Journal of Hydrology 541 (2016) 90–98

Contents lists available at ScienceDirect

Journal of Hydrology

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

Reducing uncertainty in flood frequency analyses: A comparison of local and regional approaches involving information on extreme historical floods

K. Halbert, C.C. Nguyen [1], O. Payrastre *, E. Gaume

LUNAM Université, Ifsttar, GERS, CS4, 44344 Bouguenais Cedex, France

ARTICLE INFO

Article history:
Available online 14 January 2016

Keywords:
Flash floods
Frequency analysis
Ungauged floods
Historical floods
Regional analysis
Bayesian analysis

SUMMARY

This paper proposes a detailed comparison of local and regional approaches for flood frequency analyses, with a special emphasis on the effects of (a) the information on extreme floods used in the analysis (historical data or recent extreme floods observed at ungauged sites), and (b) the assumptions associated with regional approaches (statistical homogeneity of considered series, independence of observations). The results presented are based on two case studies: the Ardèche and Argens rivers regions in south-east of France. Four approaches are compared: 1 – local analysis based on continuous measured series, 2 – local analysis with historical information, 3 – regional index-flood analysis based on continuous series, 4 – regional analysis involving information on extremes (including both historical floods and recent floods observed at ungauged sites). The inference approach used is based on a GEV distribution and a Bayesian Monte Carlo Markov Chain approach for parameters estimation. The comparison relies both on (1) available observed datasets and (2) Monte Carlo simulations in order to evaluate the effects of sampling variability and to analyze the possible influence of regional heterogeneities. The results indicate that a relatively limited level of regional heterogeneity, which may not be detected through homogeneity tests, may significantly affect the performances of regional approaches. These results also illustrate the added value of information on extreme floods, historical floods or recent floods observed at ungauged sites, in both local and regional approaches. As far as possible, gathering such information and incorporating it into flood frequency studies should be promoted. Finally, the presented Monte Carlo simulations appear as an interesting analysis tool for adapting the estimation strategy to the available data for each specific case study.

1. Introduction

A large number of applications in hydrology require flood frequency analyses, either for flood risk assessment or for design of water related structures. Considering that the local measured discharge series are rarely long enough to provide reliable flood quantiles estimates of high return periods (typically 50 to 1000 years), large research efforts have been devoted to the proposal of methods for reducing uncertainties in flood frequency analyses. Some proposed approaches aim at deriving the shape of peak discharge distributions based on rainfall statistical properties and assumptions on watershed physical behaviors (Guillot and Duband, 1967; Eagleson, 1972; Arnaud and Lavabre, 2002; Gaume, 2006; Paquet et al., 2013; Rogger et al., 2013). But most of the past works aimed at integrating additional information in flood frequency analyses: either historical information, i.e. temporal extension of the data set by incorporation of historical and paleoflood data (Hosking and Wallis, 1986; Cohn and Stedinger, 1987; Neppel et al., 2010; Payrastre et al., 2011; Kjeldsen, 2014), or regional information, i.e. spatial extension of the data set by merging statistically homogeneous series to build a large regional data sample (Hosking and Wallis, 1997; Fill and Stedinger, 1998; Kjeldsen et al., 2002; Seidou et al., 2006; Ribatet et al., 2007a; Kjeldsen and Jones, 2009; Renard, 2011; Alobaidi et al., 2015). Some of these studies suggested to combine both a spatial and temporal extension (Jin and Stedinger, 1989; Merz and Bloeschl, 2008; Viglione et al., 2013) showing that depending on the datasets used, this method was providing either little benefit or a substantial improvement in the estimation accuracy at sites where historical

---

* Corresponding author.
E-mail addresses: olivier.payrastre@ifsttar.fr (O. Payrastre), eric.gaume@ifsttar.fr (E. Gaume).

[1] Present address: University of Science and Technology, 54 Street Nguyen Luong Bang, 555704 Da Nang City, Viet Nam.

http://dx.doi.org/10.1016/j.jhydrol.2016.01.017

information was already available. In the specific context of flash flood prone areas, where extreme flood events often affect ungauged watersheds of limited spatial extent, Gaume et al. (2010) and Nguyen et al. (2014) proposed to incorporate information on extreme floods observed at ungauged sites in regional analyses, in addition to gauged series. Given that these floods are often known to be the largest ones over a relatively long period, they can be considered in a similar manner as historical information. Again Gaume et al. (2010) concluded that the added value of such approaches may highly depend on the characteristics of the regional available information on ungauged extremes.

Bayesian Monte Carlo Markov Chain (MCMC) statistical inference procedures, providing reliable estimates of inference credibility intervals, may now be used to evaluate the added value of any additional information introduced in flood frequency analyses: i.e. added value revealed by the reduction of the computed credibility intervals (Castellarin et al., 2005; Reis et al., 2005; Seidou et al., 2006; Ribatet et al., 2007b; Castellarin, 2007; Payrastre et al., 2011; Viglione et al., 2013). But Nguyen et al. (2014) showed that computed inference credibility intervals could be biased and hence misleading in case of regional frequency analyses because some major asumptions on which their computation is based, especially regional homogeneity, could be not totally valid in real world applications. Regional heterogeneities challenge the available statistical analyses methods, but they also may significantly reduce the relevance of regional flood frequency analyses: the performance of such analyses will depend on a subtile balance between the increase of size of the analyzed data sets and the disruptive effect due to the merging of non-perfectly homogeneous or independent data. The real added value of regional analyses can therefore only be evaluated for each case study considering data availability, but also the possible level of regional heterogeneity and dependency between observations.

Considering these difficulties, a methodology is proposed herein to compare the performances of local and regional approaches in an as objective as possible way. It is based on Monte Carlo simulations of series with characteristics similar to those of the observed data sets. It has been applied here for the comparison of the following methods: 1 – Local analysis of the gauged series, 2 – Local analysis complemented by historical information, 3 – Classical regional analysis based on the model proposed by Hosking and Wallis (1997) using the gauged series, and 4 – Modified regional analysis using the estimated values of extreme discharges at ungauged sites (Gaume et al., 2010; Nguyen et al., 2014).

Two case studies are presented: the Ardèche river and the Argens river regions, both located in the south-east of France. This article analyses and compares the performances of the aforementioned approaches, depending on the characteristics of the available datasets and on the possible heterogenities of the considered regions. Despite the effects of possible dependency of regional observations are not explicitly considered herein, they will also be observed and commented regarding the real dataset of the Argens river case study.

The article is organized as follows: the first section includes a description of the methods involved (inference approaches and Monte Carlo simulations); a presentation of the two case studies is then proposed in the second section; the results are finally presented and discussed in the third section.

2. Description of methods

2.1. The tested inference procedures

The inference procedures applied herein are directly derived from Gaume et al. (2010), Payrastre et al. (2011), and Nguyen et al. (2014). They are included in the nsRFA package of the R statistical software: BayesianMCMC and BayesianMCMCreg functions. As in numerous previous works (Reis et al., 2005; Renard et al., 2006), they are based on the likelihood of the available data sets and on a Bayesian MCMC algorithm for the estimation of the posterior distributions of the parameters given the observed data sets.

Let us consider a regional data sample D including s different sites, all supposed to follow the index flood principle (Dalrymple, 1960), i.e. the flood peak discharge distributions at each site of the region are identical apart from a site-specific scale factor: the index flood μ_i (i being the index of the site). In this region the dataset available at each site i may include:

- a series of gauged annual maximum peak discharges $Q_{i,j}$ ($j = 1,..,n_i$) being the indices of the year.
- and/or the peak discharges $Q_{i,k}$ of h_i extreme floods ($k = 1,..,h_i$), each of these floods being the largest one observed over a period of $t_{i,k}$ years. These floods can be either historical floods at gauged sites or recent extreme discharges recorded at ungauged sites.

The expression of the likelihood of this dataset may be expressed as follows:

$$\ell(\mathbf{D}|\theta) = \prod_{i=1}^{s}\left[\prod_{j=1}^{n_i}\left[\frac{1}{\mu_i}\cdot f_\theta\left(\frac{Q_{i,j}}{\mu_i}\right)\right]\cdot\prod_{k=1}^{h_i}\left[\frac{1}{\mu_i}\cdot f_\theta\left(\frac{Q_{i,k}}{\mu_i}\right)\cdot F_\theta\left(\frac{Q_{i,k}}{\mu_i}\right)^{(t_{i,k}-1)}\right]\right] \tag{1}$$

This equation combines (i) the probability of the gauged series (first term), (ii) the probability of each extreme flood (historical or extreme floods) and the probability of having not exceeded this flood during a $t_{i,k}$ year period over which this flood is considered to be the largest one (second term). The four approaches which will be compared hereafter differ mainly by the content of the data set D. These four approaches correspond to the following situations:

- Case 1: Local approach, based only on the local continuous series recorded at one site in the considered region. This implies $s = 1$ (single site), $\mu_1 = 1$ (no normalization of the series with the index flood) and $h_1 = 0$ (no historical flood). Eq. (1) is therefore reduced to its first term.
- Case 2: Local approach with historical data. This approach is similar to the previous one, with the addition of local historical floods ($h_1 > 0$).
- Case 3: Classical regional approach. This approach is the one developed by Hosking and Wallis (1993, 1997). The data set D is here limited to the continuous gauged series available at s different sites, with no information on extreme floods ($h_i = 0$ for all i). The index flood values μ_i correspond to the average of each series $Q_{i,j}$ ($j = 1,..,n_i$).
- Case 4: Regional approach with extreme discharges. This approach combines both flood series at gauged sites with associated historical data and/or extreme discharges available at ungauged sites. To enable the use of information at ungauged sites, in this last approach the computation of the index flood μ_i is based on an index flood relation of the form $\mu_i = S_i^\beta$, with S_i being the surface of the upstream catchment at site i and β an additional parameter to be calibrated.

Detailed descriptions of these different approaches and of the Bayesian MCMC algorithm can be found in Payrastre et al. (2011, 2013) regarding the local approaches with or without the historical data, and lastly in Gaume et al. (2010) and Nguyen et al. (2014) regarding the regional approaches.