Journal of Hydrology 508 (2014) 385–396

Contents lists available at ScienceDirect

Journal of Hydrology

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

Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations

C.C. Nguyen [a,b], E. Gaume [a,*], O. Payrastre [a]

[a] LUNAM Université, IFSTTAR, GERS, F-44344 Bouguenais, France
[b] Danang university, Vietnam

ARTICLE INFO

Article history:
Received 22 March 2013
Received in revised form 4 August 2013
Accepted 23 September 2013
Available online 13 November 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Purna Chandra Nayak, Associate Editor

Keywords:
Floods
Regional analysis
Statistics
Extremes
GEV

SUMMARY

Flood frequency analyses are often based on recorded series at gauging stations. However, the length of the available data sets is usually too short to provide reliable estimates of extreme design floods. Hence, hydrologists have tried to make use of alternative sources of information to enrich the datasets used for the statistical inferences. Two main approaches were therefore proposed. The first consists in extending the information in time, making use of historical and paleoflood data. The second, spatial extension, consists in merging statistically homogeneous data to build large regional data samples. Recently, a combination of the two techniques aiming at including estimated extreme discharges at ungauged sites of a region in the regional flood frequency analyses has been proposed. This paper presents a consolidation of this approach and its comparison with the standard regional flood frequency approach proposed by Hosking & Wallis. A modification of the likelihood function is introduced to enable the simultaneous calibration of a regional index flood relation and of the parameters of the regional growth curve. Moreover, the efficiency of the proposed method is evaluated based on a large number of Monte Carlo simulated data sets. This work confirms that extreme peak discharges estimated at ungauged sites may be of great value for the evaluation of large return period (typically over 100 years) flood quantiles. They should therefore not be neglected despite the uncertainties associated to these estimates.

1. Introduction

Although a large number of statistical inference methods have been progressively developed, the question of estimating extreme design floods is still problematic due to the generally limited amount of available data. Continuous discharge series at gauged sites are generally too short to provide reliable estimates of extreme quantiles - typically the 100-year or higher return period quantiles (NERC, 2000). To cope with this difficulty, hydrologists have tried to complement the available data sets, either through a "temporal extension", incorporating data on historical and paleofloods (Hosking and Wallis, 1986a; Hosking and Wallis, 1986b; Stedinger and Cohn, 1986; Cohn and Stedinger, 1987; Gary and Stedinger, 1987; Sutcliffe, 1987; Minghui and Stedinger, 1989; Sheffer et al., 2003; Reis et al., 2005; Neppel et al., 2010; Payrastre et al., 2011), or through a "spatial extension", merging data sets in regions considered as statistically homogeneous, "trading space for time" according to the words of Hosking & Wallis (Hosking and Wallis, 1997; Heinz and Stedinger, 1998; Charles and Stedinger, 1999; Ouarda et al., 2001; Kjeldsen et al., 2002; Merz and Blöschl, 2003; Seidou et al., 2006; Ribatet et al., 2007; Norbiato et al., 2007; Wallis et al., 2007; Kjeldsen and Jones, 2009; Haddad and Rahman, 2011).

Recently, Gaume et al. (2010) observed that estimated extreme peak discharges at ungauged sites are often available, but never really used in flood frequency studies and proposed a method to incorporate such information in regional flood frequency analyses.

The proposed approach is based on the index flood principle (Dalrymple, 1960), assuming that, within a statistically homogeneous region, all local statistical distributions are identical apart from a site-specific scaling factor: the index flood. Usually, the index flood corresponds to the mean of the local series (Hosking and Wallis, 1997). The approach proposed by Gaume et al. (2010) is based on the calibration of an index flood relation linking the characteristics of the watersheds and the index flood value. Although this relationship represents an additional homogeneity requirement that may limit the extent of the region used for the statistical analysis, it also enables to estimate the index flood at ungauged sites, and thus to incorporate the corresponding ungauged extremes in the regional sample.

Based on several case studies, Gaume et al. (2010) showed the possible great value of such an approach, depending on the characteristics of available extreme flood inventories. The index flood relationship proposed, of the form S^{β} (S being the area of the watershed and β a parameter to be calibrated), appeared satisfactory

* Corresponding author. Tel.:+33(0)2 40 84 58 84.
E-mail address: eric.gaume@ifsttar.fr (E. Gaume).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.09.058

in the test regions. The presented inference results were based on a Bayesian MCMC framework (Castellarin, 2005; Reis et al., 2005; Seidou et al., 2006; Ribatet et al., 2007; Castellarin et al., 2007; Payrastre et al., 2005; Payrastre et al., 2011) to adjust the regional growth curve with associated 90% credibility intervals. The results showed that the incorporation of ungauged extremes could lead to a significant reduction of the width of the computed credibility intervals.

In the initial version of the method (Gaume et al., 2010), the index flood relation was adjusted prior to the calibration of the regional growth curve, and the uncertainties associated with its calibration were not taken into account. This led certainly to underestimate the credibility intervals and over-rate the added value of the ungauged extremes and of the proposed method. The effects of possible variations (heterogeneities) in the average relation calibrated in a given region should also be considered for a fair comparison with other statistical methods. This paper proposes an extension of the initial method to account for uncertainties in the calibrated index flood relation. It also tests the effect of possible regional variations in the average relation on the efficiency of the proposed statistical inference approach.

The performances (i.e. widths and correctness of computed credibility intervals) of the proposed approach and of the standard regional frequency approach proposed by Hosking and Wallis (1997) are first compared in the case where gauged data only are considered. The comparison is based on samples generated through Monte Carlo simulations in order to be able to verify the accuracy of the calculated credibility intervals and to introduce controlled heterogeneities in the samples. In a second step, both approaches are applied to the statistical analysis of a data set from the Ardèche region in France composed of 168 records at 5 gauging stations and 18 estimated ungauged extremes.

The paper is organised as follows: Section 2 presents the basics and adaptations of the two regional flood frequency methods: Hosking & Wallis and the proposed approach. The performances of the two approaches are compared based on simulated samples of random variables in Section 3. In Section 4, the methods are applied to the real-world case study. Conclusions are drawn in Section 5.

2. Tested regional flood frequency analysis methods

2.1. The index flood hypothesis

The two approaches considered in this paper are based on the same fundamental simple scaling hypothesis or index-flood principle (Dalrymple, 1960): in a statistically homogeneous region, all the local annual maximum peak discharge distributions are supposed to be identical apart from a site-specific scaling factor. This hypothesis is summarized in Eq. 1:

$$Q_i(F) = \mu_i q(F) \tag{1}$$

Where F is the probability of non-exceedance, i is the index of the site $(i = 1, \ldots, s)$, s the total number of sites in the homogeneous region, $Q_i(F)$ is the discharge quantile, $q(F)$ is the regional dimensionless (i.e. reduced) quantile and μ_i is the index flood (or scaling factor).

The index flood may be any constant value proportional to the expectancy of the local distribution. Usually, when only data from gauged sites are considered, the index flood is estimated by the at-site sample mean (Hosking and Wallis, 1997; Castellarin, 2005; Castellarin et al., 2007). A regional flood frequency method where the index flood is computed as the average of the local series of annual maxima will be called hereafter method of Hosking & Wallis. This, even if a likelihood based Bayesian MCMC procedure rather than a L-moment based procedure, as suggested by Hosking and Wallis (1997), is used to calibrate the parameters of the regional growth curve. Gaume et al. (2010) suggested an alternative approach to account for extreme discharge estimates that may be available at ungauged sites. An inventory of ungauged extremes may include h extreme peak discharges Q_k $(k = 1, .., h)$, each Q_k corresponding to the largest flood at site k during a period of length n_k. In order to include this additional information in the regional dataset, Gaume et al. (2010) proposed to use an index flood relation linking the index flood value to the catchment area S, since an average annual peak discharge can obviously not be computed at ungauged sites:

$$\mu_i = S_i^{\beta} \quad and \quad \mu_k = S_k^{\beta} \tag{2}$$

Where S_i and S_k are the catchment areas at the corresponding sites, and β a coefficient to be calibrated. More complex relations based on various climatic and physio-geographic characteristics may be tested in the future, but at the price of an increased number of parameters to be calibrated. In the initial version of the method, the value of β was adjusted through a regression between the log transform of the average annual peak discharges and the watershed areas at gauged sites.

It is proposed here to calibrate β along with the parameters of the regional growth curves using a modified likelihood formulation as described below.

2.2. Likelihood of the observed sample

The inference approach applied herein is directly derived from Gaume et al. (2010) and inspired by numerous previous works (Reis et al., 2005; Renard et al., 2006; Payrastre et al., 2011): i.e. based on the likelihood of the available data sets and a Bayesian MCMC algorithm for the estimation of the parameters of the growth curve and of their posterior distribution according to the observed data set.

Considering the regional sample **D** described above, including both (i) s series of gauged annual maximum discharges $Q_{i,j}, j = 1, \cdots, n_i$ being the index of the year, and (ii) the h estimated largest peak discharges Q_k over n_k years at h ungauged sites, the standard expression of the likelihood of the regional sample **D** would be the following:

$$\ell(\mathbf{D} \mid \theta) = \prod_{i=1}^{s}\left[\prod_{j=1}^{n_i} f_\theta\left(\frac{Q_{i,j}}{\mu_i}\right)\right]\prod_{k=1}^{h}\left[f_\theta\left(\frac{Q_k}{\mu_k}\right)\right]\prod_{k=1}^{h}\left[F_\theta\left(\frac{Q_k}{\mu_k}\right)\right]^{(n_k-1)} \tag{3}$$

Where f_θ and F_θ are respectively the probability density function and the cumulative probability function of the selected statistical distribution for the regional growth curve, and θ corresponds to the vector of parameters to be estimated. The GEV distribution, often used to describe peak discharge growth curves (Lu and Stedinger, 1992; Stedinger and Lu, 1995; Coles and Powell, 1996; Coles and Tawn, 1996; Heinz and Stedinger, 1998; Seidou et al., 2006), was selected here (Eq. 4 and 5). The vector θ comprises the position, scale and shape parameters (ξ, α, κ) of the GEV distribution.

$$F_\theta(Q) = \exp\left[-\left(1 - \frac{\kappa(Q-\xi)}{\alpha}\right)^{1/\kappa}\right]_{\alpha>0} \tag{4}$$

$$f_\theta(Q) = \frac{1}{\alpha}\left(1 - \frac{\kappa(Q-\xi)}{\alpha}\right)^{1/\kappa-1}\exp\left[-\left(1 - \frac{\kappa(Q-\xi)}{\alpha}\right)^{1/\kappa}\right]_{\alpha>0} \tag{5}$$

In Eq. 3, the first term corresponds to the probability of the gauged series. It is the only necessary term if continuous series of measured annual maximum discharges are used. The second term is the probability of the ungauged extremes. The third complementary term is the probability associated to the fact that the ungauged extreme value has not been exceeded during the remaining $(n_k - 1)$ years at