Environmental Modelling & Software 69 (2015) 55–62

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A network-based analysis of spatial rainfall connections

Bellie Sivakumar [a, b, *], Fitsum M. Woldemeskel [a]

[a] School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia
[b] Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 27 November 2014
Received in revised form
24 February 2015
Accepted 28 February 2015
Available online

Keywords:
Rainfall variability
Spatial dynamics
Connections
Complex networks
Clustering coefficient
Degree distribution
Australia

ABSTRACT

This study presents an analysis based on recent developments in network theory to examine the spatial dynamics of rainfall. The concepts of *clustering coefficient* and *degree distribution* are employed to study the spatial connections in a rainfall network across Australia. The clustering coefficient is a measure of local density, while the degree distribution is a measure of spread. Monthly rainfall data over a period of 68 years from a network of 230 gaging stations across Australia are analyzed, and different correlation thresholds are considered. The clustering coefficient results help identify *actual neighbors* and *actual links* in the network as well as stations/regions with high and low connectivities. The results from both methods also suggest that the network is not a purely random graph but may be an exponentially truncated power-law network. The connectivity and type of the raingage network are also influenced by the correlation threshold used for identifying connections.

1. Introduction

Rainfall forms a key input in numerous studies associated with water and environmental systems, including streamflow forecasting, soil moisture estimation, water quality analysis, and sediment transport modeling. Therefore, adequate understanding of its spatial/temporal variability is crucial for reliable investigations and outcomes. However, such is also an extremely challenging problem, as rainfall is highly variable in space/time, due to a combination of factors, including climatic conditions, rainfall generating mechanisms, topographic characteristics, land use, and proximity to sea and other water surfaces. This is particularly the case for large countries, such as Australia, where rainfall is highly variable in space/time, due to the influence of different climates in different regions/periods.

During the last century, numerous approaches have been proposed and applied to study the spatial/temporal variability of rainfall. Such approaches are based on correlation, scale, pattern, similarity, dimensionality, entropy, and many other properties. Extensive details of these can be found in, for example, Zawadzki (1973), Berndtsson (1988), Gupta and Waymire (1990), Krstanovic and Singh (1992), Mishra and Coulibaly (2009), Niu (2013), and Sivakumar et al. (2014). The approaches and the associated methods have certainly helped advance our understanding of the spatial variability of rainfall. Notwithstanding this advancement, our ability to reliably represent the rainfall variability in space remains far from satisfactory. There is, therefore, a need to find new ways to improve studies on spatial rainfall representation. As rainfall in space can be represented in the form of *connections* in a *network* (e.g. connections in rainfall observed across a network of monitoring stations), developments in *network theory* and other concepts in the field of *complex systems science* can offer new avenues. This offers motivation for applying the concepts of *complex networks* to study spatial rainfall dynamics.

The concept of networks is not new. Its origin can be traced back to the eighteenth century (Euler, 1741), with important theoretical developments and applications since then (see Listing, 1848; Cayley, 1857; Erdös and Rényi, 1960; Bollobás, 1998), largely under the umbrella of *graph theory*. However, new discoveries (e.g. small-world networks, scale-free networks, network motifs, community structure) in the past two decades within the context of the *science of complex networks* (e.g. Watts and Strogatz, 1998; Barabási and Albert, 1999; Girvan and Newman, 2002; Milo et al., 2002) have helped put the concept at a whole different level. As a result, theoretical studies and practical applications of the ideas of complex networks are among the most fascinating scientific endeavors

* Corresponding author. School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia.
E-mail addresses: s.bellie@unsw.edu.au, sbellie@ucdavis.edu (B. Sivakumar).

http://dx.doi.org/10.1016/j.envsoft.2015.02.020
1364-8152/

at the current time; see, for example, Barabási (2002) and Estrada (2012). While their applications in hydrology and closely-related fields are still in the state of infancy (e.g. Suweis et al., 2011; Boers et al., 2013; Scarsoglio et al., 2013; Sivakumar and Woldemeskel, 2014), the results are certainly encouraging. Indeed, Sivakumar (2015) argues that the science of networks can offer a generic theory for hydrology, one that based on *connections*.

To our knowledge, there have thus far been only three studies (Malik et al., 2012; Boers et al., 2013; Scarsoglio et al., 2013) that have applied the ideas of complex networks for examining spatial variability of rainfall. Malik et al. (2012) studied the spatial characteristics of extreme (summer) monsoonal rainfall over South Asia, through analysis of daily gridded rainfall data from 1951 to 2007. Boers et al. (2013) investigated the spatial characteristics of extreme rainfall synchronicity of the South American Monsoon System (SAMS), through analysis of a 15-year long (January 1998–December 2012) gridded daily rainfall data with a spatial resolution of $0.25^\circ \times 0.25^\circ$, obtained from the Tropical Rainfall Measuring Mission (TRMM) 3B42 V7 satellite product. Scarsoglio et al. (2013) examined the spatial dynamics of annual precipitation around the globe, through analysis of a 70-year long (January 1941–December 2010) gridded precipitation data from the Global Precipitation Climatology Centre (GPCC) Database. These studies and the reported outcomes are clearly interesting, as they offer important insights regarding the utility of different measures of complex networks for spatial dynamics of rainfall and their effectiveness. Nevertheless, two key points need to be highlighted here: (1) the above studies focused on different scales of data for different purposes – Malik et al. (2012) and Boers et al. (2013) focused on daily data for extreme rainfall analysis, while Scarsoglio et al. (2013) focused on annual data for long-term patterns; and (2) the studies were based on gridded rainfall data, which, although often are at finer resolutions and obtained by merging space-based and ground-based data, generally have a greater degree of uncertainty when compared to raingage data.

As for the scale of data for analysis, it is important to recognize that, despite the possible existence of scaling behavior, the dynamics of rainfall are often significantly different at different temporal scales, due to various factors. For instance, the properties of rainfall at the annual scale (studied by Scarsoglio et al., 2013) are more related to long-term climatic patterns and at the daily scale (studied by Malik et al., 2012 and Boers et al., 2013) are related to within season and event variability, while rainfall properties at the monthly scale are more related to seasonal, annual, and decadal variability. This explains the obvious need for the study of rainfall dynamics at the monthly scale. In addition, from the perspective of medium-term (from few years to few decades) water resources planning and management, including for water supply, reservoir operation, agriculture, and environmental flows, monthly scale is far more appropriate than daily or annual scale. Therefore, in the present study, we apply the ideas of complex networks to examine the spatial dynamic characteristics of monthly rainfall, and we also consider data measured using raingages. We analyze monthly rainfall data recorded over a period of 68 years (1940–2007) at a network of 230 raingage stations across Australia. We use the *clustering coefficient* (which quantifies the tendency of a network to cluster – a measure of local density) and *degree distribution* (which expresses the fraction of nodes in a network with a certain number of links – a measure of spread) as measures to examine the spatial connections in rainfall. We also study the influence of different rainfall correlation thresholds.

The rest of this paper is organized as follows. Section 2 presents a brief description of the procedure for calculation of clustering coefficient and degree distribution in networks. Section 3 offers details of the study area and rainfall data. Section 4 presents the analysis, results, and their discussion. Some closing remarks are made in Section 5.

2. Network methodology

A *network* is a set of points connected together by a set of lines. The points are called as *nodes* or *vertices* and the lines are called as *links* or *edges*. Mathematically, a network can be represented as $G = \{P,E\}$, where P is a set of N nodes ($P_1, P_2, \ldots, P_N$) and E is a set of n links. There are many different ways to study the characteristics of networks. For instance, networks can be studied by their clustering, topology, adjacency, centrality, and entropy properties. Similarly, there are also different measures and methods to represent these properties. These include clustering coefficient, degree distribution, average shortest path length, and degree centrality, among others. This study uses clustering coefficient and degree distribution to examine the spatial dynamics of rainfall. A brief description of these is below.

2.1. 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 procedure for the calculation of the clustering coefficient is as follows. Let us assume a node i in a network and that it has k_i links which connect it to k_i other nodes, as shown in Fig. 1. These k_i nodes are the *neighbors* of node i, and can be identified based on some criterion, such as correlation between node i and the other nodes in the network. 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 (Fig. 1, right). 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 (solid lines on Fig. 1, right) and the total number $k_i(k_i - 1)/2$ (i.e. all lines on Fig. 1, right),

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \tag{1}$$

The procedure is repeated for each and every node of the network. The average of the clustering coefficients C_i's of all the individual nodes is the clustering coefficient of the whole network C.

The clustering coefficient of the individual nodes and of the entire network can be used to obtain information about the type of network, grouping of nodes, and identification of dominant nodes, among others. For instance, a clustering coefficient of 1.0 indicates a completely ordered network, while a very low clustering coefficient indicates a random network.

2.2. Degree distribution

In a network, different nodes may have different number of links. The number of links (k) of a node is called as *node degree*. The degree is an important characteristic of a node, as it allows one to derive many measurements for the network.

The spread in the node degrees is characterized by a distribution function, $p(k)$, which expresses the fraction of nodes in a network with degree k. This distribution, called *degree distribution*, is often a reliable indicator of the type of network. For instance, in a purely random graph, since the links are placed randomly, the majority of nodes have approximately the same degree, and close to the average degree $\overline{k}$ of the network. Therefore, the degree distribution of a completely random graph is a Poisson distribution with a peak at $P(\overline{k})$, and is given by:

$$p(k) = \frac{e^{-\overline{k}}\overline{k}^k}{k!} \tag{2}$$

Similarly, depending upon the properties of networks, degree distribution can also be, for example, Gaussian

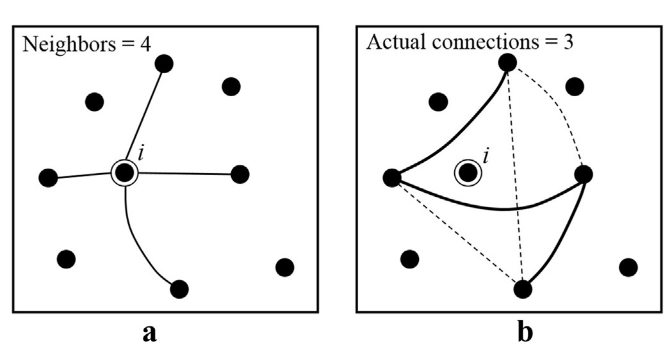

Fig. 1. Network connections and calculation of clustering coefficient: (a) neighbors of node i; and (b) all links and actual links (solid lines).