Stochastic Environmental Research and Risk Assessment (2021) 35:579–595
https://doi.org/10.1007/s00477-020-01931-9

ORIGINAL PAPER

Study of temporal streamflow dynamics with complex networks: network construction and clustering

Nazly Yasmin[2] · Bellie Sivakumar[1,2]

Accepted: 7 November 2020 / Published online: 18 November 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Applications of the concepts of complex networks for studying streamflow dynamics are gaining momentum at the current time. The present study applies a coupled phase space reconstruction–network construction method to examine the clustering property of the temporal dynamics of streamflow. The clustering of the temporal streamflow network is determined using clustering coefficient, which quantifies the tendency of a network to cluster (a measure of local density). Monthly streamflow time series observed from each of 639 stations (i.e. 639 networks) in the United States are studied. The presence of links between nodes (i.e. phase space reconstructed vectors) in each streamflow network (i.e. station) is identified using the Euclidean distance. Different distance thresholds are used to examine the influence of threshold on the clustering coefficient results and to identify the critical threshold. The results indicate that the distance threshold has significant influence on the clustering coefficient values of the temporal streamflow networks. With the critical distance threshold values, the clustering coefficients for the 639 stations are found to be between 0.15 and 0.81, suggesting very different types of network connections and dynamics. The clustering coefficient values are found to provide useful information on the influence of a given month (i.e. timestep) of the year on the temporal dynamics. Reliable interpretations of the clustering coefficient values in terms of catchment characteristics and flow properties are also possible.

Keywords Streamflow · Temporal dynamics · Complex networks · Nonlinear dynamics · Clustering coefficient · Distance threshold

1 Introduction

Adequate understanding of the temporal (and spatial) dynamics of streamflow is important for a wide range of purposes in hydrology and water resources, including for forecasting of floods and droughts and estimation of soil erosion. The past decades have witnessed the development and applications of numerous scientific concepts and mathematical techniques to study the temporal dynamics of streamflow. In recent years, applications of the concepts of complex networks (A network is a set of points, called nodes, connected by a set of lines, called links) to study the temporal dynamics of streamflow have been gaining momentum (e.g., Braga et al. 2016; Serinaldi and Kilsby 2016; Han et al. 2018; Yasmin and Sivakumar 2018); see Sivakumar and Woldemeskel (2014, 2015), Halverson and Fleming (2015), Fang et al. (2017), Konapala and Mishra (2017), Naufan et al. (2018), Agarwal et al. (2019, 2020), Tiwari et al. (2019), Alarcòn and Lozano (2019), and Han et al. (2020) for applications of complex networks-based methods for studying the dynamics of streamflow and other hydrologic processes, more broadly; see also Sivakumar (2015).

In the application of the complex networks-based methods, a fundamental step is the construction of the network, i.e. nodes and links. Studies employing the complex networks-based methods to examine the temporal dynamics of streamflow have largely used the streamflow

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00477-020-01931-9) contains supplementary material, which is available to authorized users.

✉ Bellie Sivakumar
b.sivakumar@iitb.ac.in

[1] Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

[2] UNSW Water Research Centre, School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia

data in their original form for network construction (e.g., Braga et al. 2016; Serinaldi and Kilsby 2016). That is, given a streamflow time series from a monitoring station, each timestep (e.g., daily) is considered as a node and all the possible connections between the different timesteps as the links for network construction. The actual links are identified based on some criterion, such as the geometric height in streamflow values between the different timesteps.

This type of network construction is indeed useful to examine the temporal dynamics of streamflow, as has been demonstrated by past studies (e.g., Braga et al. 2016; Serinaldi and Kilsby 2016). However, it also has an important limitation in that it offers only a "one-dimensional" view, i.e. the construction considers links in streamflow values only between any two (i.e. a pair of) timesteps. Since the evolution of streamflow often possesses patterns across multiple timesteps, this type of network construction is inadequate. Rather, a "multi-dimensional" perspective, where more than two timesteps can be considered to construct the network, is necessary. In the "multi-dimensional" perspective of data/network construction, concepts of nonlinear dynamics and chaos are useful. In particular, the phase space reconstruction concept (Packard et al. 1980) allows reconstruction of a multi-dimensional phase space using only a single-variable time series through some techniques, such as delay embedding (Takens 1981). Numerous studies have explored the utility of the phase space reconstruction concept to represent the temporal dynamics of streamflow and other hydrologic processes, and the reader is referred to Hossain and Sivakumar (2006), Kim et al. (2009), Sivakumar (2009, 2017), Kyoung et al. (2011), and Tongal et al. (2013) for details.

In light of the "multi-dimensional" view offered by the phase space reconstruction concept, Yasmin and Sivakumar (2018) have recently explored the utility of the concept in the complex networks-based analysis of streamflow. Instead of treating each timestep of streamflow data as a node in the network, as was done by the earlier studies (e.g., Braga et al. 2016; Serinaldi and Kilsby 2016), Yasmin and Sivakumar (2018) treated each point (vector) in the multi-dimensional reconstructed phase space as a node. With this type of network construction, they determined the strength of each node (and of the entire network, i.e. nodes of streamflow series from any given station) using a distance measure (i.e. Euclidean distance between any two nodes). Yasmin and Sivakumar (2018) applied this coupled phase space reconstruction–network construction approach independently to monthly streamflow time series observed from 639 stations (i.e. 639 networks) in the contiguous United States. For any given streamflow network or station (i.e. streamflow series), they used the distribution of the strength of the nodes to recognize the network type. They also used the node strengths for the different stations to identify the similarities and differences among the 639 streamflow stations.

Encouraged by the results reported by Yasmin and Sivakumar (2018), the present study applies the coupled phase space reconstruction–network construction approach to study another important property of the temporal streamflow network, i.e. clustering. The clustering property of the streamflow network is examined by determining the clustering coefficient. The clustering coefficient quantifies the tendency of a network to cluster (Watts and Strogatz 1998) and, therefore, is basically a measure of local density. The coupled phase space reconstruction–network construction approach with clustering coefficient calculation is applied independently to monthly streamflow time series observed over a period of 53 years (January 1950–December 2002) from each of 639 stations in the contiguous United States. The presence of links between the different nodes in the streamflow network (i.e. any given station) is identified using the Euclidean distance between any two reconstructed vectors of the streamflow time series (using phase space reconstruction). If the distance between any two vectors is below a given threshold, then there is a link; otherwise, there is none. Different distance thresholds are also used to examine the influence of threshold on the clustering coefficient results and to identify the critical threshold. The clustering coefficients for the 639 streamflow networks (i.e. stations) are used to identify the similarities and differences among them. The relationship between the clustering coefficient of the streamflow network and catchment/flow properties is also examined.

The outcomes from this study would be useful for a variety of purposes associated with streamflow (and other hydrologic) studies, including identification of the appropriate complexity of streamflow models, prediction of streamflow (especially those based on "nearest neighbor" techniques—for example, clustering coefficient values as the basis to identify the nearest neighbors), estimation of missing data or temporal data-filling, and formulation of a streamflow/catchment classification framework (i.e., based on clustering coefficient to identify "similarity" among catchments).

The rest of this paper is organized as follows. The concept of clustering coefficient and its calculation in the context of the phase space reconstruction-based network construction are described in Sect. 2. Details of the study area and the streamflow data used for analysis in the present study are provided in Sect. 3. Results and discussion are presented in Sect. 4, followed by conclusions in Sect. 5.

2 Methodology

In the present study, the temporal dynamics of streamflow is examined through the clustering property, measured in terms of clustering coefficient, of the streamflow network. The methodology involves three major steps: (1) multi-dimensional phase space reconstruction using only the single-variable streamflow time series; (2) network construction based on phase space reconstruction vectors (rather than the original time series); and (3) calculation of clustering coefficient. The methodology is described below. For an easier understanding of the concepts and calculations, the presentation is ordered as follows. First, details of the clustering coefficient method, in a general sense, are provided. Next, the coupled phase space reconstruction–network construction approach is described. Finally, the calculation of the clustering coefficient for the phase space reconstruction-based network construction is described.

2.1 Clustering and clustering coefficient

One of the most fundamental characteristics of networks is their tendency to cluster. The tendency of a network to cluster is quantified by the clustering coefficient. Several definitions and associated methods of clustering coefficient exist in the literature (e.g., Watts and Strogatz 1998; Barrat and Weigt 2000; Newman 2001). However, the method proposed by Watts and Strogatz (1998) has been widely used in many different fields, including in hydrology. This method measures the local density.

Let us assume a node i in a network and that it has k_i links (identified based on some criterion, such as distance or correlation) which connect it to k_i other nodes. If the neighbors of the original node i were part of a cluster, then there would be $k_i(k_i - 1)/2$ links between them. With this, the clustering coefficient of node i is calculated as the ratio between the number E_i of links that actually exist between these k_i nodes and the total number $k_i(k_i - 1)/2$ in the network, as:

$$CC_i = \frac{2E_i}{k_i(k_i - 1)}. \quad (1)$$

The procedure is then repeated for each and every node of the network. The clustering coefficient of the whole network CC is the average of the clustering coefficients CC_i's of all the individual nodes. An example for the calculation of the clustering coefficient of a network can be seen in Sivakumar and Woldemeskel (2014).

The clustering coefficient of the individual nodes and of the entire network can be used to obtain important information about the type of network, grouping (or classification) of nodes, and identification of dominant nodes (e.g., super nodes), among others. For instance, a high clustering coefficient (close to 1.0) indicates a regular network, since all the nodes are connected among themselves and in an equal manner, while a very low clustering coefficient (close to 0) indicates a random network. The clustering coefficient for a small-world network (e.g., Watts and Strogatz 1998) or a scale-free network (e.g., Barabási and Albert 1999) is generally not only smaller than that of the regular network, but also considerably larger than that of a comparable random network (i.e., having the same number of nodes and links).

2.2 Network construction using phase space reconstruction

Let us assume a (monthly) streamflow time series X_i, where $i = 1, 2, \ldots, N$, as shown in Fig. 1a (with $N = 6$) and that the objective is to study the temporal dynamics using complex networks-based methods. A common procedure to construct the streamflow network is to treat each timestep i (of X_i) as a node in the network and identify the links between them, as shown in Fig. 1b, in a purely "one-dimensional" sense, as has been done in many past studies (e.g., Braga et al. 2016; Serinaldi and Kilsby 2016). In the "multi-dimensional" perspective for network construction, proposed by Yasmin and Sivakumar (2018), the concept of phase space reconstruction is used to construct a multi-dimensional phase space using only a single-variable time series, through delay embedding (e.g., Takens 1981), as:

$$\boldsymbol{Y}_j = (X_j, X_{j+\tau}, X_{j+2\tau}, \ldots, X_{j+(m-1)\tau/\Delta t}) \quad (2)$$

where $j = 1, 2, \ldots, N - (m - 1)\tau/\Delta t$; m is the dimension of the vector $\boldsymbol{Y}_j$ (embedding dimension); and τ is an appropriate delay time taken to be some suitable integer multiple of the sampling time Δt. For instance, with the time series in Fig. 1a, the reconstruction of the phase space in two dimensions ($m = 2$) with a delay time of 1 ($\tau = 1$), i.e. the projection of the attractor on the plane $\{X_i, X_{i+1}\}$, is shown in Fig. 1c, with a total five vectors: $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, $\boldsymbol{Y}_3$, $\boldsymbol{Y}_4$, and $\boldsymbol{Y}_5$. In this reconstructed phase space, the first vector ($\boldsymbol{Y}_1$) holds the information of months 1 and 2, the second vector ($\boldsymbol{Y}_2$) holds the information of months 2 and 3, and so on. Each of these points in the phase space is considered as a node in the network (i.e. there is a total of five nodes, rather than the six with the original time series), with each node having possible links with the rest of the nodes. Figure 1d shows all the possible links in this network of five nodes.

2.3 Clustering coefficient calculation of phase space reconstruction-based network

With the above multi-dimensional phase space reconstruction for network construction, for a given m, the

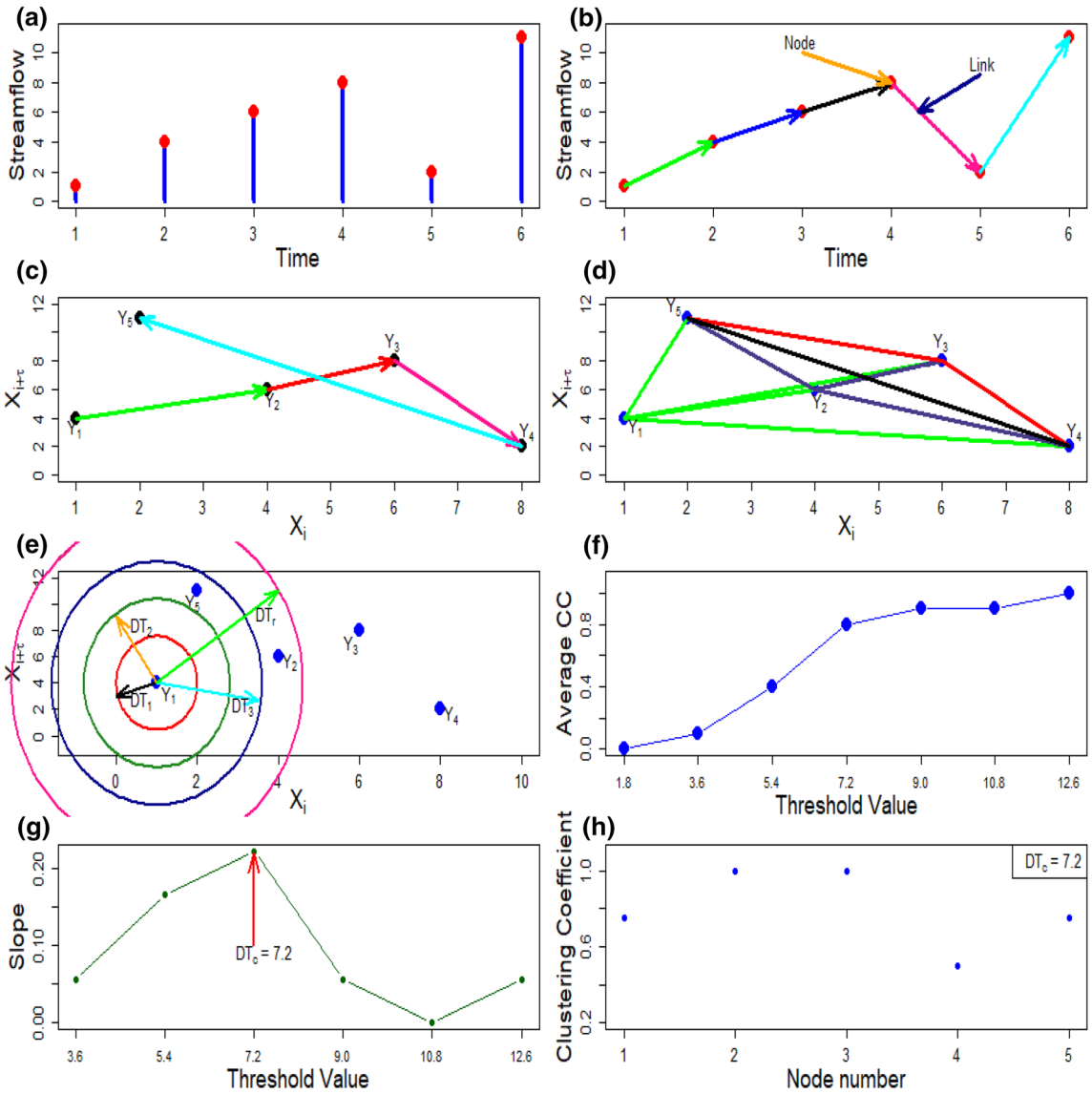

Fig. 1 Streamflow time series and network construction: **a** time series; **b** one-dimensional network construction (i.e., m = 1); **c** phase space reconstruction in higher dimensions (with dimension $m = 2$); **d** network construction using phase space reconstruction (with $m = 2$); **e** example of actual connections; **f** average clustering coefficient versus distance threshold (DT); **g** slope versus distance threshold; and **h** clustering coefficient value of each node corresponding to the critical threshold (DT_C)

distance (e.g., the Euclidean distance) between any two node pairs i and j, and denoted as d_{ij}, can be calculated. To define the distance threshold, distances (i.e. difference in streamflow values) from one node to the other nodes are considered as a reasonable measure. Once the distances d_{ij} are obtained, the existence/non-existence of links in the network can be identified by considering a distance threshold(s), DT. If $d_{ij} \leq DT$, then there is a link; otherwise, there is no link. Once the existence/non-existence of links is identified, the clustering coefficient for any node is calculated. However, since the appropriate distance threshold is not known a priori, different distance threshold values (DT_r) are considered, as indicated through different radii (r) in Fig. 1e, and the corresponding clustering coefficient values are calculated. As may be realized, when the assumed distance threshold value is large enough to have all the node–node distances below such a threshold, further increasing the distance threshold will not have any effect. The procedure is repeated for each of the nodes in the network. Then, the average clustering coefficient for the entire network for each of the different distance thresholds are also obtained. Figure 1f shows an example of the relationship between the average clustering coefficient of all the nodes in the network (i.e., the clustering coefficient of the network) and the distance threshold value. In general, the clustering coefficient increases with

an increase in the distance threshold value, as one would normally expect.

While Fig. 1f shows that the clustering coefficient increases with the distance threshold value and indicates the maximum distance threshold necessary, it is not really useful to identify the most appropriate threshold for network connections, i.e., critical threshold. One way to address this issue is by calculating the 'extent of difference' in the clustering coefficient value with respect to the distance threshold and identifying the distance threshold that yields the highest difference. This can be done by calculating the slope of the clustering coefficient of the network for different intervals of the distance threshold and identifying the slope that is the highest. Figure 1g presents the outcome of such a procedure. As can be seen in Fig. 1g, the value of the slope seems to increase with the threshold value up to a certain maximum point at about $DT = 7.2$ and then decreases beyond that. Therefore, this point can be considered as the critical threshold, i.e., the optimum distance that preserves the key properties of the network (Gao and Jin 2009), such as strong/weak connections between one node to every other node in the network. Figure 1h presents the clustering coefficient value of each node depending on the critical threshold value.

As may be realized, the results presented in Fig. 1 are basically for a two-dimensional phase space reconstruction, i.e. $m = 2$. To find the optimal embedding dimension, the procedure needs to be repeated for different embedding dimension values. There exists a variety of methods, especially in the context of nonlinear dynamic and chaos theories, to identify the optimal embedding dimension; see Sivakumar (2017) for details. In the present study, the optimal embedding dimension for phase space reconstruction is obtained using the false nearest neighbor algorithm (Kennel et al. 1992).

3 Study area and data

In the present study, to examine the temporal dynamics of streamflow using the coupled nonlinear dynamics-complex networks approach and the clustering coefficient, streamflow data from the United States (US) are studied. Specifically, monthly streamflow data observed over a period of 53 years (January 1950–December 2002) at each of 639 stations in the contiguous United States are considered. Figure 2 presents a map of the contiguous United States, with the locations of the 639 streamflow stations considered in this study.

The contiguous United States includes 48 states and the District of Columbia. The region occupies an area of about 8,080,464 km^2. Of this area, about 7,663,941 km^2 is contiguous land, and about 416,522 km^2 is water area (https://www.cia.gov/library/publications/the-world-factbook/fields/279.html#as). Water area is the sum of the surfaces of all inland water bodies, such as lakes, reservoirs, and rivers, as delimited by international boundaries and/or coastlines. The contiguous US and the rivers cover a wide range of hydroclimatic, topographic, geomorphic, and land use properties, and other factors that influence the streamflow process and, thus, represent a wide range of streamflow dynamic properties (Poff et al. 2006). For instance, the Great Plains are semi-arid, and the mountains are alpine in the western United States (Robinson and Dietz 2019). The climate is arid in the Great basin, desert in the southwest, Mediterranean in the coastal California, and oceanic in the coastal northwest (e.g., Encyclopaedia Britannica 2019; Antevs 1952). The climate varies from humid continental in the north to humid subtropical in the south in the eastern United States.

Over the past few decades, extreme hydrologic events have been reported to be increasing across much of the contiguous United States. However, these trends vary from region to region. Some areas, such as the southwest US, have experienced prolonged droughts (e.g. Cook et al. 2015), whereas regions in the central US have experienced an increase in the frequency of heavy precipitation (e.g. Pryor et al. 2009). Areas in the eastern US have also observed an increase in the frequency of heavy precipitation and streamflow events (e.g. Groisman et al. 2001).

The contiguous United States is home to about 250,000 rivers, with a total length of about 5,600,000 km (Kammerer 1990). The rivers include several large and major rivers, including the Missouri River, the Mississippi River, the Rio Grande, the Arkansas River, and the Colorado River. The main stems of 38 rivers are at least 800 km (500 miles) long. The longest river is the Missouri River, which is a tributary of the Mississippi River, but the biggest in terms of water volume is the deeper Mississippi River; see Kammerer (1990) for additional details.

As a result of the hydroclimatic variability, topography, river length, basin area, and many other factors, streamflow dynamics in the contiguous United States also significantly vary by region. Streamflow values have also generally increased during the twentieth century in the contiguous United States. These increases, however, have not been uniform across the range of annual streamflow, from the minimum to the maximum, nor have they been uniform geographically or seasonally. However, some level of human activity influences the flow of water in most rivers in the United States. The most indicated activities are dams and diversions for irrigation, but groundwater pumping and land use change are also significant factors. Human activities can considerably affect the variability of streamflow in response to weather and climatic conditions. Indeed, dams are constructed in part to reduce flooding and to increase flows during droughts (Chen et al. 2016).

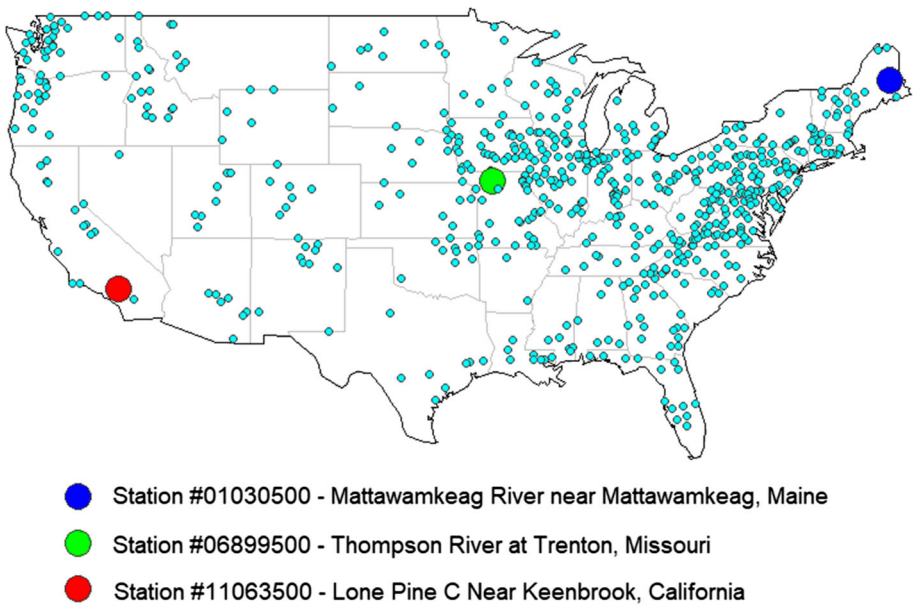

Fig. 2 Locations of 639 streamflow monitoring stations in the United States, with focus on three stations that are used for illustration of the temporal network analysis

The streamflow data considered for analysis in this study are from the US Geological Survey (USGS) database for surface water (https://nwis.waterdata.usgs.gov/nwis/sw), and specifically from the Hydro-Climatic Data Network (HCDN). The HCDN was originally developed by Slack and Landwehr (1992), and subsequently updated at various times. The HCDN is a sub-set of all USGS streamflow gaging stations. Details of the HCDN can be seen in Lins (2012) and Kiang et al. (2013), among others. The HCDN streamgage stations have been screened to eliminate sites where human activities influence the natural flow of the waterway. As of now, the HCDN database provides streamflow data of 743 stations until the year 2009, but the period and length of the data for these stations also vary. Details of the HCDN data are available in Lins (2012); see also http://water.usgs.gov/osw/hcdn-2009.

A number of studies have examined the HCDN streamflow data set or a sub-set, for various purposes and using different methodologies. Among such studies, the ones by Sivakumar (2003), Sivakumar and Singh (2012), Sivakumar and Woldemeskel (2014), Vignesh et al. (2015), and Yasmin and Sivakumar (2018) are particularly relevant to the present study, for their focus on the application of nonlinear dynamic concepts (including phase space reconstruction) to study the temporal dynamics of streamflow or complex networks to study spatial/temporal connections in streamflow dynamics. These studies have generally analyzed monthly (average) streamflow data that were observed before the year 2003. To be consistent with these studies in the period of streamflow data analyzed, especially with those of Vignesh et al. (2015) and Yasmin and Sivakumar (2018), which are highly relevant to the present study, monthly streamflow data observed during January 1950–December 2002 from each of the above 639 streamflow stations are considered. For additional details about these streamflow data, including some of their important statistical characteristics, the reader is directed to Vignesh et al. (2015) and Yasmin and Sivakumar (2018).

4 Results and discussion

4.1 Analysis

To examine the temporal streamflow connections, the methodology proposed in Sect. 2 is employed independently to each of the 639 monthly streamflow time series (i.e. streamflow networks) from the contiguous United States. The clustering coefficient analysis is performed on the normalized streamflow time series. The purpose of normalization of the streamflow time series here is to have a common (and standard) range of streamflow time series for all the 639 stations. This helps to avoid possible distortions due to the differences in the ranges of streamflow values between the individual stations (especially since the streamflow values in the 639 stations have differences of several orders of magnitude), including in the range of distances between the reconstructed vectors and in the distance threshold values to be considered for identification of the optimal distance threshold (i.e. number and intervals of distance threshold values). As mentioned earlier, each streamflow time series is considered as a network, and the reconstructed vectors using the phase space reconstruction concept are considered as the nodes and the connections between the nodes are considered as the links. For each station, three different delay time values, $\tau = 1$, 3, and 12,

are used for phase space reconstruction, i.e., monthly, seasonal, and annual separation of elements in the phase space reconstruction. For each of these three τ values, the optimal embedding dimension for each station is chosen by employing the false nearest neighbor (FNN) method (Kennel et al. 1992). The FNN dimension estimates for the 639 streamflow time series have also been presented in the study by Vignesh et al. (2015), which also discusses the influence of delay time on dimension estimation. Here, for the sake of simplicity and considering space constraints, the presentation and discussion are limited only to $\tau = 1$. However, for the benefit of the interested reader, the results for $\tau = 3$ and $\tau = 12$ are presented in the Supplementary Information.

For any given node i in the network, the clustering coefficient (CC) is calculated based on the distance threshold (DT). To study the influence of the distance threshold and, hence, to identify the optimal threshold and the clustering coefficient, 26 different distance threshold values (at equal intervals) are considered, to have 25 intervals of increments. The interval between the thresholds is identified based on the difference between the maximum and the minimum distances between the reconstructed vectors. The threshold that provides the largest gradient in the clustering coefficient is considered as the optimal threshold.

4.2 Clustering coefficient results

For the purpose of illustration of the clustering coefficient analysis for temporal connections in the streamflow network, results for only three (of the 639) streamflow time series are presented here. The three stations are chosen from different regions in the contiguous US. These stations are: (1) Station #01030500 (Mattawamkeag River near Mattawamkeag, Maine) in the northeast; (2) Station #06899500 (Thompson River at Trenton, Missouri) in the mid-west; and (3) Station #11063500 (Lone Pine C near Keenbrook, California) in the southwest. The locations of these stations are shown in Fig. 2, and their streamflow time series (original non-normalized time series) are presented in Figs. 3a, 4a and 5a, respectively. The three streamflow time series reveal that: (1) Station #01030500 exhibits flow variations over the entire period, with a clear annual cycle throughout; (2) Station #06899500 shows some slight increases in flow, including one or more significant peaks at certain times, in the latter periods when compared to the entire period; and (3) Station #11063500 shows no specific trends in terms of flow evolution, but significant peaks in flow appear at certain times. Table 1 presents some basic statistics for these three streamflow time series, including the mean, minimum, and maximum flows. The table also presents the delay time considered for presentation here (i.e., $\tau = 1$) and the corresponding FNN dimensions (i.e., optimal dimension for phase space reconstruction) obtained for the three time series, which are 8, 5, and 11, respectively.

Figures 3, 4 and 5 present the results of the clustering coefficient analysis of streamflow time series from the above three stations, respectively. Each of Figs. 3, 4, and 5 includes the following: (a) monthly streamflow time series (original non-normalized); (b) phase space reconstruction (of original non-normalized time series) in two dimensions (i.e., $m = 2$), with $\tau = 1$; (c) clustering coefficient (CC) value of each of the available nodes (for normalized time series) for six different distance threshold (DT) values (selected out of the 26 considered); (d) average clustering coefficient (CC) value of all nodes for different threshold (DT) values; (e) identification of the critical threshold (DT_C), using the slope of the average clustering coefficient and threshold value; and (f) clustering coefficient value of each of the available nodes at the critical threshold value. Overall, the three streamflow time series and, correspondingly, the phase space reconstruction and clustering coefficient results show distinct behaviors. Some specific observations from the results are as follows:

1. The phase spaces for the three streamflow time series show some kind of an "attractor." However, the extent of the spread of the trajectories in the phase space does not seem to indicate a very low-dimensional system (which is also reflected by the FNN dimensions). The one obtained for Station #01030500 (Fig. 3b) seems a relatively better attractor (i.e., clearer and more concentrated in the phase space) when compared to the other two.
2. The clustering coefficient values of nodes (for six chosen distance thresholds, which are also different for the three stations) for the three time series show distinct behaviors (see Figs. 3c, 4c, 5c), as follows. All nodes of Station #01030500 have clustering coefficient values that are somewhat evenly distributed from 0 to 1 for all the threshold values. Most of the nodes of Station #06899500 have high clustering coefficient values from 0.8 to 1 for all threshold values and very few nodes have lower clustering coefficient values from 0 to 0.79. The clustering coefficient values of Station #11063500 show similar results to those obtained for Station #06899500, but with some notable exceptions (especially at lower node numbers, i.e., earlier periods in time).
3. The relationship between the clustering coefficient of the network (i.e., average of the clustering coefficients of all nodes) and the distance threshold values for the three time series exhibit interesting behavior. Overall, as expected (and highlighted in Sect. 2, Fig. 1), the

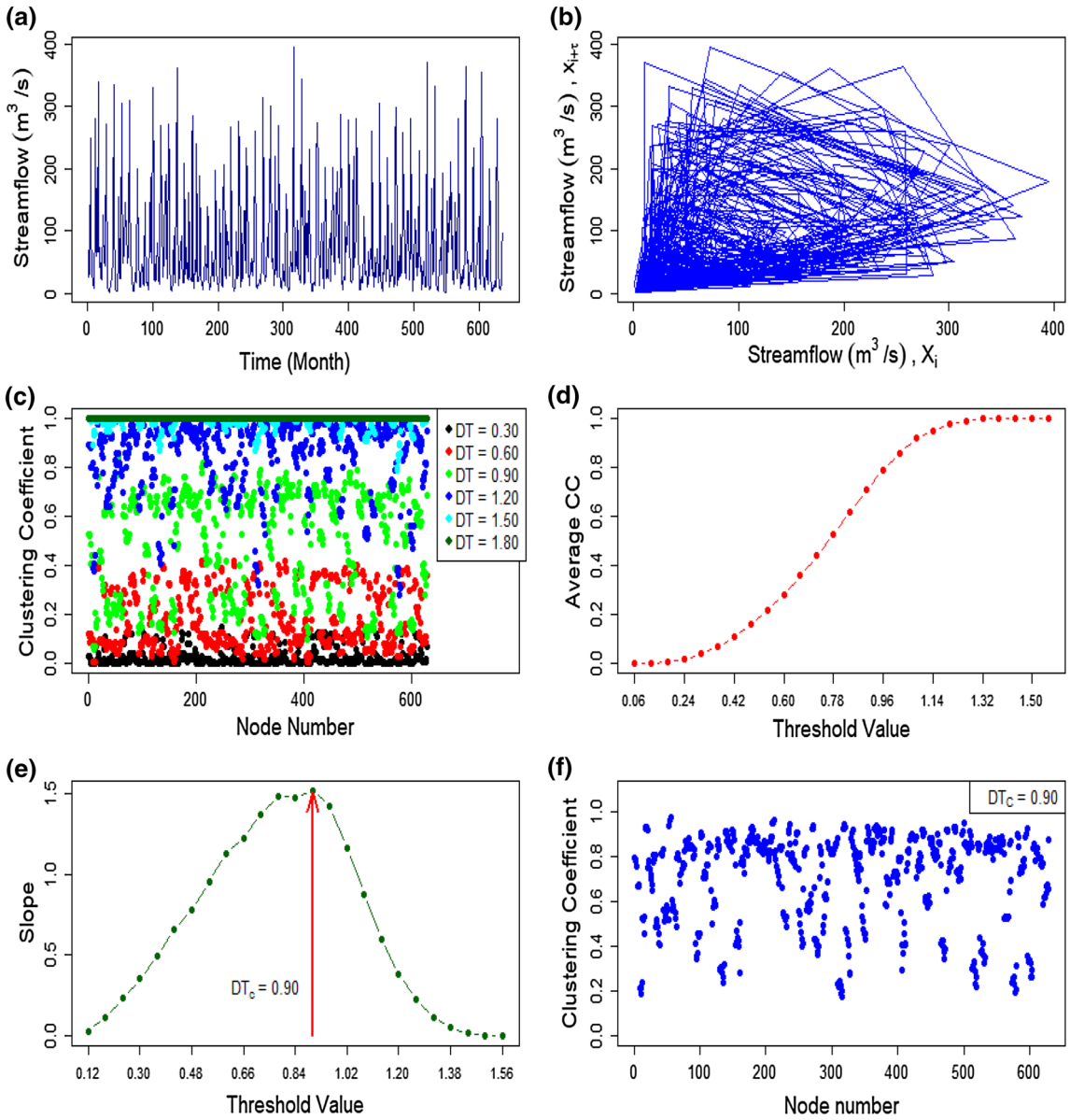

Fig. 3 Streamflow time series and clustering coefficient results for Station #01030500: **a** time series; **b** phase space reconstruction in two dimensions, with delay time $\tau = 1$; **c** clustering coefficient versus node number; **d** average clustering coefficient versus distance threshold (DT); **e** slope versus distance threshold; and **f** clustering coefficient values of all available nodes corresponding to the critical threshold (DT_C)

clustering coefficient value increases with an increase in the threshold value and indicates the maximum distance threshold for all the three stations (Figs. 3d, 4d, 5d). However, the shape of the curve is different for the three stations: for example, the clustering coefficient value for Station #01030500 (Fig. 3d) steadily increases from 0 to 1 with an increase in the threshold value up to 1.32 and saturates beyond this threshold value. The clustering coefficient value for Station #06899500 (Fig. 4d) seems to increase significantly from 0 to 0.89 with the threshold value up to 0.35, then increases somewhat steadily from 0.93 to 0.98 with the threshold value up to 0.5, and finally increases and saturates at some points from 0.93 to 1 beyond this threshold value. For Station #11063500 (Fig. 5d), the clustering coefficient rapidly increases from 0 to 0.8 with threshold value up to 0.42, and then somewhat evenly increases from 0.8 to 1 with the threshold value up to 1.33 and saturates beyond this threshold value.

4. There is notable difference in the critical distance threshold (*DTc*) values identified for the three streamflow time series. The critical threshold value for Station #01030500 is 0.90 (Fig. 3e), for Station

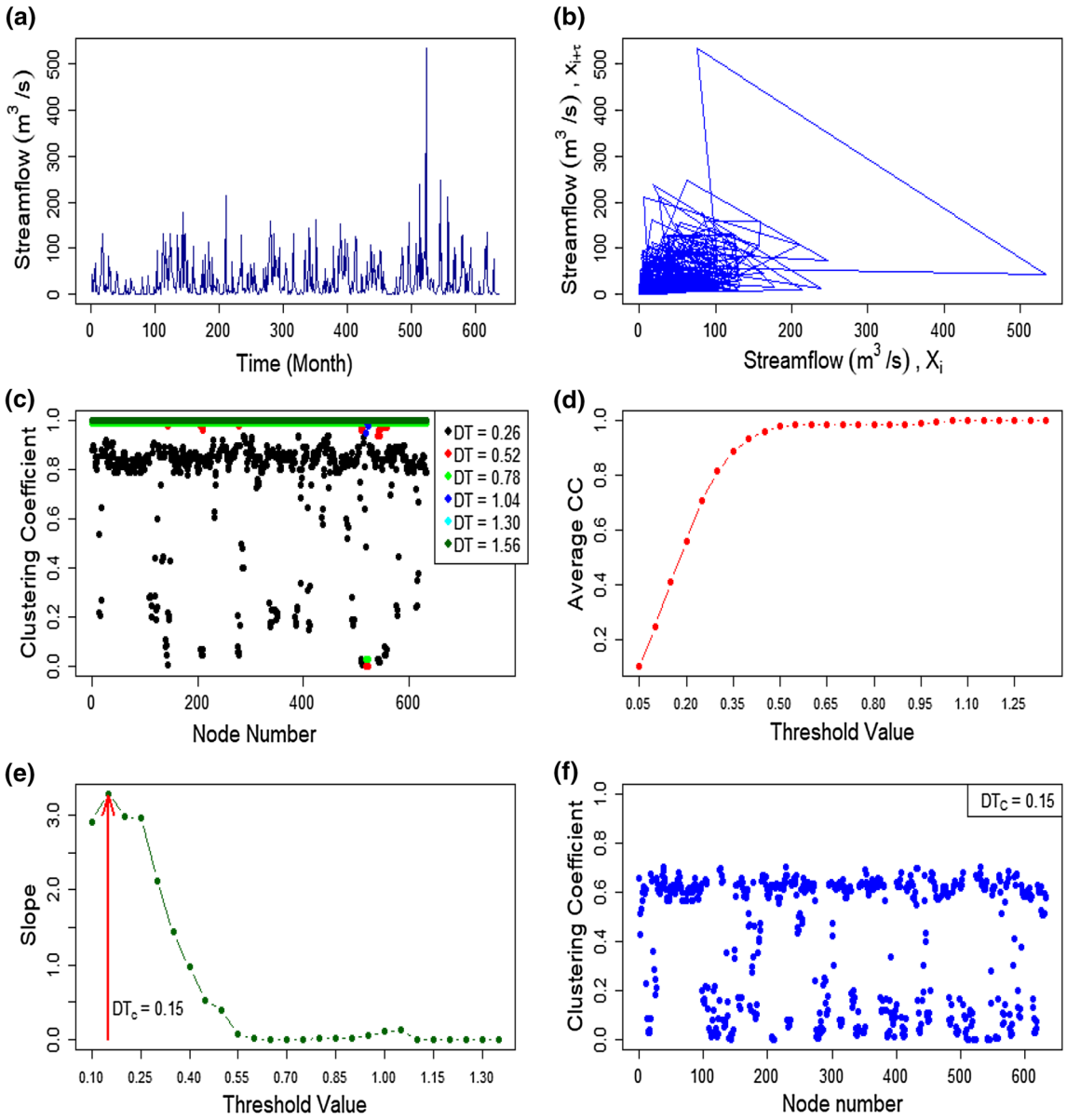

Fig. 4 Streamflow time series and clustering coefficient results for Station #06899500: **a** time series; **b** phase space reconstruction in two dimensions, with delay time $\tau = 1$; **c** clustering coefficient versus node number; **d** average clustering coefficient versus distance threshold (DT); **e** slope versus distance threshold; and **f** clustering coefficient values of all available nodes corresponding to the critical threshold (DT_C)

#06899500 is 0.15 (Fig. 4e), and Station #11063500 is 0.14 (Fig. 5e).

5. There is also notable difference in the clustering coefficient values obtained for the three networks (stations). The clustering coefficient values for the three streamflow time series (i.e., corresponding to the critical threshold values) are 0.71, 0.41, and 0.42, respectively; see Table 1 for additional details.

Figure 6 presents the clustering coefficient values for the 639 streamflow time series, corresponding to the critical threshold values obtained. In this figure, six specific ranges of clustering coefficient (i.e., < 0.26, 0.26–0.35, 0.36–0.45, 0.46–0.55, 0.56–0.65, and > 0.65) are considered for better visualization and interpretations. The exact number of stations falling within each clustering coefficient range is presented in Table 2.

The results reveal the following observations, among others: (1) stations in the far northwest and far northeast have generally high clustering coefficient values; (2) stations in the mid-west and south have generally low clustering coefficient values; (3) many regions have a good mix of high clustering coefficient values and low clustering coefficient values; (4) some distant stations have similar

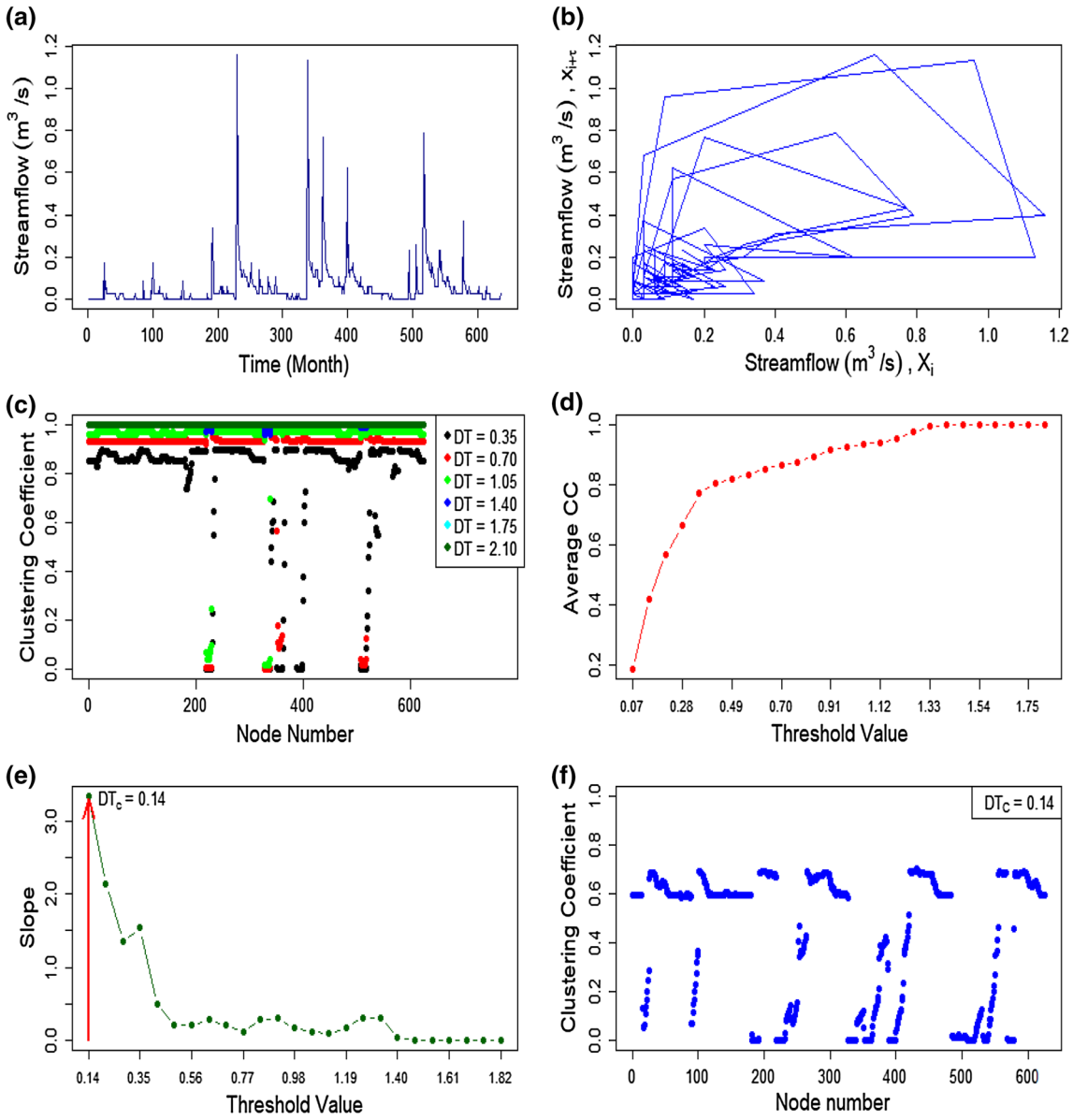

Fig. 5 Streamflow time series and clustering coefficient results for Station #11063500: **a** time series; **b** phase space reconstruction in two dimensions, with delay time $\tau = 1$; **c** clustering coefficient versus node number; **d** average clustering coefficient versus distance threshold (DT); **e** slope versus distance threshold; and **f** clustering coefficient values of all available nodes corresponding to the critical threshold (DT_C)

clustering coefficient values; and (5) some nearby stations have significantly different clustering coefficient values. These observations suggest that while geographic proximity may result in catchment similarity and, correspondingly, flow similarity in certain regions, it need not always be the case for all regions. Indeed, as seen in Fig. 6, some geographically closer stations are found to exhibit very high or very low clustering coefficient values (e.g., the catchments in the state of Florida in the southeast region have clustering coefficient values in the range as high as > 0.65 and also in the lower range of 0.26–0.35), while some catchments that are geographically distant (e.g., catchments in the far northwest and far northeast) are found to exhibit similar (e.g., very high) clustering coefficient values, i.e. > 0.65. Therefore, the clustering coefficient values can be useful for understanding not only the temporal connections in streamflow at a given station but also the spatial connections (see Sivakumar and Woldemeskel (2014) for spatial analysis).

Since the purpose in this study is to examine the temporal dynamics of streamflow using clustering coefficient analysis, it would be useful to investigate the importance of each node in any streamflow network. To this end, Fig. 7a, b, for example, presents the actual connections for two

Table 1 Streamflow characteristics and clustering coefficient results for three selected stations from the United States*

Station	Flow (m^3/s)	Delay time (τ)	Embedding dimension (m)	Critical threshold (DT_C)	Clustering coefficient (CC)
#01030500 (Maine)	Mean = 73.06 Minimum = 1.10 Maximum = 395.02	1	8	0.9	0.71
#06899500 (Missouri)	Mean = 30.41 Minimum = 0.14 Maximum = 534.05	1	5	0.15	0.41
#11063500 (California)	Mean = 0.06 Minimum = 0 Maximum = 1.16	1	11	0.14	0.42

*Figure 2 indicates the locations of these three stations, and Figs. 3, 4, and 5 present the streamflow time series and the results from the clustering coefficient analysis for these stations

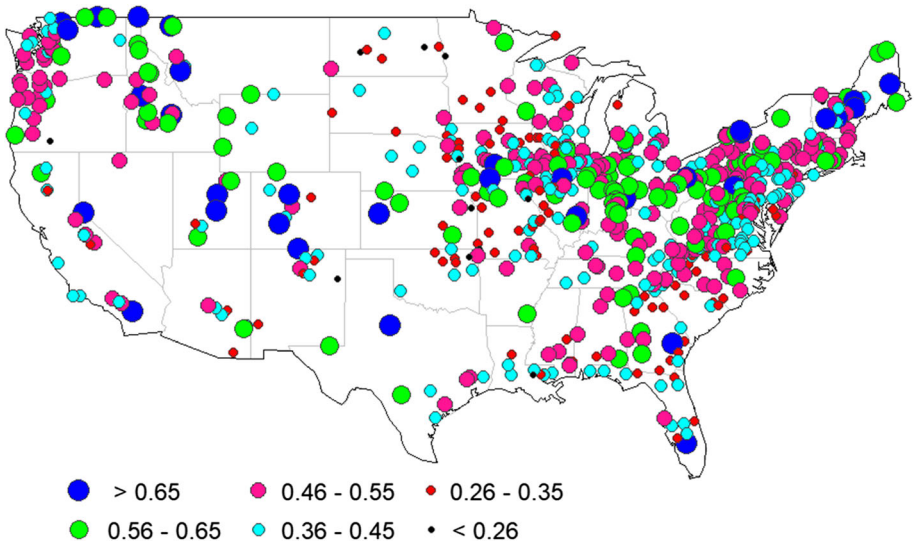

Fig. 6 Clustering coefficient values for streamflow time series from 639 stations in the contiguous United States, for delay time = 1

Table 2 Number of stations falling within six different ranges of clustering coefficient [for critical threshold (DTc)] for 639 streamflow time series in the United States

Clustering coefficient range	Number of stations for critical threshold (DT_C)
< 0.26	14
0.26–0.35	94
0.36–0.45	184
0.46–0.55	214
0.56–0.65	100
> 0.65	33

different nodes, i.e., node #17 and node #410, of the network constructed for Station #06899500 for the critical threshold value, i.e., $DTc = 0.15$. As seen, node #17 has a total of 26 actual connections below or equal to the critical threshold value of 0.15, while node #410 has a total of only 14 actual connections below or equal to the critical threshold value of 0.15. For both nodes, the connections are scattered across in time; for instance, node #17 has connections with certain nodes starting from node #120 to node #617, without any specific order (or pattern); similarly, node #410 has connections with certain nodes starting from node #14 to node #613, again without any specific order (or pattern). These observations indicate that different nodes may have different "connections" and "neighbors" and that they are not necessarily those that can be identified from the chronological history of the time series.

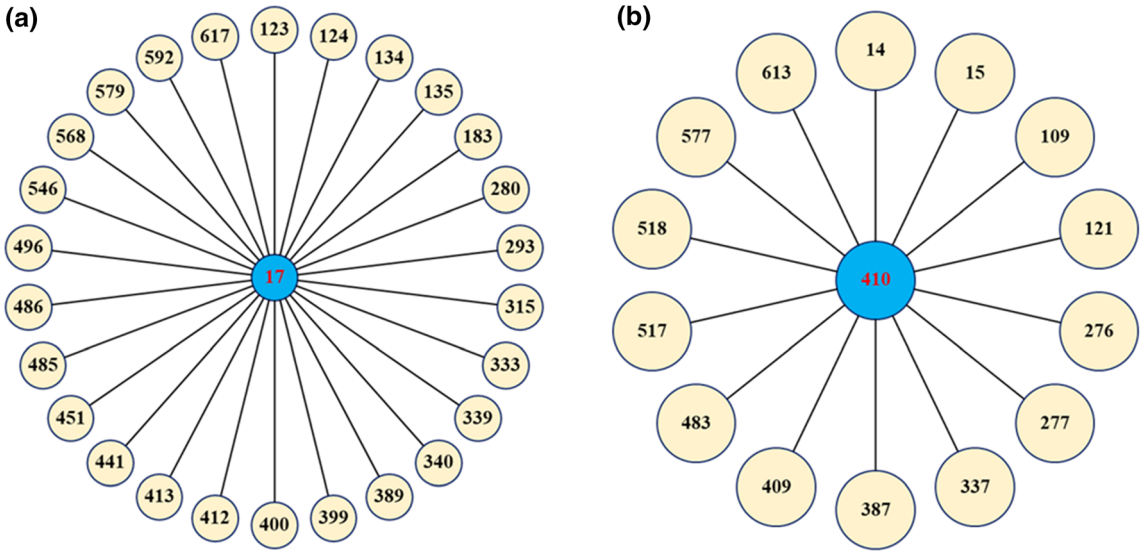

Fig. 7 Actual connections for two different nodes in the network constructed from Station #06899500, for delay time = 1: **a** node #17; and **b** node #410

This has important implications for a range of streamflow studies, including for predictions and interpolation, especially those that are based on "similarity" and "nearest neighbor" techniques. Further support to this can also be provided, in a way, based on the influence of timestep (i.e. month in this case) on the clustering coefficient results, as presented next.

4.3 Influence of time (month) on clustering coefficient analysis

Since the purpose of this study is to examine the temporal dynamics in streamflow using clustering coefficient analysis, it would also be useful to investigate the importance of each timestep in the streamflow time series. One way to do this is by simply looking at the clustering coefficient values of all nodes that have "similar structures". In the present study, since the nodes correspond to the vectors in the reconstructed phase space, identification of vectors of "similar structure" should offer a reliable means. Since monthly streamflow series are studied here, this task becomes simpler. For any given station, considering that successive months form the first elements of the successive vectors in the phase space reconstruction (m is fixed, i.e., the FNN dimension, for example) and that the successive elements of each vector are separated by a delay time that is also fixed ($\tau = 1$, i.e., monthly separation), there is a total of 12 different types of nodes, corresponding to the 12 months that form the first element of the reconstructed vector. Therefore, the average of the clustering coefficients of all the nodes in each vector structure should indicate the clustering coefficient of the given month.

Figure 8a–c presents the clustering coefficients of the 12 individual months for the three selected stations considered for illustration above. The results offer some interesting observations:

1. For Station #01030500, there seems to be a clear "cyclic" pattern in the clustering coefficient value, every 4 months. The first 4 months of the year have relatively lower clustering coefficient values (less than 0.65); the next 4 months have relatively higher clustering coefficient values (greater than 0.80); and the last 4 months have relatively lower clustering coefficient values (less than 0.70).
2. For Station #06899500, there seems to be no clear pattern in the clustering coefficient value in terms of specific number of months, but some patterns in terms of how the clustering coefficient changes with respect to each month. For example, the clustering coefficient value decreases uniformly from January to March, then increases somewhat uniformly from April to October, and finally decreases in November and December.
3. For Station #11063500, there does not seem to be any clear pattern, but the situation is somewhat opposite to the one observed for Station #06899500. The clustering coefficient value increases in the first 2 months (January and February), then decreases over the next 4 months (March to June) and saturates during June to July, and finally increases over the next 5 months (August to December).

These observations suggest that the networks-based clustering coefficient measure, with the network constructed using phase space reconstruction, provides useful

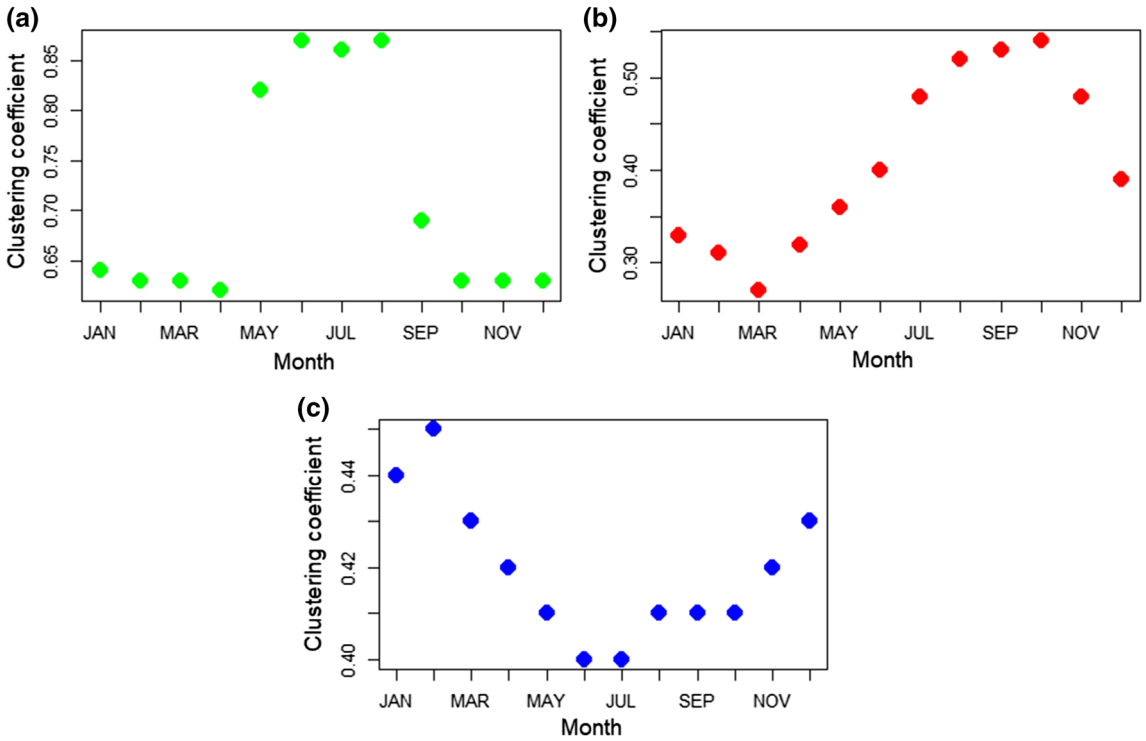

Fig. 8 Clustering coefficient of streamflow network versus month of the year, for delay time = 1: **a** Station #01030500; **b** Station #06899500; and **c** Station #11063500

information about the dominance of any month (or any other timestep, for that matter) on the evolution of streamflow and, thus, explains the seasonal and other influences.

4.4 Clustering coefficient versus catchment/flow properties

With the clustering coefficient values for the monthly streamflow time series observed at 639 stations across the United States varying from 0.15 to 0.81, it would be interesting to see how the clustering coefficient value of each network (i.e., station) compares with the catchment characteristics and streamflow statistics. To this end, Fig. 9 compares the clustering coefficient values with station drainage area (Fig. 9a), station elevation (Fig. 9b), flow mean (Fig. 9c), and flow coefficient of variation (Fig. 9d) for all the 639 stations. As seen, very low clustering coefficient values are observed mainly for catchments with very small drainage areas (less than 100 km^2), very low elevations (less than 2 m), very low flow means (less than 2 m^3/s), and very low CV values (less than 1.5). The results also generally indicate that high (and also medium) clustering coefficients are observed for all ranges (i.e., low, medium, and high) of drainage area, elevation, flow mean, and flow coefficient of variation. For instance, very high clustering coefficient values (about 0.7 or higher) are associated with very small and very large drainage areas (less than 300 km^2 as well as more than 20,000 km^2), very low and very high elevations (less than 200 m as well as more than 2000 m), and very low and very high CV values (less than 1 as well as more than 5). An exception to such is in terms of flow mean, as high clustering coefficient values are mainly associated with very low flow means (less than 1 to less than 50 m^3/s). Further, while very high clustering coefficient values are mainly observed for very few catchments, most of them also happen to have very small and very large drainage areas, very low and very high elevations, very low flow means, and very low and very high CV values, as can be seen on the right half of Fig. 9a–d.

Table 3, for instance, presents the details of 16 such stations, corresponding to the 16 points on the far right region in Fig. 9a–d (roughly marked with a box). These stations are situated in different states/regions across the United States, from northwest to southwest to south to mid-west to northeast, as presented in Table 3. Therefore, the networks-based clustering coefficient analysis, with phase space reconstruction for network construction, seems to be a suitable tool to represent the temporal dynamics of streamflow regardless of the region, with due consideration to catchment characteristics and flow statistical properties, with important implications for interpolation and other purposes.

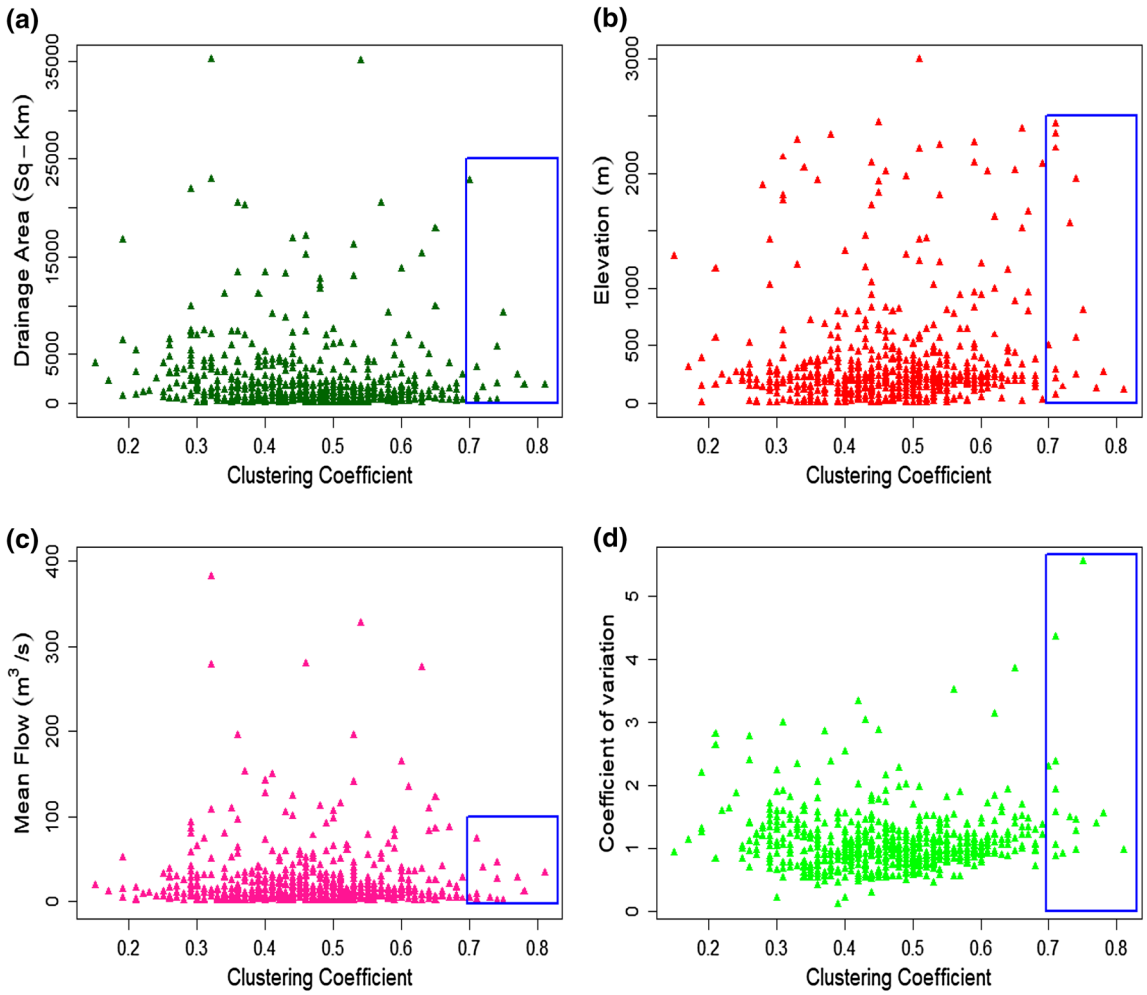

Fig. 9 Relationship between network (station) clustering coefficient and catchment/flow properties for 639 stations, for delay time = 1: **a** drainage area; **b** elevation; **c** flow mean; and **d** flow CV

5 Closing remarks

This study examined the clustering property of the temporal streamflow network through a coupled phase space reconstruction–network construction approach. Monthly streamflow time series from 639 stations (i.e. 639 streamflow networks) in the United States were studied. For each network, the presence of links between different nodes (phase space reconstructed vectors) was identified using the Euclidean distance between the nodes, and a critical distance threshold was determined based on a sensitivity analysis. The 639 streamflow networks were found to exhibit many different types of clustering behaviors, with some having very strong temporal connections and some others having very weak connections. The clustering coefficient values were found to provide useful information on the influence of any month (or timestep) on the temporal dynamics in the network and also could be reliably interpreted in terms of catchment characteristics and flow properties.

The outcomes of the present study offer further evidence on the usefulness of clustering coefficient for examining temporal streamflow dynamics and the utility of the coupled phase space reconstruction–network construction approach in complex networks-based methods. The outcomes have important implications for a wide range of studies and purposes associated with the streamflow process (and hydrologic processes, more broadly), including (1) in the identification of the appropriate type and complexity of streamflow (catchment) models; (2) in the prediction of streamflow; (3) in the estimation of missing data; and (4) in the formulation of a streamflow/catchment classification framework (i.e., based on clustering coefficient to identify "similar" catchments).

The encouraging outcomes from this study and past ones (e.g. Braga et al. 2016; Serinaldi and Kilsby 2016; Han et al. 2018; Yasmin and Sivakumar 2018) on the temporal

Table 3 Clustering coefficient values and catchment/flow properties for 16 selected stations in the United States*

Station no.	Station name	CC	Drainage area (km^2)	Elevation (m)	Mean flow (m^3/s)	CV
#01030500	Mattawamkeag River near Mattawamkeag, Maine	0.71	3672.61	66.14	73.06	1.06
#01076500	Pemigewasset River at Plymouth, New Hampshire	0.72	1610.97	139.31	39.02	0.92
#01144000	White River at West Hartford, Vermont	0.81	1787.09	114.16	33.93	0.97
#01562000	Raystown Branch Juniata River at Saxton, Pennsylvania	0.74	1958.03	242.55	26.25	0.98
#03117500	Sandy Creek at Waynesburg, Ohio	0.71	655.27	291.08	7.79	0.87
#03379500	Little Wabash River Below Clay City, Illinois	0.77	2929.28	119.57	27.44	1.4
#06860000	Smoky Hill River at Elkader, Kansas	0.75	9207.41	799.37	0.59	5.56
#06898000	Thompson River at Davis City, Iowa	0.78	1815.58	266.41	11.1	1.56
#08247500	San Antonio River at Ortiz, Colorado	0.71	284.9	2429.26	0.69	2.37
#09059500	Piney River near State Bridge, Colorado	0.71	223.26	2216.61	2.13	1.57
#09310500	Fish Creek Above Reservoir near Scofield, Utah	0.71	155.66	2337.82	1.39	1.93
#09330500	Muddy Creek near Emery, Utah	0.74	271.95	1950.72	1.07	1.27
#10258500	Palm Cyn C near Palm Springs, California	0.71	241.13	213.36	0.14	4.36
#12401500	Kettle River near Ferry, Washington	0.74	5697.98	559.86	45.42	1.46
#13240000	Lake Fork Payette River AB Jumbo CR near Mccall, Idaho	0.73	126.65	1566.67	4.01	1.49
#08080500	Double Mountain Fork Brazos River near Aspermont, Texas	0.7	22,781.55	495.24	3.81	2.3

*These stations correspond to the stations in the far-right region of Fig. 9 (marked with a box)

streamflow dynamics and similar outcomes reported by studies on the spatial streamflow dynamics (e.g. Sivakumar and Woldemeskel 2014; Halverson and Fleming 2015; Fang et al. 2017; Han et al. 2020) also provide opportunities and motivation to apply the complex networks-based methods for spatio-temporal streamflow dynamics. In this regard, efforts to develop a coupled phase space reconstruction–network construction approach for spatio-temporal streamflow dynamics is currently underway. We hope to report the details of such a study in the near future.

It is also appropriate to note, at this point, that the present study considered only a single-variable (i.e., streamflow) time series for studying the temporal dynamics of streamflow, through reconstruction of a multi-dimensional phase space and network construction. While such a single-variable approach is certainly useful, it does not offer a complete representation of the temporal streamflow dynamics, since the actual and dominant variables that govern the streamflow dynamics were not considered. This is an important limitation of this study. This limitation can be easily overcome by considering data of the actual influencing variables (e.g. rainfall and potential evapotranspiration) and employing a multi-variable phase space reconstruction approach for representation of the system dynamics. Methods for multi-variable phase space reconstruction already exist (e.g. Cao et al. 1998) and have also been applied in hydrology within the context of nonlinear dynamic and chaos studies (e.g. Porporato and Ridolfi 2001; Sivakumar et al. 2005). We will attempt the multi-variable phase space reconstruction for temporal network construction in a future study.

Acknowledgements This study was supported by the Australian Research Council (ARC) Future Fellowship Grant (FT110100328). Bellie Sivakumar acknowledges the financial support from ARC through this Future Fellowship Grant. Nazly Yasmin acknowledges the financial support of the Australian Post Graduate Award (University of New South Wales). The authors thank the two reviewers and the Associate Editor for their constructive comments and useful suggestions on an earlier version of the manuscript.

References

Agarwal A, Caesar L, Marwan N, Maheswaran R, Merz B, Kurths J (2019) Network-based identification and characterization of teleconnections on different time scales. Sci Rep 9, Article Number 8808

Agarwal A, Marwan N, Maheswaran R, Ozturk U, Kurths J, Merz B (2020) Optimal design of hydrometric station networks based on complex network analysis. Hydrol Earth Syst Sci 24(5):2235–2251

Alarcòn RR, Lozano S (2019) A complex network analysis of Spanish river basins. J Hydrol 578:124065

Antevs E (1952) Cenozoic climates of the Great basin. Geol Rundsch 40:94–108

Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512

Barrat A, Weigt M (2000) On the properties of small-world networks. Eur Phys J B 13:547–560

Braga AC et al (2016) Characterization of river flow fluctuations via horizontal visibility graphs. Phys A 444:1003–1011
Cao L, Mees A, Judd K (1998) Dynamics from multivariate time series. Physica D 121:75–88
Chen J, Shi H, Sivakumar B, Peart MR (2016) Population, water, food, energy and dams. Renew Sustain Energy Rev 56:18–28
Cook BI, Ault TR, Smerdon JE (2015) Unprecedented 21st century drought risk in the American Southwest and Central Plains. Sci Adv 1(1):e1400082. https://doi.org/10.1126/sciadv.1400005
Encyclopædia Britannica (2019) Great Basin. https://www.britannica.com/place/Great-Basin
Fang K, Sivakumar B, Woldemeskel FM (2017) Complex networks, community structure, and catchment classification in a large-scale river basin. J Hydrol 545:478–493. https://doi.org/10.1016/j.jhydrol.2016.11.056
Gao Z, Jin N (2009) Complex network from time series based on phase space reconstruction. Chaos 19(3):033137. https://doi.org/10.1063/1.3227736
Groisman PY, Knight RW, Karl TR (2001) Heavy precipitation and high streamflow in the contiguous United States: trends in the twentieth century. B Am Meteorol Soc 82(2):219–246. https://doi.org/10.1175/1520-0477
Halverson MJ, Fleming SW (2015) Complex network theory, streamflow, and hydrometric monitoring system design. Hydrol Earth Syst Sci 19(7):3301–3318. https://doi.org/10.5194/hess-19-3301-2015
Han X, Sivakumar B, Woldmeskel FM, Guerra de Aguilar M (2018) Temporal dynamics of streamflow: application of complex networks. Geosci Lett. https://doi.org/10.1186/s40562-018-0109-8
Han X, Ouarda TBMJ, Rahman A, Haddad K, Mehrotra R, Sharma A (2020) A network approach for delineating homogeneous regions in flood frequency analysis. Water Resour Res 56(3):e2019WR025910
Hossain F, Sivakumar B (2006) Spatial pattern of arsenic contamination in shallow wells of Bangladesh: regional geology and nonlinear dynamics. Stoch Environ Res Risk Assess 20(1–2):66–76
Kammerer JC (1990) Largest Rivers in the United States, US Geological Survey Fact Sheet, Open File Report 87-242
Kennel MB, Brown R, Abarbanel HDI (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A 45(6):3403–3411
Kiang JE, Stewart DW, Archfield SA, Osborne EB, Eng K (2013) A national streamflow network gap analysis. US Geological Survey Scientific Investigations Report 2013-5013, Reston, Virginia, USA
Kim HS, Lee KH, Kyoung MS, Sivakumar B, Lee ET (2009) Measuring nonlinear dependence in hydrologic time series. Stoch Environ Res Risk Assess 23:907–916
Konapala G, Mishra AK (2017) Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental USA. J Hydrol 555:600–620
Kyoung MS, Kim HS, Sivakumar B, Singh VP, Ahn KS (2011) Dynamic characteristics of monthly rainfall in the Korean peninsula under climate change. Stoch Environ Res Risk Assess 25(4):613–625
Lins HF (2012) USGS Hydro-climatic data network 2009 (HCDN–2009). US Geological Survey Fact Sheet 2012-3047, US Geological Survey, Reston, VA, USA
Naufan I, Sivakumar B, Woldemeskel FM, Raghavan SV, Vu MT, Liong SY (2018) Spatial connections in regional climate model rainfall outputs at different temporal scales: application of network theory. J Hydrol 556:1232–1243
Newman MEJ (2001) The structure of scientific collaboration networks. Proc Natl Acad Sci USA 98:404–409
Packard NH, Crutchfield JD, Farmer JD, Shaw RS (1980) Geometry from a time series. Phys Rev Lett 45(9):712–716
Poff NL, Bledsoe BP, Cuhaciyan CO (2006) Hydrologic variation with land use across the contiguous United States: geomorphic and ecological consequences for stream ecosystems. Geomorphology 79:264–285
Porporato A, Ridolfi R (2001) Multivariate nonlinear prediction of river flows. J Hydrol 248(1–4):109–122
Pryor SC, Howe JA, Kunkel KE (2009) How spatially coherent and statistically robust are temporal changes in extreme precipitation in the contiguous USA? Int J Climatol 29(1):31–45. https://doi.org/10.1002/joc.1696
Robinson EB, Dietz JL (2019) Great plains. Encyclopædia Britannica. https://www.britannica.com/place/Great-Plains
Serinaldi F, Kilsby CG (2016) Irreversibility and complex network behavior of stream flow fluctuations. Phys A 450:585–600. https://doi.org/10.1016/j.physa.2016.01.043
Sivakumar B (2003) Forecasting monthly streamflow dynamics in the western United States: a nonlinear dynamical approach. Environ. Modell. Softw. 18:721–728
Sivakumar B (2009) Nonlinear dynamics and chaos in hydrologic systems: latest developments and a look forward. Stoch Environ Res Risk Assess 23(7):1027–1036. https://doi.org/10.1007/s00477-008-0265-z
Sivakumar B (2015) Networks: a generic theory for hydrology? Stoch Environ Res Risk Assess 29:761–771
Sivakumar B (2017) Chaos in hydrology: bridging determinism and stochasticity. Springer, Dordrecht
Sivakumar B, Singh VP (2012) Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. Hydrol Earth Syst Sci 16(11):4119–4131. https://doi.org/10.5194/hess-16-4119-2012
Sivakumar B, Woldemeskel FM (2014) Complex networks for streamflow dynamics. Hydrol Earth Syst Sci 18(11):4565–4578. https://doi.org/10.5194/hess-18-4565-2014
Sivakumar B, Woldemeskel FM (2015) A network-based analysis of spatial rainfall connections. Environ Model Softw 69:55–62
Sivakumar B, Berndtsson R, Persson M, Uvo CB (2005) A multivariable time series phase space reconstruction approach to investigation of chaos in hydrological processes. Int J Civ Environ Eng 1(1):35–51
Slack JR, Landwehr VM (1992) Hydro climatic data network (HCDN): a US Geological Survey streamflow data set for United States for the study of climate variations, 1847-1988. US Geological Survey Open File Report, pp 92–129
Takens F (1981) Detecting strange attractors in turbulence. In: Rand DA, Young LS (eds) Dynamical systems and turbulence, vol 898. Lecture notes in mathematics. Springer-Verlag, Berlin, pp 366–381
The World Fact book. Central Intelligence Agency (2019). https://www.cia.gov/library/publications/resources/the-world-factbook/docs/notesanddefs.html#
The World Factbook. Central Intelligence Agency (2019). https://www.cia.gov/library/publications/the-world-factbook/fields/279.html#as
Tiwari S, Jha SK, Sivakumar B (2019) Reconstruction of daily rainfall data using the concepts of networks: accounting for spatial connections in neighborhood selection. J Hydrol 579:124185
Tongal H, Demirel MC, Booij MJ (2013) Seasonality of low flows and dominant processes in the Rhine River. Stoch Environ Res Risk Assess 27:489–503
Vignesh R, Jothiprakash V, Sivakumar B (2015) Streamflow variability and classification using false nearest neighbor

method. J Hydrol 531:706–715. https://doi.org/10.1016/j.jhydrol.2015.10.056

Watts DJ, Strogatz SH (1998) Collective dynamics of small world networks. Nature 393(6684):440–444

Yasmin N, Sivakumar B (2018) Temporal streamflow analysis: coupling nonlinear dynamics with complex networks. J Hydrol 564:59–67

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.