Interactive comment on

"Optimal Design of Hydrometric Station Networks Based on Complex Network Analysis" by

Ankit Agarwal et al.

Editor's comments

Two referees have evaluated the paper "**hess-2018-113: Optimal Design of Hydrometric Station Networks Based on Complex Network Analysis".** Both referees mentioned that the paper presents an interesting and attractive idea, namely the transposition of complex network analysis methods to evaluate and support the optimal design of hydrometric networks. The authors proposed a ranking measure which helps apparently identifying the less influential stations. However both referees had made very important and fundamental critics, mainly doubts concerning the adequacy of the proposed method and the soundness of the conclusions.

We thank the editor for a constructive summary of our manuscript and also for his critical and supportive suggestions. Your feedback is vitally important to increase the readability of the work.

We have responded (in black) to each reviewer comment (in red).

Both referees had asked for additional explanations, justifications and analysis concerning:

The advantages of the method.

The advantages of the proposed method can be summarized under the following points and have been added to the revised manuscript:

- It has wide applications in all other relevant fields where identifying key nodes (influential nodes) in the network is vital for efficient and effective functioning of the system. Several methods have been proposed to evaluate the importance of nodes (Section 3.2). Our case studies show that the proposed measure WDB has an even higher discrimination power compared to existing node ranking measures and that it effectively ranks the nodes in the network. Additionally, WDB is 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 (P3/L8-13).
- Furthermore, our study emphasizes that the proposed approach for designing optimal observation networks, based on the complex network paradigm, is potentially more promising than existing ones (P18/L13-15).

The hypotheses used: e.g. the subjective cutoff value for the Qs to define the network, the stations to be removed first, the ranking method depending on the final objective.

In the revised version we have rectified all the above-mentioned issues (section 4.2, see response to below-mentioned reviewer-2 comments).

Improvements regarding the relative kriging errors.

We have carefully considered these critiques and responded correspondingly in the reviewer comments (Section 4.4).

The need for more tests for the validation;

As suggested, we have provided another test to check the interpolation error based on the leave-one-out sampling method providing more realistic estimates of real interpolation error (P16-17).

Anonymous Referee #1

I am not convinced that Fig.2 and Table 2 really help to understand what the result of the implementation of the proposed ranking method gives if applied to a real observation network. The example in Fig.2 is extremely simple and can hardly be extrapolated to a raingauge network including hundreds of gauges. It is unclear what node 5 in figure 2, which illustrates the added value of the proposed ranking method, really stands for.

We agree with the reviewer that the discussion of this simple example is not sufficient for explaining large observation networks, such as raingauge networks with hundreds of stations. However, our intention was to motivate the development and to illustrate the idea behind the new node ranking measure using a simple example. The differences to standard measures, such as degree and betweenness centrality, becomes more evident when looking at a small number of network nodes. Independently from this example, the proposition that our new measure is useful for large networks of raingauges is discussed and evaluated using the decline rate of network efficiency and the kriging error (section 4.3, 4.4).

In the revision, we have made clear (P8/L5) that this simple example is used for easy understanding and motivation, and that we do not derive that our new measure works for the large raingauge network as well.

The role of node 5 in Fig. 3 (Fig.2. In earlier version) can be understood with the followings simple theoretical example (P8-9/L22).

In climate networks, local centers correspond to nodes which are important for local climate phenomena, while bridges correspond to nodes which connect different climatic subsystems, such as Indian monsoon and El-Nino, leading to teleconnections (Paluš, 2018). Bridge nodes spread a process to the entire network with more force than a local center in a community, even when the local center has more spreading power, i.e. more connections, than the bridging node (Lawyer, 2015). By deducing the global bridge node in a spatial hydrometeorological network, most of the process flow information can be captured.

What about raingauges that are poorly correlated with other gauges that could rather appear as dead-ends or even isolated nodes in the build network?

We interpret this question as a modified question of the reviewer in his/her first review: *"If I understand correctly stations with strong similarities with other stations will have a large number of connections and hence a high WDB value and conversely".* As we mentioned in our previous response letter the ranking measure that we propose (WDB) is not simply related to the number of links. Hence, the ranking of the nodes will depend on their information contribution to the network, see for example, on page 14, line 15: *"… it awards stations which provide unique information which cannot be generated from other stations in the network …"* However, we agree with the reviewer that such a situation needs to be considered carefully. We have added the following discussion on such situations in the revision (P10-11/L19).

There is one situation where our method would require additional care: Let's imagine a node that is unrelated to other nodes (no links). Physically, one might imagine this scenario in a meteorological sub-region characterized by fine-scale convective thunderstorms with sparse rain gauge coverage. Hence, precipitation event synchronization across rain gauges in that sub-region would be poor. In that case, indeed, this station would not be the part of the constructed network, and would not be ranked. This station should be treated carefully as it provides unique information.

Some extracted maps from figure 4 showing on a limited size area, the topography along with the location and resulting ranks of the raingauges and maybe also the location of the 10% higher ranked removed gauges could improve a lot the presentation of the method.

We thank the reviewer for her/his excellent suggestion. In the revised version we have incorporated a new figure showing 10% high ranking stations, 10% lowest ranking stations along with topography (Fig.9).

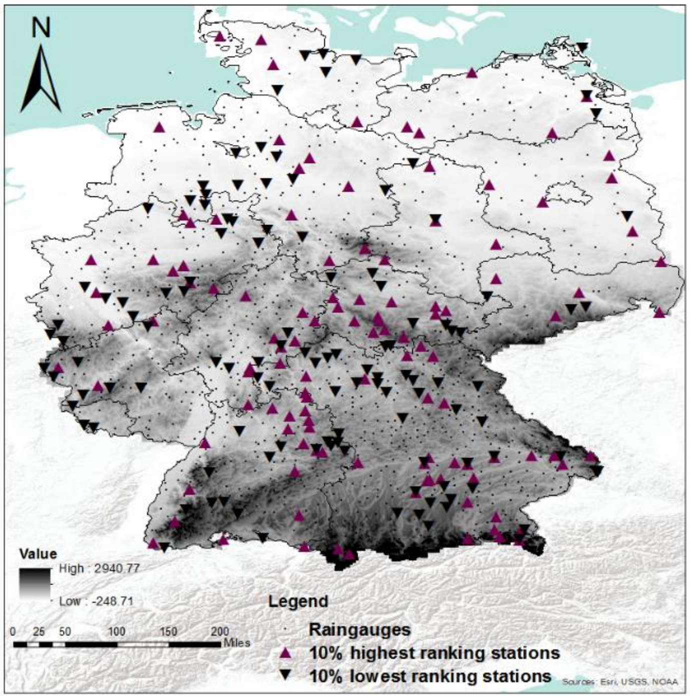


Figure 9: Location of 10% highest ranking rain gauges (red) and 10% lowest ranking stations (black) showing on Germany along with topography of the area.

Moreover, I confirm that additional validation is needed and this is acknowledged by the authors. At least the resulting variance of the rainfall fields corresponding to the various tested networks in table 4 should be provided. If significantly modified by the gauge selection method – probably moderately for the random selection method – it could have a major impact on the figures in table 4. This should be considered in the interpretations. But evaluations based on a leave-one-out approach should also be conducted.

As recommended by the reviewer in her/his review letter we have computed a new validation scheme based on leave-one-out approach and resulting variance of rainfall fields of corresponding various tested networks (Appendix A1) in the revised version.

"….We further check the interpolation error based on the leave-one-out sampling method providing another estimates of the interpolation error. For the three cases we eliminate successively stations from the network and study the impact on the relative kriging error. The leave-one-out approach is applied before the network reduction, i.e. on the original network of 1229 stations. For example, while removing highly influential stations from the entire network of 1229 stations, in the very first step, rank 1 station is eliminated and the relative kriging error is calculated for five characteristics, i.e., mean, 90%-, 95%-, 99%-percentile, and number of wet days (Fig. 8(a)-(e)). In the second step, rank 1 and rank 2 stations are eliminated from the network. We repeat the procedure up to 10 stations since the calculation is computationally very expensive.

We further repeat the procedure for the other two cases of expendable stations and random stations. All the results are plotted in the Fig. 8 (a)-(e). We observe that eliminating high ranking stations (influential) induce positive and high ($\mathfrak{R} > 5\%$) error in the network as compared to removing expendable stations and random stations from the network.

Removing influential nodes in our 10 trials shows a smoothly increasing pattern of the error whereas the pattern for removing randomly selected stations is rather volatile. Here, we see sudden changes in the relative kriging error which is expected as stations with very different ranks can be removed from step to step. These results support our finding that eliminating highly ranked stations will have major impacts on the network, whereas low-ranking stations could be removed or relocated without inducing significant errors…."

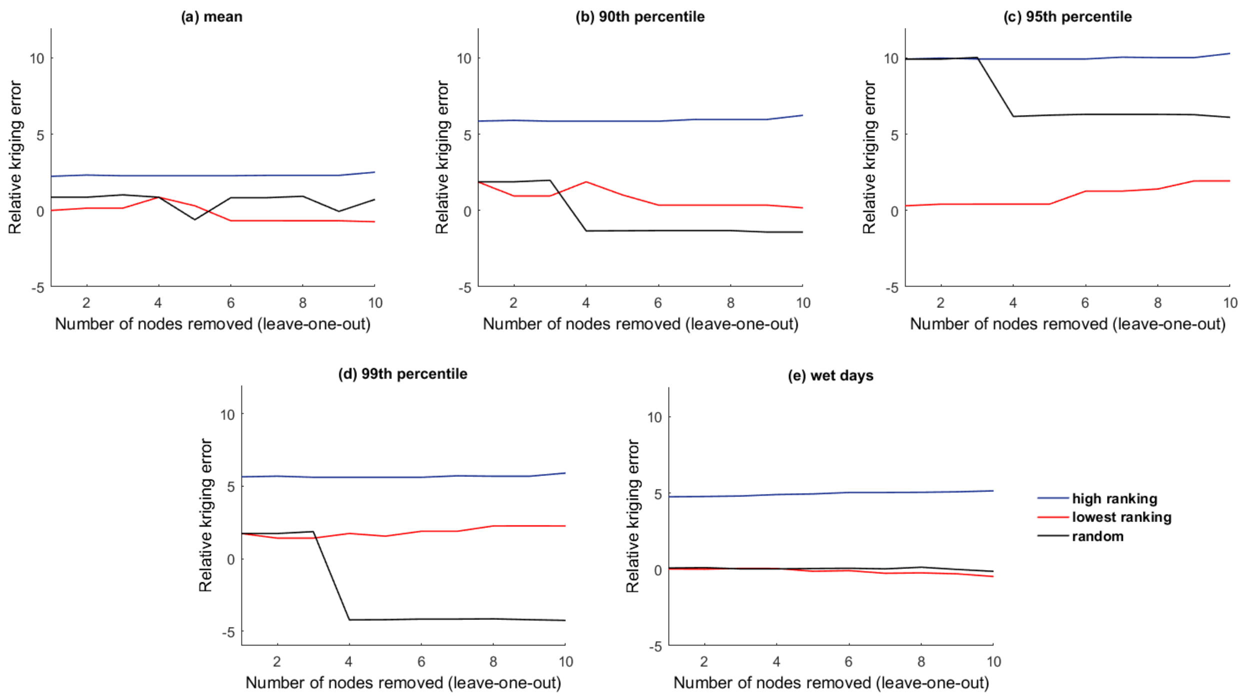


Figure 8: Relative kriging error as a function of the number of stations removed from the network. In the first step we remove the highest ranking station, followed by the two highest ranking stations and so on. We then calculate the relative kriging error in the mean (a), 90^{th} percentile (b), 95^{th} percentile (c), 99^{th} percentile (d) and wet days (e). We apply the same procedure to expendable stations and randomly selected stations.

Remarks: The real leave-one-out approach is different and used mainly for validation of the models. However, here the above done analysis (leave-one-out) only mimic what has been already showed by removing a bunch of raingauge from the network. Hence authors strongly suggest not to include this analysis in the revised version. Otherwise it might affect the readability of the manuscript.

We further provide additional details on kriging which has been added to the appendix A in the revised version.

Kriging variogram modelling

The kriging modelling mandates a theoretical variogram function that is to be 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) and is given by (Adhikary et al., 2015)

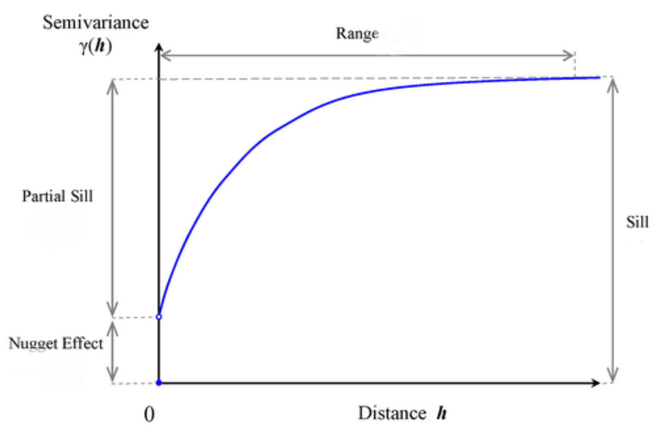
$$\gamma(\boldsymbol{h}) = \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{N}(\boldsymbol{h})} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}(\boldsymbol{h})} [(\boldsymbol{Y}(\boldsymbol{i}) - \boldsymbol{Y}(\boldsymbol{j}))^{2}] \tag{A1}$$

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

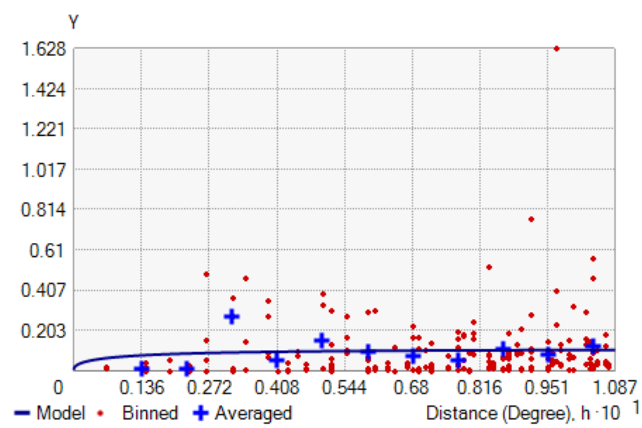
The variogram models are a function of three parameters, known as the range, the sill, and the nugget (Fig.A1 (a)). The range is typically the level of h at the correlation between point values is zero (i.e., there is no longer any spatial autocorrelation). The value of γ at range is called the sill. The variance of the sample is used as an estimate of the sill. Nugget represents measurement error and/or microscale variation at spatial scales that are too fine to detect 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 spatial. The difference between the sill and the nugget is known as the partial sill.

The value of all the parameters and resultant variogram for mean, 90th percentile, 95th percentile, 99th percentile and wet days has been reported in the Table A1 and Figure A1 (b-d) respectively. The variogram has been kept constant during network reductions.

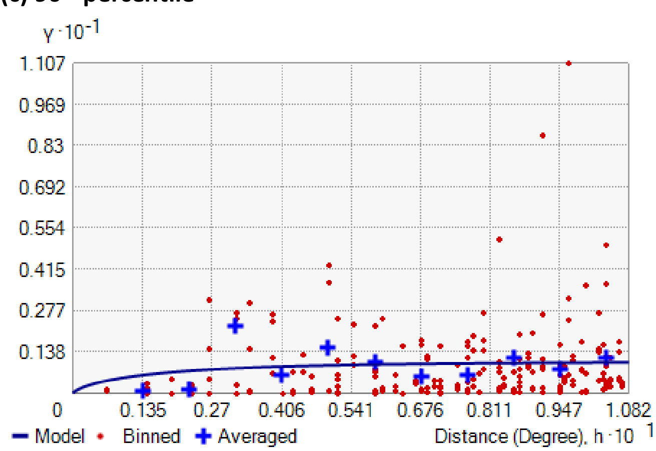
(a) **Typical variogram model**



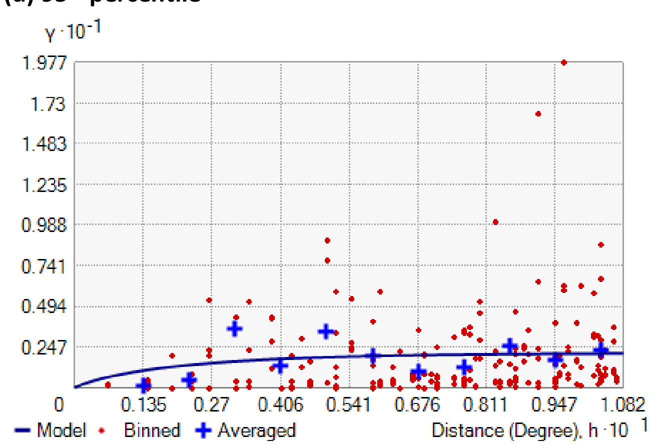
(b) Mean



(c) 90th percentile



(d) 95th percentile



(e) 99th percentile | (f) wet days

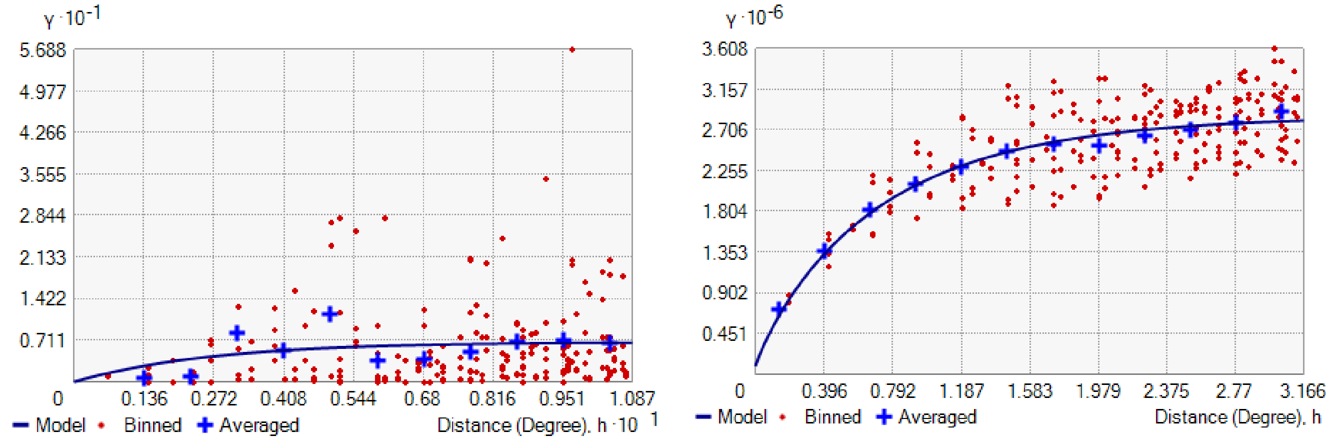


Figure A1: Typical variogram models (a) and fitted variogram models for mean (b), 90th percentile (c), 95th percentile (d), 99th percentile (e) and wet days (f).

Table A1: Parameters values for the fitted variogram.

Parameters	Mean	90th percentile	95th percentile	99th percentile	Wet days
Nugget	0.0058	0	0	0	0.905
Range	0.0782	0.0782	0.0782	0.0782	2.363
Partial sill	0.103	1.055	2.140	6.808	2.771

Anonymous Referee #2

This manuscript introduced the use of complex network analyses for designing optimal hydrometric networks. I find the concept interesting, but the authors somewhat fail to explain **what the advantage of this method is,** and to make me really understand **what the network analyses will mean in the case of hydrometeorological observations.**

It is clear how a linear network can be defined, as in Figure 1, but I find it difficult to imagine the network that is built from the event synchronization.

We are sorry that we have not been clear enough in explaining the importance and usefulness of networks for climate data analysis. Here, we briefly explain the meaning of network analysis and its advantages in hydrometeorological observations. The explanation has been added in the revised version P2/L27.

Recently, complex network theory has emerged as a very powerful tool for understanding large spatiotemporal complex systems such as climate. In which a node represent a rainfall station, grid station, streamflow station etc. interconnected with links in a non-trivial manner. The links are computed based on statistical associations between climate parameter time series at different points on Earth. Hence, network analysis of hydrometeorological system have yielded new insights in the investigation of climate dynamics (Tsonis and Swanson 2008).

For example, Malik et al., (2012) used complex networks and event synchronization to investigate the spatiotemporal dynamics of the Indian summer monsoon. The analysis provided valuable insights into the spatial organization, scales, and structure of the rainfall events during the Indian summer monsoon (June–September). Marwan and Kurths, (2015) showed that complex networks can be successfully deployed to study the transitions in past climate or to develop a prediction scheme for extreme rainfall events. Climate networks have provided also insights regarding further important questions in climate sciences, such as stability of multidecadal oscillations, regional impact of teleconnections, heat transport in the climate systems, etc. (Tsonis et al., 2004, Gozalchiani et al., 2005, Donges et al., 2009, Palus et al., 2011, Berezin et al., 2012 ,Guez et al., 2012, Deza et al., 2013, Boers et al., 2013, Tupikina et al., 2014 and Feng and Dijkstra.,2014). Not only in the area of climate science, several other studies have demonstrated that complex network theory greatly contributes to the spatiotemporal analysis of brain dynamics, friction networks, traffic flow, turbulent heated jets and multiphase flow system, etc. Furthermore, Steinhaeuser et al., (2012) have shown that complex networks have the ability to extract the information content in multiple variables relevant to a variable of interest, thus leading to better understanding of the underlying process dynamics.

To explain the network construction using ES, we have added a new schematic figure in the revised version (Fig.2).

We have inserted the following schematic figure with modifications in a revised version to better explain the network construction using event synchronization (Section 2.2). All the equations and symbols have been explained in the main text of manuscript.

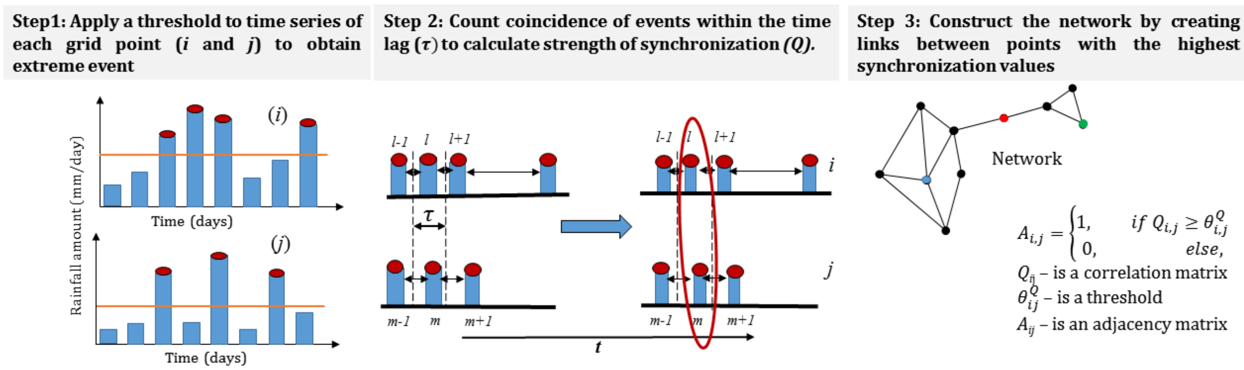


Figure 2: Schematic of network construction using event synchronization (ES). All the equations and symbols have been explained in the main text.

Major comments

Additional synthetic case study for expandable stations: Maybe the authors could show a small example where only a few (imaginary) stations are analyzed with the network methodology. *Then it can be shown how and why some stations are redundant and can be removed.* The real case example from Germany is interesting, but with such a high number of stations, it is challenging to understand what actually happens.

The specific application to use the WDB measure for ranking raingauges in Germany may indeed be difficult to understand. However, for easy understanding and visualization we have already included small examples (Fig.3 & 4) and that an additional small example would not bring additional benefit.

Threshold cutoff justification: I am not convinced by the use of a somewhat subjective cutoff value for the Qs to define the network, without at least a much deeper discussion around the effect. This will to a large degree ignore the level of similarity, it is just a yes/no transformation. Increasing or decreasing the threshold could drastically change the importance of the nodes in the network. Two stations with similarity just above the threshold will be treated the same way as two stations which are almost identical. On the other hand, two stations just under the threshold are treated completely different than the stations just above, even if their similarities are almost the same.

We thank the reviewer for raising the concern with the subjectivity of threshold. In the revised version we have rectified this limitation by selecting threshold objectively (Section 4.1).

"...Here, $\boldsymbol{\theta^Q_{i,j}}$ is chosen in such a way to capture only highly synchronized stations. Two stations are significantly synchronized, if the synchronization value exceeds the 95th percentile of the synchronization obtained by two synthetic variables that have the same number of events positioned randomly in the time series. We calculate synchronization for 100 pairs of random time series from which we derive the 95th percentile of synchronization. We use 95% as our threshold, since it satisfies both necessary conditions: a high synchronization and a sufficient number of extreme events for comparison (Agarwal et al., 2017; Rheinwalt et al., 2016; Stolbova et al., 2014).

Threshold cutoff (two stations just under the threshold are treated completely different than the stations just above, even if their similarities are almost the same) is a major common issue in all the methods for example, peak over threshold etc. Also, for instance, two person earning almost same amount might end up in paying different amount of taxes depending lying just above or below the prescribed threshold.

Global bridge node: The authors do several times mention the importance of global bridge nodes, and the possibilities these give in analyses of complex networks. For example*: "For instance, in climate networks an early warning signal could be generated by capturing the flow of information at such points."* This might be explained better in some of the references, but it should anyway be better explained what a local center and global bridge node really means in the climate network, and what kind of information we could particularly capture from this node.

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 (Jensen et al., 2016), such as the Asian monsoon and El Niño/Southern Oscillation, leading to teleconnections (Paluš, 2018). Bridge nodes spread a process to the entire spatial region globally whereas the effect of a local center is confined to a region (community) (Lawyer, 2015, refer Fig.2).

In temperature based climate networks it is the energy that is transported, and with this, some kind of information about the atmospheric state in a region (Hlinka et al., 2017). For rainfall networks in general, the links reflect the major propagation path ways of moisture, for extreme precipitation it is even more specific and reflects certain weather conditions, e.g. a specific "Großwetterlage" in central Europe. Ozturk et al., 2018 proposed a complex network based approach to estimate the tendency of extreme rainfall movement over Japan during typhoons. They iteratively approximated likely tracks of the extreme precipitation for each grid cell, many of which present redundant information, and hence the computation is time inefficient (several days). We suggest that by applying their method only on global bridges and local centers, we can deduce the likely track of extreme efficiently.

The same information has been added in the revised version in the P8-P9.

Kriging: I noticed that also the other reviewers asked for some improvements regarding the relative kriging errors. In addition to what they wrote, I was not sure whether the variogram is recomputed when stations are removed. If this is done, then variogram fitting is a science in itself, whether done manually or automatically, and this can lead to changes in the kriging error, making small changes more a result of random changes. The kriging error should normally not decrease when you remove stations, so the reduction in table 4 for the mean is most likely because the variogram has been fitted differently. When kriging error is used to estimate network modifications, the variogram is therefore usually kept constant, to avoid having to also analyze the variogram fitting. The larger changes are still significant.

We thank the reviewer for highlighting this important piece of information which is essential for the replicability of the work. We confirm that the variogram has been kept constant during the network modification and added this information to the manuscript (P16/L8).

Some smaller issues:

P2L25 the sentence is somewhat contradictory to the previous one, try to rephrase.

Yes, the statements were contradictory. We have modified them in the revised version.

P13 - Fig 6 Remove 10% from the x-label

Changed.

P16 Eq. A3 explain why the numbers are 1 and ½ in the equation.

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 it double counts in $C(x|y)$ and $C(y|x)$. We have added this explanation in the revision (Section 2.2).

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

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, 2017.

Hlinka, J., Jajcay, N., Hartman, D. and Paluš, M.: Smooth information flow in temperature climate network reflects mass transport, Chaos Interdiscip. J. Nonlinear Sci., 27(3), 035811, doi:10.1063/1.4978028, 2017.

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.

Lawyer, G.: Understanding the influence of all nodes in a network, Sci. Rep., 5(1), doi:10.1038/srep08665, 2015.

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.

Marwan, N. and Kurths, J.: Complex network based techniques to identify extreme events and (sudden) transitions in spatio-temporal systems, Chaos Interdiscip. J. Nonlinear Sci., 25(9), 097609, doi:10.1063/1.4916924, 2015.

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.

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.

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.

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], Kurths Jürgen [1, 2], and Merz Bruno[1, 3]
[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 that identifies critical nodes of station networks, based on complex network analysis. 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, 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). 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 in hydrometeorology.

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 (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), optimization techniques such as simulated annealing (Mishra and Coulibaly, 2009; Pardo-Igúzquiza, 1998; Yoo et al., 2003), physiographic analysis (Laize, 2004), multivariate factor analysis (Hargrove and Hoffman, 2004), 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).

Most of these studies inherently assume that a more optimal network is achieved through expanding the network with supplementary stations. However, increasing the number of stations does not necessarily decrease the uncertainty (Stosic et al., 2017). 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. Hence, a network can also be optimized by eliminating expendable stations from the network.

Against this background, this study aims to identify influential and expendable stations based on their relative information content by developing a new node ranking measure for hydrometric station networks. We use complex networks which are a collection of nodes interconnected with links in a non-trivial manner. Complex network techniques have recently 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 (Schultz et al., 2014), streamflow networks (Halverson and Fleming, 2015) and climate networks (Agarwal et al., 2018b). However, the application of complex networks in hydrology is still in its infancy.

In this study, we approach a complex network based method to identify the influential and expendable stations in a rainfall network. The novelty of this study is twofold: 1) We propose a new measure for identifying the most influential nodes in a network, and 2) we use event synchronization as a similarity measure. 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., 2016; Kitsak et al., 2010; Zhang et al., 2013 and Hu et al., 2013). Degree (k), betweenness centrality (B) and closeness centrality (CC) are the methods 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 called weighted degree-betweenness (WDB), combining degree (k) and betweenness centrality (B), which combines the advantages of both. Our case studies show that the proposed measure WDB has an even higher discrimination power compared to other existing methods and that it effectively ranks the nodes in the network. Additionally, 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.

Further, we use event synchronization as a similarity measure. In a complex network, links are set up between each pair of nodes based on how the nodes interact with each other. This interactions is measured through statistical measures such as zero-lag correlation or time-delayed correlation (Agarwal et al., 2018b). However, these measures are limited by the underlying assumptions, e.g. measuring linear relations. Further, they give equal weight to high and low rainfall values, whereas the main information content in a rainfall time series is embedded in the larger values. In contrast, event synchronization (ES) is a suitable measure for event-like, non-Gaussian data such as precipitation (Stolbova et al., 2014; Tass et al., 1998). It has advantages over other time-delayed correlation techniques (e.g., Pearson lag correlation), as it allows us to define the event time series by determining the threshold, and as it uses a dynamic time delay (not fixed). 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.

Therefore, the main objective of the study is to 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) , streamflow stations creating a hydrological network (Halverson and Fleming, 2015) or climate stations describing a climate network (Agarwal et al., 2018a).

Formally, a network or graph is defined as an ordered pair $Z = \{\mathrm{N}, \mathrm{E}\}$; containing a set $N = \{N_1, N_2, \dots . N_N\}$, 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 (Fig.1)

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

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), synchronization (Agarwal et al., 2017; Conticello et al., 2017) or mutual information (Paluš, 2018).

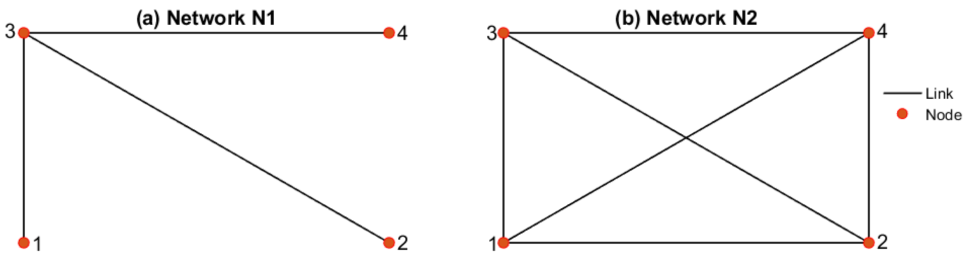


Figure 1: Topology of two sample networks to explain network structures and measures. (a) Network N1 with four nodes and three links; (b) network N2 with four nodes and six links.

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 define extreme event series of the signal, depending on the kind of extreme, and as it uses a dynamic time delay. 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. 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 the last decade, various modifications have been proposed, related to, for instance, boundary effects (Rheinwalt et al., 2016) and bias toward the number of events which can be explained as, let us say an event above threshold α percentile occurs in the signal $x(t)$ and $y(t)$ (Fig. 2, step 1) 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 (Fig. 2, step 2)

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

where S_x and S_y is the total number of such events (greater then threshold α) that occurred in the signal $x(t)$ and $y(t)$, respectively. The above definition of the time lag helps to 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 achieved by defining quantities $C(x|y)$ and $C(y|x)$ where

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

and

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

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

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

Q_{xy} is a normalized measure of the strength of event synchronization between signal $x(t)$ and $y(t)$. This implies $Q_{xy} = 1$ for perfect synchronization and zero if no events are synchronized.

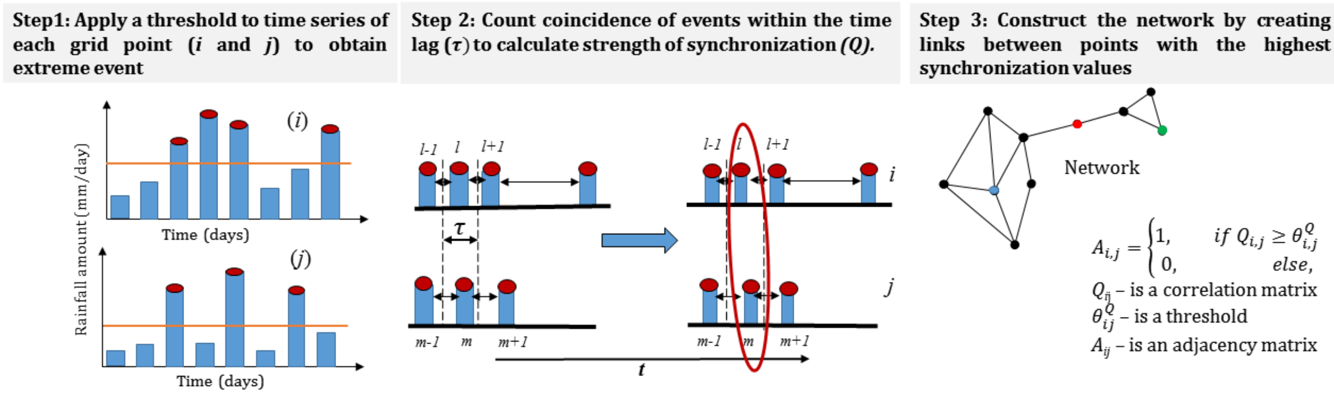


Figure 2: Schematic of network construction using event synchronization (ES). Equations and symbols have been explained in the main text.

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 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., 2016b) 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. The degree of any i node is calculated as

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

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, *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 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 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 a network:

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

where the line between node i and j is e_{ij} and its importance is defined as $I_{e_{ij}} = \frac{U}{\lambda}$ where $U = (k_i - p - 1).(k_j - p - 1)$ 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). 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 $I_{e_{ij}} = 0$).

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

$$WDB_i = B_i + I_i \quad (10)$$

$where\ B_i$ 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

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

$where\ k_i$ is the degree of node i, k_j is the degree of the nodes j which are directly connected to node i_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 network 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 limitations which are explained using a simple network, the undirected and unweighted network $Z = (N, E)$ with 8 nodes and 11 edges shown in Figure 3. The calculated network measures k_i, B_i and WDB_i of each node are given in same Fig. 3 along with the node number.

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 climatic subsystems, such as Indian monsoon and El Niño/Southern Oscillation, leading to teleconnections (Paluš, 2018). Bridge nodes spread a process to the entire region globally whereas the effect of a local center is confined to a region (community) (Lawyer, 2015).

In temperature based climate networks it is the energy that is transported, and with this, some kind of information about the atmospheric state in a region (Hlinka et al., 2017). For rainfall networks in general, the links reflect the major propagation path ways of moisture, for extreme precipitation it is even more specific and reflects certain weather conditions, e.g. a specific “Großwetterlage” in central Europe. Ozturk et al., (2018) proposed a complex network based approach to estimate the tendency of extreme rainfall movement over Japan during typhoons. They iteratively approximated likely tracks of the extreme precipitation for each grid cell, many of which present redundant information, and hence the computation is time inefficient (computing took several days). We suggest that by applying their method only on global bridges and local centers, we can deduce the likely track of extreme efficiently.

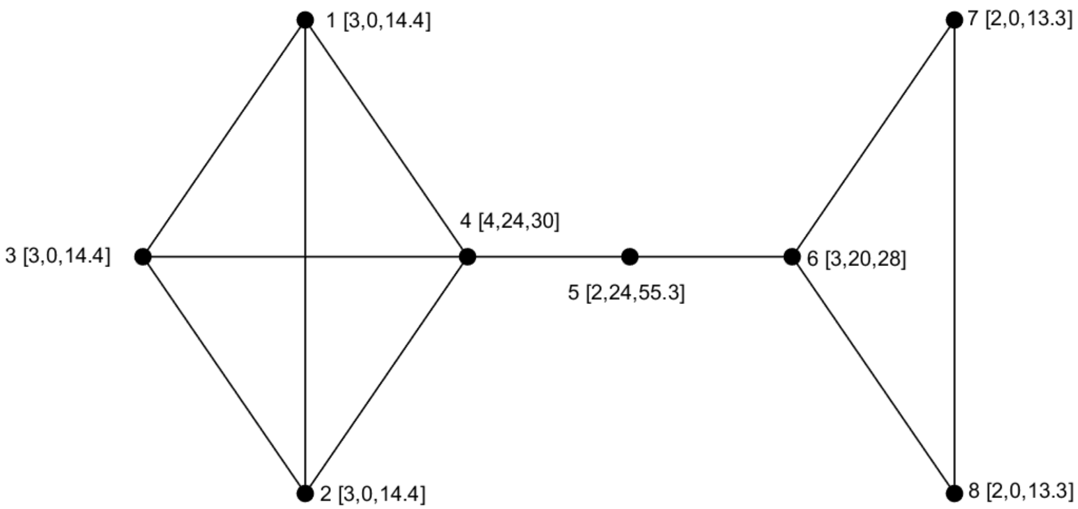


Figure 3: Synthetic network to explain the degree *(k)*, betweenness (B) and weighted degree-betweenness (WDB) measures, with node number (1 to 8) followed by the degree, betweenness value and WDB value in brackets. Degree is limited in distinguishing the role of different nodes in the network. Betweenness does not distinguish centers from bridges, as it attributes the same value to the local centre (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.

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 we need to sequentially rank 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 Fig. 4. The corresponding network measure values are also given in Fig. 4.

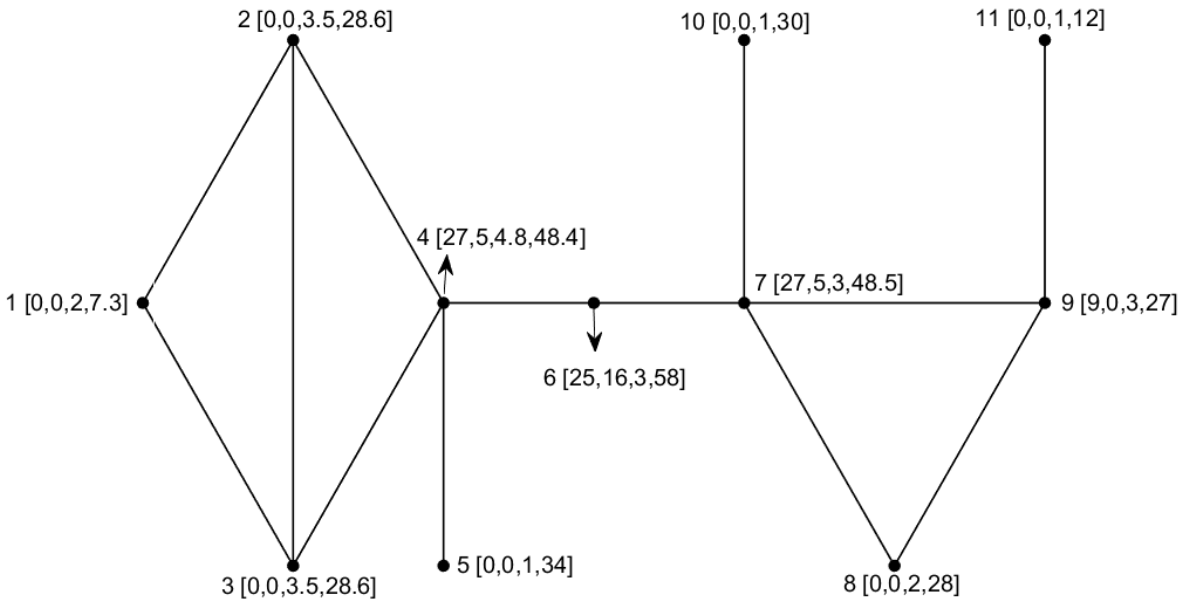


Figure 4: Synthetic network used to compare the network measures betweenness, bridgeness, and DIL with the proposed measure WDB. Numbers 1 to 11 are node counts, and values in brackets represent the network measure values in order of $[\boldsymbol{B}, \boldsymbol{Bri}, \boldsymbol{DIL}, \boldsymbol{and}\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

Fig. 4 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. For example, WDB differentiates between nodes 4 and 7 for which the bridgeness measure provides equal scores.

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), or non-hub (low degree), of the individual nodes.

There is one situation where our method would require additional care: Let's imagine a node that is unrelated to other nodes (no links). One might imagine this scenario in a meteorological sub-region characterized by fine-scale convective thunderstorms with sparse rain gauge coverage. Hence, precipitation event synchronization in that sub-region would be poor. This station would not be the part of the constructed network and would not be ranked. However, this station should be treated carefully as it provides unique information.

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

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

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. 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

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

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 reduces more strongly if higher 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). 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_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).

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

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

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 entire 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 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, 90^{th}, 95^{th} and 99^{th} 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 5). The data covers 110 years at daily resolution (1 January 1901 to 31 December 2010). The 1229 rain stations inside Germany (blue dots in Fig. 5) 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 (red dots in Fig. 5) were included in the analysis to minimize spatial boundary effects in the network construction, however, these stations were excluded from the node ranking analysis.

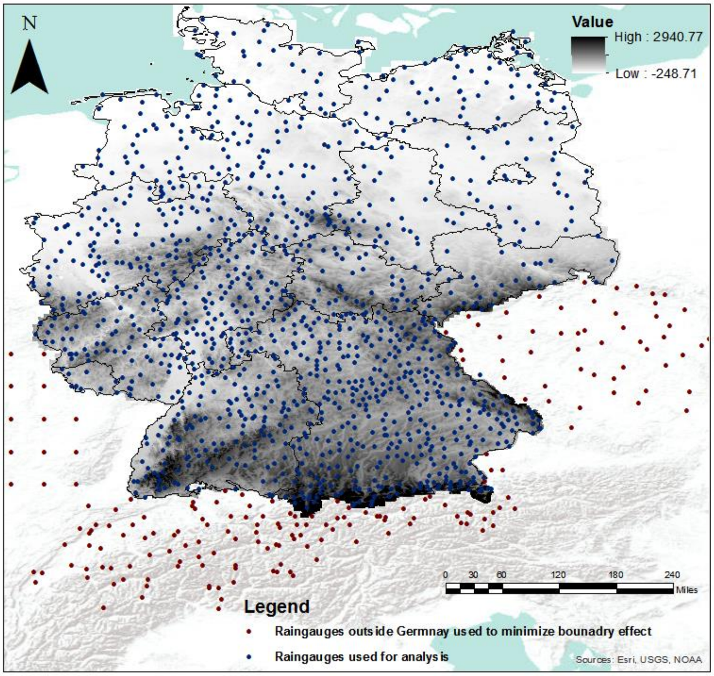


Figure 5: Location of rain stations in Germany and adjacent areas. Blue dots indicate stations lying inside Germany that are used in the analysis. Red dots indicate stations outside of Germany that are used for network construction only to minimize the boundary effect.

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 (Fig. 2). 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 (Fig. 2)

$$A_{i,j} = \begin{cases} 1, & if\ Q_{i,j} \geq \theta_{i,j}^{Q} \\ 0, & else, \end{cases} \qquad (16)$$

Here, $\boldsymbol{\theta}_{i,j}^{Q}$ is chosen in such a way to capture only highly synchronized stations. Two stations are significantly synchronized, if the synchronization value exceeds the 95th percentile of the synchronization obtained by two synthetic variables that have the same number of events positioned randomly in the time series. We calculate synchronization for 100 pairs of random time series from which we derive the 95th percentile of synchronization.

$A_{i,j} = 1$ 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. Fig. 6 confirms this behavior. μ is high for high-ranking stations and decays with node ranking. Interestingly, $\mu < 0$ 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.

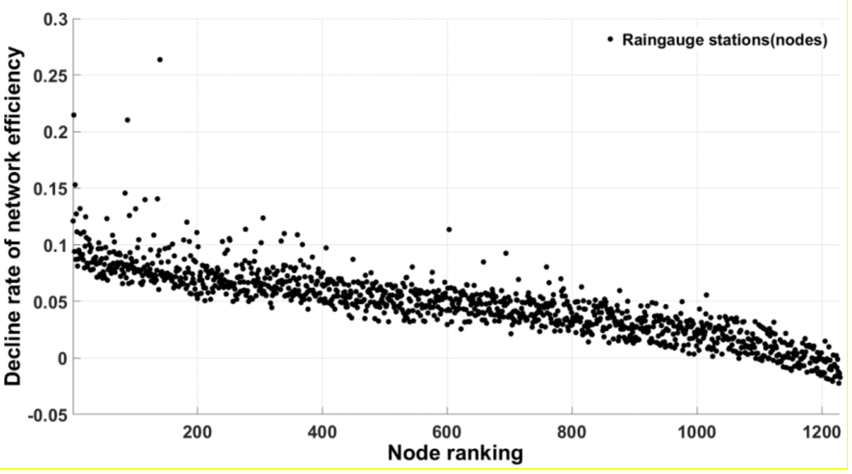


Figure 6: Decline rate of network efficiency corresponding to the removal of each node in the rainfall network. In each implementation, only one node is removed from the network according to the ranking with replacement (bootstrapping).

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 7 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 Fig. 7 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 and within one trial is caused by randomness. For example, μ rises instantaneously when the algorithm picks up a high ranking station.

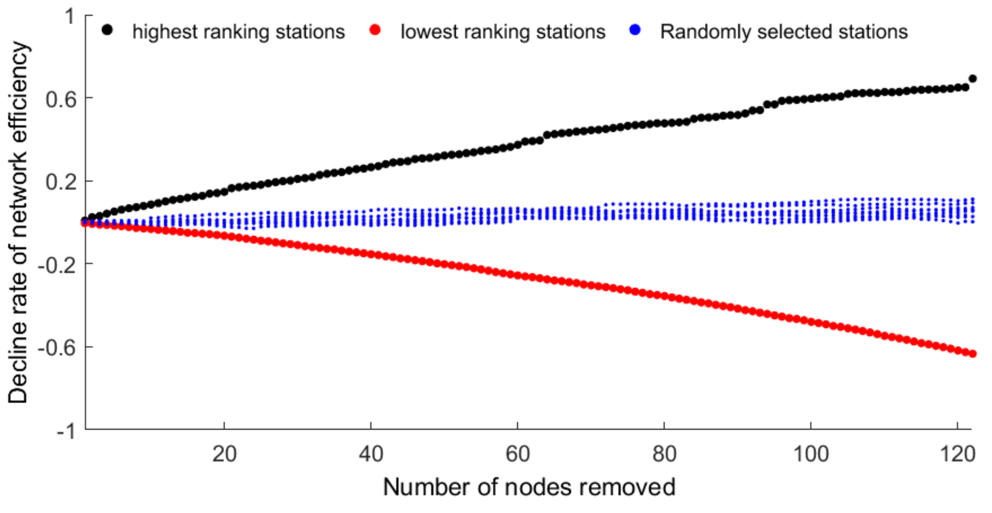


Figure 7: Decline rate of network efficiency as a function of the number of stations removed from the network. Case I: up to the 10% highest ranking stations are removed (black), case II: up to the 10% lowest ranking stations are removed (red), case III: up to 10% randomly drawn stations are removed (10 trials) (blue).

4.4 Relative Kriging Error

As the second approach to assess the suitability of WDB for identifying influential and expendable stations, we analyse the change in the kriging error when stations are removed from the network. The variogram is kept constant during the network modifications. 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 1).

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 mostly positive and high ($\Re > 5\%$) values, because high ranking nodes are removed as well which leads to higher rates of $\Re(\%)$.

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	**90^{th} percentile**	**95^{th} percentile**	**99^{th} percentile**	**Wet days**
I	**10% highest ranking**	9.3*	32.9*	72.3*	57.1*	69.1*
II	**10% lowest ranking**	-2.1	4.4	3.1	11.1*	-1.7
III	**10% randomly selected**	5.4*	27.3*	52.3*	42.6*	4.1

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 expendable hydrometric stations as shown in Fig.8. For example, removing low ranking stations in the German rain gauge network (Fig.8) 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 expendable 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.

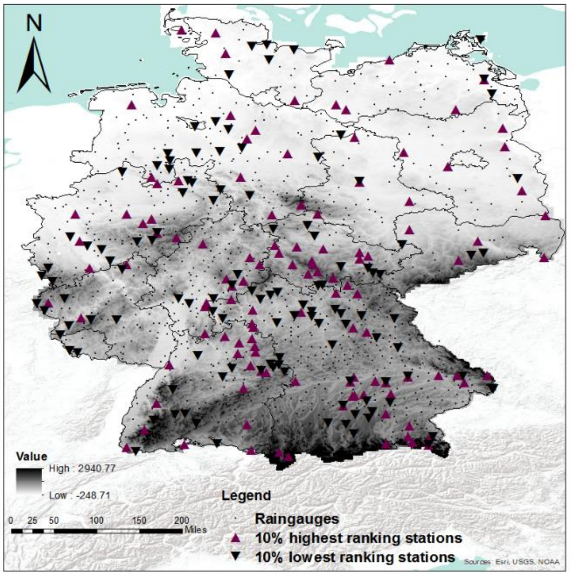


Figure 8: Location of 10% highest ranking rain gauges (red) and 10% lowest ranking stations (black) showing on Germany along with topography of the area.

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.

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 pre-processed by the Potsdam Institute for Climate Impact Research (Conradt et al., 2012).

Appendix

A. Kriging variogram modelling

The kriging modelling mandates a theoretical variogram function that is to be 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) and is given by (Adhikary et al., 2015)

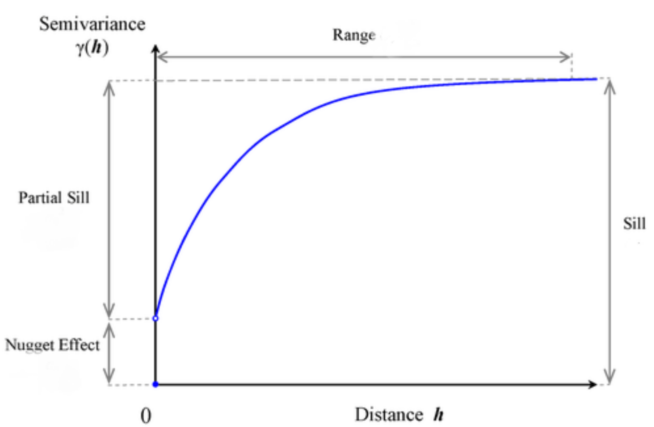
$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [(Y(i) - Y(j))^2] \tag{A1}$$

where $N(h)$ is the number of sample data points separated by a distance h; , i and j represent sampling locations separated by a distance 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 required for kriging interpolation.

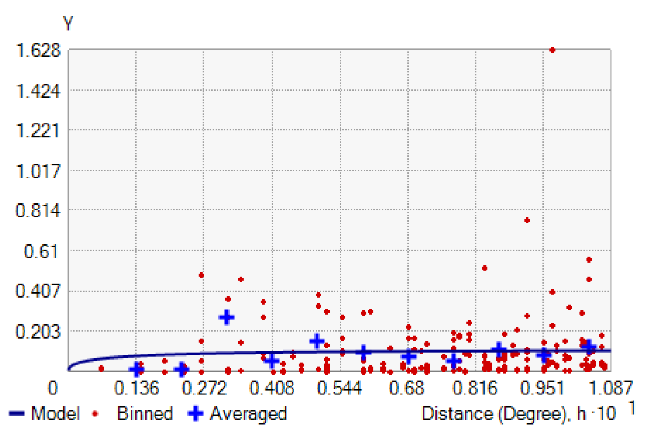
The variogram models are a function of three parameters, known as the range, the sill, and the nugget (Fig.A1 (a)). The range is typically the level of h at the correlation between point values is zero (i.e., there is no longer any spatial autocorrelation). The value of γ at range is called the sill. The variance of the sample is used as an estimate of the sill. Nugget represents measurement error and/or microscale variation at spatial scales that are too fine to detect 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 spatial. The difference between the sill and the nugget is known as the partial sill.

The value of all the parameters and resultant variogram for mean, 90th percentile, 95th percentile, 99th percentile and wet days has been reported in the Table A1 and Figure A1 (b-d) respectively. The variogram has been kept constant during network reductions.

(a) Typical variogram model

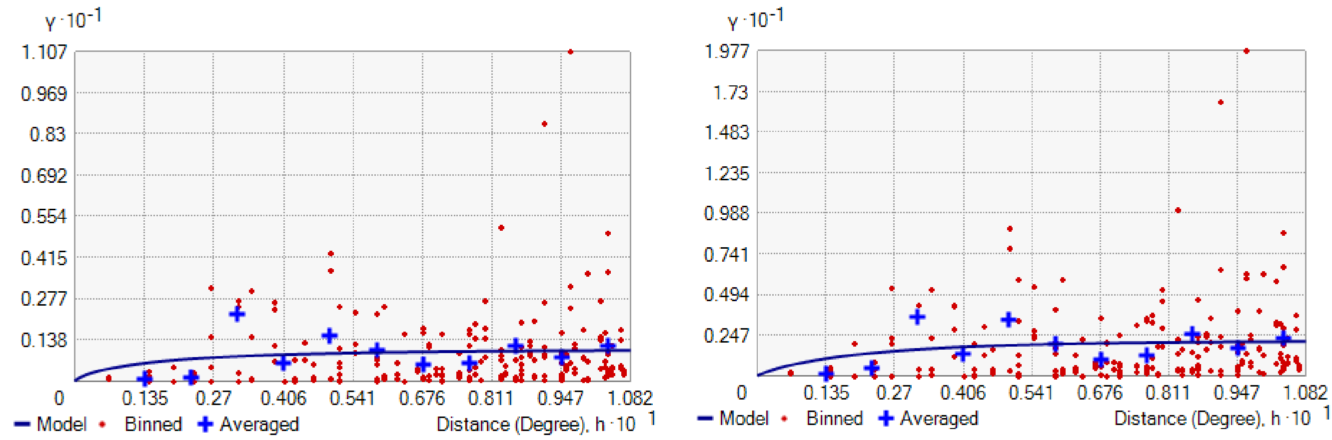


(b) Mean



(c) 90th percentile

(d) 95th percentile



(e) 99th percentile **(f) wet days**

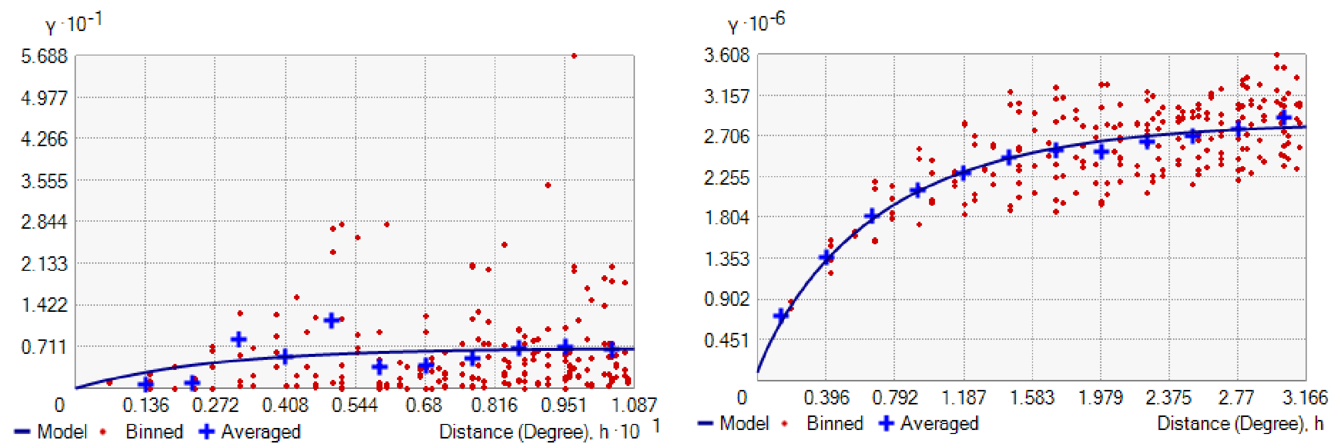


Figure A1: Typical variogram models (a) and fitted variogram models for mean (b), 90th percentile (c), 95th percentile (d), 99th percentile (e) and wet days (f).

Table A1: Parameters values for the fitted variogram.

Parameters	**Mean**	**90th percentile**	**95th percentile**	**99th percentile**	**Wet days**
Nugget	0.0058	0	0	0	0.905
Range	0.0782	0.0782	0.0782	0.0782	2.363
Partial sill	0.103	1.055	2.140	6.808	2.771

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. Ugur Ozturk was partly funded by the Federal Ministry of Education and Research (BMBF) within the project CLIENT II-CaTeNA (FKZ 03G0878A).

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.

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.

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.

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: 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.

Hlinka, J., Jajcay, N., Hartman, D. and Paluš, M.: Smooth information flow in temperature climate network reflects mass transport, Chaos Interdiscip. J. Nonlinear Sci., 27(3), 035811, doi:10.1063/1.4978028, 2017.

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, 2016a.

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, 2016b.

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.

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.

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.

Lawyer, G.: Understanding the influence of all nodes in a network, Sci. Rep., 5(1), doi:10.1038/srep08665, 2015.

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.

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

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.

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.

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.

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.

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.

Zhang, X., Zhu, J., Wang, Q. and Zhao, H.: Identifying influential nodes in complex networks with community structure, 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.