Journal of Hydrology 492 (2013) 151–162

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation

Fredline Ilorme [a], Veronica Webster Griffis [b,*]

[a] Dept. of Civil & Environ. Engr., Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea
[b] Dept. of Civil & Environ. Engr., Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931-1295, United States

ARTICLE INFO

Article history:
Received 19 November 2012
Received in revised form 22 February 2013
Accepted 23 March 2013
Available online 10 April 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Attilio Castellarin, Associate Editor

Keywords:
Regional L-moments
Flood frequency
Hydrologic homogeneity

SUMMARY

Regional flood frequency techniques are widely used to estimate flood quantiles when flow data is unavailable for the basin under study or the record length is insufficient for reliable analyses. Data from nearby gauged sites are pooled to compensate for the lack of at-site data. This requires the delineation of hydrologically homogeneous regions in which the flood regime is sufficiently similar to allow the spatial transfer of information. It is generally accepted that hydrologic similarity results from similarity among basins' physiographic characteristics, and thus these characteristics can be used to delineate regions and classify ungauged sites. However, as currently practiced, the delineation is highly subjective and dependent on the similarity measures and classification techniques employed. Herein, a novel procedure for region delineation is proposed and evaluated using data for sites across the Southeastern United States. Key components of this procedure are a new statistical metric to identify physically discordant sites and a new methodology to identify the physical attributes that are the most indicative of extreme hydrologic response. The novel approach for region delineation is shown to produce regions which are more homogeneous and more efficient for quantile estimation at ungauged sites than those delineated using alternative physically-based procedures typically employed in practice. In addition, the identified physical attributes can be used to infer the flood regime and estimate quantiles at sites outside the extent of the area used for model development.

1. Introduction

Standard procedures for at-site flood frequency analysis involve assembling the annual maximum flood record at the site of interest and fitting an analytic probability distribution to the data (e.g., IACWD, 1982). The fitted distribution is then used to estimate flood quantiles associated with a given return period, such as the flood magnitude expected to be equaled or exceeded once every 100 years (i.e., the 100-year event). However, in most cases the at-site record length is too short to accurately estimate flood quantiles for return periods of interest: estimation of the 100-year event often requires extrapolating beyond the observed flood record. In other cases, flood data are unavailable at the site of interest, making at-site flood frequency analysis impossible. One solution is to "trade space for time" (Stedinger et al., 1993) using a regional flood frequency analysis, such as the Index Flood or Regional L-moments method (e.g. Dalrymple, 1960; De Michele and Rosso, 2001; Fill and Stedinger, 1998; Hosking and Wallis, 1988, 1997; Kjeldsen and Rosbjerg, 2002; Stedinger and Lu, 1995), wherein the characterization of flood flows at the site of interest is derived using information pooled from nearby hydrologically similar gauged sites (NRC, 1988).

An important component of the Regional L-moments method is the delineation of hydrologically homogeneous regions (or groups of sites) in which the flood regime is deemed sufficiently similar to allow the spatial transfer of information from gauged sites to ungauged sites (see for example, Stedinger and Lu, 1995, and citations therein). Tests for hydrological homogeneity are based on statistical measures of the similarity among at-site flood statistics which reflect the shape of the flood distribution, such as the product-moment coefficient of variation, CV, (e.g., Dalrymple, 1960) or the L-moment analogue, L-CV (e.g., Hosking and Wallis, 1997). Therefore, hydrological homogeneity really represents statistical homogeneity of the flood distributions (or shape thereof) throughout the region. Sites within a hydrologically homogeneous region can therefore be assumed to share the same parent (or non-dimensional regional) flood distribution with a common shape parameter, but each watershed has a site-specific scale factor (the location, a.k.a. "index flood," parameter) to represent possible changes in flood magnitude across the region. For application at

* Corresponding author. Tel.: +1 906 487 1079; fax: +1 906 487 2943.
E-mail addresses: filorme@mtu.edu (F. Ilorme), vgriffis@mtu.edu (V.W. Griffis).

0022-1694/$ - see front matter
http://dx.doi.org/10.1016/j.jhydrol.2013.03.045

gauged sites, the mean of the at-site flood series is used as the scale factor; however, for ungauged sites, this factor must be related to physiographic characteristics of the watershed, such as drainage area.

Multivariate statistical techniques such as cluster analysis, principal component analysis, canonical correlation analysis, and multiple discriminate analysis are commonly employed to delineate homogeneous regions, i.e. group sites with similar hydrologic response, and subsequently classify ungauged locations (e.g., Burn, 1997; Burn et al., 1997; Chiang et al., 2002a,b; Rao and Srinivas, 2006; Srinivas et al., 2008; Zrinji and Burn, 1994). Application of these methods requires the selection of appropriate similarity measures to characterize the hydrologic response, or flood regime, at individual sites. Possible similarity measures include at-site flood statistics, such as the flood magnitude corresponding to a specified return period (T years), the timing and/or duration of flood events, the product-moment CV, or L-moment ratios (e.g., L-CV or L-Skewness). However, use of flood statistics to delineate regions may compromise the validity of the hydrological homogeneity test as the latter is based on the same data (Burn, 1990, 1997). As a result, the delineated regions may appear homogeneous, but would be inefficient for developing regional quantile estimators. In addition, unless the delineated regions are sufficiently contiguous, ungauged basins cannot be classified within a delineated region.

It is generally assumed that statistical homogeneity arises in response to physical homogeneity, as similar hydrologic response should be observed in watersheds with comparable basin characteristics (e.g. physiographic characteristics and meteorologic inputs). (See for example, Chiang et al., 2002a,b; GREHYS, 1996a; Hosking and Wallis, 1997; Isik and Singh, 2008; Rao and Srinivas, 2006.) Thus, hydrologically homogeneous regions can be delineated using clustering procedures on physical attributes instead of the flood statistics on which subsequent tests for homogeneity are based, and physical characteristics can be used to classify an ungauged site within a delineated region. Some studies have observed, however, that similar hydrologic response is not guaranteed by similarity among basin characteristics because of complex interactions among those characteristics (Burn, 1997; Burn et al., 1997; Zrinji and Burn, 1994). In addition, delineated regions are highly dependent on the choice of similarity measures used to infer homogeneity (e.g., Burn, 1990, 1997; Burn et al., 1997; Castellarin et al., 2001; GREHYS, 1996b; Hosking and Wallis, 1997) and once regions are obtained, the identification of and decision to remove discordant sites is highly subjective. Typically, efforts to improve regional homogeneity involve moving hydrologically discordant sites from one region to another without any clearly defined physical basis (Rao and Srinivas, 2006; Srinivas et al., 2008). These are major limitations of the Regional L-moments method, and significantly impair its ability to provide accurate quantile estimates at ungauged sites.

The goals of this paper are to evaluate to what extent hydrological homogeneity is explained by physical homogeneity, and to decrease the subjectivity associated with the delineation of hydrologically homogeneous regions for regional flood frequency analyses. Herein, a novel procedure for region delineation is proposed and evaluated using data for sites across the Southeastern United States. Key components of this procedure are a new statistical metric to identify physically discordant sites, and a methodology to identify the physical variables that are the most indicative of extreme hydrologic response. In addition, results presented herein are used to validate the hypothesis that the identified physical attributes can be used to infer the flood regime and successfully estimate quantiles in basins with comparable physical characteristics, but which lie outside the extent of the area used for model development.

2. Methods and assumptions

The Regional L-moments method involves three main steps: (1) quantile estimation at gauged sites, (2) delineation of hydrologically homogeneous regions, and (3) development of the regional flood distribution for extrapolation of quantile estimates to ungauged sites. Herein, the Regional L-moments method described by Hosking and Wallis (1997) is used for development of the regional flood distribution. While the procedure of Hosking and Wallis includes use of L-moment ratio diagrams to select the appropriate parent distribution for annual maximum flood series in a given area, only the generalized extreme value (GEV) distribution will be employed herein. Vogel and Wilson (1996) use L-moment ratio diagrams to demonstrate that the GEV, log-Pearson type III (LP3), and lognormal (LN) distributions are all reasonable models of annual maximum flood series in the United States. Further, when compared to the uncertainty in flood quantile estimates, the differences between reasonable choices of distributions are negligible (Hosking and Wallis, 1997; Stedinger, 1980; Stedinger and Griffis, 2008), and the GEV distribution is most frequently used in the context of regional index flood modeling (e.g., Chowdhury et al., 1991; Hosking and Wallis, 1988, 1997; Lettenmaier et al., 1987; Madsen and Rosbjerg, 1997; Stedinger and Lu, 1995; Wallis and Wood, 1985). Hosking and Wallis (1997) and Stedinger et al. (1993) provide detailed discussions and relevant equations for defining the at-site and regional GEV distributions by the method of L-moments. Readers are referred to these sources and the references therein for more detail regarding these aspects of the Regional L-moments method. As the focus of this paper is on reducing the subjectivity associated with the delineation of hydrologically homogeneous regions, only those aspects of the procedure as typically applied in practice will be expounded upon below.

2.1. Delineation of homogeneous regions

A variety of multivariate statistical techniques have been employed to delineate homogeneous regions, i.e. group sites with similar hydrologic response, and subsequently classify ungauged locations (e.g., Burn, 1997; Burn et al., 1997; Chiang et al., 2002a,b; Rao and Srinivas, 2006; Srinivas et al., 2008; Zrinji and Burn, 1994). As discussed in Section 1, similarity measures based on physical or meteorological characteristics should be used in place of similarity measures based on gauged flood data so that subsequent tests for hydrological homogeneity are not compromised, and ungauged sites can be classified within a delineated region. Still, a major complication remains in that the regions formed are dependent on the choice of basin characteristics to employ as indicators of hydrologic similarity, as well as the statistical technique(s) selected for region delineation (Castellarin et al., 2001; GREHYS, 1996b). This concern will be addressed by the novel method proposed in Section 3; the sub-sections below describe the multivariate statistical techniques employed therein, as well as in practice today in the context of Regional L-moments.

2.1.1. Cluster analysis

Cluster analysis (CA) groups sites on the basis of a statistical distance measure reflecting the similarity (or dissimilarity) among the set of attributes (similarity measures) selected to represent each gauging station. Several clustering techniques are available in the statistical literature (Johnson and Wichern, 2007, p. 671), and have all been used in the delineation of hydrologically homogeneous regions (e.g., Baeriswyl and Rebetez, 1997; Bhaskar and O'Connor, 1989; Burn, 1989, 1997; Castellarin et al., 2001; Chiang et al., 2002a; Dinpashoh et al., 2004; Hosking and Wallis, 1997; Rao and Srinivas, 2006). The most commonly used technique is Ward's method as it tends to delineate regions roughly equivalent in size,