Journal of Hydrology 394 (2010) 101–117

Contents lists available at ScienceDirect

Journal of Hydrology

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

Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites

E. Gaume [a,*], L. Gaál [b], A. Viglione [c], J. Szolgay [b], S. Kohnová [b], G. Blöschl [c]

[a] *Laboratoire Central des Ponts et Chaussées, BP 4129, 44341 Bouguenais Cedex, France*
[b] *Slovak University of Technology, Radlinského 11, 813 68 Bratislava, Slovakia*
[c] *Vienna University of Technology, Karlsplatz 13/222, A-1040 Wien, Austria*

ARTICLE INFO

Keywords:
Flash floods
Regional flood frequency analysis
Bayesian MCMC
Ungauged extremes
Slovakia
France

SUMMARY

This paper proposes a method for using major flash flood events occurred at ungauged catchments to reduce the uncertainties in estimating regional flood quantiles. The approach is based on standard regionalization methods assuming that the flood peak distribution rescaled by a site-dependent index flood is uniform within a homogeneous region. A likelihood formulation and a Bayesian Markov Chain Monte Carlo (MCMC) algorithm are used to infer the parameter values of the regional distributions. This statistical inference technique has been selected for its rigorousness – various hypotheses are explicitly formulated in the likelihood function, its flexibility as for the type of data that can be treated, and its ability to compute accurate estimates of the confidence intervals for the adjusted parameters and for the corresponding flood quantiles.

The proposed method is applied to two data sets from Slovakia and the South of France that consist of series of annual peak discharges at gauged sites and estimated peak discharges of extreme flash flood events that have occurred at ungauged sites. The results suggest that the confidence intervals of the quantiles can be significantly narrowed down provided that the set of ungauged extremes is the result of a comprehensive sampling over the selected region. This remains valid, even if the uncertainties in the estimated ungauged extreme discharges are considered. The flood quantiles estimated by the proposed method are also consistent with the results of site specific flood frequency studies based on historic and paleoflood information.

Introduction

A large part of our knowledge on extreme flood discharge values is based on inventories of data regarding extraordinary events (Gaume et al., 2009; Pekárová, 2009; Solín, 2008; Costa and Jarrett, 2008; Herschy, 2005; Alcoverro et al., 1999; Svoboda and Pekárová, 1998; Costa, 1987; Mimikou, 1984; Rodier and Roche, 1984; UNESCO, 1976; Pardé, 1961). Often, these extremes affect ungauged watersheds, especially in flash flood prone areas, and the question of the valuation of this collected data beyond the simple inventory remains largely open. Even if estimates of such extraordinary events are important source of information on flood extremes, they are seldom really included in formal flood statistical analyses.

The most common practice consists in gathering extreme discharge values in a given area to build the so-called envelope curves (e.g., Castellarin, 2007; Jarvis, 1924). This is a simple and pragmatic approach, which gives an idea of the possible peak discharge values. But it is not completely satisfactory for various reasons. (i) Regions may often be defined a priori within administrative or geographical boundaries. The resulting envelope may therefore only be representative of a homogeneous sub-part of the selected region with the risk of large overestimations of possible flood extremes in the other parts. (ii) The position of the envelope depends on the size (number of station-years) of the considered sample. (iii) Even if not impossible, it is difficult to assign a return period to the envelope peak discharge and in any case it will be based on strong assumptions (e.g., Castellarin, 2007). (iv) Finally, the envelope curve only characterizes a given quantile of the flood peak discharge distribution and not the whole distribution. There is no continuity and consistency between the envelope curve approach and the statistical distribution adjustments based on series of observed discharges in the same region.

At the same time, various methods have been proposed to reduce the uncertainties of at-site flood frequency analyses and produce more robust flood quantile estimates based on larger sample sizes. Two main families of approaches can be distinguished: (i) 'spatial extension' of information on floods can be obtained through regional flood frequency methods consisting in aggregating statistically homogeneous data to build large regional data

---

[*] Corresponding author. Tel.: +33 2 40845884.
E-mail address: gaume@lcpc.fr (E. Gaume).

0022-1694/$ - see front matter
doi:10.1016/j.jhydrol.2010.01.008

samples (e.g., Wallis et al., 2007; Merz and Blöschl, 2003; Hosking and Wallis, 1997); (ii) 'temporal extension' of information on floods can be performed in at-site flood frequency studies on gauged streams incorporating historical or paleoflood peak discharge estimates (Reis and Stedinger, 2005; Parent and Bernier, 2002; Hosking and Wallis, 1986; Stedinger and Cohn, 1986). None of these methods enables, until now, the inclusion of data concerning extremes occurred in ungauged watersheds, for which information has been collected in ad hoc inventories.

This is obtained in this paper by combining techniques developed for spatial and temporal extension. The main idea is to use the methods initially applied for including past extreme values in flood frequency analyses – i.e., the Bayesian Markov Chain Monte Carlo (MCMC) framework (Reis and Stedinger, 2005; Payrastre, 2005; Kuczera, 1999) – substituting the historical peak discharges by the extremes observed in ungauged catchments. This idea is in line with the general philosophy of regional analyses, which is to 'trade space for time' (Hosking and Wallis, 1997). Let us here recall rapidly the principles of the inclusion of historic data in at-site flood frequency analyses, the principles of the incorporation of ungauged extremes in regional flood frequency analyses being very similar.

Consider a typical case of temporal extension of a discharge series based on historic extreme floods. For one river section, 50 years of systematic measurements of discharge are available and eight major historical events were recorded in the 150-year period preceding the systematic measurements. In order to properly account for the historical information, besides assuming of stationarity in time, the evaluation of the eight peak discharges is not sufficient. It is also important to consider the number of years n in which these eight events were the eight major floods, and to evaluate the threshold S which has certainly not been exceeded during this period by the other floods. In other words, the historical information consists not only in the eight extreme discharge values but also in $n - 8$ years of non-exceedance of the threshold S. The choice of n and S should meet the criterion of 'exhaustiveness' (i.e., no other major flood should have exceeded S in the period of time n), which is a necessary condition for a proper statistical inference with censored data (Leese, 1973). Fig. 1 presents two statistical adjustments not including (Fig. 1a) or including (Fig. 1b) the historic period obtained using the Bayesian MCMC procedure (Payrastre, 2005). Fig. 1 represents the maximum likelihood adjusted distribution (continuous line) and the estimated 5–95% confidence limits for this adjusted distribution according to the available data set, computed through the Bayesian MCMC procedure. The historic extreme values appear as brackets on Fig. 1b to indicate that uncertainties in the estimation of the extremes were taken into account. The highest and lowest possible estimates for historic extremes were taken into account to obtain the adjustment presented on Fig. 1b (see "Delineation of homogeneous regions" for details). As it can be seen on this example, the Bayesian MCMC procedure is flexible. It can account for uncertainties in estimated extremes. It also provides estimates of confidence bounds (credibility intervals using the Bayesian vocabulary) for the statistical adjustments and the estimates of the quantiles. The inclusion of the historic period leads to a clear reduction of these credibility intervals in this example revealing its informative value despite the uncertainties in the discharge estimation.

Consider now a different situation: in a region, series of measured discharges are available at various gauged river sections, for example over 30 years on average, and eight major flood events happened, and were surveyed in ungauged catchments over the last 50 years (e.g., eight localized flash floods). Assume that the region is hydrologically homogeneous, which corresponds to the stationarity assumption made in the temporal extension example. This situation can also be seen as a case in which censored data are available: in the ungauged part of the region, eight extreme events were recorded in the last 30 years. The idea developed in this paper is to pool the gauged systematic discharges and the extremes in ungauged basins together in an index flood framework (Hosking and Wallis, 1997; Dalrymple, 1960) so that the pooled data can be represented in a way similar to Fig. 1. Due to the particularities of the regional data sets composed of gauged and ungauged sites, this necessitates adaptations of both: the index

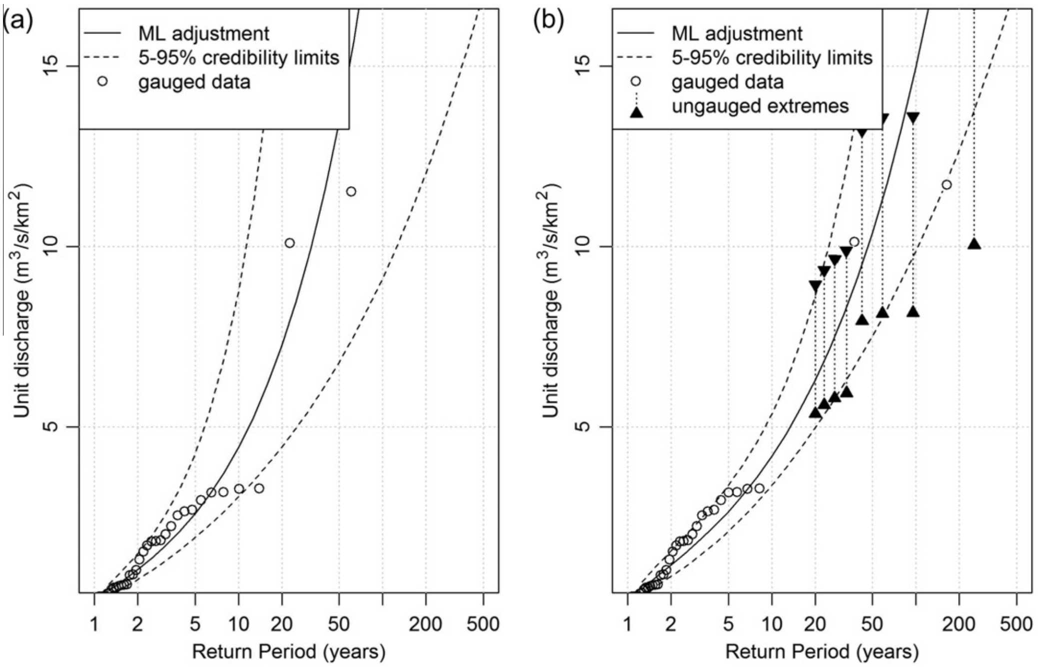

Fig. 1. Example of a statistical inference for the Lauquet River (Aude region, France) (a) based on a series of annual peak discharges, and (b) taking into account historic extremes (taken from Payrastre (2005)).