaluate the ranking of stations derived from the proposed WDB measure using the decline rate of network efficiency. The rain gauges are ranked in decreasing order according to their WDB values. Highly ranked rain gauges are interpreted as the most influential stations, and low ranked as expendable stations.

25 Firstly, we analyze the decline rate of network efficiency μ when one station is removed from the network. In each trial, we remove only one station (starting with the highest rank). After *n*=1229 (number of nodes) trials, we investigate the relationship between μ and the node ranking measured by WDB. We expect an inverse relationship between μ and WDB: the higher the node ranking, the more important is that node, leading to a higher loss in network efficiency (Fig. 5). μ is high for high-ranking stations and decays with node ranking. Interestingly, μ < 0 for very low

ranking stations, i.e., the network efficiency increases when single, low ranking stations are removed. This is explained by the decrease of the redundancy in the network when such stations are removed.

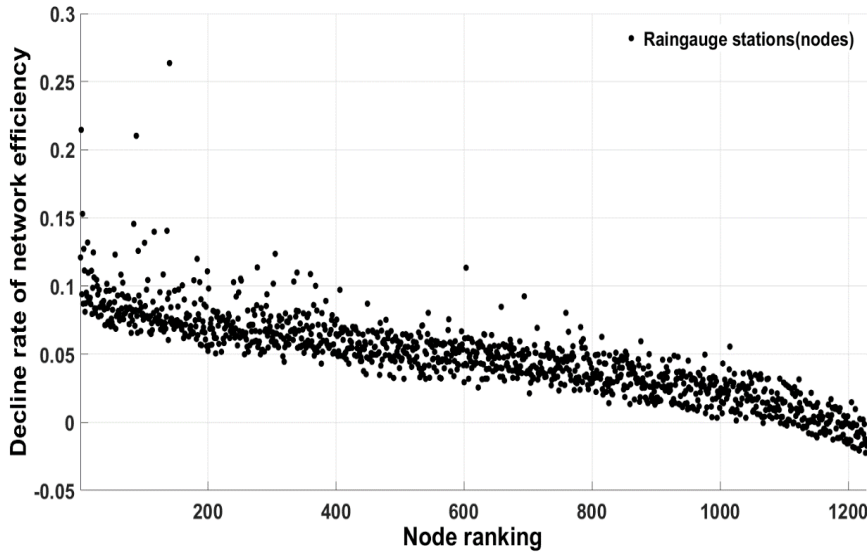



Figure 5: Decline rate of network efficiency corresponding to the removal of each node in the rainfall
5 **network. In each implementation, only one node is removed from the network according to the ranking with replacement (bootstrapping).**

Secondly, we remove successively a larger number of stations, from 1 to 123 stations (10%), considering three cases. In case I, we remove up to the 10% highest ranking stations. This implies that in the first iteration, we remove the top-ranked station, and in the second iteration, we remove the top two stations and so on. Fig. 6 shows an apparent
10 increase in μ when more and more influential stations are removed. In case II, up to the 10% lowest ranking stations are successively removed. The efficiency increases when the lowest ranking stations are removed. In case III, up to 10% stations are randomly removed. Case III is repeated ten times to understand the effect of random sampling. In general, μ increases with removing random stations. However, the effect is much lower (in absolute terms) compared to the effect of removing high or low ranking stations, respectively. The variation in μ between the ten trials and
15 within one trial is caused by randomness. For example, μ rises instantaneously when the algorithm picks up a high ranking station.

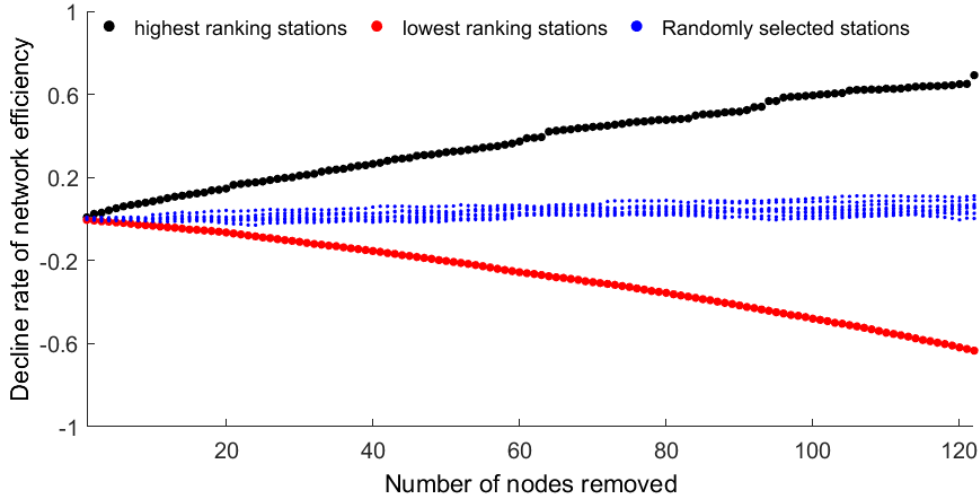



Figure 6: Decline rate of network efficiency as a function of the number of stations removed from the network. Case I: up to the 10% highest ranking stations are removed (black), case II: up to the 10% lowest ranking stations are removed (red), case III: up to 10% randomly drawn stations are removed (10 trials) (blue).

4.4 Relative Kriging Error ($\mathfrak{R}$)

As the second approach to assess the suitability of WDB for identifying influential and expendable stations, we analyse the change in the kriging error ($\mathfrak{R}$) when stations are removed from the network. We first estimate the kriging standard error KSE_{old} across the study area for all 1229 stations. Then, we measure the kriging standard error across the study area when stations are removed (K_{new}), and calculate the change in the error (eq.15). The variogram is kept constant during the network modifications. Similar to the evaluation using the decline rate of network efficiency in section 4.3, three cases are investigated: removing the 10% highest ranking stations, removing the 10% lowest ranking stations, and ten trials of removing 10% of the stations randomly.

The change in the kriging error is calculated for five characteristics, i.e., mean, 90%-, 95%-, 99%-percentile, and number of wet days (Table 1). For each case and rainfall characteristics, we run the model 100 times; the mean value of $\mathfrak{R}$ is reported in Table 1.

Removing the 10% high-ranking stations (case I) leads to positive and high (between 12 and 73%) relative kriging errors for all five statistics considered, i.e., the kriging error increases substantially when these stations are removed. In contrast, when the 10% lowest ranking stations (case II) are not considered, the $\mathfrak{R}$ values are small. The relative

errors in estimating the mean, percentile rainfall characteristics (90th and 95th) and number of wet days at ungauged locations are lower than 5%, suggesting that these stations do not contribute much information. Case III, i.e., removing stations randomly, results in rather high errors (between 5 and 51%), however, they are much smaller than case I.

5 **Table 1: Relative kriging error for the three different cases. The relative kriging error for case III is the average across ten trials. Stars indicate a high relative error >5%.**

Case	Removal of stations	Relative kriging error $\Re(\%)$				
		Mean	90th percentile	95th percentile	99th percentile	Wet days
I	10% highest ranking	11.7*	29.9*	73.3*	58.1*	62.1*
II	10% lowest ranking	0.09	4.2	3.7	8.1*	2.9
III	10% randomly selected	6.4*	23.3*	51.3*	46.6*	4.7

5 Discussion

Building on the young science of complex networks, a novel node ranking measure, the weighted degree-betweenness WDB, is proposed. The proposed method based on degree and betweenness centrality does not only
10 account for the local (captured by degree) and global (captured by betweenness centrality) characteristics of nodes but also for the cumulative contribution of the directly connected (localized) nodes. We compared WDB with other traditional (i.e. degree and betweenness centralities) and contemporary (i.e., Bridgeness and DIL) measures by applying it to prototypical situations. The results show that degree and betweenness centrality are unable to differentiate between different roles of a node in a network. Although the contemporary network measures
15 Bridgeness and DIL showed higher power in discriminating different roles, they do not provide a nuanced picture of marginal differences, for example, between a local centre and a global bridge. Hence, our tests with synthetic networks suggest that *WDB* is superior in distinguishing different roles, compared to existing measures, and provides a unique value to each node depending on its importance and influence in our test networks.

Besides this methodological development, this study proposes to use WDB for supporting the optimal design of large hydrometric networks. Its preliminary application to the German rain gauge network shows its ability to rank the nodes in such large hydrometric networks. For example, removing low ranking stations does not have an adverse impact on network efficiency, and kriging errors are hardly increasing. This is explained by the redundancy in the

5   information that those stations provide, which in turn is attributed to the similarity between the gauges due to common driving mechanisms or spatial similarity, as advocated by Tobler's Law of Geography (Tobler, 1970). Our analysis suggests that *WDB* identifies the expendable nodes correctly, as shown by the decline rate of efficiency and the insignificant change in the relative kriging error. On the other hand, WDB awards stations that provide unique information as it considers different aspects of the spatio-temporal relationships in the observation network.

10 We further analyse the characteristics of the stations with the highest ranks. We plot the network (Fig. 7a) corresponding to the 10% (~122) highest ranking stations, i.e., all the links originating only from these 122 stations. The size and colour of each diamond-shaped rain gauge mark their degree and betweenness centrality. All other stations are plotted in the background without highlighting their degree and betweenness. We further plot the connections corresponding to two high ranking stations (Fig. 7b) and two low ranking stations (Figure 7c) to ease

15 interpretation. Although the degree of these four stations is roughly the same, the connections of low ranking stations are regionally confined, and they rather reflect the similarity in rainfall variability within (homogenous) regions. Highest ranked stations are not governed by only local or global features but rather by the combination of both (Figure 7a). This observation could reflect the critical nodes in pathways of atmospheric moisture transport, extreme rainfall propagation or in case of high betweenness centrality, it indicates to a handful of stations, which are

20 positioned in-between the large communities and unlike most stations, they tend to possess intercommunity connections (Halverson and Fleming, 2015; Molkenthin et al., 2015; Tupikina et al., 2016). We plot the median (Fig. 7d) and 95th percentile (Fig. 7e) of the geographical distance between all the connected rain gauges to test whether the long-range connections of the selected nodes in Fig. 7b are a typical feature of high ranked stations. There is a clear association between rank and distance: High ranked stations tend to show longer connections, implicitly

25 affirming that the WDB measure has the potential to capture highly influential nodes in the network.

The results presented (Fig. 7) supports the conclusion derived from the kriging error analysis in section 4.4. Removing an influential station (Fig. 7b) fosters higher kriging errors than removing a random low ranking station (Fig. 7c). Hence, the new measure could support the optimal design of large hydrometric networks or redesign of existing hydrometric networks by ranking nodes. The influence of the similarity measure, number of stations present

30 in the network, spatial boundary, data length, and threshold has to be further investigated before the method can become fully operational. Acknowledging the infancy state of complex network science in hydrology, we emphasize

the need for more intensive application, new interpretable network measures and visualization tools to find the modern solutions of traditional hydrological problems.

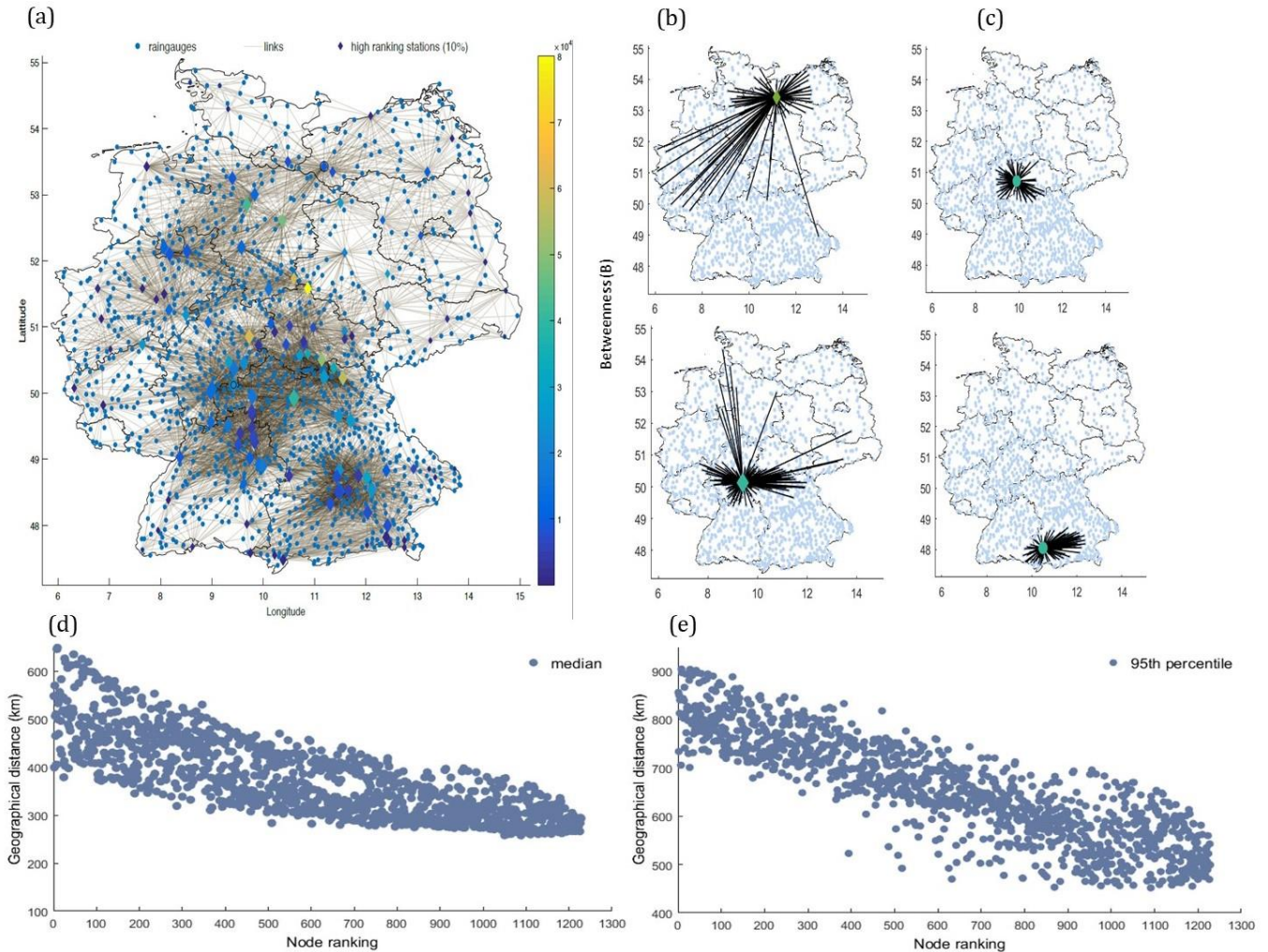



Figure 7: Connections and location of 10% (~122) highest ranking rain gauges (a). The size and colour of the diamond marker indicate the degree and betweenness centrality of the rain gauges, respectively. Connections corresponding to two high ranking stations (b, station ID: 21320, 16149) and two low ranking stations (c, station ID: 26132, 20356). Median (d) and 95th percentile (e) geographical distance plotted against node ranking.

6 Conclusions

This study proposes to apply complex networks to the optimization of hydrometric monitoring networks. In addition, it proposes a novel node ranking measure for identifying influential and expendable nodes in a complex network. The new network measure, weighted degree-betweenness (*WDB*), combines the measures degree and betweenness

5   centralities. It does not only account for the local and global characteristics of nodes but also the cumulative contribution of the directly connected (localized) nodes. Its comparison to existing measures demonstrates that *WDB* is more sensitive to the different roles of nodes, such as global connecting nodes or local centres as it considers various aspects of the Spatio-temporal relationships in observation network.

We propose to use *WDB* for ranking rain gauges in hydrometric networks. Applying *WDB* to a network of 1229 rain

10   gauges in Germany allows identifying influential and expendable stations. Two criteria, the decline rate of network efficiency and the kriging error, are used to evaluate the performance of the proposed node ranking measure. The results suggest that the proposed measure is indeed capable of effectively ranking the stations in large hydrometric networks.

We suggest that the proposed measure is not only useful for rain gauge networks but also has the potential to support

15   the selection of an optimal number of stations for the prediction in ungauged basins (PUBs) and estimating missing values by identifying influential stations in the region. Similarly, the proposed method can be applied to gridded satellite data (e.g. rainfall, soil moisture), to locate the strategic points where stations should be installed to ensure a highly efficient observation network. However, acknowledging the rarity of complex network studies in hydrology and the preliminary work of our study, the advantages and disadvantages of this new measure need to be further

20   investigated. This includes addressing threshold and spatial boundary issues of the network, developing new physical interpretable measures, and visualization tools. More studies are needed to prove the benefits of complex networks science in hydrometric network design.

Data availability

The precipitation data was provided by the German Weather Service. The data is publicly accessible at

25   https://opendata.dwd.de/. The data was pre-processed by the Potsdam Institute for Climate Impact Research (Conradt et al., 2012).

Appendix

A.   Spatially embedded network construction

We randomly select 11 rain gauge stations in Germany to illustrate the network construction (section 2.1) from observations (Fig. A1). We first compute the cross-correlation between each pair of stations (Table A1) and apply the 90th percentile threshold (0.44), i.e., only links between stations with values higher than 0.44 are shown.

We compute the WDB score for each station using Eq. 10. Station 3 shows the highest WDB score (Fig. A1). This station accounts for the local and global characteristics of the network, besides the cumulative effect of its direct neighbors, i.e., stations 2, 5, 7, 8, and 10. We infer two groups (stations 1, 2, 3, 6, 8, and 3, 4, 5, 7, 9, 10, 11) in the network that are bridged by station 3. This node is particularly crucial in the context of measuring process, process identification, or interpolation of measurements (Jensen et al., 2016).

Table A1: Cross-correlation values along with the geographical location of eleven rain gauges selected for illustrative purposes.

Nodes	Long.	Lat.	1	2	3	4	5	6	7	8	9	10	11
1	6.55	50.42	1.00	0.46	0.50	0.32	0.33	0.59	0.41	0.42	0.27	0.32	0.24
2	8.83	50.52	0.46	1.00	0.58	0.38	0.38	0.43	0.39	0.54	0.30	0.40	0.27
3	9.42	50.13	0.50	0.58	1.00	0.41	0.51	0.45	0.49	0.48	0.35	0.50	0.36
4	10.73	51.28	0.32	0.38	0.41	1.00	0.45	0.27	0.30	0.31	0.27	0.41	0.29
5	11.57	50.12	0.33	0.38	0.51	0.45	1.00	0.30	0.41	0.33	0.40	0.64	0.46
6	6.27	49.93	0.59	0.43	0.45	0.27	0.30	1.00	0.39	0.44	0.24	0.30	0.22
7	8.52	48.62	0.41	0.39	0.49	0.30	0.41	0.39	1.00	0.39	0.52	0.45	0.41
8	8.03	49.88	0.42	0.54	0.48	0.31	0.33	0.44	0.39	1.00	0.29	0.37	0.25
9	10.33	48.68	0.27	0.30	0.35	0.27	0.40	0.24	0.52	0.29	1.00	0.46	0.51
10	10.9	49.72	0.32	0.40	0.50	0.41	0.64	0.30	0.45	0.37	0.46	1.00	0.50
11	12	48.97	0.24	0.27	0.36	0.29	0.46	0.22	0.41	0.25	0.51	0.50	1.00

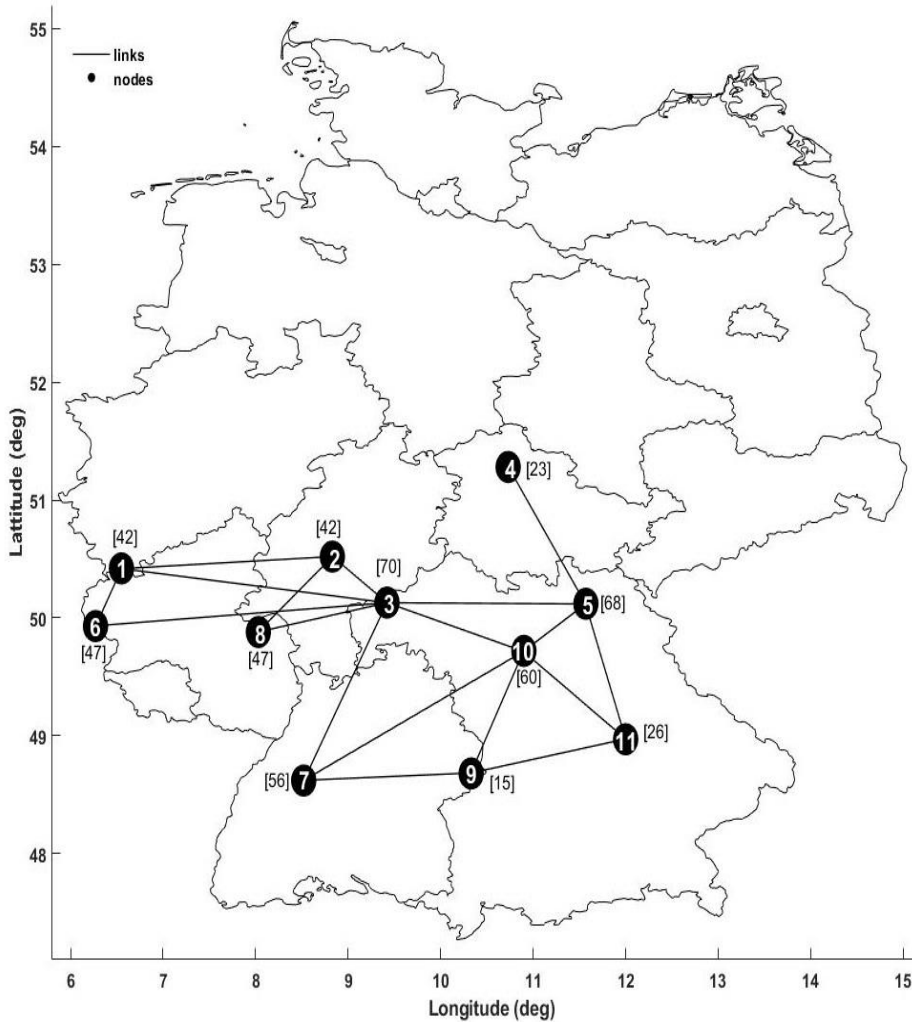



Figure A1: Location of eleven randomly selected rain stations used to construct a complex network based on the cross-correlation similarity measure and 90th percentile threshold. Diagonal values (autocorrelation) in Table 1 have been ignored in network construction. Numbers 1 to 11 are node counts, and values in brackets represent the WDB values.

B. Variogram modelling

The kriging modelling assumes a theoretical variogram function that is fitted with an experimental variogram of the observed data. The experimental variogram ($\gamma(h)$) is calculated from the observed data as a function of the distance of separation (h) (Adhikary et al., 2015), and is given by

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[(Y(i) - Y(j))^2\right], \tag{A1}$$

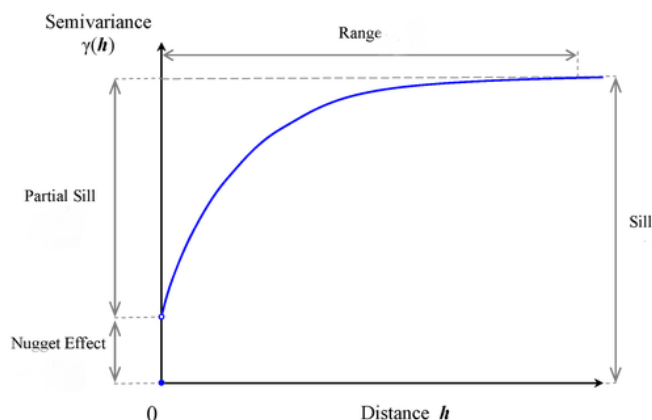
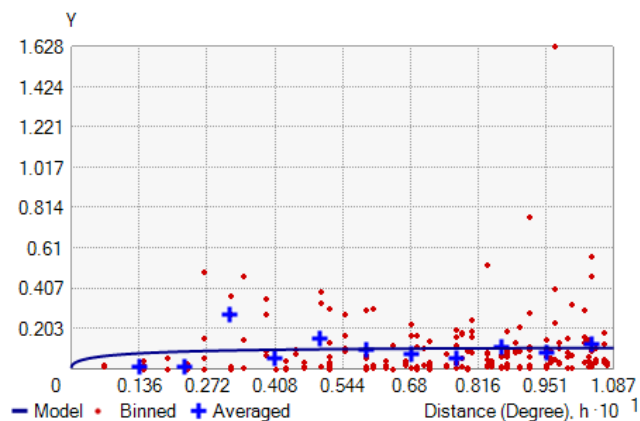
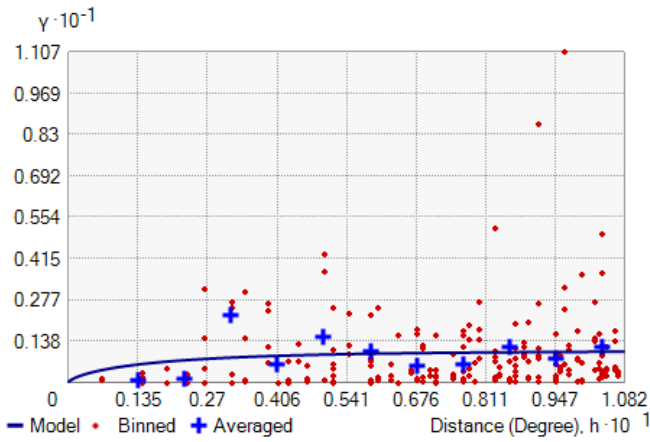
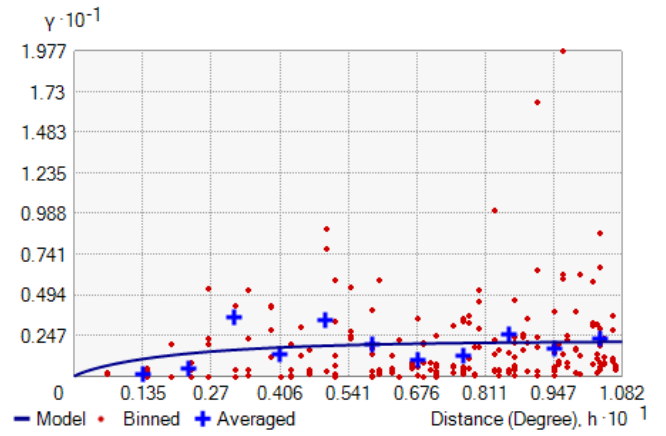
where $N(h)$ is the number of sample data points separated by distance h; i and j represent sampling locations separated by h; $Y(i)$ and $Y(j)$ indicate values of the observed variable Y, measured at the corresponding locations i and j, respectively. The theoretical variogram function (γ*(h)) allows the analytical estimation of variogram values for any distance and provides the unique solution for weights with intermediate steps required for kriging

5 interpolation (Adhikary et al., 2015).

The variogram models are a function of three parameters; the range, the sill, and the nugget (Fig. A2 (a)). The range is the distance, where the models first flatten out, i.e., station locations within the range distance are spatially correlated, whereas locations farther apart are not. The value of γ at the range is called the sill, which is estimated by the variance of the sample. The nugget represents measurement errors and/or microscale variation at very small

10 spatial scales and is seen as a discontinuity at the origin of the variogram model. The ratio of the nugget to the sill is known as the nugget effect and may be interpreted as the percentage of variation in the data that is not related to space. The difference between the sill and the nugget is known as the partial sill (Adhikary et al., 2015; Keum et al., 2017).

The values of all parameters and the resulting variogram for daily mean, 90th, 95th, and 99th percentile precipitation

15 and number of wet days are reported in Table A2 and Fig. A2 (b-d), respectively. The variogram has been kept constant during network reductions.

(a) Typical variogram model   **(b) Mean**

(c) 90th percentile   **(d) 95th percentile**

(e) 99th percentile

(f) wet days

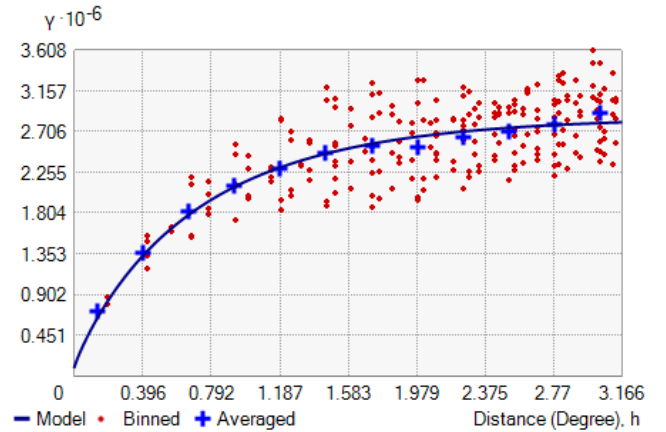
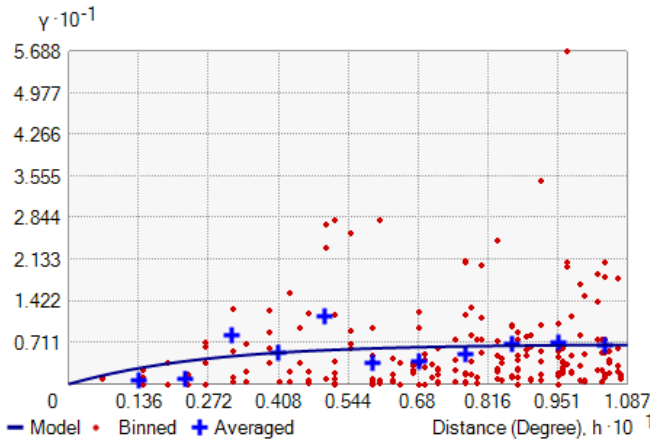



Figure A2: Typical variogram model (a) and fitted variogram models for daily mean (b), 90th (c), 95th (d), and 99th (e) percentile precipitation and number of wet days (f).

Table A2: Parameters values for the fitted variogram.

Parameters	Mean	90th percentile	95th percentile	99th percentile	Wet days
Nugget	0.0056	0	0	0	0.805
Range	0.0781	0.0782	0.0782	0.0782	2.361

Partial sill	0.102	1.055	2.140	6.808	2.761

Competing interests

The authors declare that they have no conflict of interest.

Data Availability

The authors used Germany's precipitation data which is maintain and provided by German Weather Service. The data is publicly accessible at https://opendata.dwd.de/. The further preprocessing on the data was done by Potsdam Institute of climate impact research which can be seen at (Oesterle, 2001, Conradt et al., 2012;).

Acknowledgement

This research was funded by 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). AA acknowledges the funding support provided by the Indian Institute of Technology Roorkee through Faculty Initiation Grant number IITR/SRIC/1808/F.IG. RM acknowledges the funding received from the SERB, Government of India under the project ECR/16/1721. Also, we gratefully acknowledge the provision of precipitation data by the German Weather Service. Ugur Ozturk was partly funded by the Federal Ministry of Education and Research (BMBF) within the project CLIENT II-CaTeNA (FKZ 03G0878A). The authors gratefully thank the Roopam Shukla (RDII, PIK-Potsdam) for helpful suggestion.

Author contribution

AA designed and implemented the research model. AA developed the node ranking algorithm and performed several test cases. UO tested the node ranking algorithm on various other networks. NM, JK & BM closely supervised the work and encouraged AA to investigate the node ranking algorithm for hydrological dataset. AA implemented the method and performed the analyses and tests. RM, JK & BM helped to interpret the findings. All authors discussed the results and contributed to the final manuscript.

References

Adhikary, S. K., Yilmaz, A. G. and Muttil, N.: Optimal design of rain gauge network in the Middle Yarra River catchment, Australia, Hydrol. Process., 29(11), 2582–2599, doi:10.1002/hyp.10389, 2015.

Agarwal, A.: Unraveling spatio-temporal climatic patterns via multi-scale complex networks, Universität Potsdam., 2019.

Agarwal, A., Marwan, N., Rathinasamy, M., Merz, B. and Kurths, J.: Multi-scale event synchronization analysis for unravelling climate processes: a wavelet-based approach, Nonlinear Process. Geophys., 24(4), 599–611, doi:10.5194/npg-24-599-2017, 2017.

Agarwal, A., Marwan, N., Maheswaran, R., Merz, B. and Kurths, J.: Quantifying the roles of single stations within homogeneous regions using complex network analysis, J. Hydrol., doi:10.1016/j.jhydrol.2018.06.050, 2018a.

Agarwal, A., Maheswaran, R., Marwan, N., Caesar, L. and Kurths, J.: Wavelet-based multiscale similarity measure for complex networks, Eur. Phys. J. B, 91(11), doi:10.1140/epjb/e2018-90460-6, 2018b.

Agarwal, A., Caesar, L., Marwan, N., Maheswaran, R., Merz, B. and Kurths, J.: Network-based identification and characterization of teleconnections on different scales, Sci. Rep., 9(1), doi:10.1038/s41598-019-45423-5, 2019.

Amorim, A. M. T., Gonçalves, A. B., Nunes, L. M. and Sousa, A. J.: Optimizing the location of weather monitoring stations using estimation uncertainty, Int. J. Climatol., 32(6), 941–952, doi:10.1002/joc.2317, 2012.

Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y. and Zhou, C.: Synchronization in complex networks, Phys. Rep., 469(3), 93–153, doi:10.1016/j.physrep.2008.09.002, 2008.

Boers, N., Rheinwalt, A., Bookhagen, B., Barbosa, H. M. J., Marwan, N., Marengo, J. and Kurths, J.: The South American rainfall dipole: A complex network analysis of extreme events: BOERS ET AL., Geophys. Res. Lett., 41(20), 7397–7405, doi:10.1002/2014GL061829, 2014.

Boers, N., Goswami, B., Rheinwalt, A., Bookhagen, B., Hoskins, B. and Kurths, J.: Complex networks reveal global pattern of extreme-rainfall teleconnections, Nature, doi:10.1038/s41586-018-0872-x, 2019.

Bullmore, E. and Sporns, O.: The economy of brain network organization, Nat. Rev. Neurosci., doi:10.1038/nrn3214, 2012.

Chacon-Hurtado, J. C., Alfonso, L. and Solomatine, D. P.: Rainfall and streamflow sensor network design: a review of applications, classification, and a proposed framework, Hydrol. Earth Syst. Sci., 21(6), 3071–3091, doi:10.5194/hess-21-3071-2017, 2017.

Chen, D., Lü, L., Shang, M.-S., Zhang, Y.-C. and Zhou, T.: Identifying influential nodes in complex networks, Phys. Stat. Mech. Its Appl., 391(4), 1777–1787, doi:10.1016/j.physa.2011.09.017, 2012.

Conradt, T., Koch, H., Hattermann, F. F. and Wechsung, F.: Precipitation or evapotranspiration? Bayesian analysis of potential error sources in the simulation of sub-basin discharges in the Czech Elbe River basin, Reg. Environ. Change, 12(3), 649–661, doi:10.1007/s10113-012-0280-y, 2012.

Conticello, F., Cioffi, F., Merz, B. and Lall, U.: An event synchronization method to link heavy rainfall events and large-scale atmospheric circulation features, Int. J. Climatol., 38(3), 1421–1437, doi:10.1002/joc.5255, 2018.

Donges, J. F., Zou, Y., Marwan, N. and Kurths, J.: Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods, Eur. Phys. J. Spec. Top., 174(1), 157–179, doi:10.1140/epjst/e2009-01098-2, 2009.

Donges, J. F., Petrova, I., Loew, A., Marwan, N. and Kurths, J.: How complex climate networks complement eigen techniques for the statistical analysis of climatological data, Clim. Dyn., 45(9–10), 2407–2424, doi:10.1007/s00382-015-2479-3, 2015.

Ekhtiari, N., Agarwal, A., Marwan, N. and Donner, R. V.: Disentangling the multi-scale effects of sea-surface temperatures on global precipitation: A coupled networks approach, Chaos Interdiscip. J. Nonlinear Sci., 29(6), 063116, doi:10.1063/1.5095565, 2019.

Fang, K., Sivakumar, B. and Woldemeskel, F. M.: Complex networks, community structure, and catchment classification in a large-scale river basin, J. Hydrol., 545, 478–493, doi:10.1016/j.jhydrol.2016.11.056, 2017.

Ferster, B., Subrahmanyam, B. and Macdonald, A.: Confirmation of ENSO-Southern Ocean Teleconnections Using Satellite-Derived SST, Remote Sens., 10(2), 331, doi:10.3390/rs10020331, 2018.

Gao, C., Wei, D., Hu, Y., Mahadevan, S. and Deng, Y.: A modified evidential methodology of identifying influential nodes in weighted networks, Phys. Stat. Mech. Its Appl., 392(21), 5490–5500, doi:10.1016/j.physa.2013.06.059, 2013.

Halverson, M. J. and Fleming, S. W.: Complex network theory, streamflow, and hydrometric monitoring system design, Hydrol. Earth Syst. Sci., 19(7), 3301–3318, doi:10.5194/hess-19-3301-2015, 2015.

Hohn, M. E.: An Introduction to Applied Geostatistics, Comput. Geosci., 17(3), 471–473, doi:10.1016/0098-3004(91)90055-I, 1991.

Hou, B., Yao, Y. and Liao, D.: Identifying all-around nodes for spreading dynamics in complex networks, Phys. Stat. Mech. Its Appl., 391(15), 4012–4017, doi:10.1016/j.physa.2012.02.033, 2012.

Jensen, P., Morini, M., Karsai, M., Venturini, T., Vespignani, A., Jacomy, M., Cointet, J.-P., Mercklé, P. and Fleury, E.: Detecting global bridges in networks, J. Complex Netw., 4(3), 319–329, doi:10.1093/comnet/cnv022, 2016.

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

Keum, J., Kornelsen, K., Leach, J. and Coulibaly, P.: Entropy Applications to Water Monitoring Network Design: A Review, Entropy, 19(11), 613, doi:10.3390/e19110613, 2017.

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E. and Makse, H. A.: Identification of influential spreaders in complex networks, Nat. Phys., 6(11), 888–893, doi:10.1038/nphys1746, 2010.

Konapala, G. and Mishra, A.: Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental USA, J. Hydrol., 555, 600–620, doi:10.1016/j.jhydrol.2017.10.033, 2017.

Kurths, J., Agarwal, A., Shukla, R., Marwan, N., Rathinasamy, M., Caesar, L., Krishnan, R. and Merz, B.: Unravelling the spatial diversity of Indian precipitation teleconnections via a non-linear multi-scale approach, Nonlinear Process. Geophys., 26(3), 251–266, doi:10.5194/npg-26-251-2019, 2019.

Langbein, W. B.: Overview of conference on hydrologic data networks, Water Resour. Res., 15(6), 1867–1871, doi:10.1029/WR015i006p01867, 1979.

Liu, J., Xiong, Q., Shi, W., Shi, X. and Wang, K.: Evaluating the importance of nodes in complex networks, Phys. Stat. Mech. Its Appl., 452, 209–219, doi:10.1016/j.physa.2016.02.049, 2016.

5 Malik, N., Bookhagen, B., Marwan, N. and Kurths, J.: Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks, Clim. Dyn., 39(3–4), 971–987, doi:10.1007/s00382-011-1156-4, 2012.

Mishra, A. K. and Coulibaly, P.: Developments in hydrometric network design: A review, Rev. Geophys., 47(2), doi:10.1029/2007RG000243, 2009.

Molkenthin, N., Rehfeld, K., Marwan, N. and Kurths, J.: Networks from Flows - From Dynamics to Topology, Sci. Rep.,
10 4(1), doi:10.1038/srep04119, 2015.

Okamoto, K., Chen, W. and Li, X.-Y.: Ranking of Closeness Centrality for Large-Scale Social Networks, in Frontiers in Algorithmics, vol. 5059, edited by F. P. Preparata, X. Wu, and J. Yin, pp. 186–195, Springer Berlin Heidelberg, Berlin, Heidelberg., 2008.

Ozturk, U., Marwan, N., Korup, O., Saito, H., Agarwal, A., Grossman, M. J., Zaiki, M. and Kurths, J.: Complex networks
15 for tracking extreme rainfall during typhoons, Chaos Interdiscip. J. Nonlinear Sci., 28(7), 075301, doi:10.1063/1.5004480, 2018.

Oesterle, H.: Reconstruction of daily global radiation for past years for use in agricultural models, Phys. Chem. Earth Part B Hydrol. Oceans Atmosphere, 26(3), 253–256, doi:10.1016/S1464-1909(00)00248-3, 2001.

Paluš, M.: Linked by Dynamics: Wavelet-Based Mutual Information Rate as a Connectivity Measure and Scale-Specific
20 Networks, in Advances in Nonlinear Geosciences, edited by A. A. Tsonis, pp. 427–463, Springer International Publishing, Cham., 2018.

Pfurtscheller, G. and Lopes da Silva, F. H.: Event-related EEG/MEG synchronization and desynchronization: basic principles, Clin. Neurophysiol., 110(11), 1842–1857, doi:10.1016/S1388-2457(99)00141-8, 1999.

Putthividhya, A. and Tanaka, K.: Optimal Rain Gauge Network Design and Spatial Precipitation Mapping based on
25 Geostatistical Analysis from Colocated Elevation and Humidity Data, Int. J. Environ. Sci. Dev., 124–129, doi:10.7763/IJESD.2012.V3.201, 2012.

Quiroga, R. Q., Kraskov, A., Kreuz, T. and Grassberger, P.: Performance of different synchronization measures in real data: A case study on electroencephalographic signals, Phys. Rev. E, 65(4), doi:10.1103/PhysRevE.65.041903, 2002.

Rheinwalt, A., Goswami, B., Boers, N., Heitzig, J., Marwan, N., Krishnan, R. and Kurths, J.: Teleconnections in Climate
30 Networks: A Network-of-Networks Approach to Investigate the Influence of Sea Surface Temperature Variability on Monsoon Systems, in Machine Learning and Data Mining Approaches to Climate Science, edited by V. Lakshmanan, E. Gilleland, A. McGovern, and M. Tingley, pp. 23–33, Springer International Publishing, Cham., 2015.

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, Clim. Dyn., 46(3–4), 1065–1074, doi:10.1007/s00382-015-2632-z, 2016.

5 Rossi, M., Kirschbaum, D., Valigi, D., Mondini, A. and Guzzetti, F.: Comparison of Satellite Rainfall Estimates and Rain Gauge Measurements in Italy, and Impact on Landslide Modeling, Climate, 5(4), 90, doi:10.3390/cli5040090, 2017.

Rubinov, M. and Sporns, O.: Weight-conserving characterization of complex functional brain networks, NeuroImage, 56(4), 2068–2079, doi:10.1016/j.neuroimage.2011.03.069, 2011.

Saxena, A., Malik, V. and Iyengar, S. R. S.: Estimating the degree centrality ranking, pp. 1–2, IEEE., 2016.

Sivakumar, B. and Woldemeskel, F. M.: Complex networks for streamflow dynamics, Hydrol. Earth Syst. Sci., 18(11), 10 4565–4578, doi:10.5194/hess-18-4565-2014, 2014.

Stolbova, V., Martin, P., Bookhagen, B., Marwan, N. and 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, doi:10.5194/npg-21-901-2014, 2014.

Stosic, T., Stosic, B. and Singh, V. P.: Optimizing streamflow monitoring networks using joint permutation entropy, J. 15 Hydrol., 552, 306–312, doi:10.1016/j.jhydrol.2017.07.003, 2017.

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

Tupikina, L., Molkenthin, N., López, C., Hernández-García, E., Marwan, N. and Kurths, J.: Correlation Networks from Flows. The Case of Forced and Time-Dependent Advection-Diffusion Dynamics, edited by Z.-K. Gao, PLOS ONE, 11(4), 20 e0153703, doi:10.1371/journal.pone.0153703, 2016.

Wadoux, A. M. J.-C., Brus, D. J., Rico-Ramirez, M. A. and Heuvelink, G. B. M.: Sampling design optimisation for rainfall prediction using a non-stationary geostatistical model, Adv. Water Resour., 107, 126–138, doi:10.1016/j.advwatres.2017.06.005, 2017.

Webster, R. and Oliver, M. A.: Geostatistics for Environmental Scientists, John Wiley & Sons, Ltd, Chichester, UK., 2007.

25 Xu, P., Wang, D., Singh, V. P., Wang, Y., Wu, J., Wang, L., Zou, X., Liu, J., Zou, Y. and He, R.: A kriging and entropy-based approach to raingauge network design, Environ. Res., 161, 61–75, doi:10.1016/j.envres.2017.10.038, 2018.

Yeh, H.-C., Chen, Y.-C., Chang, C.-H., Ho, C.-H. and Wei, C.: Rainfall Network Optimization Using Radar and Entropy, Entropy, 19(10), 553, doi:10.3390/e19100553, 2017.

Zhang, X., Zhu, J., Wang, Q. and Zhao, H.: Identifying influential nodes in complex networks with community structure, 30 Knowl.-Based Syst., 42, 74–84, doi:10.1016/j.knosys.2013.01.017, 2013.

Zlatić, V., Božičević, M., Štefančić, H. and Domazet, M.: Wikipedias: Collaborative web-based encyclopedias as complex networks, Phys. Rev. E, 74(1), doi:10.1103/PhysRevE.74.016115, 2006.