

Coverage probability of bootstrap confidence intervals in heavy-tailed frequency models, with application to precipitation data

Jan Kysely

Received: 4 March 2009 / Accepted: 17 July 2009 / Published online: 22 August 2009
© Springer-Verlag 2009

Abstract Bootstrap, a technique for determining the accuracy of statistics, is a tool widely used in climatological and hydrological applications. The paper compares coverage probabilities of confidence intervals of high quantiles (5- to 200-year return values) constructed by the nonparametric and parametric bootstrap in frequency analysis of heavy-tailed data, typical for maxima of precipitation amounts. The simulation experiments are based on a wide range of models used for precipitation extremes (generalized extreme value, generalized Pareto, generalized logistic, and mixed distributions). The coverage probability of the confidence intervals is quantified for several sample sizes ($n=20, 40, 60,$ and 100) and tail behaviors. We show that both bootstrap methods underestimate the width of the confidence intervals but that the parametric bootstrap is clearly superior to the nonparametric one. Even a misspecification of the parametric model—often unavoidable in practice—does not prevent the parametric bootstrap from performing better in most cases. A tendency to narrower confidence intervals from the nonparametric than parametric bootstrap is demonstrated in the application to high quantiles of distributions of observed maxima of 1- and 5-day precipitation amounts; the differences increase with the return level. The results show that estimation of uncertainty based on nonparametric bootstrap is highly unreliable, especially for small and moderate sample sizes and for very heavy-tailed data.

1 Introduction

The bootstrap, introduced by Efron (1979), is a technique for determining the accuracy of statistics in circumstances in which confidence intervals cannot be obtained analytically or when an approximation based on the limit distribution is not satisfactory (Efron and Tibshirani 1993; Davison and Hinkley 1997). Bootstrap techniques have become very popular in many areas of environmental sciences, including frequency analysis in climatology and hydrology (Dunn 2001; Hall et al. 2004; Ames 2006; Kysely and Beranová 2009; Twardosz 2009; Fowler and Ekström 2009). There are two basic approaches to the bootstrap: While the *nonparametric* bootstrap is based on resampling with replacement from a given sample and calculating the required statistic from a large number of repeated samples (it is often termed simply ‘resampling’), the idea of the *parametric* bootstrap is to randomly generate samples from a parametric model (distribution) fitted to the data and to calculate the statistic from a large number of randomly drawn samples. In both cases, one attempts to infer a distribution of the estimate of a given statistic (e.g., model parameter, quantile of a distribution) from the available data.

The nonparametric bootstrap is often applied when estimating uncertainties involved in frequency models as a simple and intuitive first guess. It has been examined in terms of simulation experiments that evaluated the utility of the methods (Hall et al. 2004; Ames 2006) and widely applied in analyses of observed datasets as well as model outputs. However, when the data samples are small, their distributions are skewed, and a suitable parametric model can be assumed (which is the usual case in a frequency analysis of precipitation amounts), the parametric approach to the bootstrap may be advantageous. Kysely (2008)

J. Kysely (✉)
Institute of Atmospheric Physics,
Academy of Sciences of the Czech Republic,
Boční II 1401,
141 31 Prague, Czech Republic
e-mail: kysely@ufa.cas.cz

quantified the performance of nonparametric and parametric bootstraps for several frequency models used in extreme value analysis, concluding that the parametric bootstrap should be preferred in most cases, more importantly for heavy-tailed distributions (typical for precipitation amounts) than light-tailed distributions (typical for air temperature data). Nevertheless, the study evaluated performance of the two bootstrap methods only for one value of the tail index (shape parameter) of heavy-tailed distributions, and possible dependence on the tail behavior was not examined.

Herein, we analyze the performance of nonparametric (NP) and parametric (P) bootstraps in terms of simulation experiments for a wide range of heavy-tailed distributions used for modeling, among other things, probabilities of extreme precipitation. Heavy-tailed distributions are not exponentially bounded, i.e., they have heavier upper tails than do exponential-like distributions. In other words, when data follow a heavy-tailed distribution, design values corresponding, e.g., to 50- or 100-year return levels may be severely underestimated if the heavy tails are not correctly represented in the model used for the estimation (the limiting Gumbel distribution for maxima, sometimes applied also for modeling precipitation extremes, is not heavy-tailed). There is a consensus that extremes of some environmental variables are heavy-tailed (Katz et al. 2002), including maxima of precipitation amounts (Buishand 1991; Egozcue and Ramis 2001; Kysely and Picek 2007) and streamflow (Farquharson et al. 1992; Anderson and Meerschaert 1998; Kochanek et al. 2008), but also less common variables like sedimentation rates in lakes (which are sensitive to extreme precipitation; Lamoureux 2000).

The paper is organized as follows: in Section 2, the methodology and settings of the simulation experiments are given. Differences between coverage probabilities of confidence intervals obtained with the NP and P bootstraps are quantified in Section 3, and their dependence on the tail index and sample size is evaluated. An application of the two bootstrap methods to confidence intervals of high quantiles of observed precipitation data is shown in Section 4, and implications for use of the bootstrap confidence intervals in heavy-tailed frequency models are discussed in Section 5.

2 Methods

Simulation experiments are carried out with a number of combinations of *true* (parent) and *fitted* probability distributions. The parent distribution is that from which random artificial data samples of a specified size are drawn in the first step of each experiment; the fitted distribution is the one that is adopted for the estimation in the artificial data.

Analogously to Kysely (2008), the size of the artificial samples n was set to 20, 40, 60, and 100 in each experiment to span a range of values typical of time windows for which climatological datasets are analyzed.

2.1 Fitted model

The generalized extreme value (GEV) distribution (Appendix 1) is applied as the fitted distribution in most experiments. It includes three models for maxima of asymptotically large samples (Gnedenko 1943), and it is widely used in frequency modeling of heavy precipitation (Semmler and Jacob 2004; Gaál et al. 2008; Vereem et al. 2008; Fowler and Ekström 2009), air temperature (Kharin and Zwiers 2000, 2005; Kysely 2002; Khaliq et al. 2005), low streamflow (Onoz and Bayazit 1999; Kroll and Vogel 2002; Hewa et al. 2007), floods (Martins and Stedinger 2000; Kumar and Chatterjee 2005; Cunderlik and Ouara 2007), durations of wet and dry spells (Kharin and Zwiers 2000; Voss et al. 2002; Lana et al. 2006), wind speed (van den Brink et al. 2004), and other variables.

Additional experiments with the generalized Pareto (GP) distribution (Appendix 2) as the parent and fitted distribution are carried out in order to highlight general tendencies for heavy-tailed data. The GP distribution is useful in the ‘peaks-over-threshold’ (POT) method for modeling excesses above a sufficiently high threshold. Such approach is preferred when whole time series of data are available, due to the increase in the amount of data entering the estimation procedure. Applications of the POT method include the frequency analysis of air temperatures (Brabson and Palutikof 2002; Katsoulis and Hatzianastassiou 2005; Kysely et al. 2008), precipitation amounts (Begueria and Vicente-Serrano 2006; Bacro and Chaouche 2006; Kysely and Beranová 2009), floods (Adamowski 2000; Cox et al. 2002; Prudhomme et al. 2003), dry spells (Lana et al. 2006), wind speeds (Dupuis and Field 2004; An and Pandey 2005), and wave heights (Pandey et al. 2004).

2.2 Description of the parent models and the simulation experiments

The settings of the individual simulation experiments (denoted E1 to E4) are summarized in Table 1. Note that the parameterization used throughout the paper is in agreement with Hosking and Wallis (1997), i.e., shape parameter $k < 0$ corresponds to a heavy-tailed distribution (cf. Appendices 1 to 3).

In experiments E1, the GEV distribution is used as the parent as well as the fitted distribution. This combination of the true and fitted model represents the case when a *correct parametric model* is adopted for the examined samples. The tail behavior of the GEV distribution is governed by the

Table 1 Summary of settings of the simulation experiments

Experiment	Parent distribution				Fitted model
	Type	Location parameter	Scale parameter	Shape parameter	Type
E1	GEV	$\xi=30$	$\alpha=10$	$k=\{-0.4, -0.3, -0.2, -0.1\}$	GEV
E2	GP	$\xi=0$	$\alpha=10$	$k=\{-0.4, -0.3, -0.2, -0.1\}$	GP
E3	GLO	$\xi=30$	$\alpha=10$	$k=\{-0.4, -0.3, -0.2, -0.1\}$	GEV
E4	Mixed GEV (75%) / GLO (25%)	$\xi(\text{GEV})=30,$ $\xi(\text{GLO})=45$	$\alpha(\text{GEV})=10,$ $\alpha(\text{GLO})=12$	$k(\text{GEV})=\{-0.4, -0.3, -0.2,$ $-0.1\}, k(\text{GLO})=-0.4$	GEV

shape parameter k , and the choice of the other two parameters (location and scale) could be rather arbitrary; their values (given in Table 1) reflect typical distributions of daily maxima of precipitation amounts (in mm) over some land areas in mid-latitudes. We examine four GEV distributions as parent, with values of k ranging from -0.4 (a pronounced heavy tail) to -0.1 (in which case the tail behavior differs little from the limiting Gumbel case). Such a range of the tail index covers typical values estimated, for example, for distributions of annual maxima of 1-day and multi-day rainfall amounts in central Europe (Kysely and Picek 2007). Although tails heavier than $k=-0.4$ may sometimes be found in practical applications, too, we note that already for GEV distributions with $k \leq -0.33$ ($k \leq -0.25$) the third (fourth) statistical moment does not exist (Hosking and Wallis 1997). The bias in the estimates of k and high quantiles from samples drawn from such heavy-tailed

distributions becomes more important (cf. the bias of the shape parameter and the 100-year return values in experiments with $k=-0.4$ in Tables 2 and 3), which also makes the application and comparison of bootstrap confidence intervals for $k \lesssim -0.4$ less straightforward.

Other experiments that make use of a correct parametric model for the estimation are E2, in which the GP distribution is the true as well as the fitted distribution. The shape parameter k is analogous in the GP and GEV distributions, so the same set of values for k is used as in E1. The location parameter is usually known in applications of the POT method and equals zero (when the threshold that delimits extremes is set), so the two-parameter version of the GP distribution is adopted (Appendix 2). In order to allow for a straightforward interpretation, return levels are inverted from the estimated quantile function under the assumption that the frequency of exceedances is one per

Table 2 True (parent) values and mean estimated values of the shape parameter (k) for individual experiments and sample sizes (n)

Experiment	k in parent distribution	Mean estimated k			
		$n=20$	$n=40$	$n=60$	$n=100$
E1	-0.4	-0.308	-0.343	-0.359	-0.372
	-0.3	-0.234	-0.264	-0.274	-0.284
	-0.2	-0.150	-0.174	-0.183	-0.190
	-0.1	-0.071	-0.082	-0.090	-0.093
E2	-0.4	-0.297	-0.337	-0.354	-0.369
	-0.3	-0.221	-0.258	-0.270	-0.281
	-0.2	-0.144	-0.168	-0.177	-0.186
	-0.1	-0.059	-0.077	-0.082	-0.090
E3	-0.4	-0.241	-0.272	-0.288	-0.303
	-0.3	-0.128	-0.157	-0.166	-0.177
	-0.2	0.000	-0.022	-0.029	-0.036
	-0.1	0.141	0.126	0.125	0.118
E4	-0.4	-0.293	-0.327	-0.339	-0.350
	-0.3	-0.247	-0.278	-0.291	-0.298
	-0.2	-0.207	-0.235	-0.245	-0.258
	-0.1	-0.165	-0.194	-0.209	-0.220

Table 3 True (parent) values and mean estimated values of the 100-year return level for individual experiments and sample sizes (n)

Experiment	k in parent distribution	100-year return value in parent distribution	Mean estimated 100-year return value			
			$n=20$	$n=40$	$n=60$	$n=100$
E1	-0.4	162.4	154.2	157.3	158.9	160.2
	-0.3	129.2	127.1	128.9	128.4	129.0
	-0.2	105.5	105.2	105.5	105.8	105.6
	-0.1	88.4	90.0	88.9	89.0	88.5
E2	-0.4	132.7	124.9	127.5	128.7	130.2
	-0.3	99.4	97.0	99.3	99.0	99.3
	-0.2	75.6	76.5	76.0	75.9	76.0
	-0.1	58.5	60.2	59.4	58.8	58.7
E3	-0.4	162.1	154.6	156.5	158.0	159.3
	-0.3	129.0	124.3	125.5	125.1	125.2
	-0.2	105.3	101.2	100.7	100.5	100.2
	-0.1	88.3	83.7	82.8	82.2	82.1
E4	-0.4	174.6	169.2	171.6	172.6	172.9
	-0.3	151.5	149.1	150.9	152.1	150.9
	-0.2	135.8	134.8	136.2	135.9	137.1
	-0.1	126.2	123.9	124.6	125.4	125.5

year, i.e., the same as in the conventional ‘annual maxima’ method with the GEV distribution. This latter setting corresponds to the value of the mean exceedance rate in the GP-Poisson process model (Coles 2001) that is equal to 1.

The other two sets of experiments, E3 and E4, are carried out to demonstrate differences between the bootstrap confidence intervals when *false parametric models* are assumed and adopted for the examined samples. This condition may be common in applications in which the true (parent) distributions are unknown, although the appropriate model may be selected according to goodness-of-fit tests; the probability of selecting a false distribution for given data increases with decreasing sample size.

In experiment E3, the generalized logistic (GLO) distribution (Appendix 3) is used as parent but the GEV distribution is utilized for the estimation. GLO is a model that has become popular in hydrology following the study of floods by Ahmad et al. (1988), and it also has been found to be a useful distribution for maxima of precipitation amounts (Shoukri et al. 1988; Asquith 1998; Lee and Maeng 2003; Fitzgerald 2005; Kysely and Picek 2007; Zin et al. 2009). According to the Flood Estimation Handbook (IH 1999), it has been recommended as the standard for flood frequency analysis in the UK. We use a reparameterized version of the log-logistic distribution of Ahmad et al. (1988), in which the parameters are analogous to those of the GEV distribution (Hosking and Wallis 1997; Appendix 3). Since the GEV and GLO distributions are closely related models that rank among distributions with

the same weight of the upper tails, the setting of experiments E3 means that a false but related parametric model is adopted for the estimation. The same set of parameters as in E1 is used for the GLO distribution, with the shape parameter k varying again between -0.4 and -0.1 . Note that high quantiles of the GEV and GLO distributions with the same parameters are very similar (Tables 2 and 3).

In the last set of experiments, E4, the samples are drawn from a double-populated GEV–GLO model, i.e., a mixture of two distributions (Fig. 1). Three quarters of data in each artificial sample are drawn from the GEV distribution while the remaining quarter originates from the GLO distribution. This may represent a condition when two mechanisms producing extremes—for example heavy precipitation—are present in a sample: most extremes arise from an ‘ordinary’ population but less frequently there also occur events from a secondary ‘extra-ordinary’ population (cf. van den Brink et al. 2004). A relatively large fraction of data from the secondary population (25%) is chosen in order to highlight the differences from experiments E1, since for decreasing fractions the results converge to those of E1. The GLO distribution with a shifted location and a heavy upper tail ($k=-0.4$) is used to represent the secondary population. The model parameters (which again span a range of values for the tail behavior of the primary GEV distribution) are summarized in Table 1, and the probability density functions of the mixed models together with both components are plotted in Fig. 1. The GEV distribution is again adopted as a model for the data. The setting of experiments

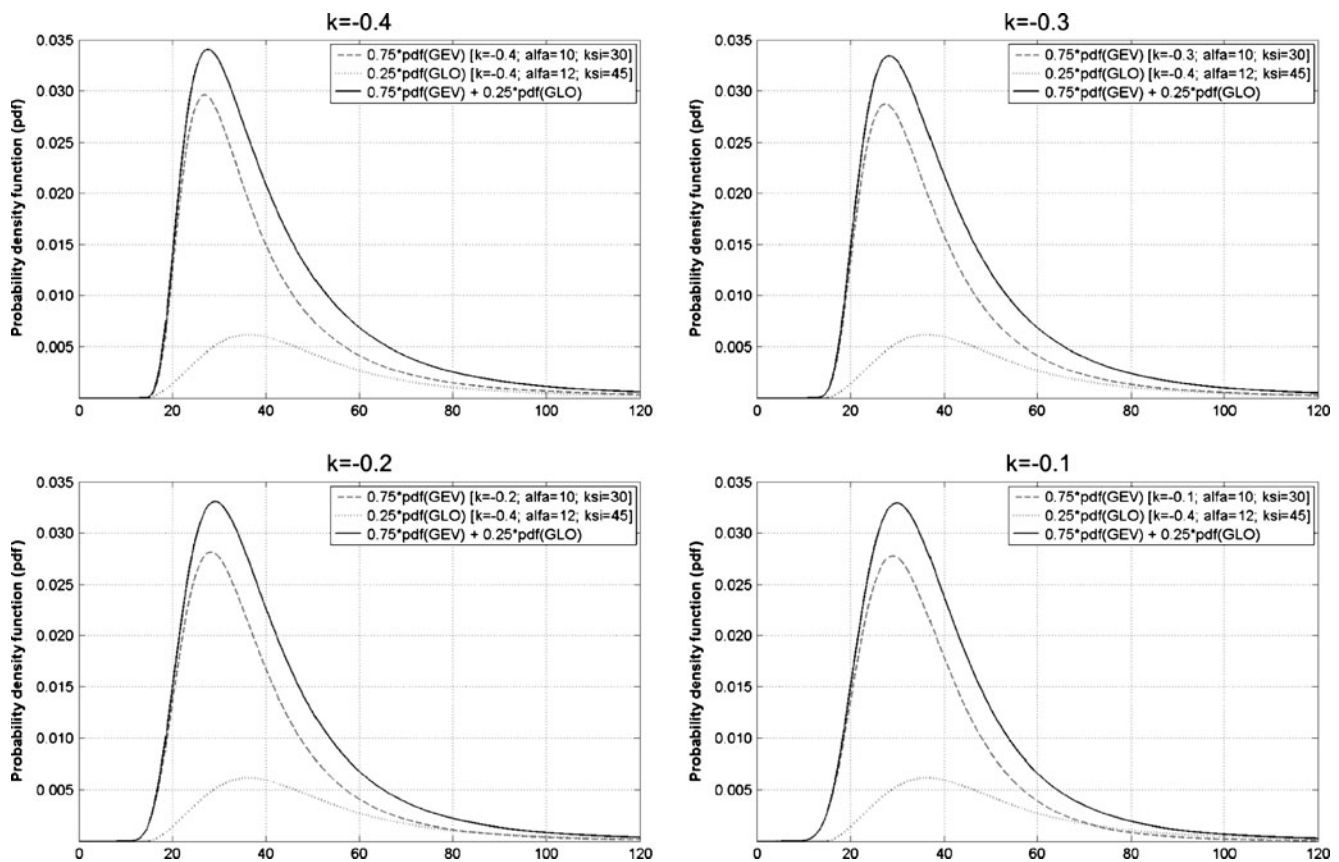


Fig. 1 Probability density functions of the two partial distributions and the double-populated parent models in experiments E4

E4 corresponds to the case when a simplified model is fitted to an examined sample.

2.3 Other settings of the simulation procedure

The simulation procedure in each experiment and each combination of sample size ($n=20, 40, 60$, and 100) and tail behavior (governed by k) is as follows (Kysely 2008):

1. Five thousand artificial samples of n values are randomly drawn from the specified parent distribution (or the mixture of the parent distributions in E4).
2. To each artificial sample, the GEV or GP distribution is fitted and its quantiles corresponding to the return levels of 5 to 200 years are estimated.
3. The 90% and 95% confidence intervals (CIs) of the model parameters and quantiles are estimated from the P and NP bootstraps. The former involves generating a large number of random samples from the *fitted* distribution (with parameters estimated from the artificial sample); the latter consists in a simple resampling with replacement of the artificial sample.

We confine our attention in this study to the most widely-used percentile CIs. For both bootstrap approaches

and all artificial samples, 1,000 iterations are carried out to estimate 2.5%, 5%, 95%, and 97.5% quantiles of distributions of the 5- to 200-year return levels, which delimit the 90% and 95% CIs. The method of L-moments (Hosking 1990) is used for estimating the parameters and quantiles of the GEV/GP distribution.

The performance of the NP and P bootstraps is evaluated in terms of empirical coverage probability of the CIs, i.e., the percentage of simulated results for which the estimated 90% and 95% CIs cover the true values of the quantiles (which are determined from parameters of the parent distribution). It is expected that an appropriate ('correct') method yields coverage close to the nominal value of 90/95% while a higher (lower) value points to CIs that are too wide (narrow) compared to the real uncertainty, provided that the quantile estimates are not biased.

3 Results

3.1 Experiments E1 (GEV fitted to GEV-distributed data)

For all values of the shape parameter k and all examined sample sizes ($n=20, 40, 60$, and 100), the P bootstrap performs considerably better in terms of the coverage

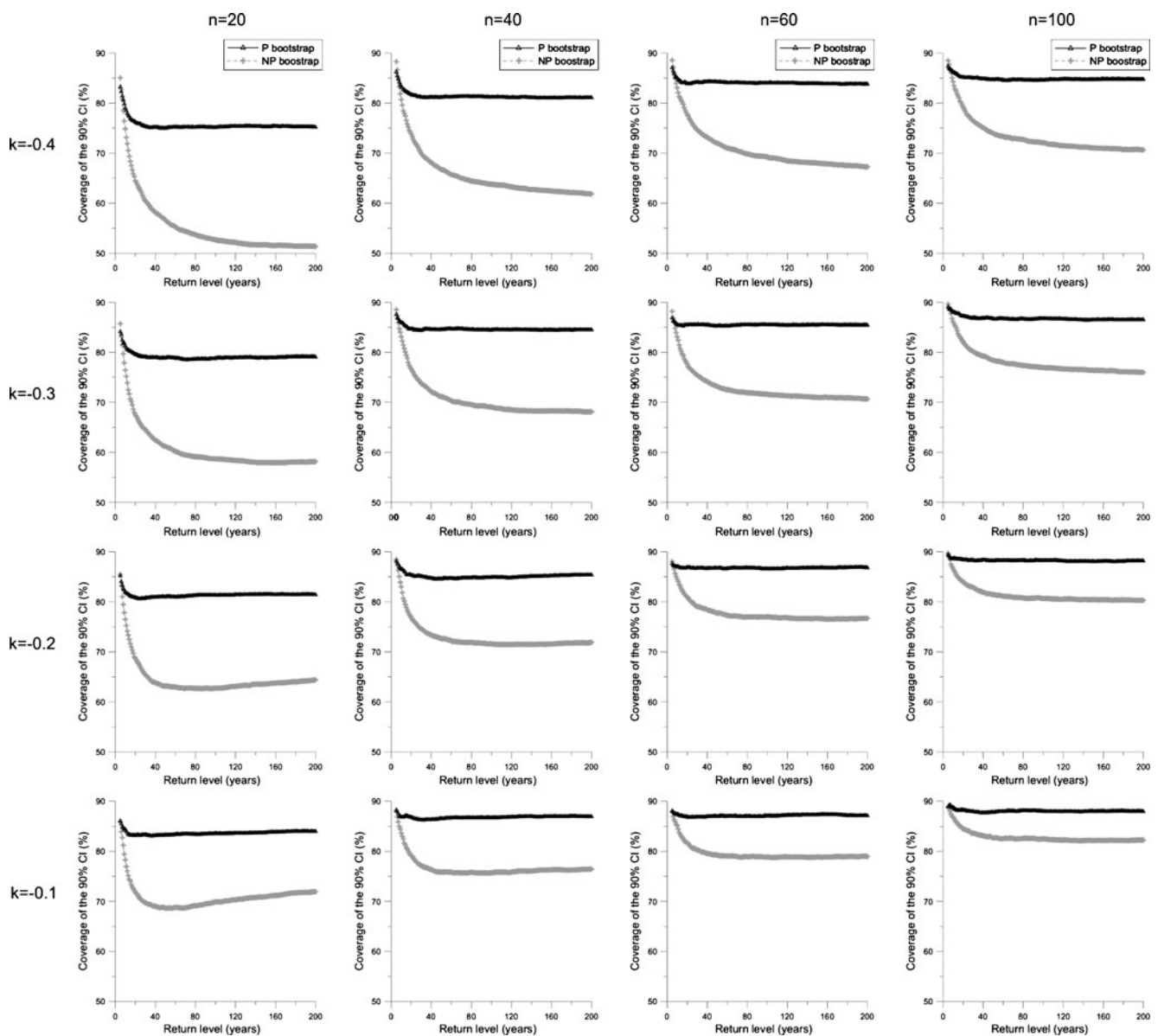


Fig. 2 Dependence of the coverage probability of the 90% CIs from the parametric (*P*) and nonparametric (*NP*) bootstraps on the *T*-year return level ($T=5$ to 200) in experiments E1, for individual sample sizes (*columns*) and values of the tail index (*rows*)

probability of the CIs (Fig. 2). The differences are particularly important for small and moderate sample sizes ($n=20, 40$) and for the very heavy-tailed GEV model ($k=-0.4$). For example, the 90% CIs of the 100-year return values estimated from samples with 40 members drawn from the GEV distribution with $k=-0.4$ cover the true value only in 64% of cases when the NP bootstrap is used. This is improved to 81% for the P bootstrap (a value still lower than the 90% that is expected). The coverage probabilities of the 90% and 95% CIs of the 100-year return levels are summarized for all experimental settings in Table 4; the findings are analogous notwithstanding whether the 90% or 95% CIs are considered.

It should be noted that the coverage probability of the 90% CIs for all values of k, n , and in the whole range of the 5- to 200-year return levels is lower than the nominal value of 90% for both the NP and P bootstraps (Fig. 2). This means that the CIs constructed using the bootstrap are always too narrow and undervalue the uncertainty involved in the estimates. However, this underestimation is much less severe when the P version of the bootstrap is employed.

Another favorable property of the P bootstrap is that the coverage probability of the CIs is almost independent on the return level. Except for very small samples ($n=20$) drawn from the GEV distribution with $k \leq -0.3$, the coverage probability of the 90% CIs constructed by

Table 4 Coverage probabilities of the 90% and 95% CIs of the 100-year return levels estimated from the nonparametric (NP) and parametric (P) bootstraps. k denotes the shape parameter of the parent distribution, n stands for the sample size

k	n	90%_NP	90%_P	95%_NP	95%_P
a. Experiments E1 (GEV fitted to GEV-distributed data)					
-0.4	20	52.7	75.1	55.8	80.8
	40	63.8	81.2	67.6	86.8
	60	69.2	84.0	74.1	88.9
	100	72.1	84.6	77.0	90.6
-0.3	20	58.7	78.8	62.0	83.4
	40	69.0	84.4	73.0	88.7
	60	71.6	85.5	76.1	90.0
	100	77.0	86.7	82.2	91.5
-0.2	20	62.7	81.3	66.5	85.6
	40	71.6	84.9	76.3	89.4
	60	77.0	86.6	81.6	91.0
	100	80.7	88.2	85.7	92.5
-0.1	20	69.9	83.6	73.8	87.0
	40	75.8	86.7	80.8	90.6
	60	78.9	87.1	84.0	90.8
	100	82.5	88.0	87.5	92.3
b. Experiments E2 (GP fitted to GP-distributed data)					
-0.4	20	58.3	76.1	62.7	81.1
	40	66.6	80.7	71.2	85.7
	60	71.3	82.8	76.6	88.0
	100	76.3	85.8	81.5	90.3
-0.3	20	62.9	79.3	67.9	83.3
	40	70.6	82.6	75.9	87.1
	60	75.1	85.5	80.7	89.3
	100	79.5	86.4	84.7	91.1
-0.2	20	68.5	81.9	72.9	85.9
	40	76.0	84.9	80.9	88.6
	60	78.4	86.0	84.0	90.0
	100	82.2	87.5	87.1	91.5
-0.1	20	73.3	84.9	78.3	88.3
	40	78.9	86.1	83.9	89.7
	60	82.2	87.8	87.4	91.5
	100	83.8	88.4	89.9	92.4
c. Experiments E3 (GEV fitted to GLO-distributed data)					
-0.4	20	52.6	74.0	55.7	79.5
	40	63.9	78.8	68.0	84.6
	60	66.2	78.6	70.8	84.5
	100	72.1	81.3	77.7	87.3
-0.3	20	57.9	74.7	61.0	79.1
	40	64.5	76.4	69.2	81.7
	60	69.9	77.8	75.0	83.0
	100	74.6	79.2	79.4	85.3
-0.2	20	61.8	74.0	66.3	78.4
	40	68.6	75.2	73.3	80.3
	60	71.5	74.3	76.1	79.8
	100	73.7	73.3	79.7	80.1

Table 4 (continued)

k	n	90%_NP	90%_P	95%_NP	95%_P
-0.1	20	66.9	74.0	72.2	78.1
	40	70.1	71.7	75.7	76.8
	60	71.5	69.3	76.0	75.2
	100	71.3	66.5	77.8	72.9
d. Experiments E4 (GEV fitted to mixed GEV/GLO-distributed data)					
-0.4	20	54.5	77.7	57.5	83.0
	40	64.6	81.3	68.8	87.3
	60	69.4	83.6	74.0	89.4
	100	73.3	85.0	78.2	90.7
-0.3	20	56.2	79.3	59.6	84.3
	40	66.8	82.9	70.4	87.8
	60	71.5	84.2	76.2	89.3
	100	76.6	86.8	81.2	91.2
-0.2	20	59.0	78.2	62.0	83.4
	40	67.5	82.1	71.7	86.7
	60	72.1	83.5	77.1	88.4
	100	76.8	84.7	81.9	90.0
-0.1	20	57.4	76.9	61.0	81.5
	40	67.6	80.1	71.3	84.8
	60	69.0	78.6	73.7	85.1
	100	73.7	79.5	79.3	86.2

means of the P bootstrap is at least 80% for all return levels, and it is close to the nominal value of 90% for moderate and large sample sizes and less heavy-tailed distributions (Fig. 2).

3.2 Experiments E2 (GP fitted to GP-distributed data)

Similar results are achieved in simulation experiments E2 in which the GP distribution is fitted to the GP-distributed data (Table 4, Fig. 3). Differences between the NP and P bootstrap are slightly less pronounced but the general tendencies remain unchanged: the P bootstrap always performs better, CIs from both the P and NP bootstraps have always too low coverage, the improvement of the P over NP bootstrap is particularly important for very heavy-tailed data and small sample sizes, and the coverage probability of the 90% CIs of quantiles corresponding to the 5- to 200-year return levels from the P bootstrap is at least 80% except for $n=20$ and $k \leq -0.3$.

Differences in coverage probabilities of CIs of high quantiles between the experiments with the GEV and GP distributions are relatively minor for the P bootstrap, which reflects the fact that the shape parameter is analogous in the two distributions. Some differences between the behavior of the coverage probabilities in the two experiments are related to the fact that the GP distribution is estimated as a two-parameter model

(Appendix 2), with the location defined by the fixed threshold also in practical applications, and a slightly different bias of estimates in the two models. The fact that the number of free parameters is smaller and skewness of the data sample (l_3/l_2) is not employed in estimating the GP distribution (in contrast to GEV and GLO—Appendix 1 and 3) is manifested in a tendency to a larger positive bias of the estimates of the shape parameter, particularly for small samples (Table 2).

3.3 Experiments E3 (GEV fitted to GLO-distributed data)

The performance of both bootstrap methods deteriorates in experiment E3 (a false model fitted to heavy-tailed data) compared with experiments E1 and E2, and the coverage probability becomes particularly low for high quantiles (Fig. 4). However, the P bootstrap performs better in most cases even though the parametric model assumed for the data is misspecified. Only for the combination of large sample sizes ($n=60, 100$) and little pronounced heavy tail ($k=-0.1$) does the NP bootstrap outperform the P bootstrap. On the other hand, the superiority of the P bootstrap is obvious even for large sample sizes ($n=100$) with pronounced heavy tails.

It should be emphasized that the coverage probability of the 90% (95%) CIs for 100-year return levels, except for large samples ($n=100$), does not exceed 79% (85%) for the

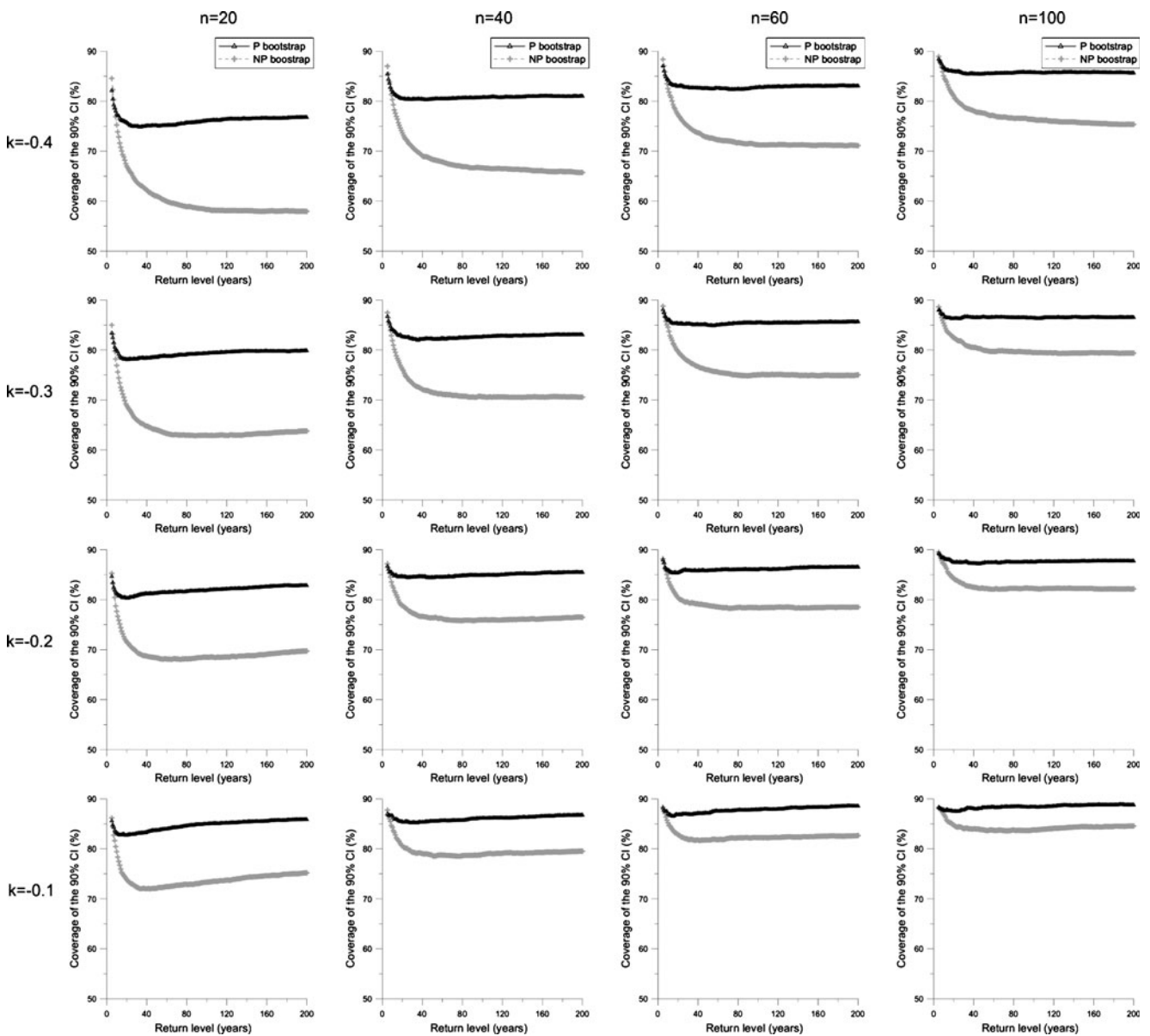


Fig. 3 Same as in Fig. 2 except for experiments E2

P bootstrap and 72% (76%) for the NP bootstrap (Table 4). That is to say that the real uncertainties of the estimates are always substantially underestimated. The low coverage is related to the bias of the estimated model, and increasing positive bias of k (manifested also in negative bias of high quantiles) for less heavy-tailed parent distributions (Tables 2 and 3) may explain the worse performance of the P bootstrap for less-pronounced heavy tails.

3.4 Experiments E4 (GEV fitted to double-populated GEV–GLO data)

The P bootstrap outperforms the NP bootstrap in all settings of experiments E4, in which a simplified (GEV) model is

applied to double-populated GEV–GLO data (Fig. 5). As in experiments E1 and E2, the differences decrease with increasing sample size but are still evident for $n=100$. The coverage probability of the 90% (95%) CIs for the 100-year return levels, except for large samples ($n=100$), does not exceed 72% (77%) for the NP bootstrap (Table 4). The performance of the NP bootstrap is particularly poor for very small samples ($n=20$), for which the coverage probability of the 90% CIs from the NP bootstrap is between 50% and 60% for return levels $T \geq 50$ years (Fig. 5). The coverage is improved considerably with the P bootstrap (75–80%).

With increasing k (towards less heavy tails) of the parent GEV distribution, the coverage probability of the CIs from

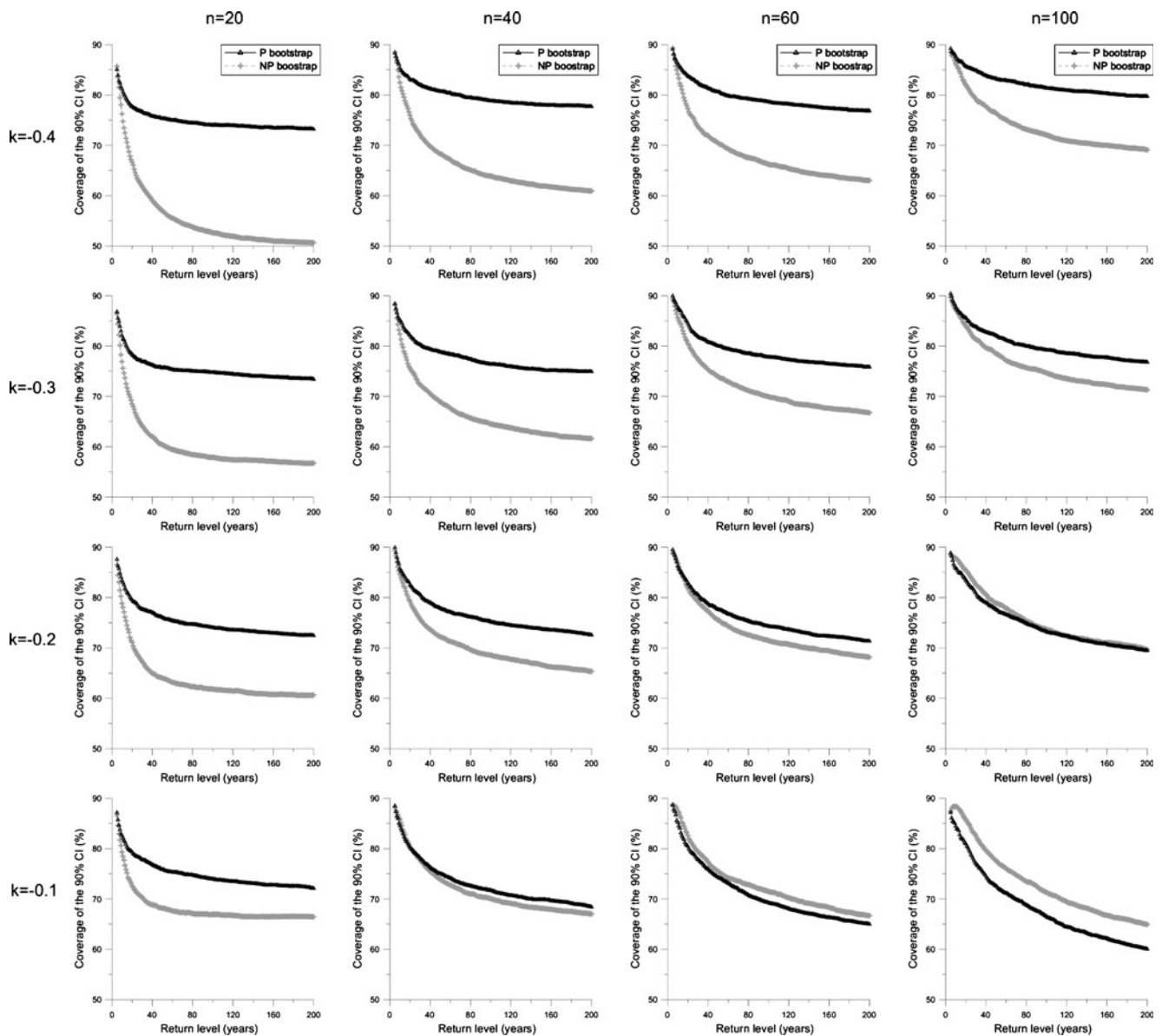


Fig. 4 Same as in Fig. 2 except for experiments E3

the P bootstrap deteriorates for high quantiles as the two parent distributions become more dissimilar (with respect to the tail behavior) and the fitted model less appropriate; the two populations that produce the samples are not differentiated in the fitted model. Another feature related to the bias of the adopted model is that for less-pronounced heavy tails, there is little improvement in the coverage probability of the CIs with increasing sample size for the P bootstrap (unlike the NP bootstrap; bottom row of Fig. 5). Nevertheless, these are not arguments against the P bootstrap: the NP bootstrap performs always worse, and the sample size of $n=100$ may be large enough to recognize in a practical application that the single-population GEV model is not suitable for such data.

4 Application to observed precipitation data

To demonstrate differences between application of the NP and P bootstraps to real climatological data, we compare CIs for high quantiles of precipitation amounts estimated by the two bootstrap approaches. The examined dataset consists of annual maxima of 1- and 5-day precipitation amounts measured at 175 rain-gauge stations covering the area of the Czech Republic, with complete series over 1961–2005. The spatial distribution of the stations is shown in Fig. 6. The dataset originates from Kysely (2009), who examined trends in characteristics of heavy precipitation in individual seasons, and it is superior in terms of spatial coverage and data quality to the one used in

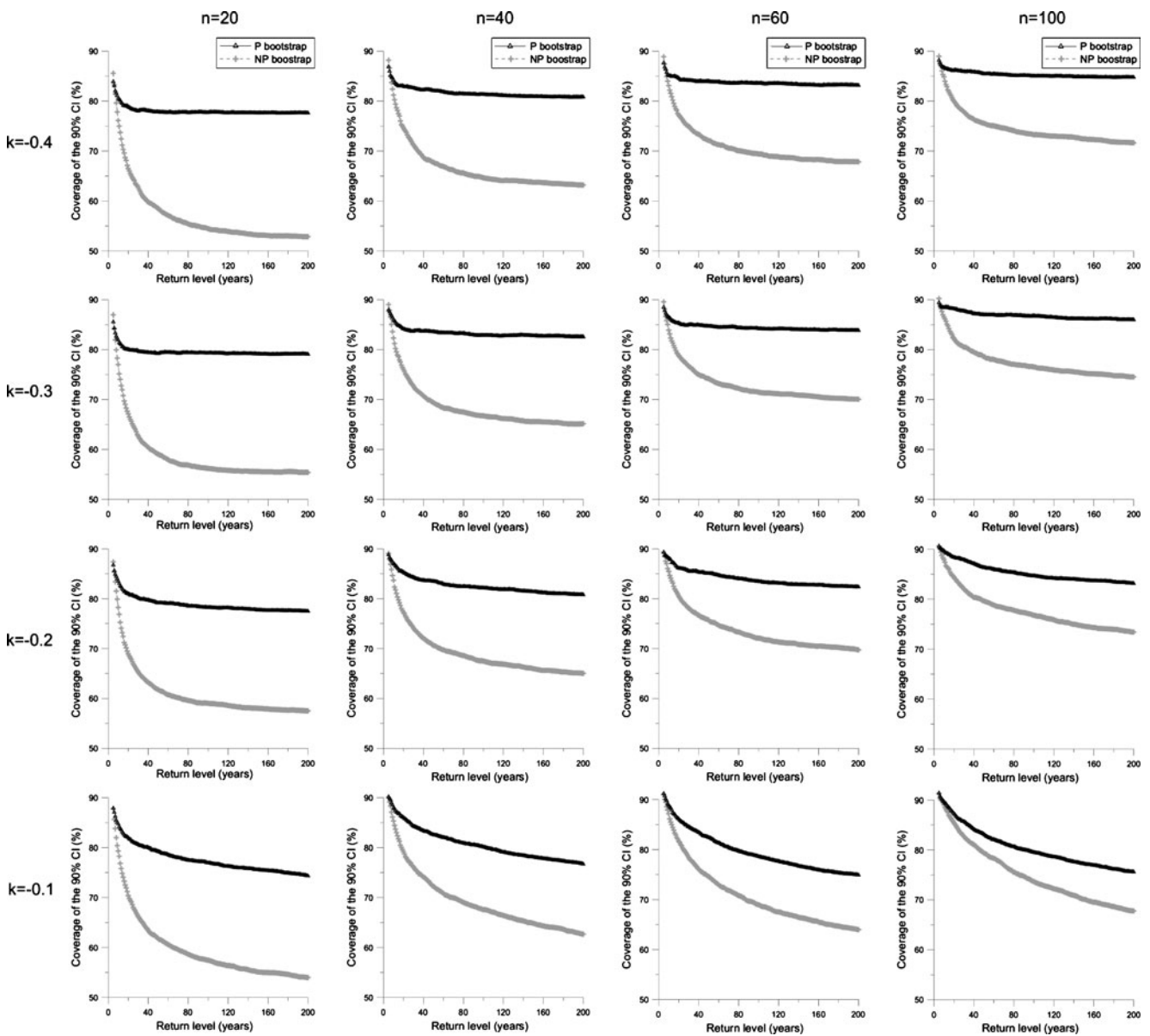


Fig. 5 Same as in Fig. 2 except for experiments E4

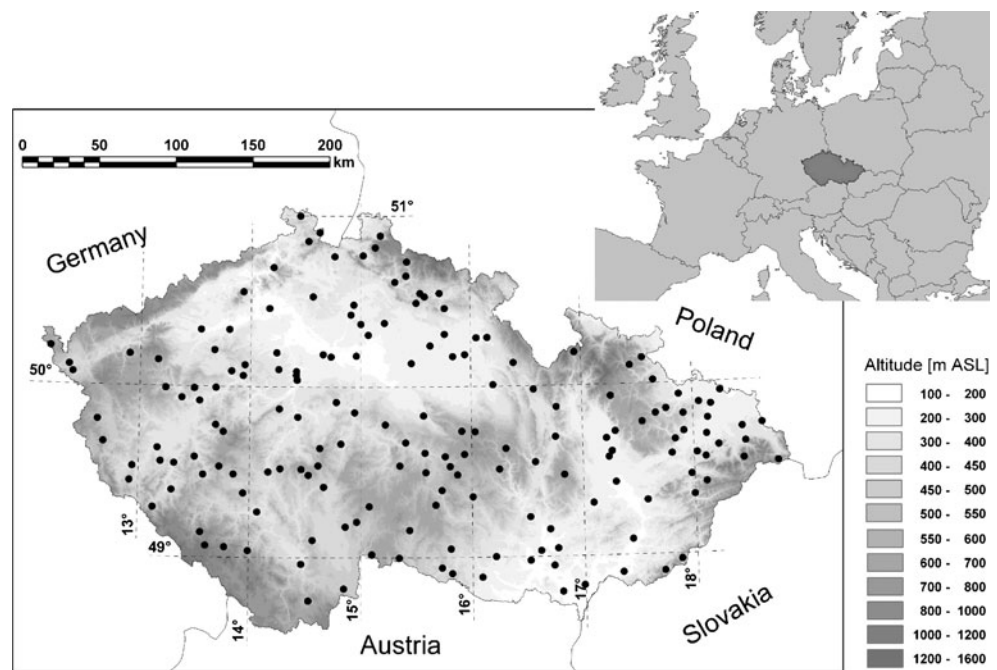
a previous study on statistical modeling of precipitation extremes in this area (Kyselý and Picek 2007). The assumption of stationarity of the examined data was checked before application of the extreme value analysis: trends significant at $p=0.05$ (according to the Mann–Kendall test) were observed at 5.7% (3.4%) of stations for annual maxima of 1-day (5-day) precipitation amounts, i.e., the percentage of significant trends at the given level is close to the nominal value of 5% in both cases.

The GEV distribution was fitted to the individual stations’ datasets using the method of L-moments, and both bootstrap approaches were used to estimate the 90% CIs of model parameters and quantiles corresponding to the return levels of 10, 20, 50, and 100 years. The number of repetitions in both NP and P bootstraps was set to 1,500.

Figure 7 shows scatter-plots of the relative width of the estimated 90% CIs against the shape parameter for individual return levels (the relative width of the CIs, i.e., the width of the CIs scaled by the value of the quantile corresponding to the return level, is plotted in order to remove variations related to the magnitude of the quantile itself, e.g., larger values at mountain stations). Although the range of the estimated values of the shape parameter is wide, the estimated GEV distribution is heavy-tailed at a large majority of the stations (151/156 for 1-day/5-day maxima).

For all return levels, there is a tendency to more liberal (narrower) CIs from the NP bootstrap. The percentage of stations at which the CIs from the NP bootstrap are narrower than those from the P bootstrap is summarized

Fig. 6 Spatial distribution of the 175 stations with precipitation data over 1961-2005



in Table 5, and the average relative widths of the CIs are shown in Table 6. As expected, the differences between the NP and P bootstraps increase with the return level; they are small for 10-year return values but become quite pronounced for 50- and 100-year return values (Table 6, Fig. 7). For

100-year return values of 1-day precipitation amounts, the average relative width of the 90% CIs is 66.3% when the P bootstrap is applied while only 49.9% when using the NP bootstrap. These values are averaged over all 175 stations, notwithstanding whether heavy-tailed or light-tailed GEV

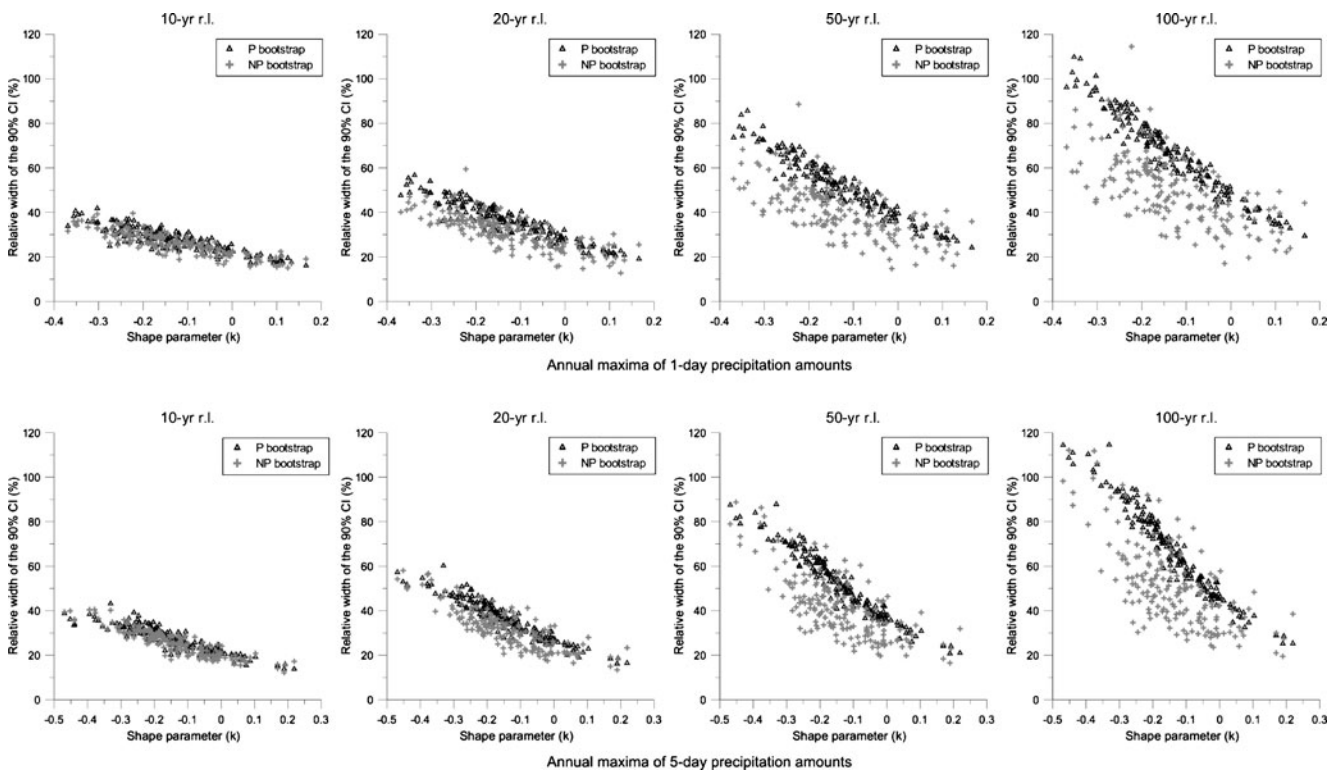


Fig. 7 Scatter-plots of the relative width of the estimated 90% CIs against the estimated shape parameter for individual return levels (*r.l.*) and annual maxima of 1-day (*top*) and 5-day (*bottom*) precipitation

amounts at 175 stations. The width of the CI is scaled by the estimated value corresponding to the given return level at each station

Table 5 Percentage of stations at which the 90% CIs estimated from the NP bootstrap are narrower than those from the P bootstrap for high quantiles of distributions of observed precipitation data (175 stations, 45 years)

Dataset	Return level (r.l.)			
	10-year r.l.	20-year r.l.	50-year r.l.	100-year r.l.
1-day annual maxima	78.9	84.6	84.6	86.3
5-day annual maxima	75.4	78.3	82.3	84.6

distribution is estimated. If only sites with an estimated heavy tail ($k < 0$) are considered, the difference is even more pronounced—the average relative width of the 90% CIs is 70.7% for the P bootstrap and 52.2% for the NP bootstrap. Since *both* NP and P bootstrap tend to yield CIs that are too narrow for heavy-tailed data, as shown above, the uncertainty of the high quantiles tends to be substantially underestimated when using the NP bootstrap while the underestimation is at least partly rectified with the P bootstrap.

Other features of the CIs that are demonstrated in Fig. 7 include dependence of the width of the CIs on k and growing width of the CIs with rising return level (the latter being increasingly important for very heavy-tailed data). Also noteworthy is the fact that the dependence of the width of the CIs on k is close to linear for the P bootstrap, represented by a much narrower band for the P bootstrap than for the NP bootstrap. Owing to the length of the examined precipitation datasets (45 years), sampling variability may strongly influence also the bounds of the estimated CIs for the NP bootstrap (while it is ‘smoothed’ with the P bootstrap). This is manifested, among other, in some outlying estimates of the relative width of the 90% CIs from the NP bootstrap in Fig. 7, particularly a large outlier in the upper row of the plots (for 1-day maxima) for the 50- and 100-year return levels. Scrutiny of the data reveals that this outlying estimate appears at a station affected by extreme rainfall on July 22, 1998 (resulting in a severe flash flood in eastern Bohemia), with 24-h precipitation amount of 163 mm, while the second largest daily amount at this site over 1961–2005 was only 83 mm. A bootstrap that consists purely in resampling with replacement of the 45 values of annual maxima puts too much weight onto the single extreme observation, and this

leads to inflated confidence bounds for the estimates in this specific case of a heavy-tailed sample. This example demonstrates that estimates based on the NP bootstrap are much more sensitive to random sampling variability and much less consistent between samples (stations in this case) than those obtained with the P bootstrap.

5 Discussion

The study compares performance of two basic variants of bootstrap—parametric and nonparametric—for estimating CIs of high quantiles in heavy-tailed data, which are typical for precipitation extremes and some other climatological and hydrological variables. When a correct parametric model is fitted to data drawn from the GEV or the GP distribution, the parametric bootstrap performs considerably better for all examined return levels (5 to 200 years), sample sizes ($n=20, 40, 60$ and 100), and tail behaviors (the shape parameter k ranging from -0.4 to -0.1). The parametric bootstrap is preferred also when a false model (GEV) is fitted to GLO-distributed data, except for the distribution with the least heavy tail ($k=-0.1$) and large sample sizes. Since probability of selecting an incorrect parametric model (by means of goodness-of-fit tests) declines with an increasing size of the data sample, the superiority of the nonparametric bootstrap in this particular case is of little practical importance.

The last-examined experiments make use of a simplified model (GEV) adopted for mixed (double-populated) data drawn from combinations of the GEV and GLO distributions. This may represent a relatively frequent case in

Table 6 Average relative widths of the 90% CIs (in %) estimated from the NP and P bootstrap for high quantiles of distributions of observed precipitation data (175 stations, 45 years). The width of the CI is scaled by the estimated value corresponding to the given return level at each station

Dataset	Bootstrap	Return level (r.l.)			
		10-year r.l.	20-year r.l.	50-year r.l.	100-year r.l.
1-day annual maxima	NP	26.0	31.3	40.9	49.9
	P	27.6	36.4	52.0	66.3
5-day annual maxima	NP	25.8	31.6	41.8	51.0
	P	27.0	36.0	52.1	67.1

extreme value analysis when the samples examined arise from populations governed by different extreme-generating mechanisms (characterized by specific distributions), which are, however, difficult to disaggregate from data records. The coverage probability of the CIs constructed from the parametric bootstrap is always better in the experiments with the mixed models, too, even for large sample sizes ($n=100$).

A tendency to more liberal (narrower) CIs from the nonparametric than parametric bootstrap is clearly demonstrated in the application to high quantiles of distributions of observed maxima of 1- and 5-day precipitation amounts, measured at 175 stations over 1961–2005. The differences increase with the return level, and the relative width of the 90% CIs of the 100-year return values of both 1- and 5-day precipitation amounts is reduced on average by 25% when the nonparametric bootstrap is used instead of the parametric bootstrap. This reduction is likely to increase if inference is based on samples from shorter time periods. Another advantage of the CIs from the P bootstrap, demonstrated in the application to real data, is that the estimates are much less influenced by random (sampling) effects.

It should also be stressed that in *all* the simulation experiments, the constructed CIs are *too narrow* and too often miss the true values of model parameters and quantiles. This means that the uncertainty of the parameters and quantiles is underestimated, more importantly so for the nonparametric bootstrap. The underestimation appears to be a general feature of bootstrap CIs for heavy-tailed data and is related to skewness in the distributions of estimates of model parameters (Tajvidi 2003). We show that the underestimation of uncertainty is more important

- For the nonparametric than parametric bootstrap
- For small sample sizes
- For higher quantiles (except when the correct model is fitted), and
- When an incorrect (although related) parametric model is used

This suggests that bootstrap should be regarded as the first guess of the uncertainty, and alternative methods—e.g. analytical expressions for the sampling variance of quantiles of the distributions (Lu and Stedinger 1992; Kjeldsen and Jones 2004) or likelihood-based confidence intervals (Tajvidi 2003)—should be considered at least for comparison. An inference relying uncritically on bootstrap may obviously be misleading.

The present simulation experiments examined behavior of bootstrap CIs for a range of frequency models. Although all possible cases encountered in analyses of precipitation data cannot be covered, the simulation results appear to be indicative of some general tendencies of CIs constructed using the bootstrap (as regards the dependence of results on

the sample size, tail behavior, and ‘correctness’ of the parametric model). We also confined ourselves to the percentile CIs since these are the most popular; see, e.g., Carpenter and Bithell (2000) or Dixon (2002) for a brief review on advanced versions of bootstrap CIs. The percentile and ‘bias-corrected and accelerated’ (BCa; Efron and Tibshirani 1993) bootstrap CIs were compared by Dupuis and Field (1998) and Kyselý (2008) for the GEV distribution, and Tajvidi (2003) for the GP distribution; the BCa CIs are usually superior, but the coverage probability is still lower than the nominal value. More sophisticated bootstrap procedures do not compensate for insufficient data, so the poor performance of the nonparametric bootstrap in small sample sizes does not much depend on the variant of bootstrap CIs.

6 Conclusions

The basic choice of bootstrap method (nonparametric vs. parametric) used for estimating uncertainties in frequency models is usually not justified in climatological applications, and respective limitations and drawbacks of the two bootstraps are not discussed and/or evaluated. We provide arguments for using the *parametric* version of the bootstrap for constructing quantile confidence intervals in heavy-tailed frequency models, provided that the suitable parametric model is known or can be assumed (which is almost always the case in modeling probabilities of precipitation extremes). Even a moderate misspecification of the distribution does not prevent the parametric bootstrap from performing better than the nonparametric one. Inasmuch as a severe misspecification of the parametric model adopted for examined data is unlikely provided that the model is supported by some goodness-of-fit tests and/or other statistical tools (such as the L-moment ratio diagram; Hosking and Wallis 1997) and the sample’s time period is not extremely short, we find it difficult to identify any reasons for using the nonparametric bootstrap. Confidence intervals constructed using the nonparametric bootstrap should be interpreted very cautiously, and especially so for small and moderate sample sizes and for distributions with very heavy tails, as they may severely undervalue the true uncertainty of the estimates. This is also the reason why the nonparametric bootstrap should be avoided when estimating uncertainty of design values for use in practical applications.

Acknowledgments The author is grateful to J. Pícek, Technical University Liberec, for assistance, discussions and comments on several versions of the manuscript. Thanks are due to the anonymous reviewers and editor for a useful suggestion to include an application to observed data, and L. Gaál and J. Hošek, Institute of Atmospheric Physics, Prague, for drawing Figs. 1 and 6. The study was supported by the Grant Agency of AS CR under project B300420801.

Appendix 1: Generalized extreme value (GEV) distribution

The cumulative distribution function of the GEV distribution with parameters ξ (location), α (scale), and k (shape) is (e.g., Hosking and Wallis 1997)

$$F(x) = e^{-e^{-y}},$$

where

$$y = -\frac{\ln\left(1 - \frac{k(x-\xi)}{\alpha}\right)}{k}, \quad k \neq 0,$$

$$y = \frac{x-\xi}{\alpha}, \quad k = 0.$$

The distribution is bounded at $\xi + \frac{\alpha}{k}$ from right (left) if $k > 0$ ($k < 0$). The quantile function is

$$x(F) = \xi + \frac{\alpha\left(1 - (-\ln F)^k\right)}{k}, \quad k \neq 0,$$

$$x(F) = \xi - \alpha \ln(-\ln F), \quad k = 0.$$

L-moments are defined for $k > -1$; the first three population L-moments are

$$\lambda_1 = \xi + \alpha \frac{1 - \Gamma(1+k)}{k},$$

$$\lambda_2 = \alpha \frac{(1 - 2^{-k})\Gamma(1+k)}{k}, \text{ and}$$

$$\lambda_3 = \alpha \frac{\Gamma(1+k)(-1 + 3 \cdot 2^{-k} - 2 \cdot 3^{-k})}{k},$$

where Γ denotes the gamma function.

The method of L-moments fits the GEV distribution by choosing its parameters so that the first three L-moments, $\lambda_1, \lambda_2, \lambda_3$, match the corresponding estimates l_1, l_2, l_3 . The resulting L-moment estimators of k, α , and ξ are given by

$$k = 7.8590z + 2.9554z^2, \alpha = \frac{l_2 k}{(1 - 2^{-k})\Gamma(1+k)}, \text{ and}$$

$$\xi = l_1 + \alpha \frac{\Gamma(1+k) - 1}{k},$$

where

$$z = \frac{2}{3 + \frac{l_3}{l_2}} - \frac{\ln 2}{\ln 3}.$$

Appendix 2: Generalized Pareto (GP) distribution

The cumulative distribution function of the GP distribution with parameters ξ (location), α (scale) and k (shape) is

$$F(x) = 1 - e^{-y},$$

where y is defined in Appendix 1. The distribution is bounded at ξ from the left, and if $k > 0$, also at $\xi + \frac{\alpha}{k}$ from the right. The quantile function is

$$x(F) = \xi + \frac{\alpha\left(1 - (1 - F)^k\right)}{k}, \quad k \neq 0,$$

$$x(F) = \xi - \alpha \ln(1 - F), \quad k = 0.$$

L-moments of the GP distribution are defined for $k > -1$; the first three population L-moments are

$$\lambda_1 = \xi + \frac{\alpha}{1+k}, \quad \lambda_2 = \frac{\alpha}{(1+k)(2+k)}, \quad \text{and}$$

$$\lambda_3 = \frac{\alpha(1-k)}{(1+k)(2+k)(3+k)}.$$

Since the location parameter ξ , corresponding to the lower bound of the GP distribution, is known in the current application ($\xi=0$), only the two remaining parameters are estimated. In such case, l_3 is not used when estimating the scale and shape parameters of the GP distribution, and the L-moment estimators of k and α are given by

$$k = \frac{l_1 - \xi}{l_2} - 2, \text{ and } \alpha = (1+k)(l_1 - \xi).$$

Appendix 3: Generalized logistic (GLO) distribution

The cumulative distribution function of the GLO distribution with parameters ξ (location), α (scale), and k (shape) is

$$F(x) = \frac{1}{1 + e^{-y}},$$

where y is defined in Appendix 1. The distribution is bounded at $\xi + \frac{\alpha}{k}$ from right (left) if $k > 0$ ($k < 0$). The quantile function is

$$x(F) = \xi + \frac{\alpha\left(1 - \left(\frac{1-F}{F}\right)^k\right)}{k}, \quad k \neq 0,$$

$$x(F) = \xi - \alpha \ln\left(\frac{1-F}{F}\right), \quad k = 0.$$

L-moments are defined for $-1 < k < 1$; the first three population L-moments are

$$\lambda_1 = \xi + \alpha\left(\frac{1}{k} - \frac{\pi}{\sin k\pi}\right), \quad \lambda_2 = \frac{\alpha k \pi}{\sin k\pi}, \quad \text{and}$$

$$\lambda_3 = -k\lambda_2.$$

The L-moment estimators of k, α , and ξ are given by

$$k = -\frac{l_3}{l_2}, \alpha = \frac{l_2 \sin k\pi}{k\pi}, \quad \text{and } \xi = l_1 - \alpha\left(\frac{1}{k} - \frac{\pi}{\sin k\pi}\right).$$

References

- Adamowski K (2000) Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *J Hydrol* 229:219–231
- Ahmad MI, Sinclair CD, Werritty A (1988) Log-logistic flood frequency analysis. *J Hydrol* 98:215–224
- Ames DP (2006) Estimating 7Q10 confidence limits from data: a bootstrap approach. *J Water Resour Plan Manage* 132:204–208
- An Y, Pandey MD (2005) A comparison of methods of extreme wind speed estimation. *J Wind Eng Ind Aerodyn* 93:535–545
- Anderson PL, Meerschaert MM (1998) Modeling river flows with heavy tails. *Water Resour Res* 34:2271–2280
- Asquith WH (1998) Depth-duration frequency of precipitation for Texas. US Geological Survey, Water-Resources Investigations Report 98–4044, Austin, Texas, pp 112
- Bacro JN, Chaouche A (2006) Uncertainty in the estimation of extreme rainfalls around the Mediterranean Sea: an illustration using data from Marseille. *J Hydrol Sci* 51:389–405
- Beguieria S, Vicente-Serrano SM (2006) Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. *Journal of Applied Meteorology and Climatology* 45:108–124
- Brabson BB, Palutikof JP (2002) The evolution of extreme temperatures in the Central England temperature record. *Geophys Res Lett* 29:2163. doi:10.1029/2002GL015964
- Buishand TA (1991) Extreme rainfall estimation by combining data from several sites. *J Hydrol Sci* 36:345–365
- Carpenter J, Bithell J (2000) Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med* 19:1141–1164
- Coles S (2001) An introduction to statistical modeling of extreme values. Springer, London, p 208
- Cox DR, Isham VS, Northrop PJ (2002) Floods: some probabilistic and statistical approaches. *Philos Trans R Soc* 360:1389–1408
- Cunderlik J, Ouarda TBMJ (2007) Regional flood-duration-frequency modeling in the changing environment. *J Hydrol* 318:276–291
- Davison AC, Hinkley DV (1997) Bootstrap methods and their application. Cambridge University Press, Cambridge p 592
- Dixon R (2002) Bootstrap resampling. In: El-Shaarawi AH, Piegorsch WW (eds) The encyclopedia of environmetrics. Wiley, New York, pp 212–219
- Dunn PK (2001) Bootstrap confidence intervals for predicted rainfall quantiles. *Int J Climatol* 21:89–94
- Dupuis DJ, Field CA (1998) A comparison of confidence intervals for generalized extreme-value distributions. *J Stat Comput Simul* 61:341–360
- Dupuis DJ, Field CA (2004) Large wind speeds: modeling and outlier detection. *Journal of Agricultural Biological and Environmental Statistics* 9:105–121
- Efron B (1979) Bootstrap methods: another look at the jackknife. *Ann Stat* 7:1–26
- Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman and Hall, New York
- Egozcue JJ, Ramis C (2001) Bayesian hazard analysis of heavy precipitation in eastern Spain. *Int J Climatol* 21:1263–1279
- Farquharson FAK, Meigh JR, Sutcliffe JV (1992) Regional flood frequency analysis in arid and semi-arid areas. *J Hydrol* 138:487–501
- Fitzgerald DL (2005) Analysis of extreme rainfall using the log-logistic distribution. *Stoch Env Res Risk A* 19:249–257
- Fowler HJ, Ekström M (2009) Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes. *Int J Climatol* 29:385–416
- Gaal L, Kysely J, Szolgay J (2008) Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia. *Hydrol Earth Syst Sci* 12:825–839
- Gnedenko B (1943) Sur la distribution limite du terme maximum d'une serie aleatoire. *Ann Math* 44:423–453
- Hall MJ, van den Boogaard HFP, Fernando RC, Mynett AE (2004) The construction of confidence intervals for frequency analysis using resampling techniques. *Hydrol Earth Syst Sci* 8:235–246
- Hewa GA, Wang QJ, McMahon TA, Nathan RJ, Peel MC (2007) Generalized extreme value distribution fitted by LH moments for low-flow frequency analysis. *Water Resour Res* 43:W06301. doi:10.1029/2006WR004913
- Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J Roy Stat Soc* 52B:105–124
- Hosking JRM, Wallis JR (1997) Regional frequency analysis. An approach based on L-moments. Cambridge University Press, Cambridge, p 224
- IH (1999) The flood estimation handbook. Institute of Hydrology, Wallingford
- Katsoulis BD, Hatzianastassiou N (2005) Analysis of hot spell characteristics in the Greek region. *Clim Res* 28:229–241
- Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. *Adv Water Res* 25:1287–1304
- Khalik MN, St-Hilaire A, Ouarda TBMJ, Bobee B (2005) Frequency analysis and temporal pattern of occurrences of southern Quebec heatwaves. *Int J Climatol* 25:485–504
- Kharin VV, Zwiers FW (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *J Clim* 13:3760–3788
- Kharin VV, Zwiers FW (2005) Estimating extremes in transient climate change simulations. *J Clim* 18:1156–1173
- Kjeldsen TR, Jones DA (2004) Sampling variance of flood quantiles from the generalised logistic distribution estimated using the method of L-moments. *Hydrol Earth Syst Sci* 8:183–190
- Kochanek K, Strupczewski WG, Singh VP, Weglarczyk S (2008) The PWM large quantile estimates of heavy tailed distributions from samples deprived of their largest element. *J Hydrol Sci* 53:367–386
- Kroll CN, Vogel RM (2002) Probability distribution of low stream-flow series in the United States. *J Hydrol Eng* 7:137–146
- Kumar R, Chatterjee C (2005) Regional flood frequency analysis using L-moments for North Brahmputra region of India. *J Hydrol Eng* 10:1–7
- Kysely J (2002) Comparison of extremes in GCM-simulated, downscaled and observed central-European temperature series. *Clim Res* 20:211–222
- Kysely J (2008) A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme value models. *Journal of Applied Meteorology and Climatology* 47:3236–3251
- Kysely J (2009) Trends in heavy precipitation in the Czech Republic over 1961–2005. *Int J Climatol*. doi:10.1002/joc.1784
- Kysely J, Beranová R (2009) Climate change effects on extreme precipitation in central Europe: uncertainties of scenarios based on regional climate models. *Theor Appl Climatol* 95:361–374
- Kysely J, Picek J (2007) Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic. *Clim Res* 33:243–255
- Kysely J, Beranová R, Picek J, Štěpánek P (2008) Simulation of summer temperature extremes over the Czech Republic in regional climate models. *Meteorol Z* 17:645–661
- Lamoureux S (2000) Five centuries of interannual sediment yield and rainfall-induced erosion in the Canadian High Arctic recorded in lacustrine varves. *Water Resour Res* 36:309–318

- Lana X, Burgueno A, Martinez MD, Serra C (2006) Statistical distributions and sampling strategies for the analysis of extreme dry spells in Catalonia (NE Spain). *J Hydrol* 324:94–114
- Lee SH, Maeng SJ (2003) Frequency analysis of extreme rainfall using L-moments. *Irrig Drain* 52:219–230
- Lu L-H, Stedinger JR (1992) Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. *J Hydrol* 138:223–245
- Martins ES, Stedinger JR (2000) Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour Res* 36:737–744
- Onoz B, Bayazit M (1999) Generalized extreme value-PWM model for distribution of minimum streamflows. *J Hydrol Eng* 4:289–292
- Overeem A, Buishand A, Holleman I (2008) Rainfall depth-duration-frequency curves and their uncertainties. *J Hydrol* 348:124–134
- Pandey MD, van Gelder PHAJM, Vrijling JK (2004) Dutch case studies of the estimation of extreme quantiles and associated uncertainty by bootstrap simulations. *Environmetrics* 15:687–699
- Prudhomme C, Jakob D, Svensson C (2003) Uncertainty and climate change impact on the flood regime of small UK catchments. *J Hydrol* 277:1–23
- Semmler T, Jacob D (2004) Modeling extreme precipitation events—a climate change simulation for Europe. *Global Planet Change* 44:119–127
- Shoukri MM, Mian IUM, Tracy DS (1988) Sampling properties of estimators of the log-logistic distribution with application to Canadian precipitation data. *The Canadian Journal of Statistics* 16:223–236
- Tajvidi N (2003) Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions. *Extremes* 6:111–123
- Twardosz R (2009) Probabilistic model of maximum precipitation depths for Kraków (southern Poland, 1886–2002). *Theor Appl Climatol*. doi:10.1007/s00704-008-0087-4
- van den Brink HW, Koennen GP, Opsteegh JD (2004) Statistics of synoptic-scale wind speeds in ensemble simulations of current and future climate. *J Clim* 17:4564–4574
- Voss R, May W, Roeckner E (2002) Enhanced resolution modelling study on anthropogenic climate change: changes in extremes of the hydrological cycle. *Int J Climatol* 22:755–777
- Zin WZW, Jemain AZ, Ibrahim K (2009) The best fitting distribution of annual maximum rainfall in Peninsular Malaysia based on methods of L-moment and LQ-moment. *Theor Appl Climatol* 96:337–344