



Regional parent flood frequency distributions in Europe – Part 1: Is the GEV model suitable as a pan-European parent?

J. L. Salinas¹, A. Castellarin², A. Viglione¹, S. Kohnová³, and T. R. Kjeldsen⁴

¹Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria

²Department DICAM, School of Civil Engineering, University of Bologna, Bologna, Italy

³Department of Land and Water Resources Management, Faculty of Civil Engineering,
SUT Bratislava, Bratislava, Slovakia

⁴Department of Architecture and Civil Engineering, University of Bath, Bath, UK

Correspondence to: J. L. Salinas (salinas@hydro.tuwien.ac.at)

Received: 2 May 2013 – Published in Hydrol. Earth Syst. Sci. Discuss.: 22 May 2013

Revised: 11 September 2014 – Accepted: 15 September 2014 – Published: 5 November 2014

Abstract. This study addresses the question of the existence of a parent flood frequency distribution on a European scale. A new database of L-moment ratios of flood annual maximum series (AMS) from 4105 catchments was compiled by joining 13 national data sets. Simple exploration of the database presents the *generalized extreme value* (GEV) distribution as a potential pan-European flood frequency distribution, being the three-parameter statistical model that with the closest resemblance to the estimated average of the sample L-moment ratios. Additional Monte Carlo simulations show that the variability in terms of sample skewness and kurtosis present in the data is larger than in a hypothetical scenario where all the samples were drawn from a GEV model. Overall, the *generalized extreme value* distribution fails to represent the kurtosis dispersion, especially for the longer sample lengths and medium to high skewness values, and therefore may be rejected in a statistical hypothesis testing framework as a single pan-European parent distribution for annual flood maxima. The results presented in this paper suggest that one single statistical model may not be able to fit the entire variety of flood processes present at a European scale, and presents an opportunity to further investigate the catchment and climatic factors controlling European flood regimes and their effects on the underlying flood frequency distributions.

1 Introduction

The first step for any assessment of the flooding potential or flood hazard is the estimation of the design flood associated with a given annual exceedance probability, often quoted in terms of a recurrence interval T measured in years. This information is most commonly obtained using flood frequency estimation techniques based on statistical analyses of series of observed flood peak discharges. Unsurprisingly, extreme flood events are seldom observed locally, and hydrologists have little or no chance of gathering data from an adequate sample of catastrophes for analysis, especially for prediction at ungauged sites, with the exception of post-event surveys (see e.g. Marchi et al., 2010; Gaume et al., 2010). It is therefore important that effective and practical procedures are available, to assist hydrologists in making inferences about flood risk, both at gauged and at ungauged sites (Blöschl et al., 2013; Salinas et al., 2013).

When only an estimate of the peak flow value is needed, at-site and regional statistical extreme value analysis of river flow data can be used, depending on data availability. However, it is well-known that such estimates can be associated with a high degree of uncertainty, and it is therefore important to ensure that decisions are robust and are made based on as much information as possible (Viglione et al., 2013; Merz and Blöschl, 2008a, b; Martins and Stedinger, 2001). The choice of method for flood frequency estimation in any particular situation is often dictated by factors such as national or institutional tradition, modeller expertise, complexity and

objective of study, legislative requirements, and data availability (Castellarin et al., 2012). It is usual though, to assume the existence of a parent flood frequency distribution within a certain region. The question of the existence itself of a parent distribution on different spatial scales has puzzled hydrologists for many years and substantial work has been done in order to verify or falsify this hypothesis. Matalas et al. (1975) found that the variance of sample coefficient of skewness was always higher for observed data than for simulated flood peaks for a set of considered parent distributions, calling this phenomenon “skew separation”. They further showed that it could not be attributed either to small sample properties of the skewness estimator or to autocorrelation of the flood peak series. Dawdy and Gupta (1995) related the magnitude of this “skew separation” to the scaling structure of flood peaks and the heterogeneity among regions on the study from Matalas et al. (1975). Other authors have also shown that different distributions than those considered in Matalas et al. (1975) were able to reproduce the skewness variability; namely, Houghton (1978) for the *Wakeby* distribution and Rossi et al. (1984) for the TCEV (*two-component extreme value distribution*), while Ashkar et al. (1992) and Bobee et al. (1993) provide criticism on using the separation of skewness for choosing the type of distribution to be used in regional flood frequency analyses and give more importance to the step of the definition of homogeneous regions in order to avoid mixing of different skewness values from different populations.

More recently, the different behaviour of the tails of the distribution functions used in environmental extremes has been investigated by El Adlouni et al. (2008), and Gaál et al. (2009) or Papalexioiu et al. (2013) performed studies on daily rainfall, similar to the present one, analysing which kind of extreme value distribution fits best the precipitation extremes

The present paper introduces the first available inventory of data and statistical methods for flood frequency estimation across Europe, compiled with the aim to homogenize and harmonize the current level of knowledge on the approach to flood frequency estimation used across Europe. The inventory has been created as part of COST (European Cooperation in Science and Technology) Action ES0901 (European Procedures for Flood Frequency Estimation – Flood-Freq), which is a European-Commission-funded project that develops a network of experts, involved in nationally funded flood frequency estimation research projects. Their main task is to undertake a pan-European comparison and evaluation of methods of flood frequency estimation under the various climatologic and geographic conditions found in Europe, and promoting a synergic approach to flood hazard assessment, as requested by the European Flood Directive (Kjeldsen, 2011; Castellarin et al., 2012).

This study addresses explicitly the question of the existence of a parent flood frequency distribution on a European scale. It presents the results of an assessment based on the analysis of a newly established pan-European database of an-

Table 1. FloodFreq streamflow database: number of sites and station-years of data for the national annual maxima sequences.

| Country | No. of sites | Station-years of data | Kind of data |
|---------------|--------------|-----------------------|---------------|
| Austria | 676 | 28 592 | Instantaneous |
| Cyprus | 9 | 382 | Daily |
| Germany | 415 | 22 516 | Daily |
| Denmark | 43 | 2789 | Daily |
| France | 1172 | 45 331 | Instantaneous |
| Ireland | 215 | 6708 | Instantaneous |
| Italy | 373 | 8207 | Instantaneous |
| Lithuania | 30 | 1953 | Instantaneous |
| Norway | 104 | 3120 | Daily |
| Poland | 39 | 3426 | Instantaneous |
| Slovakia | 174 | 7995 | Instantaneous |
| Spain | 220 | 8594 | Instantaneous |
| United Kindom | 635 | 23 200 | Instantaneous |
| FloodFreq | 4105 | 162 813 | |

nual maximum series (AMS) of flood flows and their statistical characteristics compiled for the FloodFreq project, in order to find the most suitable frequency distributions for modelling the flood frequency regimes in Europe. In the companion paper by Salinas et al. (2014), the link between catchment and climate attributes and the choice from a set of potential parent regional flood frequency distributions is investigated on a subset of the database presented in this article.

2 Inventory of data and methods and the European flood database

The first phase of the FloodFreq COST Action focused on the compilation of inventories of data sets and methods for flood frequency estimation at a European scale. An extensive survey was conducted among 15 European countries in order to assess the availability of flood data and catchment descriptors and to investigate the existence of national guidelines for flood frequency estimation. Particularly, if these guidelines existed, related to the issue of large-scale underlying parent distributions, it was of interest if any type of flood frequency distribution was recommended. The main results of the survey relevant to this paper are presented below.

2.1 European flood database

The assessment of flood data availability at national level for the 15 European countries included in the survey showed that the AMS of flood flows are the most common standard. Therefore, it was decided to focus on a collection of AMS of flood flows considering daily flows, as well as instantaneous peak flows where available.

From the 15 surveyed countries, 13 agreed to share flood data in the frame of the FloodFreq COST Action. Due to

national policies and regulations that restrict the publication of some of these data, the flood data themselves were summarized into statistical moments. In particular, the AMS for a total of 4105 sites was characterized by the number of observations n and sample L-moment ratios of orders 2–4 (i.e. sample L-coefficient of variation, sample L-coefficient of skewness and sample L-coefficient of kurtosis). Table 1 contains a summary of the national AMS data sets available in the database. The use of L-moments instead of conventional moments offers several advantages such as the possibility of characterizing a wider range of distributions, smaller bias and higher robustness of the parameter estimators when applied to short samples. For more details on L-moments see, for example, Hosking (1990) and Hosking and Wallis (1997).

2.2 National guidelines for flood frequency estimation

National guidelines for flood frequency estimation are available in 9 of the 15 surveyed countries. In Germany, reference studies are available at the level of the federal states, and in Belgium for the Flemish part only. Public agencies and institutions in six countries provide recommendations as to the most suitable parent distributions to be used for flood frequency analysis, but in general such guidance appears to be sparse. In a number of countries, the *generalized extreme value* (GEV) distribution is among the recommended choices (e.g. Austria, Germany, Italy, Spain), but a variety of two-, three- or four-parameter distributions are also used, including the *Gumbel* (GUM) distribution in Finland and Spain, the three-parameter *generalized Pareto* (GPA) distribution in Belgium, the three-parameter *generalized logistic* (GLO) distribution in the UK, or the four-parameter TCEV in Italy and Spain. The Slovenian Environment Agency uses five different distributions (*normal*, *lognormal*, *Pearson type 3*, *log-Pearson type 3* and *Gumbel*) implemented in their own software DIST. In Slovakia, the *gamma*, *three-parameter lognormal*, *log-Pearson type 3* and GEV distributions are often used. Six countries reported that they have no standard parent distribution and the choice of an appropriate model depends mostly on the region where the analysis is undertaken (Castellarin et al., 2012).

The existence of some preferred statistical models, provides a motivation for a further investigation into potential candidates for pan-European flood frequency distributions, taking advantage of the uniquely extensive flood data collection compiled in this study (see Table 1).

3 A pan-European flood frequency distribution?

3.1 L-moment ratio diagram framework

The analyses of the FloodFreq database presented in this study address the issue of probabilistic model choice in the L-moment working environment and, in particular, through the use of L-moment ratio diagrams which enable graph-

ical identification of a suitable regional parent distribution among several two- and three-parameter candidates (see e.g. Hosking and Wallis, 1997). The scientific literature seems to agree on the value of L-moment ratio diagrams for guiding the selection of a regional parent distribution (e.g. Vogel and Fennessey, 1993; Vogel and Wilson, 1996; Peel et al., 2001; Strupczewski et al., 2011). An additional advantage of the diagrams is that L-moments are particularly suitable for short samples often associated with annual flood sequences, as sample L-moments tend to be less biased than the corresponding estimators of conventional moments (Vogel and Fennessey, 1993).

Two types of L-moment ratio diagrams are commonly used in the literature to assess the goodness of fit of regional parent distributions: (i) a diagram plotting the L-coefficient of variation against the L-coefficient of skewness (or L-Cs–L-Cv diagram), and (ii) a plot of the L-coefficient of kurtosis against the L-coefficient of skewness (or L-Cs–L-Ck diagram). The former is used to assess the suitability of various two-parameter distributions, while the latter version of the diagram is more commonly used when three-parameter distributions are considered. The suitability of various candidate parents is assessed by analysing how close the cloud of sample L-moments computed for the study region lies, relative to the lines corresponding to the different theoretical models. This study only presents L-Cs–L-Ck diagrams, as all the parent distributions considered are three-parametric, while in the companion paper by Salinas et al. (2014) both L-Cs–L-Ck and L-Cs–L-Cv diagrams will be used.

3.2 Average behaviour of the FloodFreq database

Figure 1a shows the L-Cs–L-Ck diagram for the entire FloodFreq data set (see also Table 1) and includes the sample L-moment ratios for all of the catchments in the data set (light-grey circles), together with four lines illustrating the theoretical relationship between L-Cs and L-Ck for the three-parameter frequency distributions that, as highlighted by the survey commented in Sect. 2.2, were preferentially recommended in the national guidelines, namely the GEV, GLO and *three-parameter lognormal* (LN3), and *Pearson type 3* (PE3).

To reduce some of the noise that is present in the empirical data due to sampling uncertainty and better determine which of the four three-parameter distributions considered better represents the statistical properties of the sample, a record length weighted moving average (WMA) is included in Fig. 1a. In particular, the WMA is computed by taking the weighted mean of the 50 neighbouring sample L-Cs values, and plotting it against the weighted mean of the corresponding 50 sample L-Ck values. Each sample L-moment ratio is weighted proportionally to the record length to reduce the impact of sampling variability from short records. The WMA values in Fig. 1a follow closer the theoretical relationship between L-Cs and L-Ck of the GEV distribution than that of

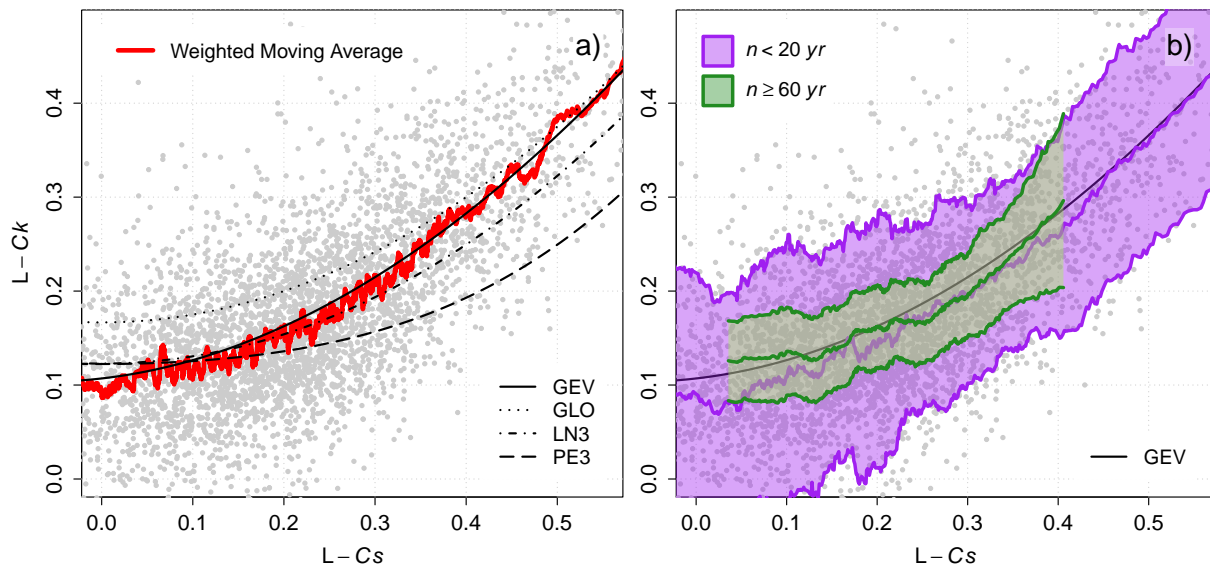


Figure 1. L-Cs–L-Ck diagram for the FloodFreq database with (a) the record length weighted moving average over 50 catchments (red line) for the whole database and (b) the record length weighted moving averages and standard deviations for stations with more than 60 yr (green lines and polygon) and less than 20 yr of data (purple lines and polygon) separately.

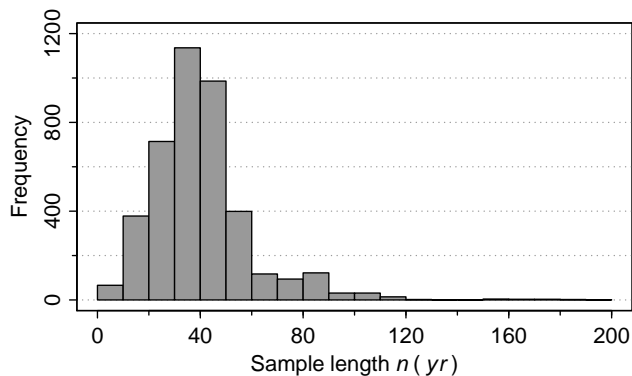


Figure 2. Sample length distribution for the entire FloodFreq database.

any of the other considered distributions. The position of the WMA indicates therefore that the GEV distribution might be a better candidate for describing the frequency regime of annual maximum floods at a pan-European level, compared to the other three extreme value distributions studied.

Given the known relationship between sampling uncertainty of the L-moment ratios and sample length n (see e.g. Viglione, 2010; Hosking and Wallis, 1997), it is necessary at this point to describe more in detail the distribution of n in the database. Figure 2 shows a histogram of the record lengths. The mean and median values are 40 and 38 yr, respectively. Later in the analysis, the database is subdivided into four n classes with the limits at 20, 40 and 60 yr. The percentage of stream gauges within each class is 10.8% (444 stations with $n < 20$ yr), 45.0% (1850 stations with $20 \text{ yr} \leq n < 40$ yr),

33.7% (1385 stations with $40 \text{ yr} \leq n < 60$ yr), and 10.4% (426 stations with $n \geq 60$ yr), respectively. This classification will ease the interpretation of the posterior analyses, trying to reduce the heterogeneity introduced by mixing stations with different sample sizes, as can be acknowledged in Fig. 1b. Here, the intervals corresponding to the moving averages plus-minus the moving standard deviations of the L-Ck values (both weighted proportionally to the record length over 50 elements) for the stations with $n < 20$ yr and $n \geq 60$ yr are depicted. One can see that, while the moving averages continue to stay closer to the GEV line in both cases, the spread of the data increases considerably with sample size, and therefore should be taken into account explicitly in the more detailed investigation performed in the following section.

3.3 Monte Carlo simulations

The GEV is the statistical model that best represents the averaged statistical properties of the entire database, if compared to the other three-parameter distributions. In order to falsify or verify the hypothesis that the GEV could actually be a valid underlying pan-European flood frequency distribution, the spread of the observed data has to be compared with the theoretical scenario where all stations represent random samples drawn from GEV distributions with a variety of sample lengths, skewness and kurtosis values. This reference scenario was set up via Monte Carlo simulations. Specifically, a total of 50 000 European-scale simulations are carried out as follows. As commented in the previous section, the effect of sample size n should be taken into account if the dispersion of the data in the L-moment ratio diagram needs to be investigated. Therefore, the database is sorted

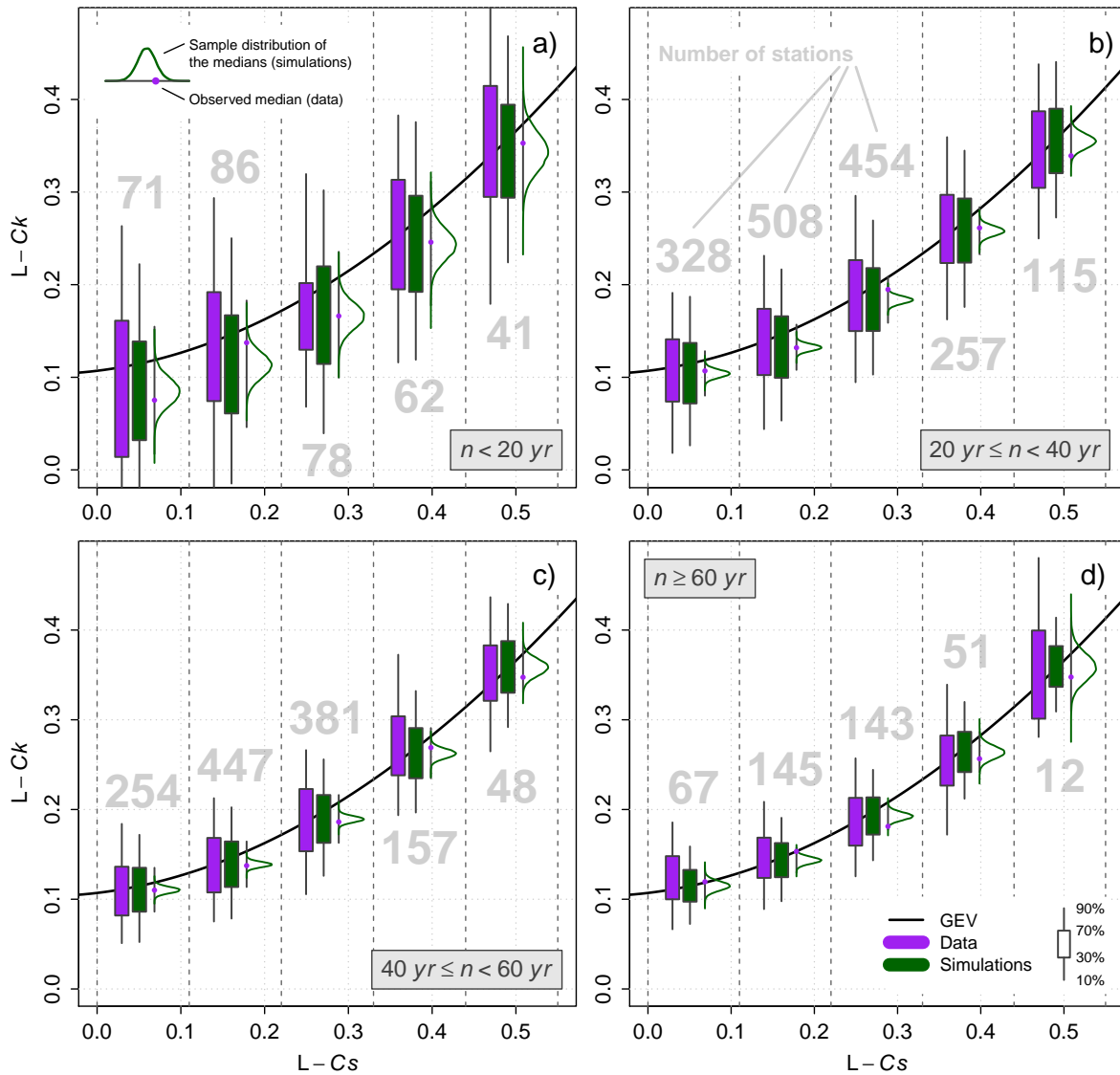


Figure 3. Monte Carlo simulations comparing the distribution of sample $L-C_k$ from 50 000 realisations at the European scale, all randomly drawn from GEV distributions (green box plots) and the distribution of sample $L-C_k$ from the FloodFreq database (purple box plots). Results are sorted by station sample length n in (a) $n < 20$ yr, (b) $20 \text{ yr} \leq n < 40$ yr, (c) $40 \text{ yr} \leq n < 60$ yr, and (d) $n \geq 60$ yr.

into the record length classes described in Sect. 3.2, i.e. 444 stations with $n < 20$ yr, 1850 stations with $20 \text{ yr} \leq n < 40$ yr, 1385 stations with $40 \text{ yr} \leq n < 60$ yr, and 426 stations with $n \geq 60$ yr. In each class, $L-C_s$ values of flood flow sequences are assumed to vary randomly between sites as described by a normal distribution with means ranging from 0.20 to 0.22 and standard deviations ranging from 0.14 to 0.17, as in all four cases the sample distributions passed a normality test. The record length distributions inside each n class are modelled by a triangular ($n < 20$ yr), uniform ($20 \text{ yr} \leq n < 40$ yr) and three-parameter gamma distributions ($40 \text{ yr} \leq n < 60$ yr and $n \geq 60$ yr). All parameters from the described distributions were estimated from the characteristics of the observed distributions of the database with the method of L-moments.

The types of distribution were chosen based on their respective probabilistic plots. Record length and $L-C_s$ are assumed independent, as no significant correlation is found between the sample values. Then, for each one of the 50 000 simulations and each n class, two samples of the corresponding length (444, 1850, 1385 or 426 depending on sample-size class) are generated from the previously defined distributions for the record lengths and population $L-C_s$ values, representing the properties of each of the synthetic European samples. The population $L-C_k$ values are then obtained from the functional relationship between skewness and kurtosis for the GEV model and, without loss of generality, the mean and $L-C_v$ values are both set to 1. For each of the virtual stations, a sample from a GEV distribution is generated, with each of

the previously simulated population properties. Finally, the sample L-moment ratios are computed for each of the generated GEV stations, which will not necessarily be located on the theoretical GEV line on the L-Cs–L-Ck diagram, due to both sample variability (finite record lengths) and biases in the sample estimators of the L-moment ratios (see e.g. Hosking and Wallis, 1997, p. 29).

To ease the interpretation of the results, the L-Cs–L-Ck values for both the observed data and the simulations are binned into five equidistant classes between the L-Cs values of 0 and 0.55, and the spread of the L-moment ratios is represented by the 10, 30, 50, 70 and 90 % quantiles of the sample L-Ck values for each bin and shown in the box plots of Fig. 3, together with the number of stations that are inside each of the bins. For the observed data, every bin is associated with only one value of the 10, 30, 50, 70 and 90 % percentiles, respectively, but in the case of the simulations we obtain 50 000 values for each of the percentiles, one for each of the European-scale simulations. This means that we obtain sample distributions for all of the quantiles, as is shown for the case of the medians in Fig. 3. The 10, 30, 70 and 90 % percentiles for the simulations have their corresponding sample distributions, but in Fig. 3 only their averaged values are represented as limits of the box plots for the sake of clarity.

In the Monte Carlo simulations, the explicit assumption that the underlying parent distribution of all stations in Europe is given by the GEV model is made, and this assumption will be verified or falsified with the classical approach of statistical hypothesis testing. We define the null hypothesis H_0 such that, for the i th bin, the j % quantile of the sample L-Ck distribution from the data is not significantly different to the j % quantile in the same bin, for all stations randomly drawn from GEV distributions. For each of the percentiles there is a sample distribution (from the simulations), and an observed value (from the observed data) available, so we can calculate the associated p values, set a significance level and reject or accept the null hypothesis for all the quantiles.

Table 2 shows these calculated p values for the plotted quantiles in Fig. 3. For almost all the cases, and independently of sample size, we cannot reject the null hypothesis that the medians are equal to those from a randomly generated GEV of Europe. This result from Monte Carlo simulations, i.e. that the mean or median L-Cs–L-Ck behaviour can be consistently explained by the GEV distribution, is more robust than the one described in Sect. 3.2 and shown in Fig. 1 on the weighted moving averages. According to, for example, Hosking and Wallis (1997, p. 29), the L-moment ratios in Fig. 1 should be shifted towards the top-right corner of the diagram if one takes into account the sample bias. This would probably cause the averages to differ slightly more from the theoretical GEV curve. In the simulations presented in this section, sample L-moments from the database are compared to sample L-moments from GEV generations and are therefore subject to the same sample bias.

The results for the dispersion of the L-Ck values show a more interesting pattern than the median values. For $n < 20$ yr, the spreads in the data and the simulations are not significantly different from each other for any of the L-Cs bins in almost all the cases (no rejections with a significance level α of 5 % and three rejections with an α of 10 %). For increasing sample size, the rate of rejection increases, especially for L-Cs values larger than 0.22. Except for two cases, the GEV distribution always fails to explain the dispersion above the median and for larger record lengths also the dispersion below the median, which is significantly higher for the observed data. It is worth noting at this point that the robustness of these results is affected by two factors. First, the number of stations inside each bin after the record length and L-Cs sorting. For example, one cannot draw any strong conclusion from the last bin in Fig. 3d, as it contains only 12 elements, while for other bins the results can be considered more consistent. Second, the biases in the L-moment ratios already commented (see e.g. Hosking and Wallis, 1997, p. 29), which are particularly large in short to medium sample sizes and higher L-Cs values, have the potential to shift some of the stations, whose L-Cs population lies in one bin, into another bin because their sample L-Cs is lower. However, checks of the number of stations per bin in the data and the simulations resulted in a difference of less than 5 %.

3.4 Discussion

Even though the results of the Monte Carlo simulations point out that selecting the GEV distribution as a pan-European parent cannot fully describe the observed variability of sample L-moments, there are some aspects that deserve a deeper discussion. In fact, it is remarkable that for all the European geographical areas considered, including catchments with very different sizes, climatic conditions, and geomorphologic characteristics covered in the FloodFreq database, there is not enough statistical evidence to reject the hypothesis that the GEV distribution is a suitable parent distribution for describing the median behaviour in terms of sample L-moment ratios. From a purely statistical point of view, this could be explained by the fact that the GEV distribution is the theoretical extreme value distribution that expresses in a closed analytical way the three possible asymptotic distributions derived from any kind of parent population, as described in the Fisher–Tippett–Gnedenko theorem (Fisher and Tippett, 1928; Gnedenko, 1943). Therefore, it offers a theoretical justification for using it to reproduce the sample frequency distribution of annual maxima series from many different hydrological and geological extreme phenomena (precipitation depths, flood flows, earthquake magnitudes, wind speeds and others) observed in different geographical contexts around the world (e.g. Robson and Reed, 1999; Castellarin et al., 2001; Thompson et al., 2007; Grimaldi et al., 2011).

Table 2. Empirical p values for the observed quantiles of L-Ck given the sample distributions generated from the GEV simulations shown in Fig. 3. In bold, values rejecting a 5 % significance test. In bold and italic, values rejecting a 10 % significance test.

| L-Ck quantile | Range of L-Cs values ($n < 20$ yr) | | | | | Range of L-Cs values ($20 \text{ yr} \leq n < 40$ yr) | | | | |
|---------------|--|--------------|--------------|--------------|--------------|--|--------------|--------------|--------------|--------------|
| | 0–0.11 | 0.11–0.22 | 0.22–0.33 | 0.33–0.44 | 0.44–0.55 | 0–0.11 | 0.11–0.22 | 0.22–0.33 | 0.33–0.44 | 0.44–0.55 |
| 90 % | 0.949 | 0.968 | 0.788 | 0.626 | 0.965 | 0.746 | 0.996 | 1.000 | 0.969 | 0.413 |
| 70 % | 0.908 | 0.945 | 0.129 | 0.825 | 0.816 | 0.806 | 0.978 | 0.975 | 0.757 | 0.375 |
| 50 % | 0.271 | 0.956 | 0.498 | 0.562 | 0.669 | 0.762 | 0.478 | 0.997 | 0.730 | 0.025 |
| 30 % | 0.121 | 0.822 | 0.846 | 0.559 | 0.515 | 0.680 | 0.784 | 0.489 | 0.469 | 0.028 |
| 10 % | 0.275 | 0.172 | 0.926 | 0.433 | 0.064 | 0.069 | 0.028 | 0.057 | 0.026 | 0.015 |
| L-Ck quantile | Range of L-Cs values ($40 \text{ yr} \leq n < 60$ yr) | | | | | Range of L-Cs values ($n \geq 60$ yr) | | | | |
| | 0–0.11 | 0.11–0.22 | 0.22–0.33 | 0.33–0.44 | 0.44–0.55 | 0–0.11 | 0.11–0.22 | 0.22–0.33 | 0.33–0.44 | 0.44–0.55 |
| 90 % | 0.990 | 0.989 | 0.978 | 1.000 | 0.731 | 0.999 | 0.999 | 0.976 | 0.966 | 0.999 |
| 70 % | 0.627 | 0.890 | 0.961 | 0.989 | 0.316 | 0.996 | 0.928 | 0.457 | 0.302 | 0.847 |
| 50 % | 0.482 | 0.342 | 0.194 | 0.893 | 0.117 | 0.813 | 0.995 | 0.005 | 0.163 | 0.238 |
| 30 % | 0.119 | 0.026 | 0.005 | 0.728 | 0.170 | 0.684 | 0.420 | 0.003 | 0.020 | 0.010 |
| 10 % | 0.385 | 0.202 | 0.000 | 0.324 | 0.010 | 0.191 | 0.046 | 0.001 | 0.000 | 0.047 |

Nevertheless, the results for the Monte Carlo simulations regarding the spread of the L-moment ratios around the GEV line show that the dispersion is bigger than expected from random sampling, particularly for the deviations above the median for non-short samples. In the terminology used by Matalas et al. (1975), a “kurtosis separation” appears, if only the GEV distribution is considered as the underlying parent across Europe. Intersite correlation is most probably present in the original annual flood flow series, from which the sample L-moment ratios have been computed, and this correlation should reduce the observed L-moment variability. In the Monte Carlo generations, this intercorrelation has been entirely neglected and still the variability of the generated L-moments is lower than the observed one. Also, sample estimation uncertainty, particularly for high values of L-Ck, could also play a role by augmenting the variability in the observed L-moments, but the systematic underestimation of the dispersion points out to the fact that the GEV distribution alone is not complex enough to fully describe the variability of sample L-Cs and L-Ck values estimated for the Flood-Freq database, being these L-moment ratios surrogates for the entire spectrum of flood generation processes occurring across Europe responsible for the diversity of shapes of flood frequency distributions. It is therefore necessary to further investigate the links of hydrological processes to the L-moment ratios, and in particular to high values of skewness and kurtosis, in order to try to explain these discrepancies.

4 Conclusions

The issue of existence of underlying parent flood frequency distributions across different processes, places and scales is directly addressed in this study. One of the most applied and recommended statistical models, the GEV distribution, has proven to capture the mean and median statistical properties

of a pan-European database annual maximum flood series, but the observed variability in the data is larger than what this model can randomly reproduce. This implies that the GEV alone cannot be considered as a candidate for a pan-European flood frequency distribution, as it is not complex enough to reproduce the entire variety of hydrological processes leading to the different shapes of flood frequency curves. This fact rises the more fundamental question if we actually need a pan-European flood frequency distribution, which does not need to be necessarily the case. The investigation of the GEV as a single model for all European catchments was performed based on a very basic inspection of the flood database, but there are many examples that show that one single analytic expression across large scales and more important, across different processes, is not valid for describing all possible local characteristics at once. Rogger et al. (2012) proved that step changes appear in the flood frequency curve when local runoff generation mechanisms are influenced by threshold processes, especially for small mountainous catchments, and these are not captured by any traditional statistical model so far. The case of several Mediterranean regions which are characterized by two distinct flood populations is also very common. These populations are referred by Rossi et al. (1984) as “ordinary floods”, generated by frontal-type rainfalls, and “extraordinary floods”, generated by highly convective rainstorms. For the modelling of these flood regimes it could be appropriate to use the TCEV model (see e.g. Francés, 1998) as there is a mixture of populations, while it could not be suitable in other regional contexts if there were not such a mixing. Merz and Blöschl (2003) showed for an Austrian data set that different typologies of floods classified after their generation mechanisms may have very different statistical properties and can therefore lead to distinct flood frequency distributions. In particular, if many flood generation processes take place in one catchment,

possibly depending on rainfall or snowfall regimes, the overall flood frequency distribution is a result of the combination of the distributions associated with the single mechanisms and is not necessarily expressed in terms of a single analytical model.

The inclusion of information on the underlying hydrological processes in the model choice is therefore of high importance. The companion paper by Salinas et al. (2014) focuses precisely on defining the controls of catchment and climate indicators on the averaged L-moment ratios and the regional flood frequency distributions.

Acknowledgements. The authors are grateful for the financial contribution made by the COST Office grant ES0901 European procedures for flood frequency estimation (FloodFreq) and for contributions made by members of Working Group 2 of the COST Action. Funding from the ERC FloodChange project (ERC Advanced Grant FP7-IDEAS-ERC-AG-PE10 291152) and from the Austrian Science Funds (FWF) as part of the Vienna Doctoral Programme on Water Resource Systems (DK-plus W1219-N22) is also acknowledged. The compilation of the Italian data set was partially supported by the Italian Government through its national grant to the programme on Characterisation of Ungauged Basins by Integrated use of hydrological Techniques (CUBIST Project) and the compilation of the Slovakian data set was partially supported by the Slovak VEGA Grant Agency under contract no. 1/0776/13. Both supports are greatly acknowledged. Referees F. Laio, R. Romanowicz, S. Grimaldi, and V. Iacobellis, as well as Günter Blöschl, are thankfully acknowledged for their constructive comments and suggestions.

Edited by: R. Woods

References

- Ashkar, F., Bobee, B., and Bernier, J.: Separation of skewness: reality of regional artifact?, *J. Hydraul. Eng.*, 118, 460–475, doi:10.1061/(ASCE)0733-9429(1992)118:3(460), 1992.
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H.: *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*, Cambridge University Press, Cambridge, UK, 2013.
- Bobee, B., Cavadias, G., Ashkar, F., Bernier, J., and Rasmussen, P.: Towards a systematic approach to comparing distributions used in flood frequency analysis, *J. Hydrol.*, 142, 121–136, 1993.
- Castellarin, A., Burn, D., and Brath, A.: Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, *J. Hydrol.*, 241, 270–285, doi:10.1016/S0022-1694(00)00383-8, 2001.
- Castellarin, A., Kohnová, S., Gaál, L., Fleig, A., Salinas, J. L., Toumazis, A., Kjeldsen, T. R., and Macdonald, N.: Review of applied statistical methods for flood frequency analysis in Europe, Milestone Report for WG2 of COST Action ES0901, Tech. rep., Centre for Ecology & Hydrology, Wallingford, UK, on behalf of COST, available at: <http://www.cost.eu/media/publications/>, 2012.
- Dawdy, D. R. and Gupta, V. K.: Multiscaling and skew separation in regional floods, *Water Resour. Res.*, 31, 2761–2767, doi:10.1029/95WR02078, 1995.
- El Adlouni, S., Bobée, B., and Ouarda, T. B. M. J.: On the tails of extreme event distributions in hydrology, *J. Hydrol.*, 355, 1–4, doi:10.1016/j.jhydrol.2008.02.011, 2008.
- Fisher, R. A. and Tippett, L. H. C.: Limiting forms of the frequency distribution of the largest and smallest member of a sample, *Math. Proc. Cambridge Phil. Soc.*, 24, 180–190, 1928.
- Francés, F.: Using the TCEV distribution function with systematic and non-systematic data in a regional flood frequency analysis, *Stoch. Hydrol. Hydraul.*, 12, 267–283, doi:10.1007/s004770050021, 1998.
- Gaál, L., Szolgay, J., Lapin, M., and Faško, P.: Hybrid approach to delineation of homogeneous regions for regional precipitation frequency analysis, *J. Hydrol. Hydromechan.*, 57, 226–249, doi:10.2478/v10098-009-0021-1, 2009.
- Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., and Blöschl, G.: Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites, *J. Hydrol.*, 394, 101–117, doi:10.1016/j.jhydrol.2010.01.008, 2010.
- Gnedenko, B. V.: Sur la distribution limite du terme maximum d'une serie aleatoire, *Ann. Math.*, 44, 423–453, 1943.
- Grimaldi, S., Kao, S. C., Castellarin, A., Papalexiou, S. M., Viglione, A., Laio, F., Aksoy, H., and Gedikli, A.: *Treatise on Water Science*, chap. Statistical hydrology, 479–517, Elsevier, Oxford, UK, available at: <http://www.sciencedirect.com/science/referenceworks/9780444531995>, 2011.
- Hosking, J. R. M.: L-moments: analysis and estimation of distributions using linear combinations of order statistics, *J. R. Stat. Soc. B*, 52, 105–124, 1990.
- Hosking, J. R. M. and Wallis, J. R.: *Regional Frequency Analysis: an Approach Based on L-Moments*, Cambridge University Press, Cambridge, UK, 1997.
- Houghton, J. C.: Birth of a parent: the Wakeby Distribution for modeling flood flows, *Water Resour. Res.*, 14, 1105–1109, doi:10.1029/WR014i006p01105, 1978.
- Kjeldsen, T. R.: COST Action ES0901: European procedures for flood frequency estimation (FloodFreq), *Geophys. Res. Abs.*, 13, EGU2011–10327, 2011.
- Marchi, L., Borga, M., Preciso, E., and Gaume, E.: Characterisation of selected extreme flash floods in Europe and implications for flood risk management, *J. Hydrol.*, 394, 118–133, doi:10.1016/j.jhydrol.2010.07.017, 2010.
- Martins, E. and Stedinger, J. R.: Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series, *Water Resour. Res.*, 37, 2559–2567, doi:10.1029/2000WR000009, 2001.
- Matalas, N. C., Slack, J. R., and Wallis, J. R.: Regional skew in search of a parent, *Water Resour. Res.*, 11, 815–826, doi:10.1029/WR011i006p00815, 1975.
- Merz, R. and Blöschl, G.: A process typology of regional floods, *Water Resour. Res.*, 39, 1340, doi:10.1029/2002WR001952, 2003.
- Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. temporal, spatial, and causal expansion of information, *Water Resour. Res.*, 44, W08432, doi:10.1029/2007WR006744, 2008a.

- Merz, R. and Blöschl, G.: Flood frequency hydrology: 2. combining data evidence, *Water Resour. Res.*, 44, W08433, doi:10.1029/2007WR006745, 2008b.
- Papalexiou, S. M., Koutsoyiannis, D., and Makropoulos, C.: How extreme is extreme? An assessment of daily rainfall distribution tails, *Hydrol. Earth Syst. Sci.*, 17, 851–862, doi:10.5194/hess-17-851-2013, 2013.
- Peel, M. C., Wang, Q. J., Vogel, R. M., and McMahon, T. A.: The utility of L-moment ratio diagrams for selecting a regional probability distribution, *Hydrol. Sci. J.*, 46, 147–156, doi:10.1080/02626660109492806, 2001.
- Robson, A. J. and Reed, D. W.: *Flood Estimation Handbook (FEH)*, chap. Statistical procedures for flood frequency estimation, p. Vol. 3, Institute of Hydrology, Wallingford, UK, 1999.
- Rogger, M., Pirkel, H., Viglione, A., Komma, J., Kohl, B., Kirnbauer, R., Merz, R., and Blöschl, G.: Step changes in the flood frequency curve: process controls, *Water Resour. Res.*, 48, W05544, doi:10.1029/2011WR011187, 2012.
- Rossi, F., Fiorentino, M., and Versace, P.: Two-component extreme value distribution for flood frequency analysis, *Water Resour. Res.*, 20, 847–856, doi:10.1029/WR020i007p00847, 1984.
- Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivalapan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, *Hydrol. Earth Syst. Sci.*, 17, 2637–2652, doi:10.5194/hess-17-2637-2013, 2013.
- Salinas, J. L., Castellarin, A., Kohnová, S., and Kjeldsen, T. R.: Regional parent flood frequency distributions in Europe – Part 2: Climate and scale controls, *Hydrol. Earth Syst. Sci.*, 18, 4391–4401, doi:10.5194/hess-18-4391-2014, 2014.
- Strupczewski, W. G., Kochanek, K., Markiewicz, I., Bogdanowicz, E., Weglarczyk, S., and Singh, V. P.: On the tails of distributions of annual peak flow, *Hydrol. Res.*, 42, 171–192, doi:10.2166/nh.2011.062, 2011.
- Thompson, E. M., Baise, L. G., and Vogel, R. M.: A global index earthquake approach to probabilistic assessment of extremes, *J. Geophys. Res.*, 112, B06314, doi:10.1029/2006JB004543, 2007.
- Viglione, A.: Confidence intervals for the coefficient of L-variation in hydrological applications, *Hydrol. Earth Syst. Sci.*, 14, 2229–2242, doi:10.5194/hess-14-2229-2010, 2010.
- Viglione, A., Merz, R., Salinas, J. L., and Blöschl, G.: Flood frequency hydrology: 3. a Bayesian analysis, *Water Resour. Res.*, 49, 1–18, doi:10.1029/2011WR010782, 2013.
- Vogel, R. M. and Fennessey, N.: L moment diagrams should replace product moment diagrams, *Water Resour. Res.*, 29, 1745–1752, doi:10.1029/93WR00341, 1993.
- Vogel, R. M. and Wilson, I.: The probability distribution of annual maximum, minimum and average streamflow in the United States, *J. Hydrol. Eng.*, 1, 69–76, doi:10.1061/(ASCE)1084-0699(1996)1:2(69), 1996.