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 © 2010 Elsevier B.V. All rights reserved.
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

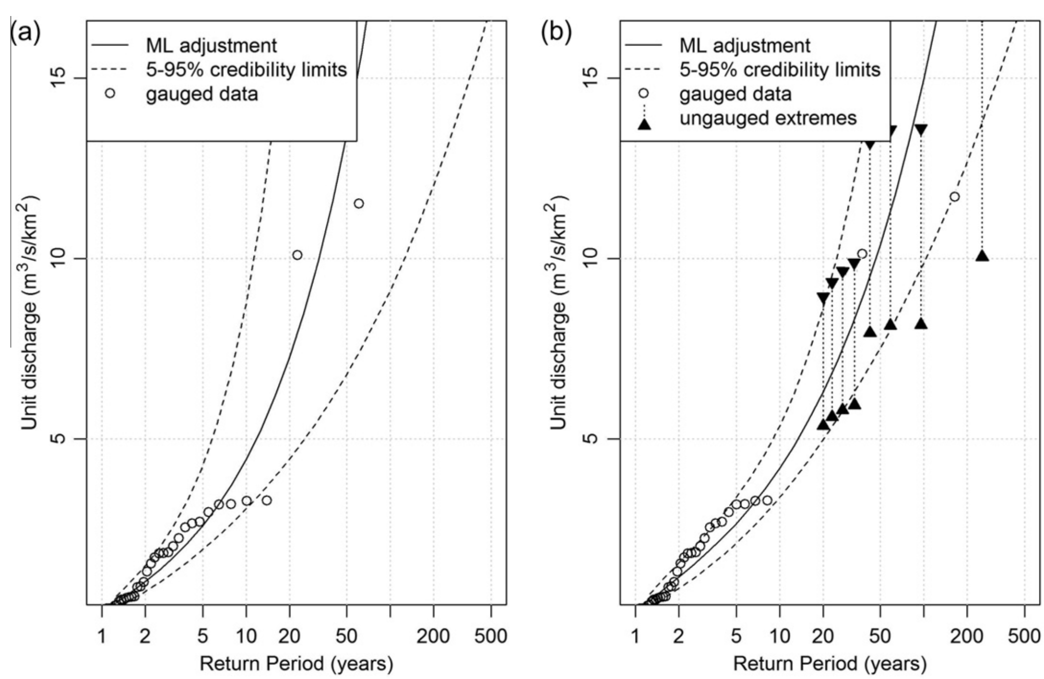

Fig. 1. Example of a statistical inference for the Lauquet River (Aude region, France) (a) based on a series of annual peak discharges, and (b) taking into account historic extremes (taken from Payrastre (2005)).

flood approach and the Bayesian MCMC implementation procedure.

The major contributions of the paper are: (i) a method to delineate the regions and assess their homogeneity enabling the incorporation of ungauged sites, which differs from the common approaches used in regional analyses (“Delineation of homogeneous regions”); (ii) a Bayesian MCMC method to perform the frequency analysis of the regional data, which implies, in particular, the non-trivial choice of threshold values and of the extent in time and space of the inventory of extremes – equivalent to n – in order to meet the criterion of exhaustiveness (“Adjustment of the regional growth curves”).

The proposed methodology is applied to two case studies with very different amounts and types of data (presented in “Presentation of the data sets”): three major flash flood events from ungauged catchments and 41 gauging stations in Slovakia, and 119 major flash flood events from ungauged catchments and 23 gauging stations in the South of France. The ungauged extremes were collected within the European project HYDRATE and are part of a European data set on extreme flash floods (Gaume et al., 2009).

Beyond the proposed procedure, this paper aims at demonstrating that ungauged extremes may bring a significant added value in regional flood frequency analyses. This is not trivial: the value of adding few single additional discharges, whose estimation is usually affected by large uncertainties, to large regional ‘gauged’ discharge samples is questionable. The Bayesian MCMC statistical inference, providing credibility intervals along with the best fitted adjustment, is particularly suitable for measuring this added value. If ‘ungauged extremes’ represent a real additional information, their inclusion should either significantly reduce the uncertainty of the estimated flood quantiles or even invalidate the extrapolations based on the ‘gauged’ data only.

The paper ends with some suggestions of improvement of the proposed method.

Presentation of the data sets

The data sets are composed of both: (i) series of annual peak discharges at gauged sites and (ii) estimated peak discharge values of extreme flash flood events occurred at ungauged sites. The Slovak and French data sets are presented briefly in this section. The ungauged discharges are part of a systematic inventory of extreme flash flood events occurred during the last 60 years that has been recently realized as part of the European research project HYDRATE (http://www.hydrate.tesaf.unipd.it/). This inventory (Gaume et al., 2009) shows a large variability of the number of reported events, related to the variable magnitudes and frequencies of damaging flash floods between the countries.

Slovakia

The selected Slovak study area covers the northern and eastern fringe of the country (Fig. 2) corresponding to the Carpathian Mountains. It is the place where the highest mean annual peak discharge values are measured and where the most destructive past flash floods occurred. The exact delineation of the study area has been based on the results of previous regionalization studies of seasonal discharge maxima (Kohnová, 1997) that delimited four hydrological sub-regions in this area: Tatras East and West and Flysch East and West. The whole area counts 41 gauging stations on watersheds smaller than 500 km^2, with average observation duration of 36 years. Larger watersheds have not been considered as they are generally not affected by flash floods.

The estimated peak discharges corresponding to the three larger flash floods that were reported in Slovakia are available. Two of them, the Štrba Creek flood (24/07/2001) and the Malá Svinka flood (20/07/1998), are relatively recent and affected the North-eastern part of the country. The Vydrňanka Creek event is an older (17/06/1939) but striking flood event that affected the North-western part of the Carpathian Mountains (Fig. 2). These three floods have affected watersheds of very limited area (Table 1). Their estimated unit peak discharges (discharge divided by the watershed area) exceed by far the highest values recorded on the gauged streams. They could appear as outliers.

Even if the information on ungauged extremes is limited, it corresponds to the type of information generally available on extreme floods in many areas and it appeared therefore interesting to test how it could be valuated within a regional flood frequency analysis. The French data set, based on a comprehensive inventory of ungauged extremes, is less problematic from this perspective (see “South of France”, “Results” and “Discussion”). One difficulty

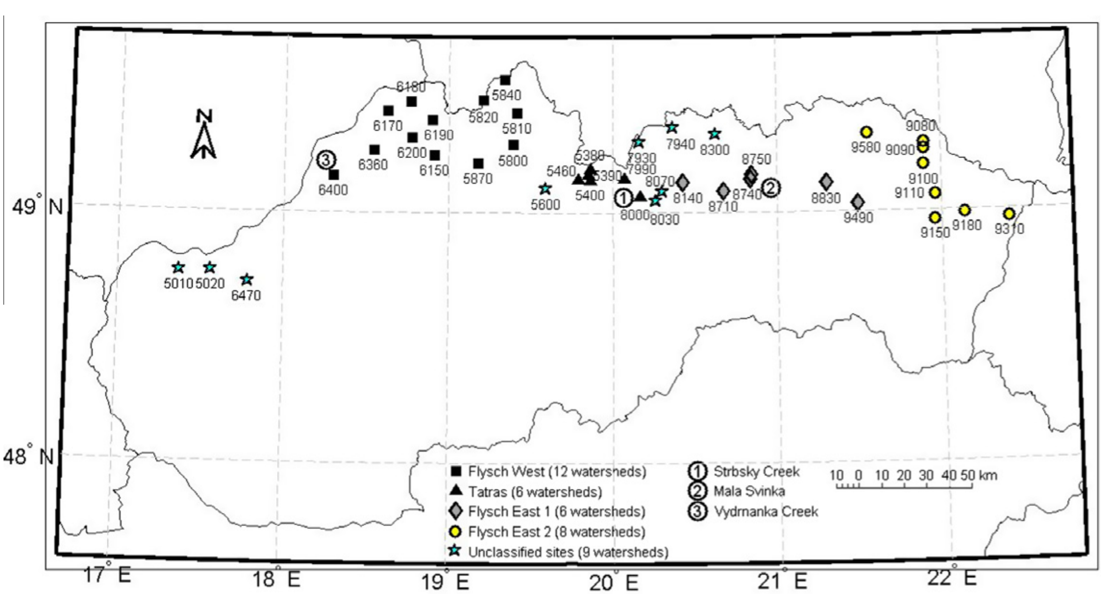

Fig. 2. Location of the selected 41 watersheds in Slovakia, their final arrangement into four sub-regions and location of three major ungauged flood events (encircled numbers).

Table 1
Estimation of the equivalent duration N_0 in station-years corresponding to the flood extremes. A_{region}: a rough estimation of the area of the region; A_{flood}: area of the watershed affected by the flood event; n_{est}: an approximate average length of the period of time covered by the inventory; N_0: equivalent number of years of observation.

Region	Major flood	A_{region} (km^2)	A_{flood} (km^2)	n_{est} (years)	N_0 (years)
Flysch West	Vydrňanka Creek	~4000	10.9	~50	~20,000
Tatras	Štrbský Creek	~500	2.5	~40	~10,000
Flysch East 1	Malá Svinka	~1000	6.5	~30	~5000

posed by the incorporation of isolated extreme values in a flood frequency computation framework is the estimation of the equivalent number of years of observation N_0 represented by these extremes: i.e., the number of similar locations implicitly taken into account in the regional inventory of extreme ungauged floods, multiplied by the number of years during which the extreme discharge has not been exceeded (see "Adjustment of the regional growth curves"). It has been roughly estimated here as the ratio between the area of the region and the area affected by the extreme flood – estimate of the number of similar watersheds in the region – multiplied by the average length of the gauged series (Table 1). Of course, such an estimate is highly questionable and a sensitivity analysis of the obtained results to N_0 has been conducted (see Table 4 and "Tatras and Flysch East 1 in Slovakia").

South of France

The French Mediterranean area and especially the Cévennes-Vivarais region is one of the regions in Europe exposed to the most frequent and severe flash floods (Gaume et al., 2009). The Cévennes-Vivarais region corresponds to the first reliefs located north from the Mediterranean Sea on the right-hand side of the Rhône River. It is covered by three French departments (administrative subdivision a little larger than a county): the Hérault, Gard and Ardèche (Fig. 3). The study is focused on these three departments.

The area counts 53 stream gauges but most of them are part of flood forecasting networks and have been only used to measure flood levels and not discharges in the past. No accurate rating curves are available for such stations. The set of gauging stations where accurate and long series of observed discharges are available is reduced to 23. The series of annual peak discharges counts about 800 records.

A comprehensive inventory of extreme flash flood events at ungauged sites has been recently realized in the three departments within the research program HYDRATE (Newinger, 2007). Thanks to the richness of information available in the region on flash floods (local studies, flood marks, flood risk maps, etc.), it has been possible to gather 236 peak discharge estimates for 119 different

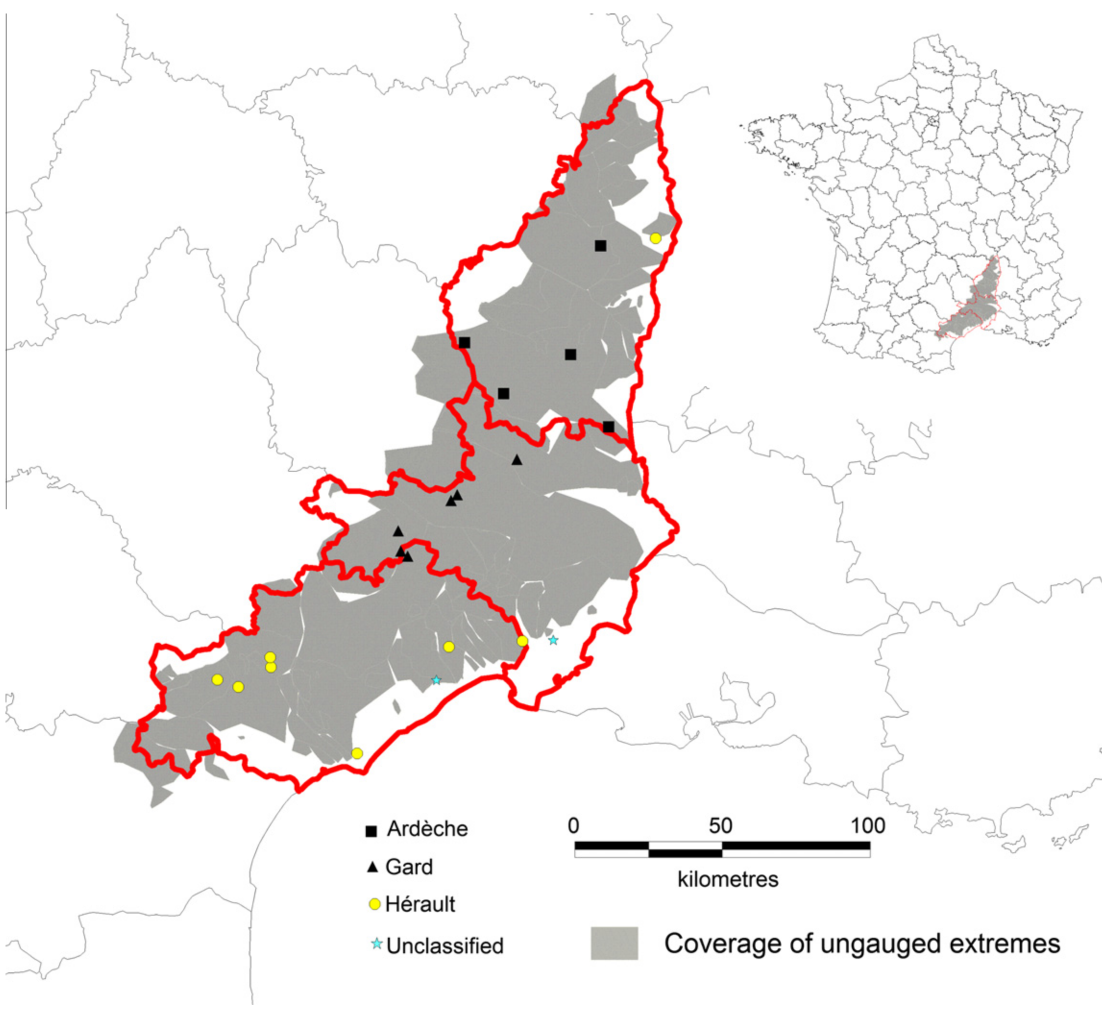

Fig. 3. Study area in France: final arrangement into three sub-regions and coverage of the inventory of extreme floods in ungauged watersheds.

locations, corresponding to the highest observed flood events at these locations during the last 50 years. The estimated discharges come from various sources; this data set has therefore been criticized to eliminate or correct obviously wrong – generally overestimated – values (Gaume et al., 2009). But hence, these values remain estimates and the effect of the possible uncertainties affecting these estimates has been evaluated hereafter. For the analyses presented hereafter, only the maximum value of each site has been selected. The final ungauged extreme set counts 119 records: 23 in the Hérault, 58 in the Gard and 38 in the Ardèche department. The masks of the corresponding watersheds have been reported on Fig. 3, indicating an excellent coverage of the three departments by this inventory of extreme discharges. It is important to note that due to the spatial extent of some extreme rainfall events, many extremes correspond to the same dates: 30/09/1958 or 09/09/2002 for example. The sets of extreme discharges had to be further refined in the inference application to get rid of the possible problems introduced by the statistical dependences in the processed data sets (see "Results").

Accuracy of the peak discharge estimates at ungauged catchments

Peak discharge estimates of extreme floods occurred on ungauged but also on gauged rivers are necessarily based on hydraulic engineering know-how and strong assumptions (Gaume et al., 2003). They are inevitably affected by significant uncertainties that have to be taken into account in statistical inferences especially if confidence intervals of discharge distributions have to be evaluated (Neppel et al., in press). The evaluation of uncertainty levels affecting extreme peak discharge estimates is not the main issue of this paper, but a realistic choice of uncertainty levels was necessary to demonstrate the potential of the ungauged extremes and the proposed method. Let us present some few arguments justifying the finally selected uncertainty levels for the HYDRATE discharges.

A large body of literature has been devoted to the evaluation of ungauged discharges (Gaume and Borga, 2008; Borga et al., 2008; House et al., 2002; Jarrett, 1990). These works generally aim at limiting the risk of large estimation errors, especially of large overestimations that can reach 100% – see for instance the studies of House and Pearthree (1995) and Jarrett (1987) who concluded after having examined and reworked past peak discharge estimates of two remarkable flash floods having affected the USA, that very large overestimation could be suspected. Nevertheless, such overestimations are probably exceptional cases and a 100% uncertainty should not be considered as the standard uncertainty level affecting extreme discharge estimates.

Concerning the HYDRATE database, it is composed of estimates coming from diverse sources (research but also technical reports) and obtained through various methods: simple hydraulic formulas in most cases (slope-area method based on the Manning–Strickler formula), sometimes 1-D or 2-D hydraulic modeling (Gaume et al., 2009; Majerčáková et al., 2004; Šťastný and Majerčáková, 2003). Some of the largest reported peak discharges in the database are

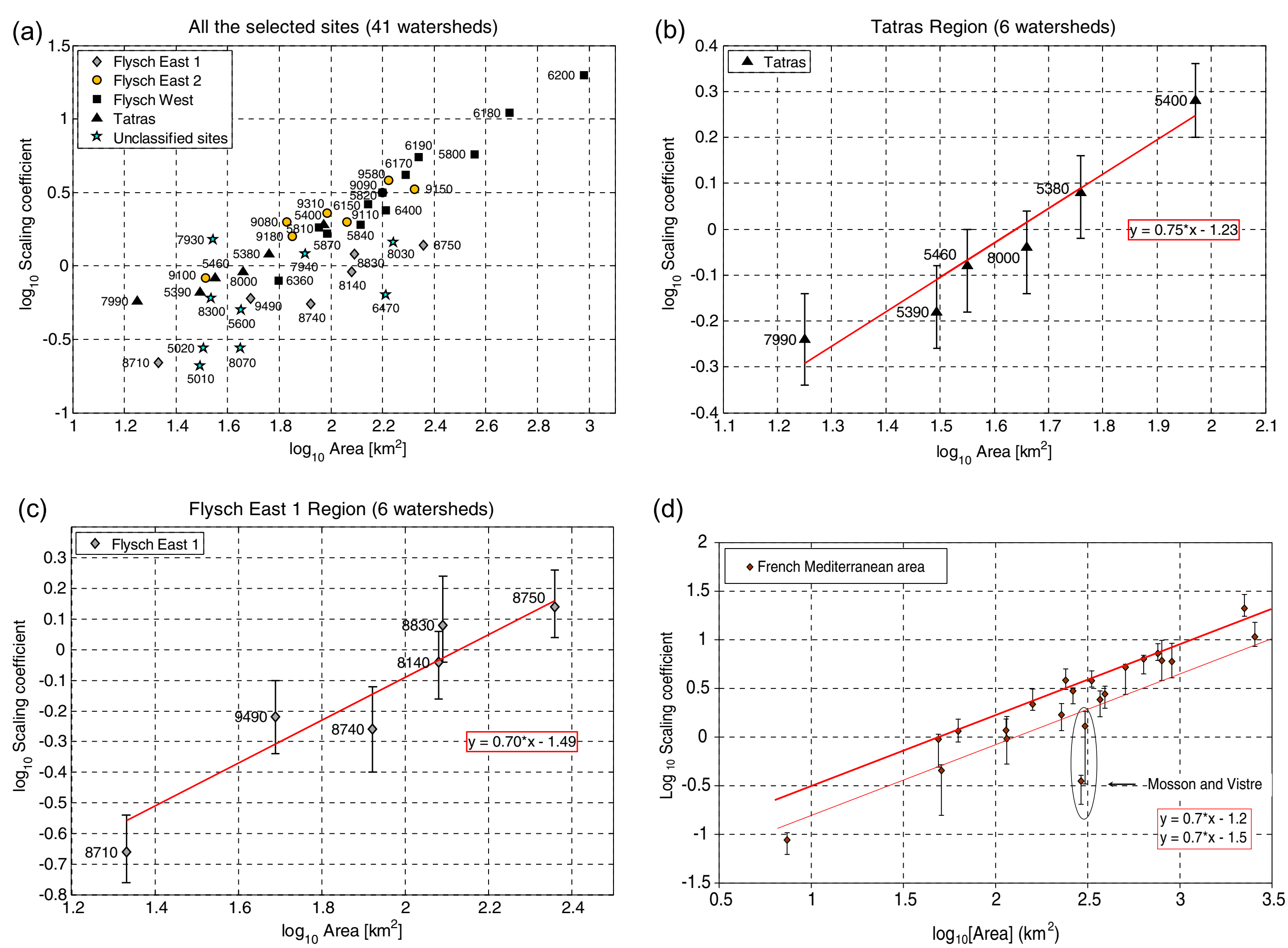

Fig. 4. Best estimated scaling coefficients according to the Wilcoxon–Mann–Whitney test for the selected 41 sites in Slovakia (a). Adjustment of the index flood relations for Tatras (b), Flysch East 1 (c) and the French Mediterranean area (d).

nevertheless the result of detailed evaluations conducted after field surveys including the estimation of water-surface velocities based on films and pictures taken by eye-witnesses and rainfall-discharge consistency tests (Gaume and Borga, 2008; Delrieu et al., 2005; Gaume et al., 2003). The corresponding peak discharges can therefore be considered as relatively accurate and at least the risk of large overestimations as very limited.

Despite the heterogeneity of the collected data, the critical analysis of the HYDRATE database revealed satisfactory features. First, the order of magnitude of the extreme peak discharges among the various countries covered was consistent and was not influenced by the estimation method used (Gaume et al., 2009). Second, within a country, especially within France, the ranking of the flood event peak discharges corresponded to the ranking of the generating rainfall events, when rainfall data was available. This illustrates that the discharge estimations are generally accurate enough to reveal differences in the magnitude of the floods: i.e., the average uncertainty is significantly lower than the range of discharge values covered by the HYDRATE inventory (a factor of 2–3 between the smallest and largest reported peak discharge, see Figs. 6 and 7). Finally, past surveys revealed that the average water velocity in the main stream during extreme flash flood events was generally comprised between 2 and 4 m/s (Gaume and Borga, 2008). Provided that the engineers have this order of magnitude in mind when estimating a discharge and considering that the flow cross-sectional area is generally known, the discharge estimation uncertainty can hardly exceed ±25 to 30%.

This state of facts, leads to conclude that a ±25% uncertainty range is a balanced evaluation for the HYDRATE database. It is not too optimistic – the uncertainty range remains large with a ratio of almost 1–2 between the lower and the higher considered estimates. A too pessimistic choice would have excessively reduced the informative content of the ungauged extremes.

Delineation of homogeneous regions

The proposed regional analysis is based on the index flood method (Dalrymple, 1960). The basic assumption of this method can be expressed by the equation

$$Q_i^{(j)} = \mu_i q^{(j)}, \quad i = 1, \cdots, M \tag{1}$$

where i denotes a given site, M is the total number of sites in the homogeneous region, $Q_i^{(j)}$ is the jth quantile at the ith site, $q^{(j)}$ is the regional dimensionless (i.e., reduced) jth quantile (i.e., the regional growth curve) and μ_i is the at-site scale factor (i.e., the index flood) that does not depend on j (simple scaling assumption).

The homogeneity of the dimensionless growth curve of possible regional samples will be tested using the heterogeneity measure H_1 proposed by Hosking and Wallis (1993; see also Castellarin et al., 2008; Viglione et al., 2007; "Homogeneity of growth curves"). But, when ungauged sites are to be included in the regional analysis, the homogeneity requirements given by Hosking and Wallis (1993) are not sufficient to define homogeneous regions. An additional condition must be fulfilled. A unique scaling relationship for the index flood ("Adjustment of the index flood relation") must hold for all the watersheds of the region, which will be used to compute the reduced peak discharges and threshold values for the ungauged watersheds of the same region. In the proposed approach we defined homogeneous regions to have a consistency from a physiographic and geographic point of view, which allowed assigning easily the ungauged sites to the identified regions. In the proposed procedure, the gauged data are first analyzed to define the index flood relations and possible homogeneous regions. The homogeneity of the growth curves, more precisely of their L-coefficient of variation, of the proposed regional sets is verified in a second step using Hosking and Wallis statistics.

Adjustment of the index flood relation

Usually, when only data from gauged sites are considered, the index flood is estimated by the at-site sample mean (Hosking and Wallis, 1997) or median (Robson and Reed, 1999). However, mean or median annual maximum floods can only be computed for gauged sites. To include ungauged extremes, the index flood has to be related to various climatic and/or physio-geographic watershed characteristics. One of the common regression models generally used in regional analyses is based on a power law relationship, with a general form

$$\mu_i = c \prod_{k=1}^{K} X_{ik}^{\beta_k} \tag{2}$$

where X_{ik} denotes the kth catchment attribute at the ith site, c and β_k are the coefficients to be found by the regression, and K is the number of the selected catchment attributes. In this paper, the simplest possible regression model has been used with a single characteristic ($K = 1$), which is the catchment area A:

$$\mu_i = cA_i^{\beta} \tag{3}$$

The results based on this simple assumption were satisfactory; however, more complex index flood relations could be tested in future developments of the method.

In the case of Eq. (3), a straightforward procedure to delineate homogeneous regions and adjust index flood relations consists in plotting the sample means (or medians) of the discharge series against the watershed areas (log–log scales) and to fit straight lines in the obtained cloud of points (see Fig. 4a). To confirm the choices done, a measure of uncertainty of the index flood estimates is necessary to evaluate whether the distance between any point and the proposed line is acceptable or not, i.e., whether the considered watershed reasonably belongs to the defined region. With no a priori knowledge on the shape of the peak discharge distributions, the confidence intervals for the computed mean or median values can hardly be estimated. It is therefore suggested herein to base the estimation of confidence bounds on a non-parametric method based on the Wilcoxon–Mann–Whitney (WMW) test (see "Application of the Wilcoxon–Mann–Whitney test" for details). The WMW test, originally proposed by Wilcoxon (1945) and extended by Mann and Whitney (1947), is generally used to assess the null hypothesis that two independent samples are drawn from the same population or more precisely to assess if 'one of two random variables is stochastically larger than the other' (Mann and Whitney, 1947). The WMW statistics is a rank sum test that is strongly controlled by the median values of the samples and is therefore often compared to the Student parametric test (Kendall and Stuart, 1979). The idea of the implementation of the WMW test to support the evaluation of index flood relations consists in selecting a reference sample and in comparing it with all the other samples rescaled by a factor μ. This factor is varied and the range of μ values for which the WMW null hypothesis cannot be rejected at a significance level of p is an estimate of the $(1 - p)$% confidence interval for the scaling factor of each sample given the choice of the reference sample and according to the WMW statistics ("Application of the Wilcoxon–Mann–Whitney test"). Each available series has been selected as reference sample in turn, to verify that the results were not sensitive to the choice of the reference sample.

Illustrations of the outcome of the proposed procedure are shown in Fig. 4b–d showing the ranges of acceptable scaling factors μ for every gauged series of the considered regions. According to the computed confidence bounds, one single relation cannot be adjusted to the whole Slovak or French data sets. An infinite number of combinations of lines and points can be proposed. The final choice has to be underpinned by further hydrological, geological or climatological considerations. For Slovakia, the isolation of homogeneous regions has been guided by the results of a previous regional analysis (Kohnová, 1997), which relied both on former geomorphologic (Mazúr and Lukniš, 1986) and hydrogeological classification of the territory of Slovakia (Fusán et al., 1980). The final four regions are globally consistent with the regions defined by Kohnová (1997) with some adjustments: (i) The Tatras region merges two initially different regions, (ii) the Flysch East region had to be divided into two sub-regions and nine watersheds out of 41 could not be classified (Fig. 4a). Another possible combination in Slovakia could have for instance consisted in merging most of the Tatras, Flysch West and East 2 regions including watersheds 8300 and 7930.

For the French data set, two relations can account for all the points except two: the Vistre and Mosson, which are rivers with very large floodplains attenuating the peak discharges and explaining their relatively low index flood values. Two other points are not exactly on the curves corresponding to the smallest and one of the largest considered watershed areas. Their distance from the adjusted curves remains limited: the distance in $\log_{10}$ scaling coefficient lower than 0.1 – which means a possible error on the scaling coefficient and the corresponding discharge quantiles of less than +25% for the smaller watershed and −25% for the largest – is almost negligible considering the other sources of uncertainties. Moreover, the two identified regions have a clear geographical consistency: the region with the highest scaling coefficients corresponds to the mountainous part of the area covering mainly the department of the Gard and Ardèche, while the other region merges the low altitude watersheds located in the Rhône River valley and the Languedoc plain essentially in the Hérault department. The distance between the adjusted log scaling coefficient relations is significant: a value of about 0.3 represents a ratio between the scaling factor values of about 2 for a given watershed area.

Homogeneity of growth curves

Before adjusting regional growth curves on the regional pooled data samples, it appeared necessary to realize a complementary homogeneity test. Since the WMW test is mostly controlled by the median of the samples, it does not guarantee that the various data samples of the regions have comparable statistical distributions or that at least their second moment (variance or coefficient of variation) are similar. Hosking and Wallis (1997) have proposed a measure, which compares the actual between-site variation in sample L_{CV} (coefficient of L-variation) for a group of sites with the variation that would be expected in the case of a homogeneous region. The test is shortly presented in "The Hosking and Wallis regional homogeneity test". According to the initial definition of the HW homogeneity measure (Hosking and Wallis, 1997), the critical H_1 values are set as follows: the region is *acceptably homogeneous* if $H_1 < 1$, *possibly heterogeneous* if $1 \leqslant H_1 < 2$, and *definitely heterogeneous* if $H_1 \geqslant 2$. However, "these heterogeneity criteria measure statistical heterogeneity from known distribution and do not account for variability that arises from other sources" (Wallis et al., 2007), variability induced for instance by the implementation and running of the measuring devices. Therefore Wallis et al. (2007) suggested selecting less restrictive thresholds for the study of annual precipitation maxima: $H_1 < 2$ may be considered as acceptably homogeneous, while $H_1 \geqslant 3$ is an indicator of heterogeneity (Wallis et al., 2007). As stream gauge measurements are at least as affected by variability as rain gauge measurements, these second less restrictive criteria have also been used in this paper.

Tables 2 and 3 present the H_1 values obtained for the proposed regional sets. The results are varied, but encouraging. Most of the finally defined regions appear as not definitely heterogeneous according to the test. The H_1 values obtained on the new Slovak regions are higher than the values obtained on the original regions defined by Kohnová (1997) indicating a certain consistency between the WMW and the HW criteria. The HW test revealed differences in annual peak discharge sample statistics within the French 'mountainous' area. This initial sample had to be split into two separate data samples corresponding to two clearly defined regions covering respectively the Ardèche and the Gard departments and corresponding to the Ardèche river system on the one hand, covering the North part of the considered region and the three other main river systems on the other (Gard, Cèze and Vidourle) located further South. The distance from the Mediterranean Sea and the corresponding variations of the climate forcing may explain these differences. The Hérault H_1 value is improved if the Lez River is retrieved from the regional set. The upstream Lez River is fed by a karstic system. This may explain its specificities. But the whole region is karstified and it is therefore difficult to determine if an ungauged watershed, partly karstic, has to be included or not in the modified Hérault region. For this reason the subsequent analyses were not conducted on the Hérault regional set. They will be focused on the four regions appearing as not 'definitely heterogeneous' according to the modified H_1 test: Tatras, Flysch East 1, Ardèche and Gard.

Table 2
Some characteristics of the final regions in Slovakia: results of the Hosking and Wallis homogeneity test (H_1) and index-flood relationships. WMW stands for Wilcoxon–Mann–Whitney.

Region	# of sites	WMW plot equation	Index-flood relationship	H_1 value	Implication of H_1 test
Flysch West	12	$y = 1.10, x - 2.00$	$\mu = 0.0100\,A^{1.10}$	3.01	Definitely heterogeneous
Tatras	6	$y = 0.75, x - 1.23$	$\mu = 0.0588\,A^{0.75}$	1.07	*Acceptably homogeneous*
Flysch East 1	6	$y = 0.70, x - 1.49$	$\mu = 0.0324\,A^{0.70}$	0.02	*Acceptably homogeneous*

Table 3
Some characteristics of the final regions in France: results of the Hosking and Wallis homogeneity test (H_1) and index-flood relationships. WMW stands for Wilcoxon–Mann–Whitney.

Region	# of sites	WMW plot equation	Index-flood relationship	H_1 value	Implication of H_1 test
Ardèche and Gard	11	$y = 0.7, x - 1.2$	$\mu = A^{0.7}$	4.52	Definitely heterogeneous
Ardèche	5	$y = 0.7, x - 1.2$	$\mu = A^{0.7}$	2.61	*Possibly heterogeneous*
Gard	6	$y = 0.7, x - 1.2$	$\mu = A^{0.7}$	0.63	*Acceptably homogeneous*
Herault	8	$y = 0.7, x - 1.5$	$\mu = A^{0.7}$	7.31	Definitely heterogeneous
Hérault (Lez excl.)	7	$y = 0.7, x - 1.5$	$\mu = A^{0.7}$	2.08	*Possibly heterogeneous*

Adjustment of the regional growth curves

Once a homogeneous region has been delineated through the procedure explained in "Delineation of homogeneous regions", the available reduced discharges rescaled according to the proposed index flood relations (Eq. (3) and Tables 2 and 3) can be pooled together. A Bayesian MCMC approach has been selected to adjust growth curves to these regional sets composed of heterogeneous and partly censored data. The approach is based on two important elements: (i) the estimation of the likelihood of the available data sample conditionally to the type of statistical distribution and to the values of its parameters and (ii) an efficient n-dimensional random variable sampling method. It is shortly presented hereafter. More in-depth presentations are provided in Reis and Stedinger (2005), in various publications devoted to its application in hydrology (e.g., El-Adlouni and Ouarda, 2009; Hassan et al., 2009; Lee and Kim, 2008; Renard et al., 2006; Kanso et al., 2003; Kuczera, 1999; Kuczera and Parent, 1998; Mailhot et al., 1997) or in specialized textbooks (Robert and Casella, 2004; Tanner, 1996).

Likelihood of the available data set

Likelihood based statistical inference has become one of the standard approaches for flood frequency analyses. The maximum likelihood estimator is unbiased and is generally asymptotically efficient. It also enables without difficulties the incorporation of censored data as historical or paleoflood estimated discharges or regional extremes.

To compute the likelihood of a sample of peak discharge values, the type of statistical distribution must be selected and the sample strategy must be clearly defined. The distribution type selected herein is the three-parameter *generalized extreme value* (GEV) distribution (Jenkinson, 1955), which is one of the popular distribution types used for hydrological applications (e.g., Koutsoyiannis, 2004). However, any other statistical distribution family could have been tested. Its cumulative distribution function F_θ and probability density function f_θ have the following form:

$$F_\theta(x) = \exp\left[-\left(1-\frac{k(x-a)}{b}\right)^{1/k}\right]_{b>0} \quad \text{with} \quad \theta = \{a,b,k\} \tag{4}$$

$$f_\theta(x) = \frac{1}{b}\left(1-\frac{k(x-a)}{b}\right)^{(1/k-1)} \times \exp\left[-\left(1-\frac{k(x-a)}{b}\right)^{1/k}\right]_{b>0} \quad \text{with} \quad \theta = \{a,b,k\} \tag{5}$$

where θ is the vector of parameters, and $k \neq 0$.

The sampling strategy of the ungauged extremes must be clearly defined to compute the likelihood of the pooled sample properly. In the case of a regional analysis including regional ungauged extremes, this sample D is composed of s annual reduced discharge maxima systematically recorded at the gauged sites $\{x_1,\ldots,x_i,\ldots,x_s\}$, of k reduced extreme peak discharges estimated at ungauged sites $\{y_1,\ldots,y_j,\ldots,y_k\}$, k depths (in years) of the inventory of extremes at the ungauged sites $\{n_1,\ldots,n_j,\ldots,n_k\}$, k numbers of extremes reported at each ungauged site $\{m_1,\ldots,m_j,\ldots,m_k\}$ and k threshold values at each ungauged site $\{S_1,\ldots,S_j,\ldots,S_k\}$. These three last terms must not be omitted. In addition, S_0 is the regional threshold which is considered not to have been exceeded at ungauged sites with no estimated extremes but accounted for in the regional analysis, and N_0 is the length in station-years of this last set of ungauged sites. If all the data are supposed to be independent realizations of the same random variable X and if the m_j extremes are the largest peak discharge values over the period n_j at the ungauged site j – this condition is necessary to ensure exhaustiveness – then the general expression of the likelihood of the sample D is the following:

$$\ell(D|\theta) = \prod_{i=1}^{s} f_\theta(x_i)\left(\prod_{j=1}^{k} f_\theta(y_j)\right)\left(\prod_{j=1}^{k}[F_\theta(S_j)]^{(n_j-m_j)}\right)[F_\theta(S_0)]^{N_0} \tag{6}$$

or

$$\ell(D|\theta) = \prod_{i=1}^{s} f_\theta(x_i)\left(\prod_{j=1}^{k}[F_\theta(y_{Uj})-F_\theta(y_{Lj})]\right)\left(\prod_{j=1}^{k}[F_\theta(S_j)]^{(n_j-m_j)}\right)[F_\theta(S_0)]^{N_0} \tag{7}$$

if the magnitudes of the ungauged extremes are known with some uncertainty and upper and lower estimates can be defined $[y_{Lj}, y_{Uj}]$, $j = 1,\ldots,k$.

The first term of the likelihood function (Eqs. (6) and (7)) corresponds to the probability of the sample of systematic gauged annual peak discharges. The second and third terms are complementary and correspond to the probability of the samples available at the ungauged sites. The information at the ungauged sites is not only composed of the estimated extreme discharges but also of the number of years for which no estimate is available because the peak discharge was not sufficiently high (i.e., did not exceed a given perception threshold S_j). This information is as important as, or even more important than the extreme peak discharge estimates as will be illustrated in the next section. With no additional information, the perception threshold is chosen equal to the smallest estimated extreme value at the ungauged site. Again, exhaustiveness is absolutely necessary for a correct inference. It must be certain that no discharge exceeded the threshold S_j during the years where no estimate is available. If it is not certain, the threshold should be increased and some estimates not taken into account. The higher the threshold, the lower the information content about the periods with no estimates in the ungauged watersheds and the lower the weight of the third term of the likelihood in the inference procedure. To ensure exhaustiveness, only one single extreme, the highest value, has been considered for each ungauged site in the applications presented hereafter. The fourth term of the likelihood (Eqs. (6) and (7)) may also be very important depending on the sampling strategy of ungauged extremes. From this perspective, the French and Slovak samples are very different. The French sample stems from a systematic inventory of the largest floods occurred in the last 50 years on various streams of the region. The sampling of sites is not random because the sites are determined by the existence of assets – typically towns or at least houses – where the extremes could have been detected and recorded, but has no a priori link with the magnitude of the past observed floods. A fourth term is not necessary in the likelihood formulation except if no information exists at some sites because no significant flood occurred, situation which has not been encountered in France. The Slovak sample is composed of extraordinary floods, considered as regional extremes. Implicitly, it is considered that no flood has exceeded these events on any other comparable watershed of the region over the given period of time of the inventory. The fourth term of the likelihood is then absolutely necessary to accurately account for the ungauged extremes with a threshold S_0 equal or close to the regional extreme and the number of station-years N_0 equal to the period covered by the inventory of extremes multiplied by the number of equivalent watersheds in the region. In such a case, it is much more difficult to be sure of the exhaustiveness of the inventory (i.e., that no other flood on an unpopulated watershed did not exceed the considered extreme), and to evaluate the number equivalent watersheds. It has been set here equal to the area of the region divided by the area of the watershed on which the extreme has been recorded. The estimates for the two regions Flysch and Tatras are provided in Table 1. A sensitivity analysis of the results to N_0 has been conducted for the Slovak case to assess the possible impact of the strong assumptions on which the estimate of N_0 is based on the results of the statistical inference (Table 4 and "Tatras and Flysch East 1 in Slovakia").

Finally, the likelihood formulation (Eqs. (6) and (7)) is based on the assumption of independence between records. Inter-site dependencies may reduce the information content of the data set and hence may have a significant impact on the uncertainties affecting the computed discharge quantiles. It is a difficult question to tackle. Some authors have suggested the computation of an equivalent number of independent annual maxima (Castellarin et al., 2005). The proposed empirical formula (Eq. (8)), based on a linear correlation assumption, could be used for instance in our case to compute an equivalent station-year length N_0' leaving aside the question of the correlation between measured discharges:

$$N_0' = \frac{N_0}{1+\rho^{\beta}(n-1)} \quad \text{with} \quad \beta = 1.4\frac{N_0^{0.176}}{(1-\rho)^{0.376}} \tag{8}$$

where n is the number of considered ungauged sites and ρ the inter-site correlation coefficient.

Considering the value of N_0 largely exceeding 1000 station-years when extreme ungauged discharges are considered and the ranges of values for ρ computed on the existing measured discharge series – between 0.1 and 0.2 – the correction of N_0 based on Eq. (8) is negligible (a few percents). According to these previous results, the influence of inter-site dependence should be limited due to the fact that N_0 is large – i.e., that the proposed method aims at valuating high return period discharges. Other tests conducted hereafter lead to a similar conclusion of an apparently moderate effect of the inter-site dependence. These are of course only indices and not demonstrations. A detailed study of the possible influence of inter-site dependence in flood frequency studies is still to be conducted but is far beyond the objectives of this paper.

Bayesian perspective and MCMC inference

The Bayesian perspective on statistical inference and the Markov Chain Monte Carlo procedures will only be shortly presented before the inference results (see Tanner (1996) or Robert and Casella (2004) for more details). The most common statistical inference approach consists in finding the parameter set θ maximizing the value of the likelihood $l(D|\theta)$ using a numerical optimization method. It is the maximum likelihood approach providing a best parameter set and an optimum statistical adjustment. Recalling the Bayes' Theorem, it is possible to write the following expression:

$$p(\theta|D) = \frac{l(D|\theta)p(\theta)}{p(D)} \tag{9}$$

where $p(\theta|D)$ stands for the marginal probability density of the parameter vector θ given the data set D (i.e., posterior distribution), $p(\theta)$ is the so called prior distribution of θ which, summarizes any a priori or alternative knowledge on the distribution of θ, and $p(D)$ is the probability of the sample D which is unknown. When no a priori information exists on the distribution of θ then $p(\theta)$ is taken equal to 1. It is the case here, which then implies that $p(\theta|D)$ is proportional to $l(D|\theta)$.

The result of the inference is not an optimum vector but its posterior density function. If the inference problem is well-posed, the spread of this posterior distribution should diminish as the information content of D increases: i.e., this spread is a measure of the information content of the sample D (Tanner, 1996; Mantovan and Todini, 2006). The spread of the posterior distributions of GEV parameter vectors and corresponding quantiles will be the criteria used hereafter to evaluate the added value of the regional ungauged extremes in the statistical inference process.

MCMC algorithms combining random walk Monte Carlo methods with Markov Chains are the last element of the proposed approach. They are a class of algorithms for sampling from multivariate random distributions efficiently (Robert and Casella, 2004; Tanner, 1996). A proportionality constant does not influence the result: i.e., the computation of $l(D|\theta)$ or $l(D|\theta)\,p(\theta)$ if a prior distribution is selected, is sufficient to sample vectors θ from a distribution with density function $p(\theta|D)$. The final result of the MCMC procedure is a set of vectors θ, typically some 10,000 vectors, with density $p(\theta|D)$. The corresponding discharge quantiles according to the GEV cumulative distribution can then be computed and uncertainty bounds – credibility intervals using the Bayesian vocabulary – estimated based on this large set of quantile values. The results presented here have been obtained by a code based on the package *nsRFA* of the *R* statistical software (Viglione, 2009).

Results

To evaluate the possible added value of regional ungauged extremes, three different runs were conducted at least in each case study: (1) inference with the systematic data only, (2) inference including the regional ungauged extremes, (3) inference including ungauged extremes imperfectly known. Uncertainty bounds of ±25% were selected: i.e., $y_{Lj} \approx 0.75y_j$ and $y_{Uj} \approx 1.25y_j$ in Eq. (7) (see also "Accuracy of the peak discharge estimates at ungauged catchments"). This also corresponds to the generally accepted error ranges for paleoflood discharge estimations (O'Connell et al., 2002). Three different case studies will be presented successively. They lead to very different results and conclusions concerning the usefulness of ungauged extremes and illustrate the richness of the Bayesian MCMC approach.

Tatras and Flysch East 1 in Slovakia

Table 4 and Fig. 5 present the results of the inference conducted on the Flysch East 1 and Tatras regions. The gauged annual peak discharge series counts 173 and 228 respectively for six gauged watersheds – average duration of 29 and 38 years of record per site. The first striking figure is the width of the credibility intervals computed on the gauged discharges only (Fig. 5a and c and Table 4). Provided that (1) the region is homogeneous, (2) the records are independent, (3) the peak discharge distribution is of the GEV type, it is only possible to say that the 100-year reduced discharge is comprised between 97.2 and 210.5 in the Flysch East 1 region with a 10% chance to be wrong. The relative magnitude is less important in the Tatras region but still remains significant.

If no uncertainty existed on the equivalent number of station years N_0, the integration of the regional extreme would lead to a clear reduction of the credibility intervals (Table 4). In both cases, the results appear not to be very sensitive to the introduction of the estimated peak discharge uncertainties. This may appear surprising, but it is in accordance with results obtained with the same method on inference including historic and paleofloods (Payrastre, 2005). It indicates that the second term of likelihood (Eqs. (6) and (7)) has a moderate influence on the inference result and that the third and fourth term are the determining ones. The number of years with non-exceedance of the identified threshold is as important, or even more important than the exact value of the ungauged extremes. Note that both are linked since the threshold is related to the estimated extreme values. If it is the case, the inference results should be very sensitive to N_0. The estimation method of N_0 is

Table 4
Estimation of the reduced discharge quantiles Q_T and their confidence intervals corresponding to the return periods $T = 100$ and 1000 years in the individual regions in Slovakia, with various assumptions concerning the threshold S_0 and the equivalent number of station-years N_0. $CI_{0.05}$ ($CI_{0.95}$) is the 5% (95%) confidence limit of the estimates Q_T, $\Delta CI = CI_{0.95} - CI_{0.05}$.

Return period	Region	S_0	N_0 (years)	Information on ungauged extremes								No ungauged extremes			
				Perfectly known				Imperfectly known							
				Q_T	$CI_{0.05}$	$CI_{0.95}$	$\Delta CI/Q_T$ (%)	Q_T	$CI_{0.05}$	$CI_{0.95}$	$\Delta CI/Q_T$ (%)	Q_T	$CI_{0.05}$	$CI_{0.95}$	$\Delta CI/Q_T$ (%)
$T = 100$ years	Flysch East 1	750	5000	127.8	100.6	166.6	51.6	129.0	99.7	166.3	51.6	135.4	97.2	210.5	83.7
		750	1000	151.4	110.0	216.5	70.3	148.6	109.5	217.6	72.7	135.4	97.2	210.5	83.7
		750	20,000	109.0	90.4	130.6	36.9	108.6	90.2	130.5	37.1	135.4	97.2	210.5	83.7
	Tatras	550	10,000	75.5	62.8	93.1	40.1	74.9	62.3	92.9	40.9	67.0	55.6	88.0	48.4
		550	2500	78.2	63.8	100.8	47.3					67.0	55.6	88.0	48.4
		550	40,000	71.1	60.4	82.9	31.6					67.0	55.6	88.0	*48.4*
$T = 1000$ years	Flysch East 1	750	5000	367.9	244.1	562.4	86.5	371.7	241.4	562.8	86.5	404.7	230.2	834.7	149.4
		750	1000	489.3	283.4	875.5	121.0	473.9	283.9	881.9	126.2	404.7	230.2	834.7	149.4
		750	20,000	278.5	203.4	368.9	59.4	276.4	203.4	369.0	59.9	404.7	230.2	834.7	149.4
	Tatras	550	10,000	168.8	123.3	237.6	67.7	166.5	121.5	236.4	69.0	136.9	98.9	216.4	85.8
		550	2500	179.7	126.9	272.7	81.1					136.9	98.9	216.4	85.8
		550	40,000	151.4	115.6	193.7	51.6					136.9	98.9	216.4	85.8

questionable. If it is divided or multiplied by four in the Flysch case, the inference outcomes are completely modified and the whole credibility interval obtained without the ungauged extreme is covered by this two tested cases. The possible added value of the extreme peak discharge is cancelled by the uncertainty affecting N_0. Even if the initial guess for N_0 is considered as a higher possible value according to the limited area of the watershed where this extreme occurred and to the fact that other extremes may have been missed, the conclusion remains the same.

The Tatras case provides slightly different conclusions. Since the available extreme is less concordant with the gauged data set, its inclusion leads to a displacement of the whole adjustment including the credibility intervals. If the initial guess for N_0 is considered as an upper possible limit, than the quantiles estimated based on the gauged data only appear as a little under-estimated.

In both cases, the inclusion of the regional extreme flood does not lead to a major modification of the estimated quantiles or of the credibility intervals. This is apparently due to the limited information available on extreme floods and to the uncertainties in the estimation of N_0. The French applications, based on richer ungauged samples, should be much more favourable in that perspective.

Ardèche region in France

The Ardèche data set is composed of 168 annual peak discharge values measured at five different gauging stations and the maximum estimated discharges during the last 50 years at 38 ungauged sites. The results obtained on the Ardèche are presented in Fig. 6 and Table 5. A specific procedure has been used to obtain correct estimates of the return periods of the sample of ungauged extremes to draw the figures. It is explained in "Plotting position for extreme ungauged floods". An adjustment obtained with the largest gauged data set of the region has been added on Fig. 6. The comparison of Fig. 6a–c gives sense to the figures reported in the Table 5. The reduction of the credibility intervals obtained, including the ungauged extremes (comparison between Fig. 6b and c), is of the same magnitude as the reduction obtained when conducting a regional rather than an at site flood frequency analysis (comparison between Fig. 6a and b).

Looking at the French ungauged set, it appears that a significant number of these extremes have been observed on the same date, induced by very intense rainfall events with large spatial extent that occur from time to time in the French Mediterranean region like the 1958 and the 2002 events (Gaume et al., 2009). The data cannot be considered as independent and as a consequence, the information content of the ungauged set may have been overestimated and the uncertainties under-estimated as already shown by Stedinger (1983). Therefore, the same analysis has been conducted leaving only the top 9 values in the ungauged set – they do not anymore occur on the same dates – and setting a constant non-exceedance threshold for the 29 remaining ungauged watersheds. Surprisingly, this does not modify significantly the results: the credibility intervals that are only marginally affected (Table 5: 'Ardèche (9 max)'), sign that dependence did not affect significantly the results and that the information content of the ungauged set has not been significantly affected by this simplification. Even in the extreme case where only the maximum ungauged discharge value is considered and its value taken as non-exceedance threshold for the other 37 watersheds, the increase of the credibility bounds remains limited (Table 5: 'Ardèche (1 max)'). This last test is very similar to the Slovak case study except that N_0 is known. But hence, the effect of including the regional ungauged extreme is more pronounced in the Ardèche case even if N_0 is much lower: 1900 station-years compared to 5000 and 10,000 in the Slovak case. The explanation for this difference seems to lie in the magnitude or more precisely in the return period of the selected threshold. Increasing the length N_0 of the ungauged series leads also to increase the threshold if only the maximum discharge is considered. Both elements have an opposite impact on the likelihood function (Eqs. (6) and (7)) – especially on the fourth term. The increase of N_0 increases the information content of the ungauged series, but the increase of the threshold reduces this information content. The second effect appears to dominate. A denser inventory of flood extremes over a limited area appears more informative than a looser one over a large region. This has, however, a limit. Since the ungauged extremes are added to gauged series, the resulting estimated ungauged peak discharges and the corresponding threshold must be of the order of magnitude or higher than the highest gauged discharges. As a conclusion, to represent really an added value to gauged data in a regional flood frequency analysis, the inventory of ungauged large floods should be sufficiently dense. A systematic inventory of extreme discharges in a sub-area of the defined homogeneous region should be preferred to a loose collation of the few most extreme flood events over the whole region.

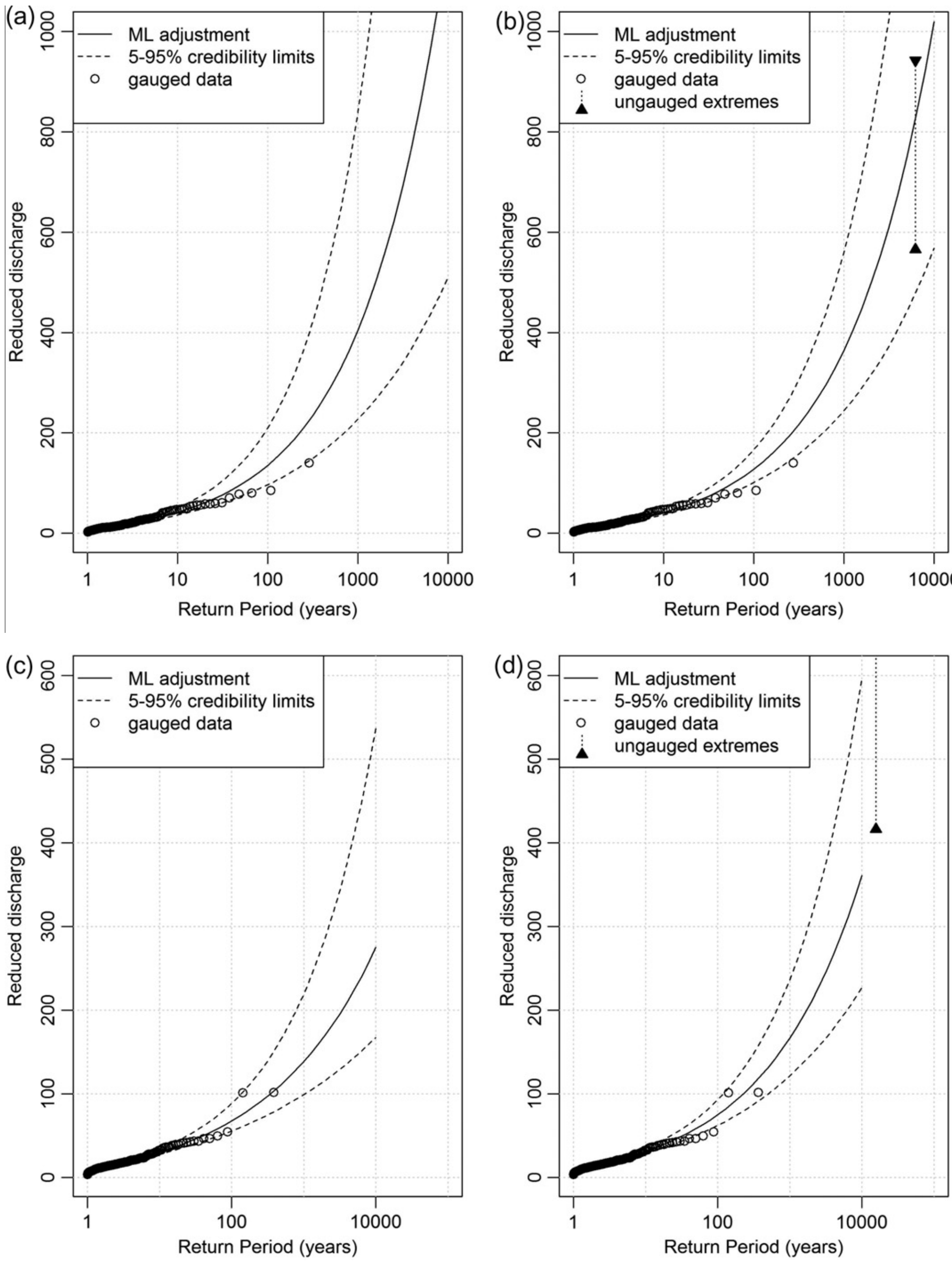

Fig. 5. Fitted GEV distributions: (a) Flysch East 1 with the gauged data only, (b) Flysch East 1 with the Malá Svinka flood included with uncertainties ($S_0 = 750$, $N_0 = 5000$), (c) Tatras with the gauged data only, and (d) Tatras with the Štrbský Creek flood included ($S_0 = 550$, $N_0 = 10{,}000$).

Table 5
Estimation of the reduced discharge quantiles Q_T and their confidence intervals corresponding to the return periods $T = 100$ and 1000 years in the Ardèche region with various numbers of ungauged extremes considered. $CI_{0.05}$ ($CI_{0.95}$) is the 5% (95%) confidence limit of the estimates Q_T, $\Delta CI = CI_{0.95} - CI_{0.05}$.

Return period	Region	S_0	N_0 (years)	Information on ungauged extremes								No ungauged extremes			
				Perfectly known				Imperfectly known							
				Q_T	$CI_{0.05}$	$CI_{0.95}$	$\Delta CI/Q_T$ (%)	Q_T	$CI_{0.05}$	$CI_{0.95}$	$\Delta CI/Q_T$ (%)	Q_T	$CI_{0.05}$	$CI_{0.95}$	$\Delta CI/Q_T$ (%)
$T = 100$ years	Ardèche (38 max)	–	1900	26.1	24.4	28.4	15.3	27.2	25.3	29.5	15.4	27.0	22.6	36.6	51.9
	Ardèche (9 max)	30	1900	26.2	24.1	28.8	17.9					27.0	22.6	36.6	51.9
	Ardèche (1 max)	40	1900	24.6	22.1	27.8	23.2	24.3	22.0	27.5	22.6	27.0	22.6	36.6	51.9
$T = 1000$ years	Ardèche (38 max)	–	1900	43.0	38.1	50.6	29.1	42.3	37.1	49.2	28.6	44.4	33.2	75.1	94.4
	Ardèche (9 max)	30	1900	42.4	37.1	49.8	30.0					44.4	33.2	75.1	94.4
	Ardèche (1 max)	40	1900	37.8	32.2	46.4	37.5	37.2	31.6	45.5	37.4	44.4	33.2	75.1	94.4

Gard region in France

This last example gives another very different illustration of the possible usefulness of ungauged extremes in flood frequency analyses. The Gard data set counts 231 annual peak discharge values measured at six different gauging stations and the maximum estimated discharges during the last 50 years at 58 ungauged sites. The same index flood relation applies in the Gard and in the Ardèche regions, but as already revealed by the HW test, the regional flood frequency distribution adjusted on the gauged data in the Gard and in the Ardèche are significantly different. The coefficient of variation of the data series is higher in the Ardèche, leading to higher values of 100-year and 1000-year quantile estimates (Fig. 7b and Table 6). One detail is nevertheless striking in the case of the Gard. The estimated 1958 peak discharge at the Mialet gauging station, not reported in the gauged data series that begin in 1960 since the previous gauging station had been destroyed by this flood but still existing and which we decided to add, appears discordant with the adjusted regional distribution. Its empirical return period does not exceed a few hundred years if the whole regional gauged set is considered but the best estimate of its return period exceeds 100,000 years according to the adjusted regional distribution (Fig. 7b). Its observation within the short period of time covered by the gauged series is very unlikely, but this isolated observation does not influence significantly the regional adjustment (Fig. 7b). The ungauged extremes confirm the relatively low return period of the Mialet 1958 flood and modify completely the regional adjustment (Table 6 and Fig. 7): the 1000-year best estimate is multiplied by a factor exceeding 2. The same figures as in the Ardèche example are observed: (1) the uncertainties in the estimated ungauged extremes have a limited effect on the adjustments and credibility intervals, and (2) to a certain extent, the ungauged set can be reduced with a limited impact. A larger set of ungauged extreme values is nevertheless necessary due to the discordance between the series of ungauged discharges and the statistical distribution adjusted on the gauged data (Fig. 7b). In this final example, the initial statistical adjustment based on gauged series is invalidated provided that a sufficient number of ungauged extreme values are available, and this despite the uncertainty in the ungauged values. Moreover, the variation of the adjustments with the length of the available data set is also a sign that the GEV distribution is not suited to the available data set. This is also confirmed by the fact that the majority of the discharges are not comprised in the 90% credibility limits of the adjusted distribution in Fig. 7c. The shape of the peak discharge distribution in the Gard seems more complex. On the semi-log scale (Fig. 7), it has the *S* shape predicted by some theoretical works on peak discharge distributions (Gaume, 2006; Sivapalan et al., 1990). But the discussion of this last point is beyond the scope of this paper.

Discussion

The results of these first tests aiming at including and evaluating the added value of ungauged extremes in a regional flood frequency analyses are promising.

(i) In the test regions, the various examples did not reveal an obvious inadequacy of the simple scaling hypothesis or of the proposed index flood relation on which the suggested approach is based. As an example, the two ungauged extreme floods in Slovakia occurred on watersheds with limited area: Malá Svinka in the Flysch East 1 region (6.5 km^2) and Štrbský Creek in the Tatras region (2.5 km^2). Their incorporation necessitated a large extrapolation of the established index flood relations towards small catchment areas, but the computed reduced discharges did not appear discordant with the gauged discharges of the same region. The Ardèche case is even more convincing. The final adjustment mostly controlled by extreme discharges estimated on ungauged watersheds of few hundred km^2, implies a non-reduced 100-year discharge at the St. Martin gauge (2240 km^2) comprised between 5300 and 6400 m^3/s. This is consistent with the 6400 m^3/s estimate based on historical floods at the same location (Naulet et al., 2005) and also with paleoflood estimates (Sheffer et al., 2003): about 5000 m^3/s for the Ardèche at Vallon Pont d'Arc (1930 km^2) where the regional approach estimates the 100-year flood between 4800 and 5800 m^3/s.

(ii) Even relatively high uncertainties in the estimated ungauged discharges have a limited impact on the final result and especially on the credibility intervals. This surprising and encouraging result – i.e., accurate ungauged extremes are not absolutely necessary – is explained by the fact that the dominant information in the ungauged series is the number of non-exceedance of perception thresholds. It is therefore essential to be able to determine accurately the various threshold values S_0 or S_j and the actual length of the ungauged series. For this reason, comprehensive and dense inventories of ungauged extremes with a controlled and predefined number of target watersheds not selected on the basis of the magnitude of the past floods (French example), should be preferred to loose collations of isolated extreme values with unknown equivalent coverage. For a given series, if there is an uncertainty on the threshold values, the highest guesses should be selected to obtain the pessimistic result concerning the credibility bounds.

(iii) The peak discharge estimation errors are considered independent. Systematic biases have a much greater impact on the computed credibility limits (Neppel et al., in press). But if a systematic bias affects the estimations of extreme

Table 6
Estimation of the reduced discharge quantiles Q_T and their confidence intervals corresponding to the return periods $T = 100$ and 1000 years in the Gard region with various numbers of ungauged extremes considered. $CI_{0.05}$ ($CI_{0.95}$) is the 5% (95%) confidence limit of the estimates Q_T, $\Delta CI = CI_{0.95} - CI_{0.05}$.

Return period	Region	S_0	N_0 (years)	Information on ungauged extremes								No ungauged extremes			
				Perfectly known				Imperfectly known							
				Q_T	$CI_{0.05}$	$CI_{0.95}$	$\Delta CI/Q_T$ (%)	Q_T	$CI_{0.05}$	$CI_{0.95}$	$\Delta CI/Q_T$ (%)	Q_T	$CI_{0.05}$	$CI_{0.95}$	$\Delta CI/Q_T$ (%)
$T = 100$ years	Gard (58 max)	–	2900	32.3	29.6	35.5	18.3	34.3	31.3	38.1	19.8	19.0	17.2	22.3	26.8
	Gard (15 max)	50	2900	31.8	28.3	35.9	23.9	31.5	28.0	35.8	24.8	19.0	17.2	22.3	26.8
	Gard (2 max)	65	2900	22.1	19.6	25.8	28.1	21.5	19.1	25.2	28.4	19.0	17.2	22.3	26.8
$T = 1000$ years	Gard (58 max)	–	2900	69.4	60.1	81.7	31.2	73.1	62.4	87.6	34.5	28.4	24.5	36.6	42.6
	Gard (15 max)	50	2900	66.3	55.2	80.2	37.7	64.8	53.8	79.7	40.1	28.4	24.5	36.6	42.6
	Gard (2 max)	65	2900	36.7	30.5	46.9	44.7	37.2	31.6	45.5	37.4	28.4	24.5	36.6	42.6

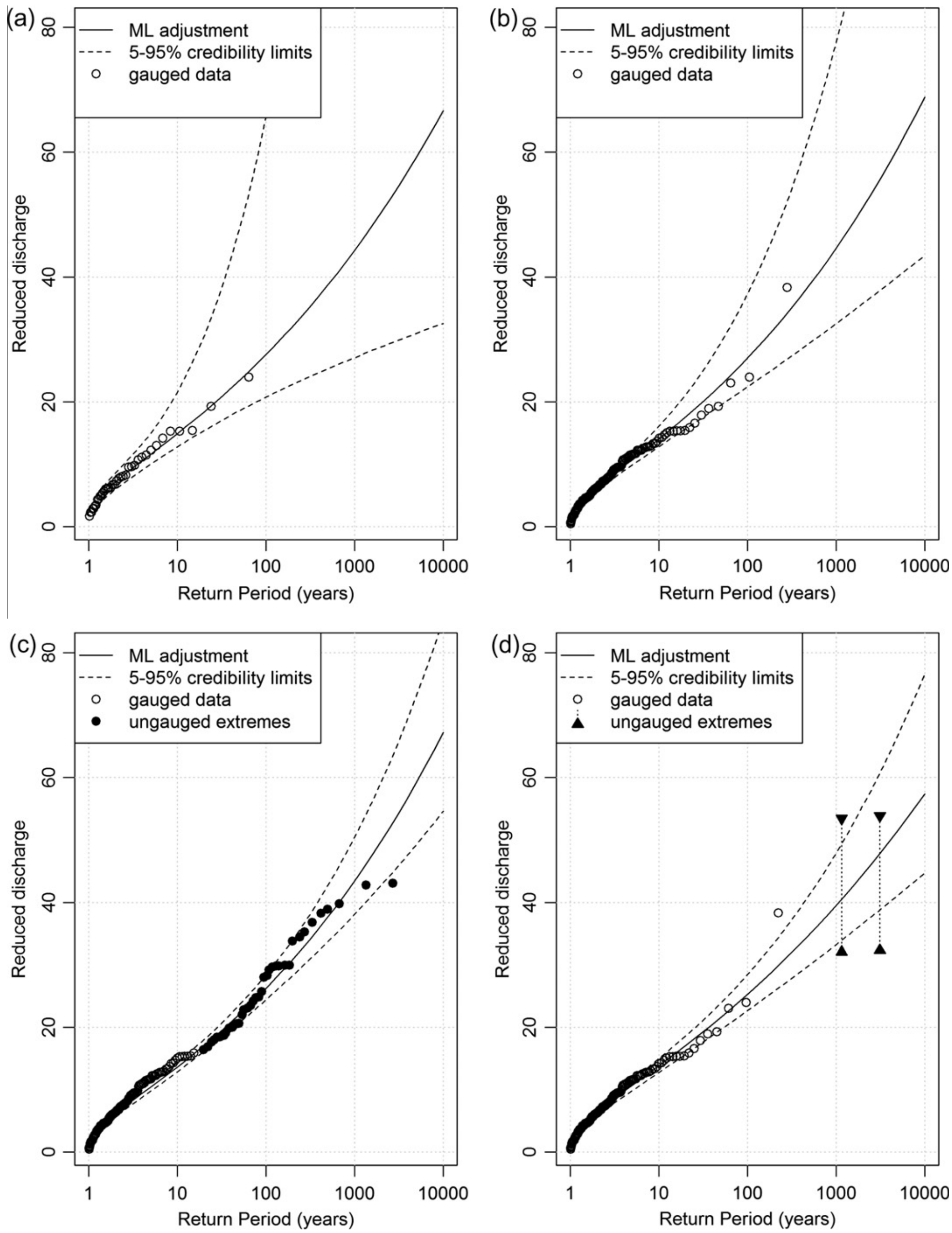

Fig. 6. Fitted GEV distributions in the Ardèche region: (a) data of the St. Martin gauging station on the Ardèche River, (b) regional gauged data set, (c) including the whole set of ungauged extremes, and (d) including the two highest ungauged extremes with uncertainty.

discharges it will affect any new estimation and will also be partly compensated when the computed flood quantiles will be used to retrieve local inundation characteristics. Biases may exist, but they are not detectable in regional extreme flood analyses and can therefore not be taken into account in the computation of credibility limits. The estimated discharge quantiles should be considered as the best guesses given the commonly accepted extreme peak discharge estimation practices.

The proposed method deserves nevertheless further improvements and tests:

(i) The index flood relation is uncertain and this has not been explicitly taken into account for the computation of the credibility intervals. The computed intervals, based on hypotheses that could not be totally verified – simple scaling, index flood relation – are certainly a little optimistic. The likelihood formulation enables the integration and calibration of both the index flood relation and the peak discharge distribution at the same time. Relations between the average of the tested distribution type (GEV here) and the area and hence between the parameters of the distribution and the area or even other possible explanatory factors could be directly introduced in the likelihood formulation

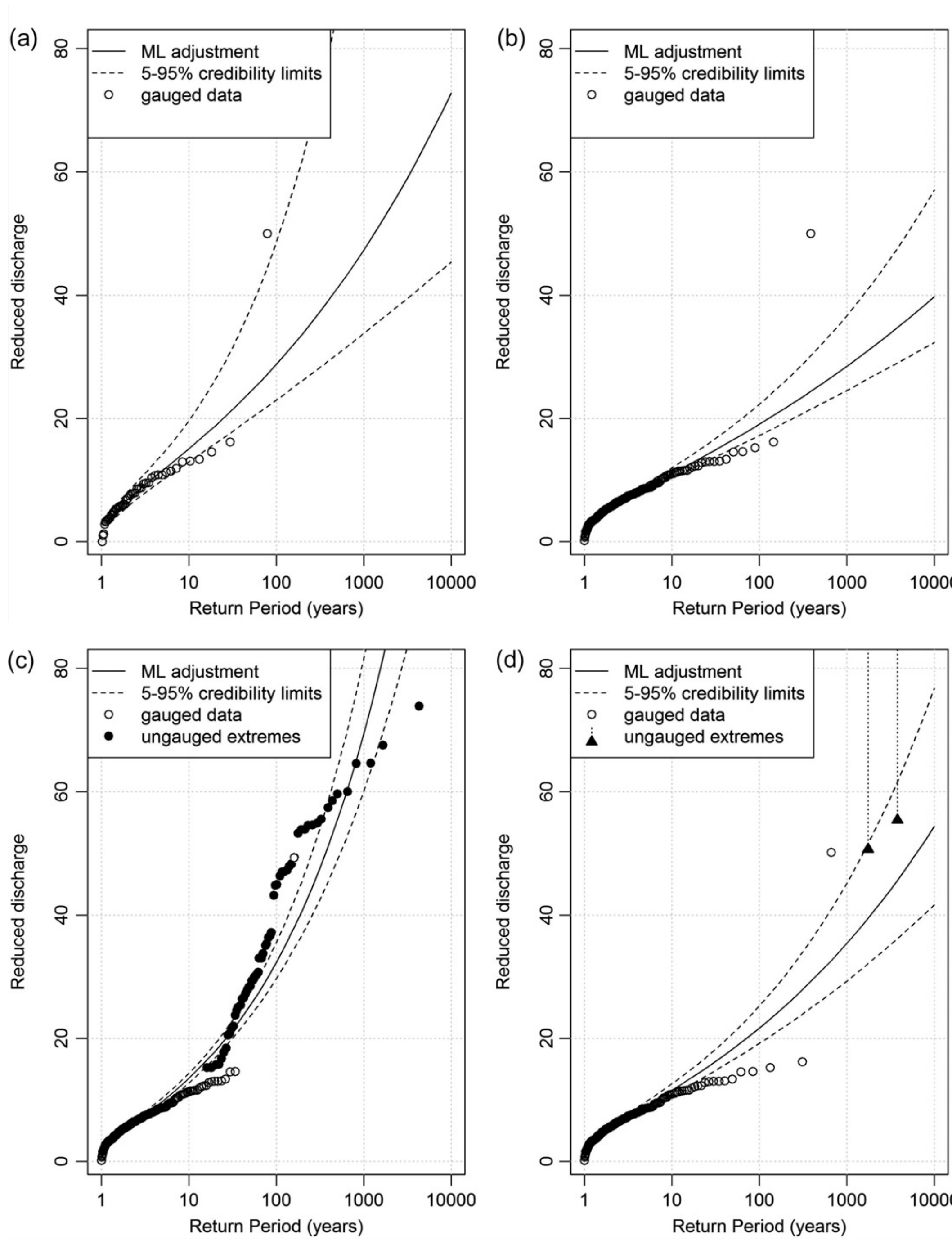

Fig. 7. Fitted GEV distributions in the Gard region: (a) data of the Mialet gauging station on the Gardon River, (b) regional gauged data set, (c) including the whole set of ungauged extremes, and (d) including the two highest ungauged extremes with uncertainty.

(Eqs. (6) and (7)). In the same line of thoughts, the likelihood formulation is sufficiently flexible to add any wished complexity to the proposed model. The evolution of the variance of the discharge distributions with the area could also be tested for instance.

(ii) The proposed approach is based on an assumption of homogeneity of the regions, in terms of both growth curves and index flood relationship, which is a more restrictive requirement then the assumption of common regional flood frequency studies. An in-depth sensitivity analysis would be helpful to determine to which extent the proposed regional analysis procedure resists to heterogeneities which necessarily exist in actual series of data.

(iii) The likelihood formulation is based on the hypothesis of independence between the records. Previous works (Castellarin et al., 2005) and some of the results presented here – especially on the French data set – seem to indicate that inter-site dependencies have a limited effect on the outcomes of the inference procedure, especially on the computed credibility limits. Explicitly accounting for dependence in statistical inference procedures is a complex and partly unsolved problem that clearly deserves additional research efforts.

(iv) Finally, it would be useful to develop statistical methods to test a posteriori if the proposed statistical model is compatible with the available data set. Such a test would have been useful in the Gard example to eliminate the GEV distribution type on the basis of objective criteria.

Conclusions

In this paper a method to incorporate extreme peak discharges evaluated on ungauged watersheds in a regional flood frequency analysis has been proposed. The method is based on a Bayesian inference where the likelihood function is built to properly handle this specific information. The application of the method to two case studies where flash floods were surveyed, in Slovakia and South of France shows the potential of using these unconventional data in such a framework. A particularly satisfying result is the relative robustness of the outcomes when the uncertainties in the estimated extreme discharges are taken into account.

Some hypotheses that are necessary to conduct the computations may be questionable, as for instance the computation of the equivalent number of years represented by the extreme flash floods in Slovakia. One of the main qualities of the proposed method is that most of the hypotheses are explicitly posed and can be discussed and modified. Sensitivity analyses can be conducted to test the influence of these hypotheses on the results as shown in this paper.

As it has been the case with historic and paleoflood information, the possibility to incorporate ungauged extremes opens new perspectives in flood frequency studies, rising new evidences and questions on flood peak distribution tails and low return period flood quantiles.

Acknowledgements

The work presented in the paper was supported by the European Commission's Sixth Framework project HYDRATE, Contract GOCE 037024.

Application of the Wilcoxon–Mann–Whitney test

In the paper, the Wicoxon–Mann–Withney (WMW; Wilcoxon, 1945; Mann and Whitney, 1947) test is adopted in the following manner:

- A reference series of annual peak discharges $\mathbf{u}_{\mathrm{ref}}$ is chosen. It is supposed that the final results does not depend on the selection of a particular reference site; however, sites with longer records are preferred.
- By means of the WMW test, a pair-wise comparison of the series of the ith site $\mathbf{u}_i$, $i = 1, \dots, M$ (where M is the total number of sites) to the reference one $\mathbf{u}_{\mathrm{ref}}$ is carried out.
- For the ith site, the WMW test statistic is computed as follows:

$$h = f_{\mathrm{WMW}}(\mu^{(i)}\mathbf{u}_{\mathrm{ref}}, \mathbf{u}_i) \tag{A1}$$

where f_{WMW} denotes the test itself, h is the test statistic (if h = 'true', the null hypothesis is rejected at the confidence level α, here $\alpha = 5$ otherwise, if h = 'false', the null hypothesis cannot be rejected), and $\mu^{(i)}$ is a properly selected constant, referred herein as scaling coefficient.

- The testing procedure is aimed at finding scaling coefficients $\mu^{(i)}$, $i = 1, \dots, M$, which results in h = 'false' value in Eq. (A1). The possible range of scaling coefficients $\mu^{(i)}$ for each site i may be characterized by the interval $I^{(i)} = (\mu^{(i)}_{\min}, \mu^{(i)}_{\max})$. It means that if any $\mu^{(i)} \in I^{(i)}$ is chosen, there is a high probability that the two samples $\mathbf{u}_i$ and $\mu^{(i)}\mathbf{u}_{\mathrm{ref}}$ come from the same distribution. $\mu^{(i)}_{\mathrm{opt}}$ is a special case of the $\mu^{(i)}$ values: it is the scaling coefficient, for which the highest p value of the WMW test is obtained.
- We interpret the modified WMW test in a graphical way where $\mu^{(i)}_{\mathrm{opt}}$ is plotted against the watershed area $A^{(i)}$ ($i = 1, \dots, M$) in log–log scale. At the same time, the intervals $I^{(i)} = (\mu^{(i)}_{\min}, \mu^{(i)}_{\max})$ are displayed as whiskers, similarly to the way that error bars are usually presented.
- The goal of the graphical interpretation of the modified WMW test is to match as many whiskers by a single straight line as possible. The position of the straight line is up to the subjective decision of the analyst. Nevertheless, we declare that the group of sites that is selected by the straight line is associated with a common index flood relationship (Eq. (3)). The exponent β of this relationship is then given by the slope of the straight line.

The Hosking and Wallis regional homogeneity test

The test is based on the weighted standard deviation V of the at-site sample L_{CV} (coefficient of L-variation):

$$V = \sqrt{\frac{\sum_{i=1}^{M} n_i \left(L_{CV}^{(i)} - L_{CV}^{R}\right)^2}{\sum_{i=1}^{M} n_i}} \tag{B1}$$

where L_{CV}^{R} is the weighted regional average of the sample L_{CV} (the weights are proportional to the record length n_i):

$$L_{CV}^{R} = \frac{\sum_{i=1}^{M} n_i L_{CV}^{(i)}}{\sum_{i=1}^{M} n_i} \tag{B2}$$

and M is the total number of sites in the region. The heterogeneity measure is then

$$H_1 = \frac{V - \mu_V}{\sigma_V} \tag{B3}$$

where μ_V and σ_V (which are the mean and the standard deviation of V) are obtained by Monte Carlo simulations (Hosking and Wallis, 1993).

The critical values of the H_1 measure, based on its initial definition (Hosking and Wallis, 1993) are determined as follows: the region is acceptably homogeneous if $H_1 < 1$, possibly heterogeneous if $1 \leqslant H_1 < 2$, and definitely heterogeneous if $H_1 \geqslant 2$. Nevertheless, in this paper we follow the modified test criteria (Wallis et al., 2007), which classify regions with $H_1 < 2$ as acceptably homogeneous and regions with $H_1 \geqslant 3$ as definitively heterogeneous ("Homogeneity of growth curves").

Plotting position for extreme ungauged floods

Let us consider the set of k sites where the same random variable Y is observed. The series of the highest peak discharges at each site is available $\{y_1, \dots, y_j, \dots, y_k\}$ as well as the series of observation durations for each site $\{n_1, \dots, n_j, \dots, n_k\}$. What empirical return period should be attributed to each y_j? The answer is not straightforward and the estimators of these return periods have been obtained through Monte Carlo simulations. But the answer does not depend on the statistical distribution of Y.

The selected procedure is based on the repeated simulation of sets of k series of dimensions $\{n_1, \dots, n_j, \dots, n_k\}$. It is not the values of the random variable Y which are drawn but directly the values of their cumulative distribution function $F(\cdot)$, that are uniformly distributed in the interval $[0, 1]$ for every statistical distribution of Y. The algorithm is the following:

(i) $M_S * k$ vectors of uniform random variables are drawn (where M_S is the number of simulations),
(ii) The maximum value of each vector is selected to create M_S vectors of variables $\{F(y_1), \dots, F(y_j), \dots, F(y_k)\}$ which are ranked,

(iii) The value of each $F(y_j)$, j corresponding to the rank, is averaged over the M_S vectors to obtain a vector $\{\overline{F_1},\cdots,\overline{F_j},\cdots,\overline{F_k}\}$. Estimates of the average values of $F(\cdot)$ for the largest maximum discharge, the second largest maximum,... are obtained through this procedure.

(iv) The vector of return periods $\{T_1,\ldots,T_j,\ldots,T_k\}$ is deduced from $\{\overline{F_1},\cdots,\overline{F_j},\cdots,\overline{F_k}\}$ using the standard relation $T_j = 1/(1-\overline{F_j})$.

References

Alcoverro, J., Corominas, J., Gómez, M., 1999. The Barranco de Arás flood of 7 August 1996 (Biescas, Central Pyrenees, Spain). Engineering Geology 51 (4), 237–255.

Borga, M., Gaume, E., Creutin, D., Marchi, L., 2008. Surveying flash flood response: gauging the ungauged extremes. Hydrological Processes 22 (18), 3883–3885. doi:10.1002/hyp.7111.

Castellarin, A., 2007. Probabilistic envelope curves for design flood estimation at ungauged sites. Water Resources Research 43, W04406. doi:10.1029/2005WR004384.

Castellarin, A., Vogel, R.M., Matalas, N.C., 2005. Probabilistic behavior of regional envelope curve. Water Resources Research 41, W06018. doi:10.1029/2004WR003042.

Castellarin, A., Burn, D.H., Brath, A., 2008. Homogeneity testing: how homogeneous do heterogeneous cross-correlated regions seem? Journal of Hydrology 360, 67–76.

Costa, J.E., 1987. A comparison of the largest rainfall-runoff floods in the United States with those of the People's Republic of China and the world. Journal of Hydrology 69, 101–115.

Costa, J.E., Jarrett, R.D., 2008. An evaluation of selected extraordinary floods in the United States reported by the U.S. geological survey and implications for future advancement of flood science. Scientific investigations report 2008-5164, US Geological Survey, Reston, Virginia, USA.

Dalrymple, T., 1960. Flood frequency analyses. Water supply paper 1543-A. US Geological Survey, Reston, Virginia, USA.

Delrieu, G., Ducrocq, V., Gaume, E., Nicol, J., Payrastre, O., Yates, E., Andrieu, H., Ayral, P.-A., Bouvier, C., Creutin, J.-D., Livet, M., Anquetin, S., Lang, M., Neppel, L., Obled, C., Parent-du-Châtelet, J., Saulnier, G.-M., Walpersdorf, A., Wobrock, W., 2005. The catastrophic flash-flood event of 8–9 September 2002 in the Gard region, France: a first case study for the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory. Journal of Hydrometeorology 6 (1), 34–52.

El-Adlouni, S., Ouarda, T.B.M.J., 2009. Joint Bayesian model selection and parameter estimation of the generalized extreme value model with covariates using birth-death Markov chain Monte Carlo. Water Resources Research 45, W06403. doi:10.1029/2007WR006427.

Fusán, O., Kodym, O., Matějka, A., Urbánek, L., 1980. Geology. In: Mazúr, E. (Ed.), Atlas of the Slovak Socialist Republic. Slovak Academy of Sciences, pp. 18–19 (in Slovak).

Gaume, E., 2006. On the asymptotic behavior of flood peak distributions. Hydrology and Earth System Sciences 10 (2), 233–243.

Gaume, E., Borga, M., 2008. Post flood field investigations after major flash floods: proposal of a methodology and illustrations. Journal of Flood Risk Management 1 (4), 175–189. doi:10.1111/j.1753-318X.2008.00023.

Gaume, E., Livet, M., Desbordes, M., Villeneuve, J.-P., 2003. Hydrological analysis of the Avene river (France) extraordinary flood, 6 and 7 October 1997. Physics and Chemistry of the Earth 28 (6–7), 263–267.

Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaškovičová, L., Blöschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutroulis, A., Marchi, L., Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., Stancalie, G., Szolgay, J., Tsanis, J., Velasco, D., Viglione, A., 2009. A collation of data on European flash floods. Journal of Hydrology 367, 70–78. doi:10.1016/j.jhydrol.2008.12.028.

Hassan, A.E., Bekhit, H.M., Chapman, J.B., 2009. Using Markov Chain Monte Carlo to quantify parameter uncertainty and its effect on predictions of a groundwater flow model. Environmental Modelling and Software 24 (6), 749–763.

Herschy, R.W., 2005. The world's maximum observed floods. Flow Measurement and Instrumentation 13, 231–235.

Hosking, J.R.M., Wallis, J.R., 1986. Paleoflood hydrology and flood frequency analysis. Water Resources Research 22 (4), 543–550.

Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. Water Resources Research 29 (2), 271–281.

Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis: an Approach Based on L-Moments. Cambridge University Press, Cambridge. 224pp.

House, P.K., Pearthree, P.A., 1995. A geomorphologic and hydrologic evaluation of an extraordinary flood discharge estimate: Bronco Creek, Arizona. Water Resources Research 31 (12), 3059–3073.

House, P.K., Webb, R.H., Baker, V.R., Levish, D.R., 2002. Ancient floods, modern hazards, principles and applications of paleoflood hydrology. American Geophysical Union, Water Science and Application, Washington, USA 5.

Jarrett, R.D., 1987. Errors in slope-area computations of peak discharges in mountain streams. Journal of Hydrology 96, 53–67.

Jarrett, R.D., 1990. Hydrologic and hydraulic research in mountain rivers. Water Resources Bulletin 26 (3), 419–429.

Jarvis, C.S., 1924. Flood flow characteristics. Proceedings of the American Society of Civil Engineers, 1545–1581.

Jenkinson, A.F., 1955. The frequency distribution of the annual maximum (minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorology Society 18, 158–171.

Kanso, A., Gromaire, M.-C., Gaume, E., Tassin, B., Chebbo, G., 2003. Bayesian approach for the calibration of models: application to an urban stormwater pollution model. Water Science and Technology 47 (4), 77–84.

Kendall, M., Stuart, A., 1979. The Advanced Theory of Statistics. Charles Griffin & Company Limited, London.

Kohnová, S., 1997. Regional analysis of specific discharge maxima on small catchments in Slovakia. PhD thesis, Slovak University of Technology in Bratislava (in Slovak).

Koutsoyiannis, D., 2004. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. Hydrological Sciences Journal 49 (4), 575–590.

Kuczera, G., 1999. Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. Water Resources Research 35 (5), 1551–1557.

Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. Journal of Hydrology 211, 69–85.

Lee, K.S., Kim, S.U., 2008. Identification of uncertainty in low flow frequency analysis using Bayesian MCMC method. Hydrological Processes 22 (12), 1949–1964.

Leese, M.N., 1973. Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series. Water Resources Research 9 (6), 1534–1542.

Mailhot, A., Gaume, E., Villeneuve, J.-P., 1997. Uncertainty analysis of calibrated parameter values of an urban storm water quality model using metropolis Monte Carlo algorithm. Water Science and Technology 36 (5), 141–148.

Majerčáková, O., Makeľ, M., Šťastný, P., Kupčo, M., Rončák, P., 2004. Selected flash floods in the Slovak Republic. GWP report. Slovak Hydrometeorological Institute, Bratislava. 34pp.

Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics 18, 50–60.

Mantovan, P., Todini, E., 2006. Hydrological forecasting uncertainty assessment: incoherence of the glue methodology. Journal of Hydrology 330, 368–381.

Mazúr, E., Lukniš, M., 1986. Geomorphological delineation of SSR and CSSR. Part Slovakia. Slovenská kartografia. Bratislava (in Slovak).

Merz, R., Blöschl, G., 2003. A process typology of regional floods. Water Resources Research 39 (12), 1340. doi:10.1029/2002WR001952.

Mimikou, M., 1984. Envelope curves for extreme flood events in north-western and western Greece. Journal of Hydrology 67, 55–66.

Naulet, R., Lang, M., Ouarda, T.B.M.J., Coeur, D., Bobée, B., Recking, A., Moussay, D., 2005. Flood frequency analysis on the Ardèche River using French documentary sources from the two last centuries. Journal of Hydrology 313, 58–78.

Neppel, L., Renard, B., Lang, M., Coeur, D., Gaume, E., Jacob, N., Payrastre, O., Pobanz, K., Vinet, F., in press. Flood frequency analysis using historical data: accounting for random and systematic errors. Hydrological Sciences Journal.

Newinger, O., 2007. Inventaire et analyse des crues éclair du Sud de la France de 1950 à 2007. Master thesis, Université Pierre et Marie Curie, Paris.

O'Connell, D.R.H., Ostenaa, D.A., Levish, D.R., Klinger, R.E., 2002. Bayesian flood frequency analysis with paleohydrologic bound data. Water Resources Research 38 (5), 1058. doi:10.1029/2000WR000028.

Pardé, M., 1961. Sur la puissance des crues en diverses parties du monde. Geographica 8, 1–293.

Parent, E., Bernier, J., 2002. Bayesian POT modeling for historical data. Journal of Hydrology 274, 95–108.

Payrastre, O., 2005. Faisabilité et utilité du recueil de données historiques pour l'étude des crues extrêmes de petits cours d'eau. Etude du cas de quatre bassins versants affluents de l'Aude. PhD thesis, Ecole Nationale des Ponts et Chaussées, Paris, France.

Pekárová, P., 2009. Flood regime of rivers in the Danube River basin. Journal of Hydrology and Hydromechanics 57 (2), 142–150 (in Slovak).

Reis, D.S., Stedinger, J.R., 2005. Bayesian MCMC flood frequency analysis with historical information. Journal of Hydrology 313 (1–2), 97–116.

Renard, B., Garreta, V., Lang, M., 2006. An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. Water Resources Research 42 (12), W12422. doi:10.1029/2005WR004591.

Robert, C.P., Casella, G., 2004. Monte Carlo statistical methods. Springer, New York. 645pp.

Robson, A.J., Reed, D.W., 1999. Statistical procedures for flood frequency estimation. Flood estimation handbook, vol. 3. Statistical procedures for flood frequency estimation. Institute of Hydrology, Wallingford.

Rodier, J.A., Roche, M., 1984. World Catalogue of Maximum Observed Floods. IASH Publication 143, IASH Press.

Sheffer, N.A., Enzel, Y., Benito, G., Grodek, T., Poart, N., Lang, M., Naulet, R., Coeur, D., 2003. Paleofloods and historical floods of the Ardèche River, France. Water Resources Research 39 (12), 1379. doi:10.1029/2003WR002468.

Sivapalan, M., Wood, E., Beven, K., 1990. On hydrologic similarity 3. A dimensionless flood frequency model using a generalized geomorphologic unit hydrograph and partial area runoff generation. Water Resources Research 26, 43–58.

Solín, L., 2008. Analysis of floods occurrence in Slovakia in the period 1996–2006. Journal of Hydrology and Hydromechanics 56 (2), 95–107 (in Slovak).

Šťastný, P., Majerčáková, O., 2003. Reconstruction of the flood in Štrba on July 2001. In: Proceedings of 'Hydrology at the Beginning of the 21st Century: Vision and

Reality', Smolenice, May 5–7, 2003. Hydrological Institute SAS, Slovak Committee for Hydrology, Slovak IGBP Committee (in Slovak).
Stedinger, J.R., 1983. Estimating a regional flood frequency distribution. Water Resources Research 19 (2), 503–510.
Stedinger, J.R., Cohn, T.A., 1986. Flood frequency analysis with historical and paleoflood information. Water Resources Research 22 (5), 785–793.
Svoboda, A., Pekárová, P., 1998. The catastrophic flood of July 1998 in the Malá Svinka catchment – its simulation. Journal of Hydrology and Hydromechanics 46 (6), 356–365 (in Slovak).
Tanner, M.A., 1996. Tools for Statistical Inference. Methods for the exploration of posterior distributions and likelihood functions. Springer, New York. 207pp.
UNESCO, 1976. World Catalogue of Very Large Floods. The UNESCO Press. 424pp.
Viglione, A., 2009. nsRFA: Non-supervised regional frequency analysis. R package. Version 0.6-9. <http://cran.r-project.org/web/packages/nsRFA> (accessed 24.08.09) (published 19.08.09).
Viglione, A., Laio, F., Claps, P., 2007. A comparison of homogeneity tests for regional frequency analysis. Water Resources Research 43, W03428. doi:10.1029/2006WR005095.
Wallis, J.R., Schaefer, M.G., Barker, B.L., Taylor, G.H., 2007. Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. Hydrology and Earth System Sciences 11 (1), 415–442.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bulletin 1, 80–83.