Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-113
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 13 March 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Optimal Design of Hydrometric Station Networks Based on Complex Network Analysis

Agarwal Ankit[1, 2, 3], Marwan Norbert[2], Rathinasamy Maheswaran[2], Ozturk Ugur [1, 2], Merz Bruno[1, 3], and Kurths Jürgen [1, 2]
[1]Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany
[2]Research Domain Transdisciplinary Concepts & Methods, Potsdam Institute for Climate Impact Research, Telegrafenberg, Potsdam, Germany
[3]GFZ German Research Centre for Geosciences, Section 5.4: Hydrology, Telegrafenberg, Potsdam, Germany

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

Abstracts

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. We propose a new measure, based on complex network analysis, to 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, 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 check its applicability in the optimal design of hydrometric networks. The proposed measure is evaluated using the decline rate of network efficiency and the kriging error. The results suggest that it effectively quantifies the importance of rain stations. The new measure is very useful in identifying influential stations which need high attention and expendable stations which can be removed without much loss of information provided by the station network.

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

1 Introduction

Hydrometric 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; Langbein, 1979). Adequate hydrometric monitoring is one of the first and primary tasks towards efficient water resources management. Information from hydrometric stations plays a crucial role in, among other things, flood estimation, water budget analysis, hydraulic design and assessing climate change. Even after the advent of remote sensing based information, such as precipitation products, in-situ observations are considered as an essential source of information on hydrometeorology.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-113
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 13 March 2018

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 monitoring stations, the more reliable the quantification of areal average estimates and point estimates at any ungauged location. However, a higher station number increases the cost of installation, operation, and maintenance, but may provide redundant information and, therefore, not increase the information content obtained from the network. Globally, there is a decreasing trend in the number of hydrometric stations in the last decades (Mishra and Coulibaly, 2009). Against the background of shrinking monetary support for hydrometric networks, their optimal design is gaining importance.

The design of hydrometric networks is a well-identified problem in hydrometeorology and has received considerable attention (Mishra and Coulibaly, 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. Several approaches have been developed for optimal network design, such as statistical analysis which include variance and dimension reduction methods (Amorim et al., 2012; Wadoux et al., 2017), spatial interpolation which includes kriging methods (Adhikary et al., 2015) and various interpolation techniques (Kassim and Kottegoda, 1991), information theory-based methods (Stosic et al., 2017; Taormina et al., 2015), optimization techniques such as simulated annealing (Barca et al., 2008; Mishra and Coulibaly, 2009; Pardo-Igúzquiza, 1998; Yoo et al., 2003), physiographic analysis (Laize, 2004), multivariate factor analysis (Hargrove and Hoffman, 2004; Yeh et al., 2006), sampling strategies (Tsintikidis et al., 2002), and user surveys or expert recommendations (Rani and Moreira, 2010). Combinations of methods have also been introduced in the last decade (Chacon-Hurtado et al., 2017; Keum et al., 2017; Mishra and Coulibaly, 2009; Sehgal et al., 2016).

Most of these studies inherently assume that a more optimal network is achieved through expanding the network with supplementary stations. However, a higher number of stations does not necessarily decrease the uncertainty (Stosic et al., 2017). Hence, a network can be optimized either by eliminating expendable stations from the network to minimize the cost or by expanding the network with the installation of additional stations to reduce the estimation uncertainty. Mishra and Coulibaly (2009) argued that the expendable stations in a network that contribute little or even nothing should be identified and removed, and at the same time, the most valuable or influential stations should be maintained and protected.

In the recent past, the science of complex networks has gained great momentum and have attracted many researchers from different disciplines and application fields, e.g., transportation networks (Bell and Lida, 1997), power grid analysis (Menck et al., 2014; Schultz et al., 2014), streamflow networks (Halverson and Fleming, 2015) and climate networks (Boers et al., 2014; Jha et al., 2015; Malik et al., 2012; Paluš, 2018; Stolbova et al., 2016). 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. The investigation of the topology of such complex networks helps in better understanding the mechanisms of highly complex systems with many components (Donges et al., 2009a; Stolbova et al., 2014).

Motivated by the encouraging results obtained by the above mentioned studies on climate networks, we develop a node ranking measure, based on complex network analysis that can be used to identify influential and expendable stations in large hydrometric station networks. Our aim is not to question the credibility of operating stations, but to propose an alternative evaluation procedure for the optimal design and redesign of observational networks.

In section 2, we introduce the basic concepts of complex networks. The proposed node ranking measure is presented and compared with existing measures in section 3 using synthetic networks. In section 4, the new measure is applied to a rain gauge network consisting of 1229 stations across Germany and compared with state-of-the-art methods.

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; Pfurtscheller and Lopes da Silva, 1999; Rubinov and Sporns, 2011), streamflow stations creating a hydrological network (Halverson and Fleming, 2015; Sivakumar and Woldemeskel, 2014) or climate stations describing a climate network (Stolbova et al., 2014; Malik et al., 2012; Rheinwalt et al., 2016).

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 vertices together with a set E of edges, $\{i, j\}$ which are 2-element subsets of N. In this work we consider undirected and unweighted simple graphs, where only one edge can exist between a pair of vertices and self-loops of the type $\{i, i\}$ are not allowed. This type of graph can be represented by the symmetric adjacency matrix

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

Figure 1 is a simple representation of such a network, i.e., one with a set of identical nodes connected by identical links. In general, (large) graphs 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 (Donges et al., 2009a; Jha et al., 2015), synchronization (Agarwal et al., 2017a; Conticello et al., 2017; Malik et al., 2012; Stolbova et al., 2016) or mutual information (Paluš, 2018).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-113
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 13 March 2018

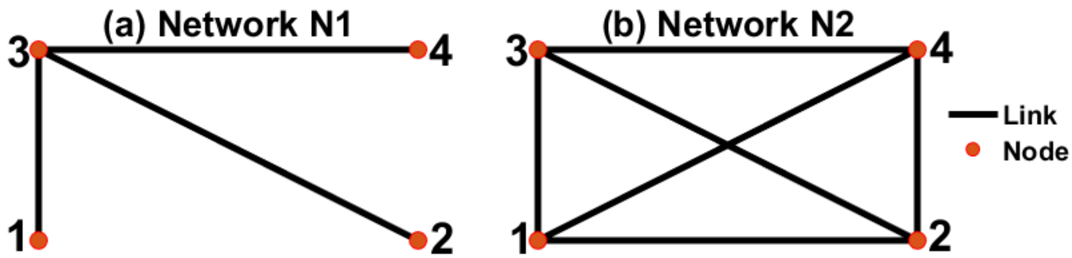

Figure 1: Topology of two sample networks to explain network structures and measures. (a) Network N1 with four nodes and three links; (b) network N2 with four nodes and six links.

2.2 Node Ranking Measures

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

The degree *k* of a node in a network counts the number of connections linked to the node directly. 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. In network N1, *B* of node 3 is 3, i.e., node 3 can disturb three pairs 1-2, 1-4, 2-4, and for other nodes $B = 0$. In the network N2, all nodes have $B = 0$ because no node can interrupt the information flow. So node 3 is a critical node in the network N1 but not in the network N2.

Jensen et al. (2016) developed the Bridgeness measure Bri to distinguish local centres, i.e. nodes that are central to a part of the network, from hybrid nodes, i.e. nodes that connect different parts of a network. *Bri* is a decomposition of betweenness centrality B into a local and a global contribution. Therefore, the Bri value of a node i is always smaller or equal to the corresponding B value and they only differ by the local contribution of the first neighbours. To calculate Bri we consider the shortest path between nodes not in the neighbourhood of node $i, N_G(i)$ (Table 1). For example, in the networks N1 and N2, all nodes have $B = 0$, hence $Bri = 0$, except node 3 in the network N1 for which all the nodes are in direct neighbourhood hence it also has $Bri = 0$.

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 network. The mathematical equation of DIL measure is presented in table 1, where the line between node i and j is e_{ij} and its importance is defined as $I_{e_{ij}} = \frac{U}{\lambda}$ where $U = (k_i - p - 1).(k_j - p - 1)$ 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 refer Liu et al. (2016)). DIL equation (Table 1) 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 $I_{e_{ij}} = 0$).

Table 1: Network measures. N: total number of nodes in a network. k_i: Degree of node i. B_i: Betweenness of node i, where $\sigma(j,k)$ represents the number of links along the shortest path between j and k; while $\sigma_i(j,k)$ is the number of links of the shortest path running through node i. In bridgeness, we consider the shortest path between nodes not in the neighbourhood of node i ($N_G(i)$).

Degree	Betweenness centrality	Bridgeness	DIL
$k_i = \frac{\sum_{j=1}^{N} A_{i,j}}{N-1}$	$B_i = \sum_{i \neq j \neq v \in \{V\}}^{N} \frac{\sigma_i(j,k)}{\sigma(j,k)}$	$Bri_i = \sum_{j \notin N_G(i) \vee k \notin N_G(i)}^{N} \frac{\sigma_i(j,k)}{\sigma(j,k)}$	$DIL_i = k_i + \sum_{j=N_G(i))} I_{e_{ij}} \cdot \frac{k_i - 1}{k_i + k_j - 2}$

3 Methodology

We propose a new node ranking measure that we call weighted degree-betweenness (WDB). We further compare the efficacy of the proposed measure with the existing node ranking methods using two synthetic networks.

3.1 Weighted Degree-Betweenness

WDB is a combination of the 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, $j = 1,2,3 \ldots .$n in proportion to their contribution to node i. Mathematically, the WDB of a node i is given by

$$WDB_i = B_i + I_i \tag{2}$$

where B_i is the betweenness centrality of node i, and I_i stands for the influence or contribution of the directly connected node j, $j = 1,2,3 \ldots .n$ to node i. It is defined for node i as

$$I_i = \sum_{j=1}^{n} \frac{B_i * k_j}{k_i + k_j} \tag{3}$$

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_i, and n is the total number of directly connected nodes 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 or betweenness centrality are used for node ranking (Gao et al., 2013; Okamoto et al., 2008; Saxena et al., 2016). However, these measures have certain disadvantages which are explained using a simple network, the undirected and unweighted network $Z = (N, E)$ with 8 nodes and 11 edges shown in Figure 2. The network measures k_i, B_i and WDB_i of each node are given in Table 2.

Degree is limited as node ranking measure since it cannot distinguish between different roles in the network. For example, nodes 5, 7, and 8 have the same degree (k_i=2), but node 5 serves as bridge node linking the two parts of the network. Information between several nodes in this network can flow through this node only. In a large complex network, such nodes have strategic relevance as most of the information can be accessed quickly just by capturing those nodes. For example, in a social network, the spreading of a disease could be slowed down or hindered by identifying these nodes. In climate networks, an early warning signal could be generated by capturing the flow of information (Donges et al., 2009a, 2009b).

Betweenness centrality has a higher power in discriminating different roles. For example, nodes 4 and 5 have the highest betweenness centrality $B = 24$ followed by node 6. Their importance for the information flow in the network is obvious, as such high B nodes can be used to control the flow of information in any network. However, betweenness B gives equal scores to local centers (nodes 4, 6), i.e., nodes of high degree central to a single region, and to global bridges (node 5), which connect different regions. This distinction is important because the roles of these nodes are different. For example, in climate networks, local centers correspond to nodes which are important for local climate phenomena, while bridges correspond to nodes which connect different subsystems of climate, such as monsoon and El-Nino, leading to teleconnections (Paluš, 2018).

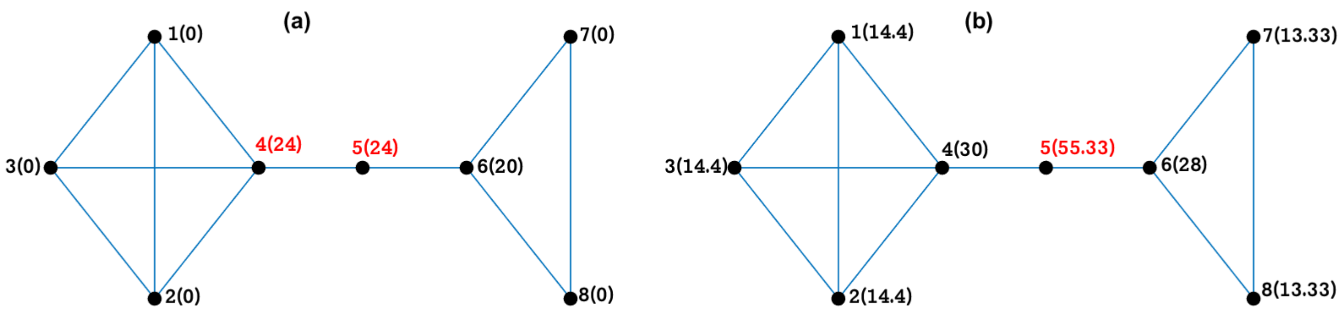

Figure 2: Synthetic network to explain the betweenness (B) and weighted degree-betweenness (WDB) measures, with node number (1 to 8) followed by the betweenness value (a) and WDB value (b) in brackets. Betweenness does not distinguish centers from bridges (marked in red), as it attributes the same value to the local center (node 4) and to the global bridge node (node 5). In contrast, *WDB* assigns the highest importance to node 5 that plays the role of a global bridge. Further, betweenness does not differentiate between the nodes 1, 2, 3, 7 and 8, while WDB provides a nuanced picture of the influence of all nodes.

Table 2: Network measures for the synthetic network shown in Figure 2. Ranking of the nodes is based on the proposed WDB measure.

Node no. (i)	k_i	B_i	WDB_i	Rank	Role of nodes
1	3	0	14.4	4	node
2	3	0	14.4	4	node
3	3	0	14.4	4	node
4	4	24	30	2	local center
5	2	24	55.3	1	global bridge
6	3	20	28	3	local center
7	2	0	13.3	5	node
8	2	0	13.3	5	node

The proposed measure *WDB* has an even higher discrimination power compared to betweenness centrality and effectively ranks the nodes in the network. Node 5 has the highest *WDB* score and is ranked as the most influential node. This reflects its role as global bridge node, as losing node 5 would disconnect the two parts of the network. *WDB* is also able to distinguish between the nodes 1, 2, 3 (*WDB* = 14.4) and the nodes 7, 8 (*WDB* = 13.3) which is important in case of we need sequentially ranking of nodes.

To further evaluate the proposed measure, we compare *WDB* with other network measures recently published, namely the bridgeness developed by Jensen et al. (2016) and degree and influence of line DIL by Liu et al. (2016). For this comparison, we use the same synthetic network as Jensen et al. (2016) shown in Figure 3. The corresponding network measure values are given in Table 3.

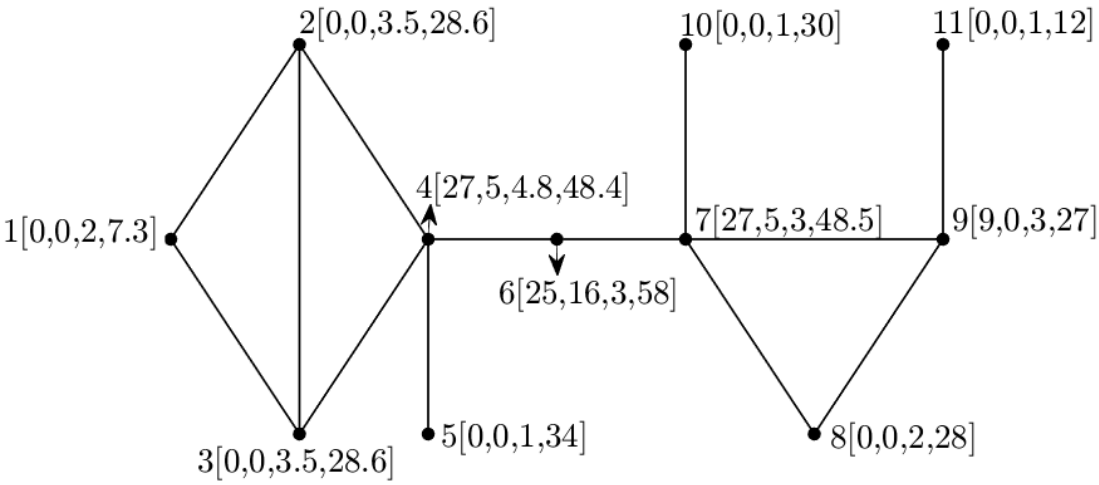

Figure 3: Synthetic network used to compare the network measures betweenness, bridgeness, and DIL with the proposed measure WDB. Number 1 to 11 are node counts, and values in brackets represent the network measure values in order of $[\boldsymbol{B}, \boldsymbol{Bri}, \boldsymbol{DIL}, \textbf{and } \boldsymbol{WDB}]$.

Fig. 3 illustrates that betweenness does not distinguish between the local centers (nodes 4, 7) and the global bridge node (node 6). It even assigns a smaller value to the global bridge node. Bridgeness expresses the higher importance of the global bridge node compared to local centers, however, it does not distinguish between all other nodes. Although *DIL* assigns different values to almost very node, these numbers do not represent the different roles of the nodes and are therefore hardly suitable as node ranking measure. *WDB* outperforms the existing measures in effectively ranking nodes in the network, such as the global bridge nodes, local centers and dead-end nodes (nodes 5, 10, 11) by assigning highest to lowest values. For example, WDB is sensitive to rank nodes 4 and 7 for which the bridgeness measure provides equally score.

This comparison of the proposed measure *WDB* with other measures that have been developed to express the importance of nodes within a network shows that WDB is able to provide a nuanced picture. The resulting node ranking reflects the different roles, such as global bridge, local center, dead-end node, hub (high degree), non-hub (low degree) (shown in Table 3) of the individual nodes.

Table 3: **Measures for the synthetic network shown in Figure 3. Ranking of the nodes is based on the proposed WDB measure.**

Node no. (i)	$\boldsymbol{B_i}$	$\boldsymbol{Bri_i}$	$\boldsymbol{DIL_i}$	$\boldsymbol{WDB_i}$	**Rank**	**Role of the node**
1	0	0	2	7.3	10	non-hub node
2	0	0	3.5	28.6	6	non-hub node
3	0	0	3.5	28.6	6	non-hub node
4	27	5	4.8	48.4	3	hub node
5	0	0	3.5	28.6	4	dead-end ode

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-113
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 13 March 2018

6	25	16	3	58	1	global bridge
7	27	5	3	48.5	2	hub node
8	0	0	2	28	7	non-hub node
9	9	0	3	27	8	non-hub node
10	0	0	1	30	5	dead-end node
11	0	0	1	12	9	dead-end node

3.3 Evaluation of the Proposed Measure for a Rain Gauge Network

In the context of hydrometric station networks, we hypothesise that higher ranking stations are more influential nodes in the network. Loosing such stations would more strongly reduce the network efficiency, i.e., the flow of information within the network, compared to lower ranking stations. Stations with the lowest ranks in the network are the least influential and are seen as expendable stations. To test this hypothesis, we apply the proposed node ranking measure to a rain gauge network consisting of more than 1000 stations in Germany. The information loss caused by removing stations is quantified via two measures: (a) decline rate of network efficiency, and (b) relative kriging error**.**

3.3.1 Decline Rate of Network Efficiency

The decline rate of network efficiency, as proposed by Liu et al. (2016), quantifies the loss in efficiency with which information flows within a network when nodes are removed from the network. Network efficiency is defined as

$$\eta = \frac{1}{N(N-1)} \sum_{n_i \neq n_j} \eta_{ij} \tag{4}$$

where 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. It is a measure of the efficiency of information or mass transport in a network. 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

$$\mu = 1 - \frac{\eta_{new}}{\eta_{old}} \tag{5}$$

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 reduces more strongly if high ranking stations are removed. This implies higher decline rates of efficiency when removing higher ranking stations from the network.

3.3.2 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). It is the best linear unbiased estimator of 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). The Ordinary Kriging technique is used in this study for interpolating rainfall data and estimating the kriging error. The kriging estimator is expressed as

$$Z^*(x_o) = \sum_{i=1}^{n} w_i Z(x_i) \tag{6}$$

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_0; and n indicates the number of observations within the domain of the search neighborhood of x_0 for performing the estimation of $Z^*(x_0)$. Ordinary Kriging is implemented through ArcGISv10.4.1 (Redlands, CA, USA) (ESRI, 2009) and its geostatistical analyst extension (Johnston et al., 2001).

To evaluate the performance of the WDB measure in identifying influential and expandable stations in a large network, we calculate the increase in the interpolation error of when stations are removed. The relative kriging error before and after removing the stations is denoted as

$$\Re(\%) = \frac{KSE_{new} - KSE_{old}}{KSE_{old}} \times 100 \tag{7}$$

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 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, 95th and 99th percentile rainfall and the number of wet days (precipitation > 2.5mm).

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 (Figure 4). The data covers 110 years at daily resolution (1 January 1901 to 31 December 2010). The 1229 rain stations inside Germany (red dots in Fig. 4) are operated by the German Weather Service. Data processing and quality control were performed according to Österle et al. (2006). 211 stations from different sources outside Germany (green 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.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-113
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 13 March 2018

Hydrology and Earth System Sciences
Open Access
Discussions

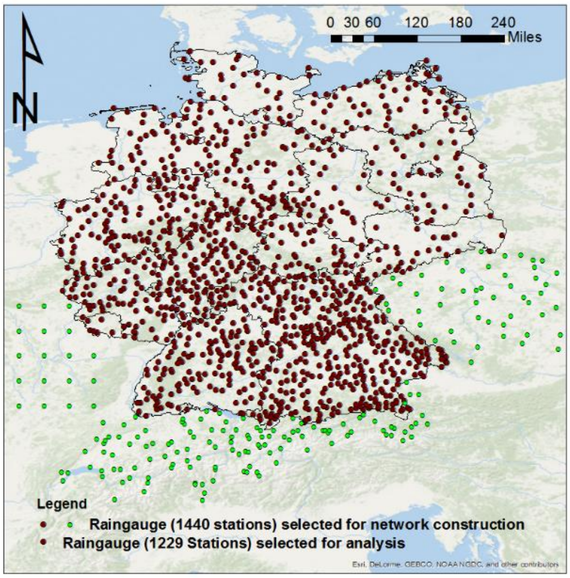

Figure 4: Location of rain stations in Germany and adjacent areas. Red dots indicate stations lying inside Germany that are used in the analysis. Green dots indicate stations outside of Germany that are used for network construction only to minimize the boundary effect.

4.2 Network Construction

We begin the network construction by extracting event time series from the 1440 daily rainfall time series. The event series represent heavy rainfall events, i.e., precipitation exceeding the 95th percentile at that station (Rheinwalt et al., 2016). The 95th percentile is a good compromise between having a sufficient number of rainfall events at each location and a rather high threshold to study heavy precipitation. All rainfall event series are compared with each other using event synchronization (see Appendix A; Agarwal et al., 2017a; Malik et al., 2012; Stolbova et al., 2014). This results in the similarity matrix Q, whereas the entry at index pair (i,j) defines synchronization in the occurrence of heavy rainfall events at station i and station j. Applying a certain threshold to the Q matrix (see Appendix A) yields an adjacency matrix

$$A_{i,j} = \begin{cases} 1, & if\ Q_{i,j} \geq \theta_{i,j}^{Q} \\ 0, & else, \end{cases} \tag{8}$$

Here, $\theta_{i,j}^{Q} = 95^{th} percentil$ is chosen in such a way to capture only highly synchronized stations. $A_{i,j} = 1$ denotes a link between the i^{th} and j^{th} station and 0 denotes otherwise. The adjacency matrix represents the connections in the rainfall network. Although the constructed network is based on all 1440 stations (to minimize the boundary effect), the subsequent topological analysis is performed only for the 1229 stations lying inside Germany.

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.

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. Figure 5 confirms this behavior. 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.

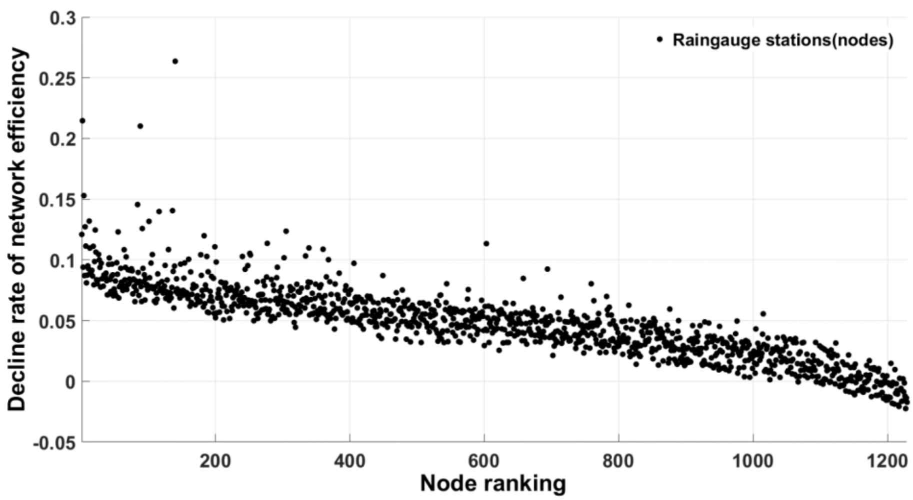

Figure 5: Decline rate of network efficiency corresponding to the removal of each node in the rainfall network. In each implementation, only one node is removed from the network according to the ranking with replacement (bootstrapping).

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. Figure 6 shows a clear increase in μ when more and more influential stations are removed. In case II, up to the 10% lowest ranking stations are successively removed. It can be seen in Figure 6 that this affects the network efficiency in a positive way: 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

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-113
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 13 March 2018

and within one trial is caused by randomness. For example, μ rises instantaneously when the algorithm picks up a high ranking station.

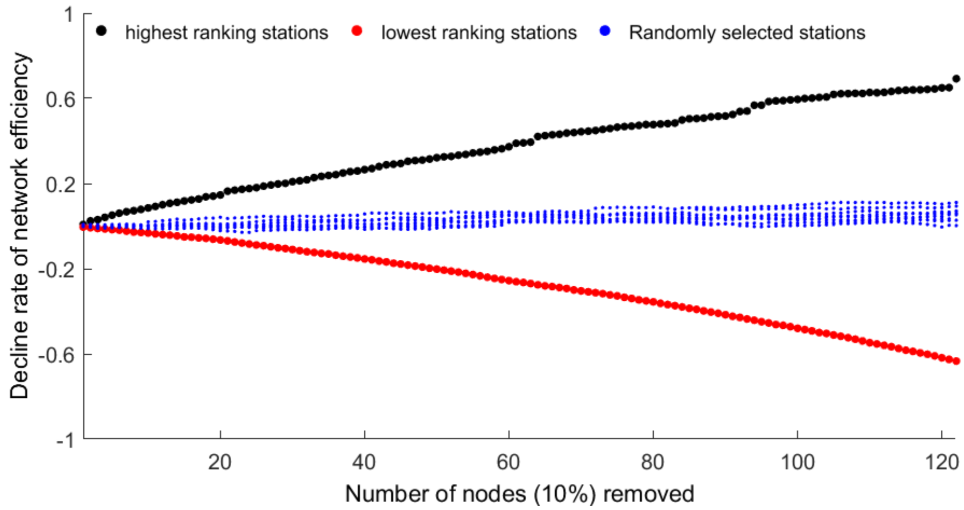

Figure 6: Decline rate of network efficiency as a function of the number of stations removed from the network. Case I: up to the 10% highest ranking stations are removed (black), case II: up to the 10% lowest ranking stations are removed (red), case III: up to 10% randomly drawn stations are removed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

4.4 Relative Kriging Error

As the second approach to assess the suitability of WDB for identifying influential and expandable stations, we analyse the change in the kriging error when stations are removed from the network. Similarly 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 4.

Removing the 10% high-ranking stations (case I) leads to positive and high ($\Re > 5\%$) values for all five statistics considered. The kriging error increases substantially when these stations are removed. When the 10% lowest ranking stations (case II) are not considered, the $\Re$ values are small compared to those obtained by removing high ranking stations. The relative errors in estimating the mean, percentile rainfall characteristics (90^{th} and 95^{th}) and number of wet days at ungauged locations are low (<5%) for the 10% lowest ranking stations, suggesting that these stations do not contribute much information. For two out of five statistics, i.e., mean and number of wet days, removing the 10% lowest ranking stations actually improves the kriging model. Case III, i.e. removing stations randomly, shows positive and high ($\Re > 5\%$) values except for the number of wet days because by doing so we remove high ranking nodes as well which lead to higher rates of $\Re(\%)$.

Table 4: 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	90^{th} percentile	95^{th} percentile	99^{th} percentile	Wet days
I	10% highest ranking	7.2*	35.9*	72.3*	59.1*	71.1*
II	10% lowest ranking	-5.1	4.1	3.8	11.4*	-1.5
III	10% randomly selected	6.6*	28.5*	60.8*	51.9*	4.8

5 Discussion

Building on the young science of complex networks, a novel node ranking measure, the weighted degree-betweenness *WDB*, is proposed. It is based on degree and betweenness centrality measure of nodes in a network. The comparison of the WDB measure with the existing node ranking measures suggests that it is more informative since it is better able to consider the different role of nodes in a complex network. The WDB measure provides a unique value to each node depending on its importance and influence in the network.

Further, this study proposes to use WDB for supporting the optimal design of large hydrometric networks. It is able to rank the nodes in a large network in relation to their importance for the flow of information, mass or energy. This ranking can be used to identify highly influential and expandable hydrometric stations. For example, removing low ranking stations in the German rain gauge network increases the network efficiency considerably, and may even decrease the error of estimating rainfall at ungauged locations. This is explained by the redundancy in the information that those stations provide, which in turn is attributed to the similarity between the gauges due to the common driving mechanisms or spatial similarity as advocated by Tobler's Law of Geography (Tobler, 1970). The results of our analysis suggests that WDB identifies the expandable nodes correctly as shown by decline rate of efficiency and relative kriging error. On the other hand, awards stations which provide unique information which cannot be generated from other stations in the network. Based on our analysis, we argue that, ranking of all nodes in large networks has the major benefit that the new measure adds to the optimal design of hydrometric networks or redesign of existing hydrometric networks.

The proposed node ranking approach differs from the existing approaches as it considers different aspects of the spatio-temporal relationships in observation networks. This measure is not only useful for optimizing observational networks, but has also potential to support 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,. For example, from a set of *N* potential stations to be used for PUBs the proposed approach can be applied to select the *M* influential stations which, when used, reduce the uncertainty (Villarini et al., 2008). Similarly, the proposed method can be applied to gridded satellite data (rainfall, soil moisture), to locate the strategic points where stations should be installed to ensure a highly efficient observation network. For instance, identifying influential grid points in the network of satellite data (rainfall, soil-moisture) will guide where to install monitoring stations.

An advantage of the proposed method is its capability to differentiate between the different roles played by individual stations. For example, global bridge nodes are able to control the flow of information, energy or mass between different parts of a network. Hence, they are of highest importance. This capability opens new possibilities for its use in complex networks. For instance, in climate networks an early warning signal could be generated by capturing the flow of information at such points (Donges et al., 2009b; Hlinka et al., 2014; Stolbova et al., 2016).

6 Conclusions

This study proposes a novel node ranking measure for identifying the influential and expendable nodes in a complex network. The new network measure weighted degree-betweenness (WDB) combines the existing measures degree and betweenness centrality and considers the neighbourhood of a node. The proposed measure is compared to other measures using synthetic networks. WDB is more sensitive to the different roles of nodes, such as global connecting nodes, hybrid nodes, and local centers, and provides a more informative ranking than the existing node ranking measures.

We propose to use this measure for the optimal design of hydrometric networks. Applying this measure to a network of 1229 rain 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 argue that the proposed measure is not only useful for optimizing observational networks, but has the potential to support the selection of an optimal number of stations (by determining influential station of the region) to be used in the prediction in ungauged basins, or to support the estimation of missing values, regionalization, and regional flood frequency analysis. When applied to gridded satellite data, it can be used to locate the strategic points where stations should be installed to ensure a highly efficient network. Furthermore, the new network measure has large potentials in other fields where science complex networks are used, such as in social networks, infrastructure networks, disease spreading networks, and brain networks.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-113
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 13 March 2018

Data availability

The precipitation data was provided by the German Weather Service. The data is publicly accessible at https://opendata.dwd.de/. The data was preprocessed by the Potsdam Institute for Climate Impact Research (Conradt et al., 2012).

Appendix

A. Event synchronization

Event synchronization (ES) has been very specifically designed to calculate nonlinear correlations among bivariate timeseries with events defined on them (Malik et al., 2012; Quian Quiroga et al., 2002; Stolbova et al., 2014). This method has advantages over other time-delayed correlation techniques (e.g., Pearson lag correlation), as it allows us to define extreme event series of the signal, depending on the kind of extreme, and uses a dynamic time delay. The latter refers to a time delay that is adjusted according to the two timeseries being compared, which allows for better adaptability to the region of interest. Another advantage of this method is that it can also be applied to a non-Gaussian and event-like data sets (Stolbova et al., 2014; Tass et al., 1998).

In last decade, various modification has been proposed in the basic algorithm, citing various issues such as boundary effect (Rheinwalt et al., 2016)., and bias toward number of events (Donges et al., 2009a), etc. thus modified basic algorithm proposed by (Agarwal et al., 2017a; Rheinwalt et al., 2016), can be explained as, let us say an event above threshold α percentile occurs in the signal $x(t)$ and $y(t)$ at time t_l^x and t_m^y where $l = 1,2,3,4 \ldots S_x$, $m = 1,2,3,4 \ldots \ldots S_y$ and within a time lag $\pm\tau_{lm}^{xy}$ which is defined as following

$$\tau_{lm}^{xy} = min\{t_{l+1}^x - t_l^x, t_l^x - t_{l-1}^x, t_{m+1}^y - t_m^y, t_m^y - t_{m-1}^y\}/2 \tag{A1}$$

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 help to separates of independent events which in turn allows to take into account the fact that different processes responsible for 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

$$C(x|y) = \sum_{l=1}^{S_x} \sum_{m=1}^{S_y} J_{xy} \tag{A2}$$

and

$$J_{xy} = \begin{cases} 1 & if \quad 0 < t_l^x - t_m^y < \tau_{lm}^{xy} \\ \frac{1}{2} & if \quad t_l^x = t_m^y \\ 0 & else, \end{cases} \tag{A3}$$

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

$$Q_{xy} = \frac{C(x|y) + C(y|x)}{\sqrt{(S_x - 2)(S_y - 2)}} \tag{A4}$$

Q_{xy} is a measure of 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 $y(t)$. (For detailed explanation refer Agarwal et al., 2017a).

Competing interests

The authors declare that they have no conflict of interest.

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). Also, we gratefully acknowledge the provision of precipitation data by the German Weather Service.

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., 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, 2017a.

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. Discuss., 1–20, doi:10.5194/npg-2017-19, 2017b.

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.

Barca, E., Passarella, G. and Uricchio, V.: Optimal extension of the rain gauge monitoring network of the Apulian Regional Consortium for Crop Protection, Environ. Monit. Assess., 145(1–3), 375–386, doi:10.1007/s10661-007-0046-z, 2008.

Bell, M. G. H. and Lida, Y.: Transportation Network Analysis: Bell/Transportation Network Analysis, John Wiley & Sons, Ltd., Chichester, UK., 1997.

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.

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.

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: AN EVENT SYNCHRONIZATION METHOD TO DETECT HEAVY RAINFALL EVENTS, Int. J. Climatol., doi:10.1002/joc.5255, 2017.

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, 2009a.

Donges, J. F., Zou, Y., Marwan, N. and Kurths, J.: The backbone of the climate network, EPL Europhys. Lett., 87(4), 48007, doi:10.1209/0295-5075/87/48007, 2009b.

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.

Hargrove, W. W. and Hoffman, F. M.: Potential of Multivariate Quantitative Methods for Delineation and Visualization of Ecoregions, Environ. Manage., 34(S1), S39–S60, doi:10.1007/s00267-003-1084-0, 2004.

Hlinka, J., Hartman, D., Jajcay, N., Vejmelka, M., Donner, R., Marwan, N., Kurths, J. and Palu?, M.: Regional and inter-regional effects in evolving climate networks, Nonlinear Process. Geophys., 21(2), 451–462, doi:10.5194/npg-21-451-2014, 2014.

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

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.

Kassim, A. H. M. and Kottegoda, N. T.: Rainfall network design through comparative kriging methods, Hydrol. Sci. J., 36(3), 223–240, doi:10.1080/02626669109492505, 1991.

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.

Laize, C. L. R.: Integration of spatial datasets to support the review of hydrometric networks and the identification of representative catchments, Hydrol. Earth Syst. Sci., 8(6), 1103–1117, doi:10.5194/hess-8-1103-2004, 2004.

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.

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.

Menck, P. J., Heitzig, J., Kurths, J. and Joachim Schellnhuber, H.: How dead ends undermine power grid stability, Nat. Commun., 5, doi:10.1038/ncomms4969, 2014.

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

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.

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

Pardo-Igúzquiza, E.: Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing, J. Hydrol., 210(1–4), 206–220, doi:10.1016/S0022-1694(98)00188-7, 1998.

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 Geostatistical Analysis from Colocated Elevation and Humidity Data, Int. J. Environ. Sci. Dev., 124–129, doi:10.7763/IJESD.2012.V3.201, 2012.

Quian Quiroga, R., 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.

Rani, D. and Moreira, M. M.: Simulation–Optimization Modeling: A Survey and Potential Application in Reservoir Systems Operation, Water Resour. Manag., 24(6), 1107–1138, doi:10.1007/s11269-009-9488-0, 2010.

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.

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.

Schultz, P., Heitzig, J. and Kurths, J.: A random growth model for power grids and other spatially embedded infrastructure networks, Eur. Phys. J. Spec. Top., 223(12), 2593–2610, doi:10.1140/epjst/e2014-02279-6, 2014.

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

Sivakumar, B. and Woldemeskel, F. M.: Complex networks for streamflow dynamics, Hydrol. Earth Syst. Sci., 18(11), 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.

Stolbova, V., Surovyatkina, E., Bookhagen, B. and Kurths, J.: Tipping elements of the Indian monsoon: Prediction of onset and withdrawal: TIPPING ELEMENTS OF MONSOON, Geophys. Res. Lett., 43(8), 3982–3990, doi:10.1002/2016GL068392, 2016.

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

Taormina, R., Chau, K.-W. and Sivakumar, B.: Neural network river forecasting through baseflow separation and binary-coded swarm optimization, J. Hydrol., 529, 1788–1797, doi:10.1016/j.jhydrol.2015.08.008, 2015.

Tass, P., Rosenblum, M. G., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., Schnitzler, A. and Freund, H.-J.: Detection of n : m Phase Locking from Noisy Data: Application to Magnetoencephalography, Phys. Rev. Lett., 81(15), 3291–3294, doi:10.1103/PhysRevLett.81.3291, 1998.

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

Tsintikidis, D., Georgakakos, K. P., Sperfslage, J. A., Smith, D. E. and Carpenter, T. M.: Precipitation Uncertainty and Raingauge Network Design within Folsom Lake Watershed, J. Hydrol. Eng., 7(2), 175–184, doi:10.1061/(ASCE)1084-0699(2002)7:2(175), 2002.

Villarini, G., Mandapaka, P. V., Krajewski, W. F. and Moore, R. J.: Rainfall and sampling uncertainties: A rain gauge perspective, J. Geophys. Res., 113(D11), doi:10.1029/2007JD009214, 2008.

Wadoux, A. M. .-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.

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.

Yeh, M.-S., Lin, Y.-P. and Chang, L.-C.: Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms, Environ. Geol., 50(1), 101–121, doi:10.1007/s00254-006-0190-8, 2006.

Yoo, C., Jung, K.-S. and Ahn, J. H.: Estimating Characteristics of Rainfall and Their Effects on Sampling Schemes: Case Study for Han River Basin, Korea, J. Hydrol. Eng., 8(3), 145–157, doi:10.1061/(ASCE)1084-0699(2003)8:3(145), 2003.

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.