Journal of Hydrology 527 (2015) 62–76

Contents lists available at ScienceDirect

Journal of Hydrology

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

Comparing three methods to form regions for design rainfall statistics: Two case studies in Australia

Khaled Haddad [a,*], Fiona Johnson [b], Ataur Rahman [a], Janice Green [c], George Kuczera [d]

[a] School of School of Computing, Engineering and Mathematics, University of Western Sydney, Australia
[b] School of Civil and Environmental Engineering, University of New South Wales, Australia
[c] Environment and Research Division, Bureau of Meteorology, Australia
[d] School of Engineering, University of Newcastle, Australia

ARTICLE INFO

Article history:
Received 28 May 2014
Received in revised form 5 February 2015
Accepted 22 April 2015
Available online 29 April 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Luis E. Samaniego, Associate Editor

Keywords:
Regionalization
Design rainfall
GLS Regression
Bayesian approach

SUMMARY

One of the fundamental steps in regional rainfall frequency analysis is deciding the method by which rainfall stations are to be grouped together to form regions. This paper compares three methods of forming regions for use in estimating design rainfalls: a fixed region approach where all the available sites are included in a single region, a Region of Influence (ROI) approach based on geographical proximity and a hybrid approach where sites with similar topographic orientations are grouped together.

The three region types were implemented in a Bayesian Generalized Least Squares Regression (BGLSR) framework which leads to regionalized regression equations that can be used to predict rainfall L-moments at ungauged sites. A leave-one-out cross validation approach was used to compare the relative accuracy, reliability and uncertainty of the derived rainfall statistics and resulting estimates of the rainfall quantiles. The study used data from two areas of Australia chosen for their highly varied topography and different climatic influences.

It was found that all three methods provided good estimates of the L-moment statistics and the rainfall quantiles. The hybrid approach produced the smallest errors in the South-East Queensland region whilst for the Tasmanian region the fixed region approach was best. The results from this study show that although there is a slight benefit in using the proposed hybrid approach for BGLSR, these benefits were minor compared to maximizing the number of stations used to calibrate the BGLSR equations. This conclusion regarding the number of stations could be tested in future work by repeating the analyses in areas with sparser station density. Another test could be to simulate reduced station coverage in the current study areas by leaving stations out of the analyses. Finally it would be interesting to see if similar results are obtained by expanding the study area so that different climatological regimes are assessed.

1. Introduction

The design and analysis of hydraulic infrastructure and flooding is often based on design rainfall because of a lack of relevant streamflow data at the locations of interest. The design rainfall data are generally presented in the form of Intensity–Duration–Frequency (IDF) curves (commonly known as IFD curves in Australia), which are derived from observations at rain gauges. Due to the relatively short record lengths of data from rain gauges compared to the frequency of the events of interest for engineering design, regional rainfall frequency analysis is generally used to reduce the uncertainty in the estimates of the rainfall quantiles (Buishand, 1991; Smithers and Schulze, 2001; Madsen et al., 2002, 2009; Brath et al., 2003; Di Baldassarre et al., 2006a,b; Castellarin et al., 2009). This is achieved by assuming that rainfall statistics from multiple stations can be combined to offset the lack of temporal information (Zrinji and Burn, 1994). The assumption that spatial data can replace long time series is based on the concept of forming homogenous regions, such that all the sites in a region share a common probability distribution (possibly with a site-specific scaling factor). Once a regional relationship is formed this can also be used to estimate statistics at ungauged sites (Madsen et al., 2002). In this paper, the regionalization is implemented through the use of Bayesian Generalized Least Squares Regression (BGLSR) and one of the key decisions in the BGLSR is the method by which rainfall stations are grouped into regions.

There are two main methods for grouping stations proposed in the literature, each of which has a number of variations. The first

* Corresponding author at: University of Western Sydney, Penrith South, NSW 1797, Australia. Tel.: +61 2 4736 0920.
E-mail address: k.haddad@uws.edu.au (K. Haddad).

http://dx.doi.org/10.1016/j.jhydrol.2015.04.043
0022-1694/

method identifies regions with fixed boundaries which are used to define the stations that will be included in the region. Traditionally these boundaries may have been based on existing divisions such as those from political or administrative borders or from geographic or physiographic boundaries (Burn and Goel, 2000). Over time the fixed region approach has evolved to use more rigorous statistical techniques such as cluster analysis to provide objective and homogenous regions (Hosking and Wallis, 1997). The boundaries of adjacent fixed regions may overlap from region to region but the important point is that each region's membership is clearly defined. The upper limit to the fixed region approach is a region that shares its boundaries with the analysis domain such that all stations are included in a single region for the analysis (Hosking and Wallis, 1997; Madsen et al., 2002; Di Baldassarre et al., 2006a). It is important to note that stations may be grouped into regions based on geographic proximity or by clustering stations that are similar in terms of other site properties, such as climatology, elevation or orientation.

The second method of grouping regions is termed the Region of Influence (ROI) approach (Burn, 1990a; Castellarin et al., 2001; Merz and Blöschl, 2005; Gaál et al., 2008; Haddad and Rahman, 2012). In this method, each station is placed at the center of a region which includes a predefined number of neighbors. Thus for each station, the ROI will potentially be different. The neighbors for the ROIs can be defined in terms of geographic proximity or according to a distance metric in the space of site properties of interest.

There have been several studies comparing different methods of forming regions for both regional rainfall analysis and also for streamflow analyses. The advantages of the ROI approach is that arbitrary regional boundaries do not need to be specified and resulting inconsistencies can therefore be avoided (Burn, 1990a). However Hosking and Wallis (1997) recommend the fixed region cluster approach as the "most practical method of forming regions from large data sets". In this study we compare both approaches, along with a newly developed hybrid approach that combines information on the orographic influences on rainfall statistics as described next.

Topography is one of the well-known drivers of spatial variation in rainfall patterns. Sharples et al. (2005) highlighted the strong relationship between rainfall and elevation. This relationship may also be affected by the slope orientation, predominant wind direction and the steepness of the terrain as noted by Daly et al. (1994). It is therefore worth investigating whether grouping stations according to topography and slope orientation when regionalizing rainfall data could lead to improvements in the results. Grouping stations according to slope orientation has been previously implemented by Daly et al. (1994) in the interpolation of rainfall data in the Parameter-elevation Regressions on Independent Slopes Model (PRISM) (PRISM Climate Group, 1998). The areas of similar slope orientation were termed "facets". Daly et al. (1994) demonstrated that facet-based interpolation had the lowest cross-validation uncertainty, bias and absolute error in rainfall quantile estimation when compared to kriging, detrended kriging and cokriging. Although such topographic/orographic information has been commonly used in spatial interpolation of daily and monthly rainfall data and climatology (Daly et al., 1994, 2002; Hutchinson, 1995; Bonnin, 2002; Sharples et al., 2005; Buytaert et al., 2006; Yang et al., 2010; Portalés et al., 2010; Hession and Moore, 2011) until now there have been few studies that have looked at using this information as a regionalization technique for large rainfalls. One exception is the recent study by Yang et al. (2010) who presented a rainfall regionalization approach which accounted for orographic effects on rainfall quantile estimation in the Pearl River Basin in China. They found that the Asian summer monsoon and typhoons are major meteorological driving forces on the precipitation regimes. Additionally, topographic features (e.g. elevation, distance to the sea, and the orientation of mountain ranges) exert controls on the spatial patterns of such regimes. Through the use of spatial regression, Hession and Moore (2011) showed that topographic variables such as elevation and slope strongly influence rainfalls in East Africa. The above studies suggest that it is important to capture the local and regional effects of topography in regionalization. We introduce a method that combines the topographic information contained in the facet analysis with a ROI approach to create a new facet-hybrid regionalization strategy, the details of which are provided in the following sections.

Regional frequency analysis is being used in the revision of the design rainfalls for Australia by the Australian Bureau of Meteorology (the Bureau) (Green et al., 2012). When compared to some other parts of the world, the topography of Australia can be considered fairly flat. Nonetheless, the effects of orography on rainfall are evident in many parts of the country, in particular along the east coast where there is a steep gradient from the coastal fringe to the adjacent hinterland. Previous studies (e.g. Sharples et al., 2005) on the orographic effects of these landforms have focused on the analysis of annual and seasonal rainfall statistics. An interesting extension is whether regionalization of the rainfall statistics required for the design rainfall data can be improved through the use of orographic information. This study therefore aims to assess if a regionalization approach based on grouping stations according to slope orientation provides superior results to fixed region and ROI analyses when using BGLSR. The statistics of interest that need to be regionalized are the sample L-moments for sub-daily rainfall durations from 60 min to 720 min.

In the remainder of the paper we describe the regionalization strategy and the details of the BGLSR modeling, the data used for the study and the statistics of interest. The results of the analysis are then presented along with discussion and conclusions.

2. Regionalization of rainfall data

This section discusses the regionalization that is proposed for the revision of design rainfalls for Australia in the context of the formation of regions for the BGLSR and how the regionalized estimates are to be used.

2.1. Formation of regions

The formation of regions in general and for BGSLR in this paper can be based on geographic and administrative boundaries or based on grouping stations that are similar in the space of physiographic and/or climatic characteristics using fixed regions defined through clustering or other objective partitioning methods (Hosking and Wallis, 1997). Alternatively the ROI approach proposes that each site is located at the center of its own region with the neighbors chosen using a similarity measure (e.g. geographic distance or the similarity of other site characteristics) (Burn, 1990a,b; Zrinji and Burn, 1994).

We extend the fixed region and ROI approaches by developing a new method that groups stations according to the orientation of the landscape (i.e. aspect). Whilst aspect could be used as a physiographic characteristic in a traditional ROI or fixed region approach, the novelty of our method is that it ensures that the stations are located within a geographically contiguous area. The facets (Daly et al., 1994) are expected to strengthen the relationships between stations within a region because all stations will be subject to the same meteorological conditions. If aspect was used in a traditional ROI then two stations could both be located on north facing land and yet be located in different areas subject