

# Confidence intervals for the coefficient of L-variation in hydrological applications

A. Viglione

Institut für Wasserbau und Ingenieurhydrologie, Technische Universität Wien, Wien, Austria

Received: 27 July 2010 – Published in Hydrol. Earth Syst. Sci. Discuss.: 10 August 2010

Revised: 22 October 2010 – Accepted: 7 November 2010 – Published: 11 November 2010

**Abstract.** The coefficient of L-variation (L-CV) is commonly used in statistical hydrology, in particular in regional frequency analysis, as a measure of steepness for the frequency curve of the hydrological variable of interest. As opposed to the point estimation of the L-CV, in this work we are interested in the estimation of the interval of values (confidence interval) in which the L-CV is included at a given level of probability (confidence level). Several candidate distributions are compared in terms of their suitability to provide valid estimators of confidence intervals for the population L-CV. Monte-Carlo simulations of synthetic samples from distributions frequently used in hydrology are used as a basis for the comparison. The best estimator proves to be provided by the log-Student t distribution whose parameters are estimated without any assumption on the underlying parent distribution of the hydrological variable of interest. This estimator is shown to also outperform the non parametric bias-corrected and accelerated bootstrap method. An illustrative example of how this result can be used in hydrology is presented, namely in the comparison of methods for regional flood frequency analysis. In particular, it is shown that the confidence intervals for the L-CV can be used to assess the amount of spatial heterogeneity of flood data not explained by regionalization models.

countered in hydrology when dealing with floods or extreme rainfall events. The coefficient of L-variation (L-CV) is another – more efficient in many cases – measure of data dispersion introduced by Hosking (1990). It has hence replaced the conventional CV in various applications of statistical hydrology. In particular, the use of the L-CV as a measure of steepness of the flood frequency curve has become a standard in regional flood frequency analysis (see e.g., Pearson, 1991; Hosking and Wallis, 1993; Stedinger and Lu, 1995; Fill and Stedinger, 1998; Hosking and Wallis, 1997; Robson and Reed, 1999; Castellarin et al., 2001). Many statistical procedures of regionalization of floods are based on the hypothesis that the L-CV is informative enough to represent the differences among the flood frequency distributions at different sites. For example, the sample L-CV is used to designate “homogeneous regions”, where it is assumed that the frequency distribution of flood peaks for different sites is the same, except for a site-specific scale factor (Dalrymple, 1960, Index-Flood method). Other studies state that the slope of the flood frequency curves (or, equivalently, their L-CV) should be taken as the statistical descriptor to be related to catchment attributes such as area or mean elevation (see e.g., Robinson and Sivapalan, 1997; Allamano et al., 2009).

Instead of estimating the L-CV by a single value, in this work we are interested to provide an interval estimation of the L-CV, i.e., to infer the range (the confidence interval) in which the population L-CV is included at a given level of probability (the confidence level). The most immediate advantage of using confidence intervals, as opposed to point estimates, is that they clearly indicate the reliability of the estimate, given by the confidence level. But the advantages are more than that. For example, in regional frequency analysis, L-CV confidence intervals could be used to compare on a single basis very different approaches such as those based on site grouping and those that allow for the continuous variability of L-CV. In the first case it is assumed that the population L-CV is constant along a homogeneous region (group

## 1 Introduction

It is well known that the sample coefficient of variation (CV), i.e., the ratio of standard deviation to the mean of a series of data, exhibits substantial bias and variance when samples are small or belong to highly skewed populations (Vogel and Fennessey, 1993). This is the problem that is normally en-



Correspondence to: A. Viglione  
(viglione@hydro.tuwien.ac.at)

of sites). In the second case it is assumed that a model to describe the variability of the population L-CV applies to the whole analysed region, for example based upon a linear or non-linear functional relationship between the L-CV and climatic/physiographic characteristics. These two assumptions can be compared by counting how often the at-site L-CV confidence intervals actually include the regionally estimated L-CV. The spatial variability of the L-CV is essentially due to sample variability and spatial heterogeneity. If the confidence intervals include the regional model in less cases than specified by the confidence level, this is because of unexplained spatial heterogeneity. Different models can therefore be compared in terms of their ability to explain the spatial heterogeneity of the data. An illustrative example is provided at the end of the paper.

The more natural way to estimate confidence intervals for a parameter is by deriving them from its sampling distribution (e.g., Kottegoda and Rosso, 1997, Sect. 5.3). In asymptotic theory, approximate estimates of the sampling distribution of the L-CV are derived for large samples and for specific underlying parent distributions (see e.g., Hosking, 1986, 1990). In the hydrological literature some attempts to extend the asymptotic results to shorter samples exist. For example, Chowdhury et al. (1991) assume that the sample L-CV is normally distributed and use a first-order estimate of its asymptotic variance corrected, for small samples, with coefficients obtained through a Monte-Carlo procedure. In their analysis a Generalized Extreme Value parent distribution is assumed. Sankarasubramanian and Srinivasan (1999) consider, in a similar work, the Generalized Normal, Log-Normal and Pearson type III distributions.

In this work we estimate the sampling distribution of the L-CV for short samples and independently of the underlying parent distribution. To meet this aim, we exploit the results of a distribution-free approach, recently proposed by the statistical community, to derive estimators for variances and covariances of sample L-moments (Elamir and Seheult, 2004). The estimators of the sample L-CV and its variance are used to fit several two-parameters candidate sampling distributions (Sect. 2). Through Monte-Carlo simulations, we check if these sampling distributions are suitable to derive confidence intervals for the L-CV, e.g., if the 90% confidence limits (i.e., the end points of a confidence interval) obtained for different samples encompass the true population parameter 90% of the times. This property is referred as “validity” of the estimator of the confidence intervals (in our case the estimator is provided by the candidate sample distribution of the L-CV). We use a graphical goodness-of-fit method that is able to show for which confidence levels the estimator can be considered valid and to evidence the reasons for the lack of validity of it, in particular the effects of bias of estimation of the sample L-CV and its variance. We also use a synthetic goodness-of-fit measure, the Anderson-Darling test statistic, which checks the validity of the estimator for all confidence levels en bloc (Sect. 3). The candidate distributions are com-

pared for a wide range of sample lengths and underlying parent distributions, whose type is selected among those more frequently used in hydrological applications: the Generalized Extreme Value, the Pearson type III, the Log-Normal with 3 parameters, the Generalized Pareto and the Generalized Logistic distributions. The results are shown in Sect. 4 where: (i) a correction for the bias of estimation of the sample L-CV and its variance is proposed (Sect. 4.1); (ii) the method is compared to the non-parametric bias-corrected and accelerated BC<sub>a</sub> bootstrap, an appealing computer-based alternative method (Sect. 4.2); and (iii) the goodness of estimation of the 90% confidence intervals for the L-CV is analysed in more detail.

The outcome of this study for hydrological applications is shown through an example (Sect. 5) in which different regionalization techniques (the index-flood technique with “fixed regions”, the “region of influence” approach and a “region-free” simple regression method) are compared against a given data-set of flood peaks. Their performance is assessed by counting how often the L-CV confidence intervals actually include the regionally estimated L-CV.

## 2 Sampling distribution of the L-CV

L-moments were introduced by Hosking (1990) and are linear combinations of the Probability Weighted Moments defined by Greenwood et al. (1979) (see also Sillitto, 1969). Sample Probability Weighted Moments, computed from the order statistics  $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ , are given by

$$b_k = n^{-1} \binom{n-1}{k}^{-1} \sum_{j=k+1}^n \binom{j-1}{k} X_{j:n}, \quad (1)$$

where  $n$  is the sample length and  $k$  is the order of the probability weighted moment. Sample L-moments are defined as

$$l_r = \sum_{k=0}^{r-1} p_{r-1,k}^* b_k, \quad (2)$$

where the coefficients

$$p_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \quad (3)$$

are those of the “shifted Legendre polynomials” (see Hosking and Wallis, 1997). Since L-moment estimators are linear functions of the sample values, they are virtually unbiased and have relatively small sampling variance.

### 2.1 Variance of the sample L-CV

The asymptotic variances and covariances of the sample L-moments  $l_1, l_2, \dots, l_n$  are derived by Hosking (1990), who also demonstrate that their distribution is asymptotically normal. Elamir and Seheult (2004) derive the exact variance structure of sample L-moments in the non-asymptotic case

without formulating assumptions on the underlying parent distributions. Let's call the variances and covariances of the sample Probability Weighted Moments  $\theta_{kk} = \text{var}(b_k)$  and  $\theta_{kl} = \text{cov}(b_k, b_l)$ , respectively. Elamir and Seheult (2004) demonstrate that a distribution-free unbiased estimator of  $\theta_{kl}$  is

$$\hat{\theta}_{kl} = b_k b_l - \frac{1}{n^{(k+l+2)}} \sum_{1 \leq i < j \leq n} [(i-1)^{(k)}(j-k-2)^{(l)} + (i-1)^{(l)}(j-l-2)^{(k)}] X_{i:n} X_{j:n}, \tag{4}$$

where  $n^{(r)} = n(n-1)\dots(n-r+1)$ . Observing that Eq. (2) can also be expressed as  $\mathbf{l} = \mathbf{b}\mathbf{C}^T$ , where  $\mathbf{l} = (l_1, \dots, l_n)$ ,  $\mathbf{b} = (b_0, \dots, b_{n-1})$  and  $\mathbf{C}^T$  is the transpose of the  $n \times n$  lower triangular matrix with entries  $p_{r-1,k}^*$  given in Eq. (3), the variance matrix of the L-moments is given by

$$\text{var}(\mathbf{l}) = \mathbf{C}\Theta\mathbf{C}^T \tag{5}$$

where  $\Theta = \text{var}(\mathbf{b})$  is the variance matrix of the probability weighted moments with elements  $\theta_{kl}$ . The matrix  $\hat{\Theta}$ , obtained with  $\hat{\theta}_{kl}$  given in Eq. (4), is then an unbiased estimator of the variance matrix  $\Theta$ . On these basis, it descends from Eq. (5) that an unbiased estimator of  $\text{var}(\mathbf{l})$  is:

$$\widehat{\text{var}}(\mathbf{l}) = \mathbf{C}\hat{\Theta}\mathbf{C}^T. \tag{6}$$

The sample L-CV, i.e., the coefficient of L-variation, is defined by the ratio of the first two sample L-moments,

$$t = l_2/l_1, \tag{7}$$

where  $l_1$  is the sample mean and  $l_2$  is a measure of the dispersion around the mean value. The variance of the L-moment ratio  $t$  can be related to the variance structure of sample L-moments given in Eq. (6) using a Taylor-series-based approximation to the variance of the ratio of two random variables (see e.g., Kendall and Stuart, 1961):

$$\widehat{\text{var}}(t) \cong \left[ \frac{\widehat{\text{var}}(l_1)}{l_1^2} + \frac{\widehat{\text{var}}(l_2)}{l_2^2} - 2 \frac{\widehat{\text{cov}}(l_1, l_2)}{l_1 l_2} \right] \left[ \frac{l_2}{l_1} \right]^2. \tag{8}$$

The functions to calculate (4), (6) and (8) are provided in the R package nsRFA (Viglione, 2010).

### 2.2 Candidate distributions for the sample L-CV

According to Hosking (1990), the L-moment ratio estimators are asymptotically normally distributed and have small bias and variance, specially if compared with the classical coefficients of variation, skewness and kurtosis (Hosking and Wallis, 1997). Keeping the hypothesis of normality, formulated in the asymptotic theory, the first candidate distribution for the sample L-CV considered here is

$$t + \sqrt{\widehat{\text{var}}(t)} \cdot N(0, 1), \tag{9}$$

where  $N(0, 1)$  is the standard normal distribution. Because for short samples the normal distribution could be too narrow, inspired by the definition of the distribution of the mean of a sample when its real variance is unknown (see e.g., Kotegoda and Rosso, 1997), we also consider

$$t + \sqrt{\widehat{\text{var}}(t)} \cdot T_{n-1} \tag{10}$$

as an alternative distribution, where  $T_{n-1}$  is a Student t distribution with  $n - 1$  degrees of freedom. Note that the variance of the distribution defined by Eq. (10) is greater than  $\widehat{\text{var}}(t)$ , since the variance of  $T_{n-1}$  is greater than 1.

Both the normal and the Student t distribution are symmetric. On the other hand, given that in hydrology random variables are typically non-negative (e.g., Koutsoyiannis, 2005), some sampling distributions should be asymmetric. Therefore, we consider also the distribution

$$G \left( k = \frac{t^2}{\widehat{\text{var}}(t)}, \phi = \frac{\widehat{\text{var}}(t)}{t} \right), \tag{11}$$

where  $G$  is a gamma distribution with shape parameter  $k$  and scale parameter  $\phi$ , whose density function is

$$f_G(x; k, \phi) = \frac{1}{\phi^k \Gamma(k)} x^{k-1} e^{-(x/\phi)}. \tag{12}$$

Among the asymmetric distributions, we also consider the log-normal and the log-Student t distributions, i.e., we assume that the logarithm of the sample L-CV is normally or Student t distributed. In the first case, the sampling distribution of the logarithm of the L-CV can be approximated by

$$\log(t) - \frac{1}{2} \log \left( 1 + \frac{\widehat{\text{var}}(t)}{t^2} \right) + \sqrt{\log \left( 1 + \frac{\widehat{\text{var}}(t)}{t^2} \right)} \cdot N(0, 1), \tag{13}$$

which follows from the definition of mean and variance of a log-normally distributed random variable. Analogously, assuming that the same relationships hold for the log-Student t distribution, which is a reasonable approximation due to the similarity of the two distributions,

$$\log(t) - \frac{1}{2} \log \left( 1 + \frac{\widehat{\text{var}}(t)}{t^2} \right) + \sqrt{\log \left( 1 + \frac{\widehat{\text{var}}(t)}{t^2} \right)} \cdot T_{n-1}. \tag{14}$$

The log-Student's t distribution is a generalization of the log-normal distribution that has heavier tails but approaches log-normality as the sample length  $n$  increases.

In the following we present the method used to verify if these approximations provide valid estimators of the confidence intervals for the L-CV for samples of length  $n$  and several underlying parent distributions. It is important to notice

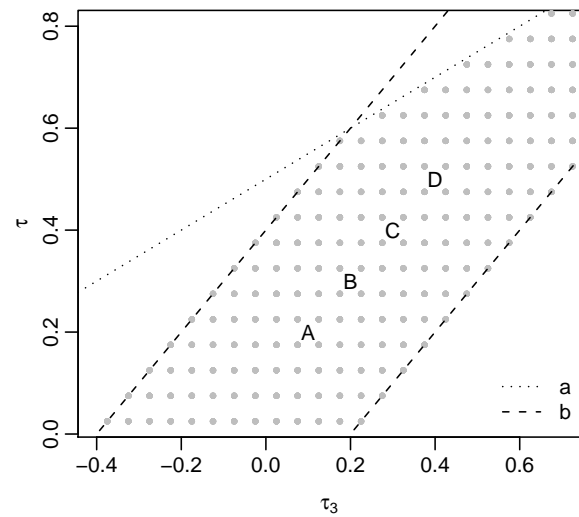
that we do not check if these distributions fit well the sampling distribution of  $t$ , but if they provide a good estimate of the range in which  $\tau$  (the population L-CV) is included at a given level of probability. For instance, if  $t$  is biased, its “real” sampling distribution is not suitable to derive confidence intervals for the population L-CV.

### 3 Validity of the confidence interval estimators

Monte-Carlo simulations are used to verify if the candidate distributions provide valid confidence interval estimators for the population L-CV. Here a confidence interval estimator is said to be valid if the  $(1 - \alpha) \cdot 100\%$  confidence limits obtained for different samples encompasses the true L-CV value in  $(1 - \alpha) \cdot 100\%$  of the times; we want the estimator to be valid for every  $\alpha \in (0, 1)$ . In other words, a confidence interval estimator is valid if it provides unbiased confidence limits. Given a parent distribution whose characteristics (and real L-CV  $\tau$ ) are known,  $N = 10\,000$  samples of length  $n$  are randomly extracted from it. For each of them, the estimates of the sample L-CV  $t$  and its variance  $\widehat{\text{var}}(t)$  are calculated as shown in Sect. 2. By applying the approximations of Eqs. (9–14), we obtain  $N$  estimates of the distribution of  $t$  for each model. The candidate distribution  $P$  of  $t$  can be consistently used to estimate the confidence intervals for  $\tau$  if the probability of non-exceedance  $P(\tau)$  is uniformly distributed between 0 and 1 (which is a consequence of the fact that  $P(\tau)$  is the probability integral transform of  $\tau$ ; see e.g., Kottegoda and Rosso, 1997, Sect. 8.2.1). Therefore we test the uniformity of  $P(\tau)$  through goodness-of-fit methods.

Many goodness-of-fit tests for the uniform distribution exist (e.g., D’Agostino and Stephens, 1986). As in Laio and Tamea (2007) we adopt here a less formal but more revealing graphical method, based on a probability plot representation. Given, for one candidate distribution of  $t$ , the sample  $P_1(\tau), P_2(\tau), P_3(\tau), \dots, P_N(\tau)$  resulting from the  $N$  Monte-Carlo simulations, the probability plot represents the  $P_i(\tau)$  values versus their empirical cumulative distribution function,  $R_i/N$ . The shape of the resulting curve reveals if the sample of probabilities is approximately uniform, in which case the  $(P_i(\tau), R_i/N)$  points are close to the bisector of the diagram.

The graphical method allows one to investigate which are the reasons for a lack of validity of a candidate distribution as a tool to estimate the confidence interval for  $\tau$ , in particular evidencing the effects of bias of estimation for  $t$  and  $\widehat{\text{var}}(t)$ . In fact, the shape of the curves in the probability plot is suggestive of the encountered problem, since the steepness of the curves is larger where more  $P_i(\tau)$  points concentrate (see Fig. 2 in Laio and Tamea, 2007). If the curve is over the bisector in the left part of the graph and under it in the right part, the  $P_i(\tau)$  points are concentrated in the vicinity of the end points 0 and 1, which corresponds to have the  $\tau$  value that falls, more frequently than expected, on the tails



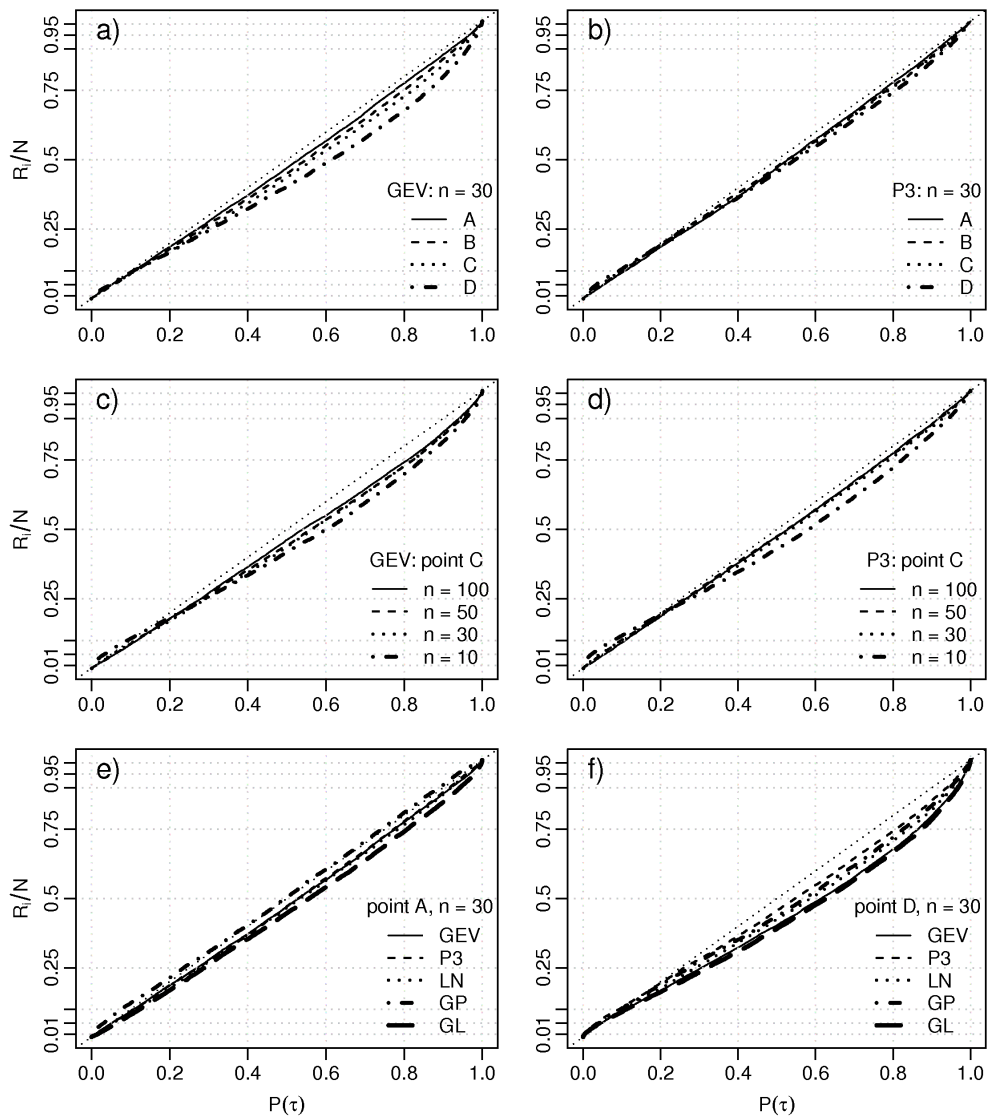
**Fig. 1.** L-CV-L-CA diagram. Lines are: (a) numerical constraint given by Hosking and Wallis (1997); (b) bisector band identified using Vogel and Wilson (1996) samples (see Viglione et al., 2007). Points A, B, C and D have L-CV  $\tau=0.2, 0.3, 0.4, 0.5$  and L-CA  $\tau_3=0.1, 0.2, 0.3, 0.4$ , respectively.

of the distributions. The  $N$  estimated variances are then too small. In the opposite case, when the curve has an S shape and crosses the bisector in the middle of the graph, the chosen variances are too high because the  $P_i(\tau)$  points are concentrated in the vicinity of the middle value 0.5. When the curves are always over or under the bisector, then the scale parameter of the distribution (which relates to  $t$ ) has been overestimated or underestimated, respectively. In the first case the real  $\tau$  value falls, more frequently than expected, on the low tail of the distributions, then the estimated  $t$  are too high, and viceversa.

In addition to the graphical method, we use the Anderson-Darling statistic in order to synthetically quantify the discrepancy between the cumulative distribution function (CDF) of  $P_1(\tau), P_2(\tau), P_3(\tau), \dots, P_N(\tau)$  and the uniform distribution  $U$  between 0 and 1. The Anderson-Darling statistic is a measure of the mean squared difference between the empirical and hypothetical CDF, in practice estimated as (e.g., D’Agostino and Stephens, 1986; Laio, 2004):

$$A^2 = -N - \frac{1}{N} \sum_{i=1}^N [(2i - 1) \ln(U[P_i(\tau)]) + (2N + 1 - 2i) \ln(1 - U[P_i(\tau)])]. \tag{15}$$

The best fit corresponds to the minimum value of  $A^2$ . The Anderson-Darling test statistic measures the validity of the estimator for all confidence levels en bloc ( $\forall \alpha \in (0, 1)$ ), which is the property we are looking for.



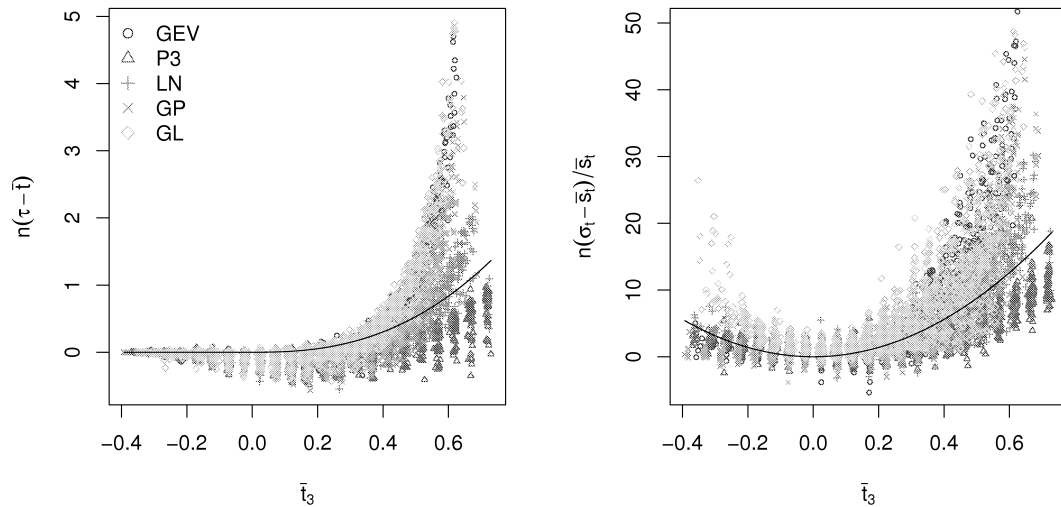
**Fig. 2.** Probability plot representation of  $P(\tau)$  when using the log-Student  $t$  distribution of Eq. (14) to estimate the confidence intervals of  $\tau$ . Different situations are considered: (a) GEV parent distributions with different asymmetry (point A, B, C, D of Fig. 1); (b) P3 parent distributions with different asymmetry; (c) GEV parent distribution and samples of different length  $n$ ; (d) P3 parent distribution and samples of different length  $n$ ; (e) different parent distributions with low asymmetry (point A); (f) different parent distributions with high asymmetry (point D).

#### 4 Results

To check the robustness of our approximations, many different situations are considered: we vary the length  $n$  of the samples, the underlying parent distribution and its parameters. The considered parent distributions are: GEV (Generalized Extreme Value), P3 (Pearson type III), LN (Log-Normal with 3 parameters), GP (Generalized Pareto) and GL (Generalized Logistic). As in Viglione et al. (2007) we choose the parameters of these distributions by reference to the L-CV-L-CA space (coefficients of L-variation and L-skewness) represented in Fig. 1 (the mean is taken equal to

1 without loss of generality) that is a reasonable space where the majority of hydrological (extreme value) samples falls (Vogel and Wilson, 1996; Viglione et al., 2007).

Some of the uniform probability plots of  $P(\tau)$  are shown in Fig. 2, where the log-Student  $t$  distribution of Eq. (14) is the candidate distribution for  $t$ . In panel (a) samples are extracted from a GEV distribution with increasing asymmetry (from point A to point D of Fig. 1). It is evident that the log-Student  $t$  approximation works better for moderate asymmetry of the parent distribution (point A) than in the highly asymmetric cases (points C and D). In all cases the curves lie below the bisector, meaning that the estimate  $t$  is too low



**Fig. 3.** Biases of  $t$  and  $s_t = \sqrt{\widehat{\text{var}}(t)}$ : (a)  $n \cdot (\tau - \bar{t})$  vs. the sample L-CA  $\bar{t}_3$ ; (b)  $n \cdot (\sigma_t - \bar{s}_t) / \bar{s}_t$  vs. the sample L-CA  $\bar{t}_3$ . Every point rises from 1 000 simulated samples and corresponds to a grey point in Fig. 1. The proposed corrections of Eqs. (16) and (17) are indicated with continuous black lines.

compared to  $\tau$ , specially in the right part of the graph, which implies that the upper tail of the distribution is too narrow. In panel (b) the parent distribution is Pearson Type III (P3). In this case the log-Student t approximation fits much better than in panel (a).

Similar conclusions can be derived from panels (c) and (d), that present the sensitivity of the log-Student t approximation to the length  $n$  of the samples (the parent distributions have, in this case, the same asymmetry, correspondent to point C of Fig. 1). As expected, for high  $n$  the goodness-of-fit of the candidate distribution increases. For very short samples ( $n = 10$ ), in some few cases (1% of the cases for GEV and 3% for P3), the estimate  $\widehat{\text{var}}(t)$  of Elamir and Seheult (2004) is negative. In those cases, the Monte-Carlo simulations have been discarded when producing the curves in Fig. 2c and d.

Panels (e) and (f) show the uniform probability plots for different parent distributions in points A (low asymmetry) and D (high asymmetry). The behavior of the curves, specially in D, visibly depends on the underlying parent distribution. Considering high asymmetries (point D), the approximation of the sample L-CV distribution with moments  $t$  and  $\widehat{\text{var}}(t)$  provides particularly bad confidence intervals for  $\tau$  for the GEV and GL distributions.

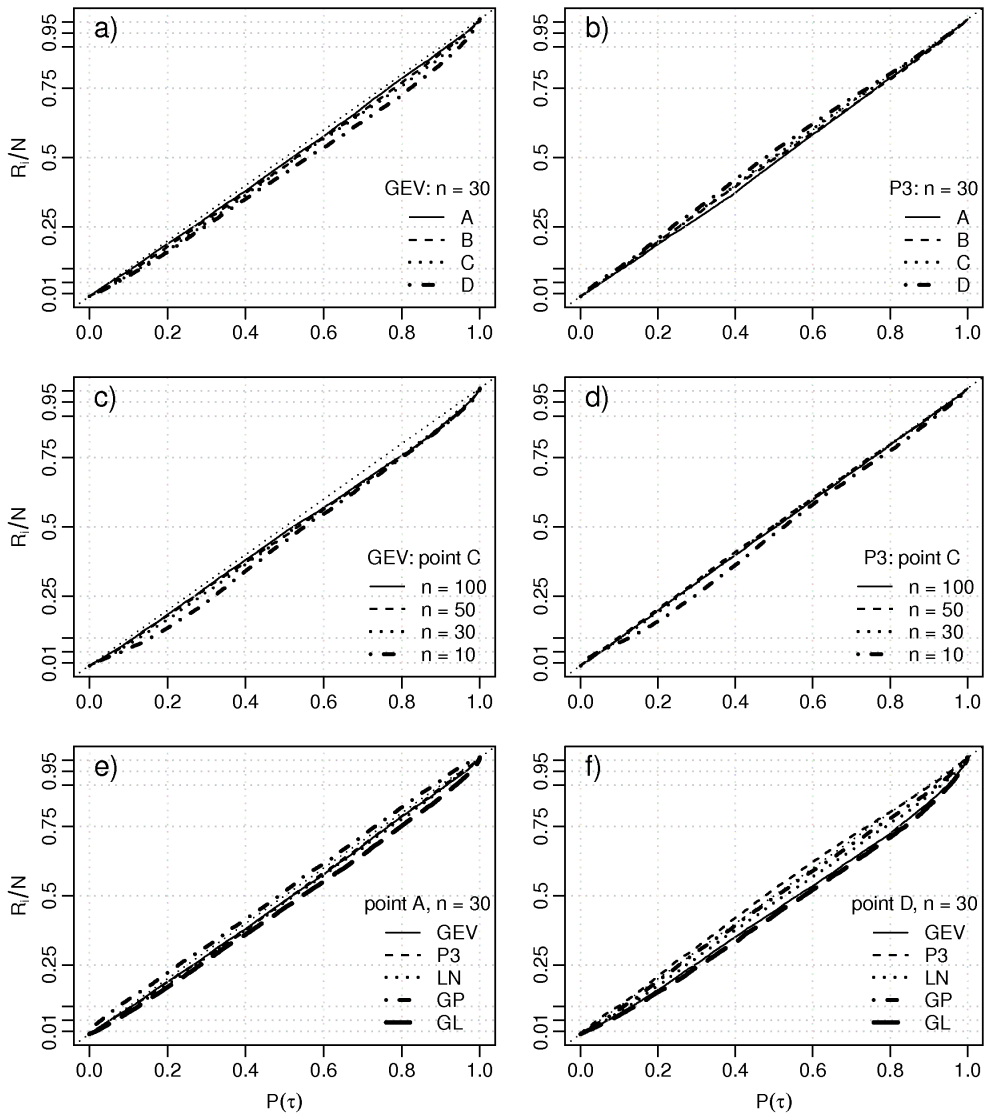
Similar plots (not shown here) have been produced for the other candidate distributions. The Normal and Student t distributions perform slightly better in the central part of the graph for small asymmetry (point A) and for the P3 parent distribution. Anyway the underestimation of the width of the upper tail is more marked than in the case of the asymmetric candidate distributions (gamma, log-Normal and log-Student). For all candidate distributions the shape of the curves are similar (below the bisector, specially for the up-

per right corner of the uniformity plots), meaning that, particularly for the GEV and GL parent distributions and for large asymmetries (point D), the estimators  $t$  and  $\widehat{\text{var}}(t)$  underestimate the true values. We consider this problem in the following subsection, where a distribution-free correction of these biases is proposed.

#### 4.1 Bias correction

No matter what type of distribution is chosen, in many cases the sampling distribution of the L-CV does not provide a valid estimator of the confidence intervals for the population L-CV. This could be expected since it is known that the sample L-CV is slightly biased for high asymmetry of the parent distribution (Hosking and Wallis, 1997, page 28) and since  $\widehat{\text{var}}(t)$  is obtained from the Taylor-series-based approximation of Eq. (8). To provide a valid estimator, these biases should be corrected.

A simple bias correction for  $t$  and  $s_t = \sqrt{\widehat{\text{var}}(t)}$  can be obtained from Monte-Carlo simulations. Samples are generated for each combination of  $\tau$  and  $\tau_3$  (grey points in Fig. 1), for each parent distribution (GEV, P3, LN, GP and GL) and for different sample lengths ( $n = 10, 15, 20, 25, 30, 40, 50, 70, 100$ ). Given one simulation ( $N=1\,000$  samples, in this case), the bias of  $t$  is estimated as  $\bar{t} - \tau$ , where  $\bar{t}$  is the arithmetic mean over  $N$  values of  $t$  and  $\tau$  is the known L-CV of the parent distribution. In panel (a) of Fig. 3, we show how the bias of  $t$ , opportunely scaled with  $n$  (see e.g., Kendall and Stuart, 1961, Sect. 17.10), increases for increasing asymmetry of the samples ( $\bar{t}_3$  is the mean of the  $N$  sample L-skewness). Analogous results are obtained



**Fig. 4.** Probability plot representation of  $P(\tau)$  when assuming the log-Student  $t$  distribution of Eq. (14) and applying the corrections of Eqs. (16) and (17). The same cases (a–f) of Fig. 2 are considered here.

for the standard deviation of the sample L-CV (panel (b) of Fig. 3). The bias of  $s_t$  is evaluated as  $\bar{s}_t - \sigma_t$ , where we assume  $\sigma_t = \sqrt{(\sum_{i=1}^N (t_i - \bar{t})^2)/N}$  as the true standard deviations of the sample L-CV.

In both cases, the bias depends on the type of distribution: the P3 distribution shows the lowest bias and the GL distribution the largest, confirming the behavior in Fig. 2. Given one distribution, the scatter of the estimated biases for a fixed  $t_3$  is mostly determined by the variability of the considered L-CV (grey points on vertical lines in Fig. 1) while the influence of the variability of sample lengths is highly reduced by the aforementioned scaling with  $n$ . Despite this effect of the variance of the parent distribution, we deem that the biases of  $t$  and  $s_t$  are substantially functions of  $n$  and  $t_3$ , along with the distribution type.

Using these Monte Carlo results in a pragmatic way, we propose the following simple corrections for the sample L-CV and its standard deviation:

$$t^{(c)} = \begin{cases} t + 3 \cdot t_3^{2.5} / n, & \text{if } t_3 > 0 \\ t, & \text{if } t_3 \leq 0, \end{cases} \quad (16)$$

and

$$s_t^{(c)} = s_t \cdot (1 + 35 \cdot t_3^2 / n), \quad (17)$$

that are obtained with a non-linear regression approach (non linear least-squares). The curves corresponding to these corrections are represented in Fig. 3 as solid black lines. These corrections intentionally do not take into account the type of distribution because this is unknown in operational applications. It is important to stress again that the objective of this

**Table 1.** Anderson-Darling statistic of discrepancy between the CDF of 10 000 values of  $P(\tau)$  and the uniform distribution  $U$ . The columns show the value of the statistic for the five candidate distributions with parameters corrected by Eqs. (16) and (17). The best fit is indicated by the bold font. Different situations are considered: **(a)** GEV parent distributions with different asymmetry (point A, B, C, D of Fig. 1); **(b)** P3 parent distributions with different asymmetry; **(c)** GEV parent distribution and samples of different length  $n$ ; **(d)** P3 parent distribution and samples of different length  $n$ ; **(e)** different parent distributions with low asymmetry (point A); **(f)** different parent distributions with high asymmetry (point D).

		normal	Student t	gamma	log-normal	log-Student t
<b>(a)</b> GEV $n = 30$	A	33.2	23.6	21.9	23.2	<b>16.6</b>
	B	53.8	42.3	43.7	46.6	<b>39.6</b>
	C	91.6	75.4	79.0	83.1	<b>73.4</b>
	D	226.2	191.4	209.2	212.9	<b>189.8</b>
<b>(b)</b> P3 $n = 30$	A	21.3	16.0	14.0	16.7	<b>13.3</b>
	B	31.1	22.0	15.9	14.0	<b>7.9</b>
	C	27.9	21.4	10.6	5.9	<b>2.5</b>
	D	50.6	41.2	29.3	16.3	<b>10.4</b>
<b>(c)</b> GEV C	$n = 100$	66.5	62.1	60.9	60.3	<b>56.7</b>
	$n = 50$	95.1	84.0	85.4	85.5	<b>77.5</b>
	$n = 30$	91.6	75.4	79.0	83.1	<b>73.4</b>
	$n = 10$	296.8	<b>154.7</b>	226.5	243.5	171.1
<b>(d)</b> P3 C	$n = 100$	7.3	6.5	3.2	2.3	<b>1.7</b>
	$n = 50$	19.5	16.1	8.3	4.5	<b>2.2</b>
	$n = 30$	27.9	21.4	10.6	5.9	<b>2.5</b>
	$n = 10$	134.9	52.0	95.2	98.3	<b>47.9</b>
<b>(e)</b> $n = 30$ A	GEV	33.2	23.6	21.9	23.2	<b>16.6</b>
	P3	21.3	16.0	14.0	16.7	<b>13.3</b>
	LN	34.0	26.4	26.9	30.2	<b>25.2</b>
	GP	45.5	39.2	40.2	39.2	<b>31.4</b>
	GL	140.4	115.4	122.2	124.2	<b>107.8</b>
<b>(f)</b> $n = 30$ D	GEV	226.2	191.4	209.2	212.9	<b>189.8</b>
	P3	50.6	41.2	29.3	16.3	<b>10.4</b>
	LN	85.5	67.4	70.4	70.4	<b>59.4</b>
	GP	46.3	33.4	28.6	23.3	<b>15.7</b>
	GL	328.2	<b>283.9</b>	309.9	315.9	286.0

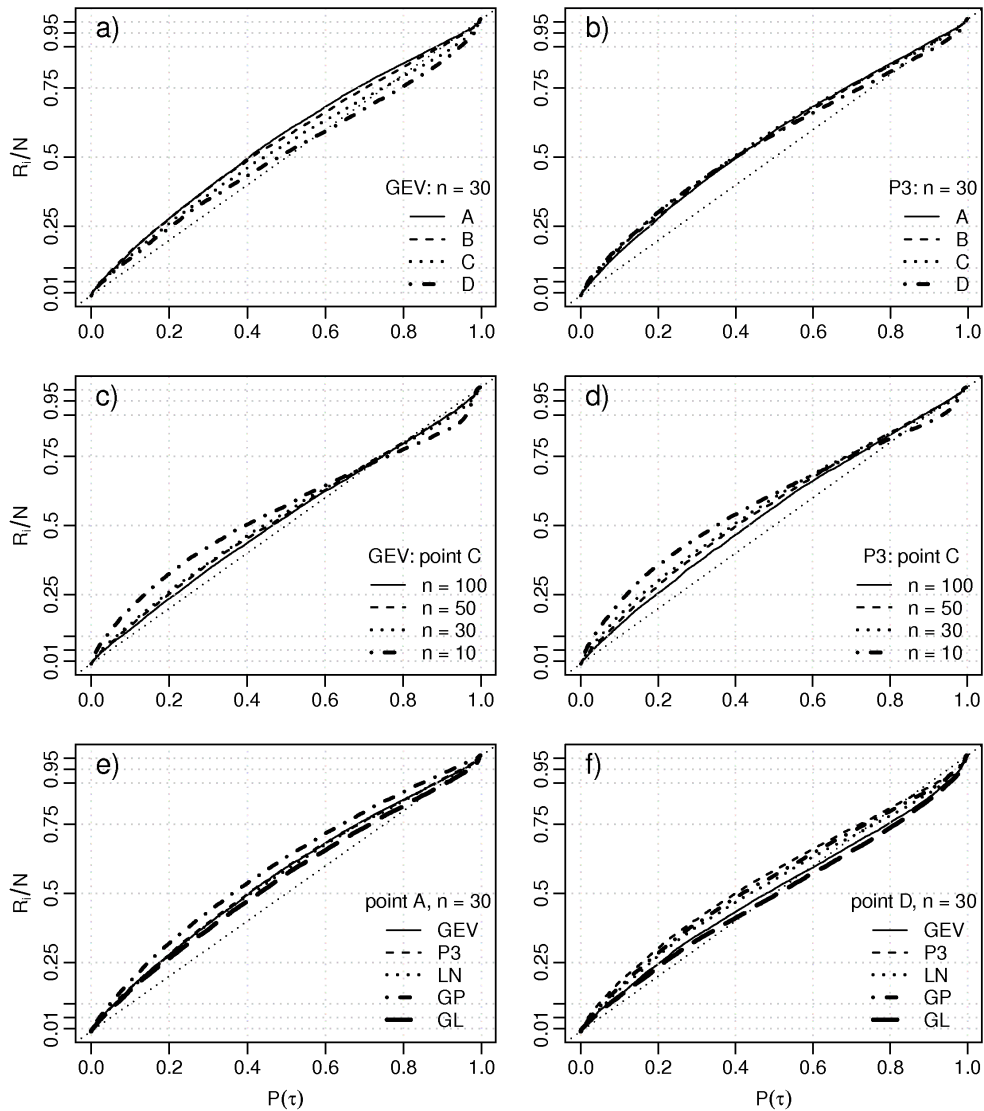
study is to derive an interval estimator of  $\tau$  which is independent of the underlying (and unknown) parent distribution of the original variable, so to provide an operational tool.

As can be seen in Fig. 4, which is analogous to Fig. 2 but considers the adjustments of Eqs. (16) and (17), the corrections generally improve the results of the uniformity tests for highly asymmetric distributions. Some problems remain for very short samples ( $n = 10$ ) and some parent distributions, but the fit generally improves. Similar results are obtained for the other candidate distributions. A synthetic comparison of these results is shown in Table 1, where the values of the Anderson-Darling statistic is shown for the cases of Fig. 4. In most of the cases, the  $A^2$  statistic is lower when the log-Student t distribution is assumed. Therefore, among the considered candidates, we select the log-Student t as the best approximated distribution to derive confidence intervals for the population L-CV.

## 4.2 The bootstrap approach

The parametric strategy is not the only possibility to construct confidence intervals. As an alternative, non-parametric bootstrap procedures can be used to make inference on  $t$  without making assumption on its distribution. The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates. It was first introduced in the context of non parametric analysis of independent and identically distributed samples (Efron, 1979), but much research into its use in more complicated settings followed (see e.g., Davison and Hinkley, 1997, as a reference). The basic idea behind the non-parametric bootstrap is to replicate the original sample many times, sampling it with replacement, and to analyse the behavior of the statistic of interest calculated on these replicates. Many aspects of the behavior of the selected statistic can be measured with bootstrap, for example its confidence intervals.





**Fig. 5.** Probability plot representation of  $P(\tau)$  for different situations if the  $BC_a$  distribution is used (Eq. 18). The same cases (a–f) of Fig. 2 are considered here.

One of the most recommended bootstrap techniques to evaluate confidence intervals is the so called bias-corrected and accelerated ( $BC_a$ ) bootstrap method. The algorithm (see e.g., Efron and Tibshirani, 1993, for a formal definition) is available in many software packages. Given the definition of  $BC_a$  intervals in Efron and Tibshirani (1993, pages 184–188), if  $P^*(\tau)$  is the probability of non exceedance of  $\tau$  on the bootstrap distribution of  $t$ , the corrected probability given by the  $BC_a$  method is

$$P(\tau) = \Phi \left[ \frac{\hat{a} \cdot \hat{z}_0^2 + (-\hat{a} \cdot \Phi^{-1}[P^*(\tau)] - 2) \cdot \hat{z}_0 + \Phi^{-1}[P^*(\tau)]}{\hat{a} \cdot \hat{z}_0 - \hat{a} \cdot \Phi^{-1}[P^*(\tau)] - 1} \right], \quad (18)$$

where  $\Phi[.]$  is the standard normal cumulative distribution function,  $\Phi^{-1}[.]$  is the percentile point on a standard normal distribution and the values of the bias correction  $\hat{z}_0$  and

the acceleration  $\hat{a}$  can be computed as indicated in Efron and Tibshirani (1993, page 186).

The uniformity plot introduced in Sect. 3 has been applied to check if the calculated  $BC_a$  distribution of Eq. (18) provides a valid confidence interval estimator for  $\tau$ . As shown in Fig. 5, the bias of the central part of the curves indicates a significant overestimation of  $\tau$ , i.e., the bias correction of the  $BC_a$  method is too strong. Moreover some of the curves, specially those correspondent to high asymmetry and small sample length, are “reverse S” shaped, meaning that the variance of the distribution is underestimated. The comparison between Figs. 5 and 4 confirms the better performance of the parametric method proposed in this paper.

**Table 2.** Coverage of the 90% confidence intervals for  $\tau$  using the following candidate distributions: corrected parametric model with normal, Student t, gamma, log-normal and log-Student t distribution, bootstrap BC<sub>a</sub> model with 2000 replicates.  $P_{05}$  and  $P_{95}$  are the percentage of trials (that are 10 000) that the indicated interval missed the true value  $\tau$  on the left or right side of the confidence interval. The same cases (a–f) of Table 1 are considered here.

		normal		Student t		gamma		log-normal		log-Student t		bootstrap BC <sub>a</sub>	
		$P_{05}$	$P_{95}$	$P_{05}$	$P_{95}$	$P_{05}$	$P_{95}$	$P_{05}$	$P_{95}$	$P_{05}$	$P_{95}$	$P_{05}$	$P_{95}$
<b>(a)</b>	A	4.2	8.0	3.7	7.5	4.9	7.3	5.2	7.0	4.8	6.5	8.6	5.2
	GEV	3.5	8.6	3.1	8.0	4.2	7.8	4.5	7.5	4.0	6.9	8.8	5.8
	$n = 30$	3.2	9.3	2.8	8.8	3.9	8.6	4.0	8.3	3.6	7.7	8.2	7.9
	D	3.1	11.9	2.7	11.2	3.9	11.0	3.9	10.6	3.4	9.9	7.7	9.4
<b>(b)</b>	A	4.0	7.2	3.5	6.6	4.8	6.4	5.1	6.1	4.6	5.5	8.5	4.9
	P3	4.4	7.9	3.9	7.2	5.0	7.1	5.3	6.7	4.7	6.2	9.5	5.0
	$n = 30$	4.7	7.4	4.1	6.8	5.3	6.7	5.3	6.3	4.8	5.8	9.8	5.8
	D	5.8	7.1	5.1	6.6	6.4	6.5	6.2	6.2	5.6	5.6	10.5	7.1
<b>(c)</b>	$n = 100$	3.7	8.9	3.5	8.6	4.2	8.3	4.4	8.1	4.2	7.9	7.0	6.8
	GEV	3.2	9.8	2.9	9.5	3.8	9.2	4.0	8.8	3.7	8.5	7.5	6.7
	$n = 50$	3.2	9.3	2.8	8.8	3.9	8.6	4.0	8.3	3.6	7.7	8.2	7.9
	C	3.4	12.4	2.4	10.5	4.1	11.2	4.0	10.6	3.2	8.5	11.8	12.2
<b>(d)</b>	$n = 100$	4.5	6.2	4.2	6.0	4.7	5.8	4.8	5.7	4.7	5.5	7.7	4.8
	P3	4.6	6.7	4.3	6.4	5.3	6.2	5.3	5.9	5.0	5.6	8.6	5.0
	$n = 50$	4.7	7.4	4.1	6.8	5.3	6.7	5.3	6.3	4.8	5.8	9.8	5.8
	C	5.6	10.1	4.6	8.4	6.5	9.0	6.1	8.5	5.0	6.6	13.2	8.7
<b>(e)</b>	GEV	4.2	8.0	3.7	7.5	4.9	7.3	5.2	7.0	4.8	6.5	8.6	5.2
	$n = 30$	4.0	7.2	3.5	6.6	4.8	6.4	5.1	6.1	4.6	5.5	8.5	4.9
	A	3.9	7.6	3.5	7.1	4.5	7.0	4.8	6.6	4.4	6.1	8.5	5.5
	GP	7.0	4.7	6.4	4.2	7.6	4.1	8.0	4.0	7.2	3.5	10.1	3.2
GL	2.8	10.5	2.5	9.9	3.7	9.7	4.0	9.3	3.6	8.7	7.9	6.6	
<b>(f)</b>	GEV	3.1	11.9	2.7	11.2	3.9	11.0	3.9	10.6	3.4	9.9	7.7	9.4
	$n = 30$	5.8	7.1	5.1	6.6	6.4	6.5	6.2	6.2	5.6	5.6	10.5	7.1
	P3	3.6	9.7	3.1	9.2	4.2	9.1	4.1	8.6	3.7	7.9	8.3	8.9
	D	4.3	8.5	3.8	8.0	4.9	7.9	4.8	7.5	4.3	7.0	9.2	7.7
	GL	3.0	13.0	2.7	12.3	3.7	12.1	3.7	11.5	3.3	10.6	7.1	10.2

### 4.3 90% confidence intervals

As was shown in Table 1, in most of the cases the log-Student t distribution gives the best approximation to derive confidence intervals of the L-CV in general, considering all confidence levels. Since the Anderson-Darling statistic gives more weight to the tails of the distribution than to the central part of it, the log-Student t distribution is expected to be the best choice when calculating the confidence intervals for the sample L-CV at high confidence levels. Here we check the goodness of the 90% confidence intervals for  $\tau$  using different candidate distributions in the corrected case, where the corrections of Eqs. (16) and (17) are applied to  $t$  and  $s_t$ , and using the BC<sub>a</sub> non-parametric bootstrap. Table 2 shows the percentage of trials (which are 10 000) in which the true value  $\tau$  was not comprised into the 90% confidence intervals on the left or right side. The target miscoverage is 5% on each side:  $P_{05}$  and  $P_{95}$  are the probabilities that  $\tau$  is lower than the 0.05 or greater than the 0.95 quantiles of the corrected candidate distributions (we expect that both

$P_{05}$  and  $P_{95}$  equal 5%). Among all the considered situations, Table 2 shows the cases considered already in Table 1 and Figs. 2, 4 and 5.

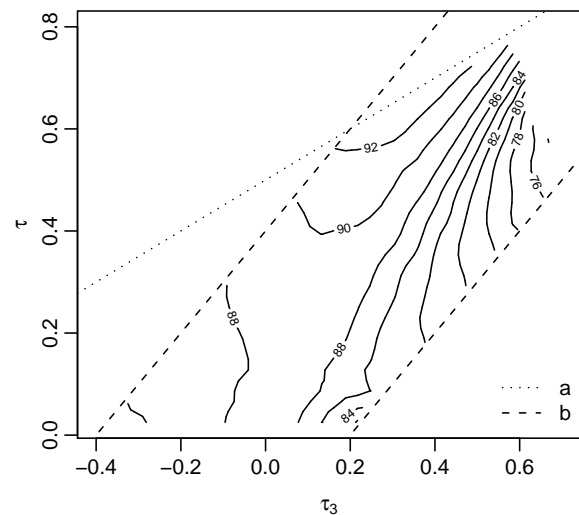
When using the parametric method, independently of the chosen candidate distribution,  $P_{05}$  is often slightly lower than 5% and  $P_{95}$  is also higher than 5%. This means that the confidence intervals for  $\tau$  are a little too large for the left tail and too narrow for the right tail. This problem is particularly evident for large asymmetries of the parent distribution and for small sample sizes, as was already evident in Fig. 4. The log-Student t approximation generally outperforms the others, i.e., the confidence intervals are better centered (both  $P_{05}$  and  $P_{95}$  are close to the ideal value 5%). The BC<sub>a</sub> intervals (2000 replicates have been used) have a different behavior:  $P_{05}$  is always higher than 5% and in many cases higher than  $P_{95}$ , which is a consequence of the overestimation of  $\tau$  shown in Fig. 5. Despite its greater computational demand, the bootstrap performs in most of the cases worse than the corrected parametric method.

Fig. 6 extends the results of Table 2 to other points of the L-CV–L-CA plain, when the log-Student  $t$  distribution is assumed. The contour lines give the coverage frequency of the 90% confidence intervals when considering 10 000 samples of length 30 for each of the grey points in Fig. 1. The coverage frequency in Fig. 6 is the average obtained by extracting samples from all the five parent distributions considered in this work (GEV, P3, LN, GP, GL). In the upper-right part of the figure the results are not optimal (i.e., the values are below 90%), due to the biases of the sample L-CV and its variance shown in Fig. 3, which still affect the results after correcting  $t$  and  $s_t$  with Eqs. (16) and (17). However, not many sample values encountered in hydrology lie in that region of the L-CV–L-CA space. For instance, if one considers the UK flood peak data available with the Flood Estimation Handbook (Robson and Reed, 1999), only 60 out of 696 series have sample L-CV and L-CA values lying in an area of the diagram where the coverage is lower than 85% and only 2 in an area where the coverage is lower than 80%.

## 5 An application

In order to demonstrate that the possibility to estimate confidence intervals for the L-CV is relevant for hydrological applications, here we show, in a simplified way, how these confidence intervals could be used to analyse and compare the outcome of different regionalization methods. Following Hosking and Wallis (1997), cluster analysis can be used to define homogeneous disjoint regions whose samples should have the same L-CV. On the other hand one can use the region of influence (ROI) approach (Burn, 1990), or allow for the continuous variability of L-CV (Robinson and Sivapalan, 1997), by estimating it with linear or non-linear regressions, or use geostatistical approaches (Merz and Blöschl, 2005; Skjøien et al., 2006). All these different approaches can be compared using confidence intervals in an intuitive way.

Here we use the UK data available with the Flood Estimation Handbook (Robson and Reed, 1999). For every regionalization technique, morphoclimatic variables should be used to pool the sites. For simplicity, we suppose that the mean annual rainfall alone (indicated as SAAR) may explain the difference of the at-site flood frequency distributions, and that this difference is completely reflected by the coefficient of L-variation. Fig. 7 shows in a simple way how four methods of estimation of the regional L-CV can be compared. Mean annual precipitation is plotted against the sample L-CV. The modeled regional L-CV is plotted as a black line. Gray circles have 90% confidence intervals that contain the modeled regional L-CV while black crosses do not – we use the log-Student  $t$  distribution with parameters corrected using Eqs. (16) and (17). In panel (a) only one region is considered with a unique regional L-CV, here called  $t^R$ , given by the arithmetic mean of all  $t$ . Assuming that the confidence intervals are correctly estimated, if the region is ho-

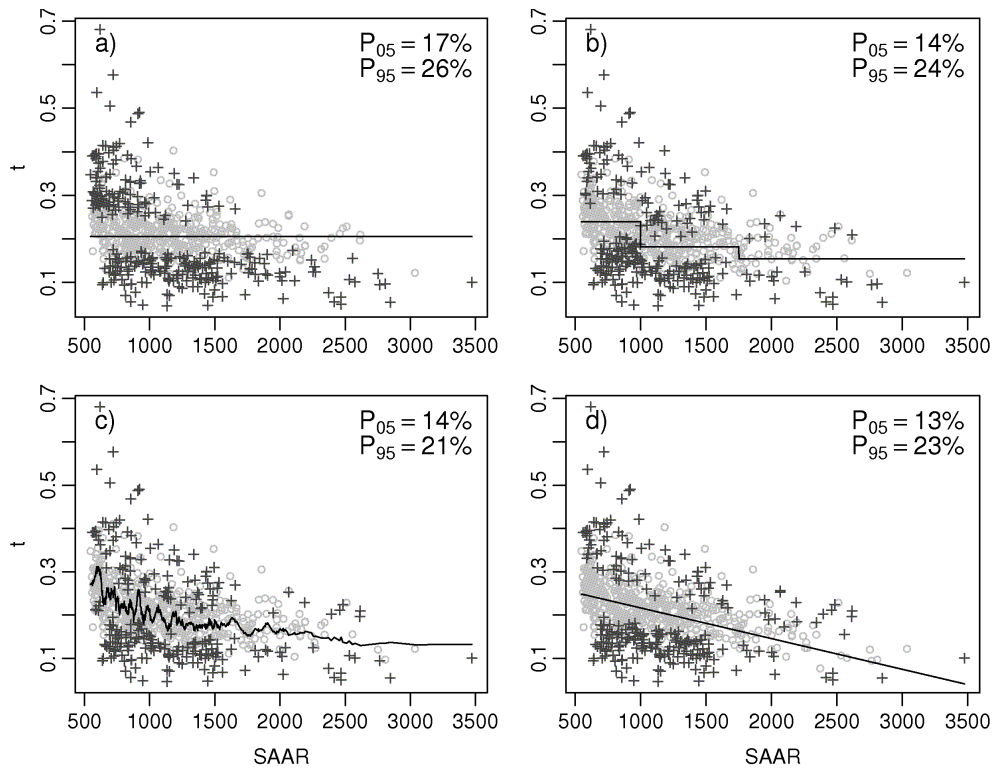


**Fig. 6.** L-CV–L-CA diagram analogous to Fig. 1. The contour lines give the average coverage frequency of the 90% confidence intervals when considering 10 000 samples of length 30, extracted from the five considered parent distributions (GEV, P3, LN, GP and GL), for each of the grey points in Fig. 1. The log-Student  $t$  distribution of Eq. (14) with corrected parameters (Eqs. 16 and 17) provides the estimator for the confidence intervals for  $\tau$ .

mogeneous the percentage  $P_{05}$  of sites for which the modeled  $t^R$  is below the confidence intervals for  $\tau$  should be equal to 5%, as well as  $P_{95}$ , i.e., the percentage of sites for which the modeled  $t^R$  is above the confidence intervals for the population L-CV. In panel (a)  $P_{05} = 17\%$  and  $P_{95} = 26\%$ , i.e.,  $100\% - (P_{05} + P_{95}) = 57\% \ll 90\%$ , which means that the region is heterogeneous.

If one subdivides the sites in 3 regions using a clustering method, as in panel (b), the values of  $P_{05}$  and  $P_{95}$  decrease ( $P_{05} = 14\%$  and  $P_{95} = 24\%$ ). Still the regional model does not explain properly the heterogeneity of the data, because of its structure and/or because the chosen parameter (SAAR) does not explain completely the variability of the L-CV. Anyway the model is more appropriate than the previous one, i.e., it can explain some more spatial heterogeneity of the data. With a region of influence approach (panel c), one obtains  $P_{05} = 14\%$  and  $P_{95} = 21\%$ . In this case, for simplicity, a simplified ROI approach has been used, that assigns to each site the average of the L-CV of the most similar 20 sites in terms of SAAR. Finally, in panel (d) a linear regression between  $t$  and SAAR is shown. The result ( $P_{05} = 13\%$  and  $P_{95} = 23\%$ ) is comparable to what we got with the three regions and ROI clustering methods.

It is important to note that the high discordance between  $P_{05}$  and  $P_{95}$  and the theoretical value of 5% is not only due to the approximation of the confidence intervals, for which we would expect uncertainties of the same order of those in Table 2 and Fig. 6 (seemingly, the fact that  $P_{95} > P_{05}$  is partly due to the underestimation of the upper confidence limit for



**Fig. 7.** Sample L-CV  $t$  versus mean annual precipitation SAAR in UK (data from the Flood Estimation Handbook, Robson and Reed, 1999). The black continuous line is the modeled regional value  $t^R$ ; grey circles ( $\circ$ ) are sites for which the modeled  $t^R$  is inside the 90% confidence interval for  $\tau$ , when assuming the log-Student  $t$  sampling distribution of Eq. (14) and applying the corrections of Eqs. (16) and (17); black crosses (+) are sites for which it is outside of it. Four models are considered: (a) one unique region where  $t^R = \bar{t}$ ; (b) three regions with different SAAR; (c) regions obtained with a simplified ROI approach; (d) linear regression model.

highly asymmetric parent distributions). This discordance is mainly due to the fact that the data are heterogeneous, that SAAR is not sufficient to explain the regional variability of L-CV, that the regional methods are approximated, etc. Lower values of  $P_{05}$  and  $P_{95}$  would be obtained by using more catchment descriptors (not only SAAR), more sophisticated techniques or combining different sources of hydrological information (Merz and Blöschl, 2008). In the extreme case, considering models with a number of parameters that approach the number of stations, one could reduce  $P_{05}$  and  $P_{95}$  also below the nominal value of 5%, which would clearly indicate overparametrisation.

The simple examples provided here intend to illustrate the method. They show that L-CV confidence intervals allow one to analyse in a consistent way very different approaches such as those based on site grouping and those that allow for the continuous variability of L-CV, for which standard techniques as homogeneity tests would be meaningless. Also, they provide a way to identify those sites that are “problematic” for the validity of the assumptions made (in the example looking at the position of crosses and circles) and that should be further checked.

## 6 Conclusions

One of the most important concerns of Flood Frequency Analysis is the underlying distribution (or the lack of knowledge on its form, to be more precise) and the desire to lose the “distribution fetters” among hydrologists is really strong. Anyway the distribution-dependent methodologies dominate in practical hydrology, where distribution-free methods are seldom used. Recently Elamir and Seheult (2004) have provided a method to estimate the variance of sample L-moments without formulating assumptions on their parent distributions. In this study, we use this result to provide estimates of the confidence intervals for the L-CV. In summary, the key steps of the work are:

- in order to provide the basis for the assessment of the method, a Monte-Carlo procedure is used to draw many samples from distributions commonly used in hydrology (GEV, P3, LN, GP, GL);
- candidate sampling distributions of the L-CV are selected among five distribution types (normal, Student  $t$ , gamma, log-normal and log-Student  $t$ ) whose

parameters are estimated from the sample L-CV and its variance, this latter given by the distribution-free method of Elamir and Seheult (2004);

- the validity of the confidence intervals for the L-CV provided by these distributions is checked through uniformity plots and the Anderson-Darling test statistic;
- no matter which candidate distribution is used, the experiments show that the validity of the confidence intervals is affected by biases of the sample L-CV and its variance, particularly when the parent distribution of the samples is highly skewed and the samples are short;
- therefore distribution-free corrections for these biases are proposed, based on sample length and asymmetry;
- the validity of the new estimators of the confidence intervals for the L-CV, obtained using the corrected sample L-CV and variance, is checked again and the log-Student t distribution proves to be the best distribution type in most of the cases.

Even if in some circumstances the validity of the confidence interval estimator still depends on the underlying parent distribution, this paper shows that the proposed approximation is in most of the cases reasonably good (and better than the bias-corrected and accelerated non-parametric bootstrap) and can be recommended for practical hydrological applications. For example, the confidence intervals for the L-CV have been used here to compare different regional models in terms of their capability to explain the spatial heterogeneity of flood peak data in UK.

*Acknowledgements.* Financial support for the project “Mountain floods – regional joint probability estimation of extreme events” from the Austrian Academy of Sciences is acknowledged. I would like to thank Björn Guse and an anonymous reviewer for their useful comments on the manuscript in HESSD. I would also like to thank José Luis Salinas and Francesco Laio for their precious help on the revision of the final paper.

Edited by: F. Laio

## References

- Allamano, P., Claps, P., and Laio, F.: An analytical model of the effects of catchment elevation on the flood frequency distribution, *Water Resour. Res.*, 45(1), W01402, doi:10.1029/2007WR006658, 2009.
- Burn, D. H.: Evaluation of regional flood frequency analysis with a region of influence approach, *Water Resour. Res.*, 26(10), 2257–2265, 1990.
- Castellarin, A., Burn, D. H., and Brath, A.: Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, *J. Hydrol.*, 241, 270–285, 2001.
- Chowdhury, J. U., Stedinger, J. R., and Lu, L.: Goodness-of-fit tests for regional generalized extreme value flood distributions, *Water Resour. Res.*, 27(7), 1765–1776, 1991.
- D’Agostino, R. and Stephens, M.: Goodness-of-fit techniques, *Statistics: textBooks and monographs*, 68, Department of Statistics, Southern Methodist University, Dallas, Texas, 1986.
- Dalrymple, T.: Flood frequency analyses, *Water Supply Paper*, 1543-A, US Geological Survey, Reston, Va., 1960.
- Davison, A. C. and Hinkley, D. V.: *Bootstrap methods and their application*, Cambridge University Press, Cambridge, 1997.
- Efron, B.: Bootstrap methods: another look at the jackknife, *Ann. Stat.*, 7, 1–26, 1979.
- Efron, B. and Tibshirani, R.: *An Introduction to the Bootstrap*, Chapman & Hall, New York, 1993.
- Elamir, E. A. H. and Seheult, A. H.: Exact variance structure of sample l-moments, *J. Stat. Plan. Inf.*, 124, 337–359, 2004.
- Fill, H. and Stedinger, J. R.: Using regional regression within index flood procedures and an empirical bayesian estimator, *J. Hydrol.*, 210(1–4), 128–145, 1998.
- Greenwood, J., Landwehr, J., Matalas, N., and Wallis, J.: Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form, *Water Resour. Res.*, 15, 1049–1054, 1979.
- Hosking, J.: The theory of probability weighted moments, *Tech. Rep. RC12210*, IBM Research, Yorktown Heights, NY, 1986.
- Hosking, J.: L-moments: analysis and estimation of distributions using linear combinations of order statistics, *J. Royal Statistical Soc.*, 52, 105–124, 1990.
- Hosking, J. and Wallis, J.: Some statistics useful in regional frequency analysis, *Water Resour. Res.*, 29(2), 271–281, 1993.
- Hosking, J. and Wallis, J.: *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press, 1997.
- Kendall, M. and Stuart, A.: *The Advanced Theory of Statistics*, Charles Griffin & Company Limited, London, 1961.
- Kottegoda, N. T. and Rosso, R.: *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, international ed., 735 pp., McGraw-Hill Companies, 1997.
- Koutsoyiannis, D.: Uncertainty, entropy, scaling and hydrological stochasticity. 1. marginal distributional properties of hydrological processes and state scaling, *Hydrolog. Sci. J.*, 50(3), 381–404, 2005.
- Laio, F.: Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters, *Water Resour. Res.*, 40, W09308, doi:10.1029/2004WR003204, 2004.
- Laio, F. and Tamea, S.: Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrol. Earth Syst. Sci.*, 11, 1267–1277, doi:10.5194/hess-11-1267-2007, 2007.
- Merz, R. and Blöschl, G.: Flood frequency regionalisation-spatial proximity vs. catchment attributes, *J. Hydrol.*, 302, 283–306, 2005.
- 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, 2008.
- Pearson, C.: New zealand regional flood frequency analysis using l moments, *J. Hydrol. (NZ)*, 30, 53–64, 1991.
- Robinson, J. and Sivapalan, M.: An investigation into the physical causes of scaling and heterogeneity of regional flood frequency, *Water Resour. Res.*, 33(5), 1045–1059, 1997.
- Robson, A. J. and Reed, D. W.: Statistical procedures for flood frequency estimation, in *Flood Estimation Handbook*, Institute of Hydrology Crowmarsh Gifford, Wallingford, Oxfordshire,

- vol. 3, p. 338, 1999.
- Sankarasubramanian, A. and Srinivasan, K.: Investigation and comparison of sampling properties of l-moments and conventional moments, *J. Hydrol.*, 218, 13–34, 1999.
- Sillitto, G.: Derivation of approximants to the inverse distribution function of a continuous univariate population from the order statistics of a sample, *Biometrika*, 56(3), 641–650, doi: 10.1093/biomet/56.3.641, 1969.
- Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging – geostatistics on stream networks, *Hydrol. Earth Syst. Sci.*, 10, 277–287, doi:10.5194/hess-10-277-2006, 2006.
- Stedinger, J. R. and Lu, L.: Appraisal of regional and index flood quantile estimators, *Stoch. Hydrol. Hydraul.*, 9(1), 49–75, 1995.
- Viglione, A., Laio, F., and Claps, P.: A comparison of homogeneity tests for regional frequency analysis, *Water Resour. Res.*, 43(3), W03428, doi:10.1029/2006WR005095, 2007.
- Viglione, A.: nsRFA: Non-supervised Regional Frequency Analysis, R package, <http://cran.r-project.org/web/packages/nsRFA/>, 2010.
- Vogel, R. M. and Fennessey, N.: L moment diagrams should replace product moment diagrams, *Water Resour. Res.*, 29(6), 1745–1752, 1993.
- Vogel, R. M. and Wilson, I.: Probability distribution of annual maximum, mean, and minimum streamflows in the united states, *J. Hydrol. Eng.*, 1(2), 69–76, 1996.