Acta Geophysica
vol. 55, no. 4, pp. 652-678
DOI 10.2478/s11600-007-0027-1

On the informative value of the largest sample element of log-Gumbel distribution

Witold G. STRUPCZEWSKI[1], Krzysztof KOCHANEK[1] and Vijay. P. SINGH[2]

[1]Water Resources Department, Institute of Geophysics
Polish Academy of Sciences, Warszawa, Poland
e-mails: wgs@igf.edu.pl; kochanek@igf.edu.pl

[2]Department of Biological and Agricultural Engineering
Texas A&M University, College Station, Texas, USA
e-mail: vsingh@tamu.edu

A b s t r a c t

Extremes of stream flow and precipitation are commonly modeled by heavy-tailed distributions. While scrutinizing annual flow maxima or the peaks over threshold, the largest sample elements are quite often suspected to be low quality data, outliers or values corresponding to much longer return periods than the observation period. Since the interest is primarily in the estimation of the right tail (in the case of floods or heavy rainfalls), sensitivity of upper quantiles to largest elements of a series constitutes a problem of special concern. This study investigated the sensitivity problem using the log-Gumbel distribution by generating samples of different sizes (n) and different values of the coefficient of variation by Monte Carlo experiments. Parameters of the log-Gumbel distribution were estimated by the probability weighted moments (PWMs) method, method of moments (MOMs) and maximum likelihood method (MLM), both for complete samples and the samples deprived of their largest elements. In the latter case, the distribution censored by the non-exceedance probability threshold, F_T, was considered. Using F_T instead of the censored threshold T creates possibility of controlling estimator property. The effect of the F_T value on the performance of the quantile estimates was then examined. It is shown that right censoring of data need not reduce an accuracy of large quantile estimates if the method of PWMs or MOMs is employed. Moreover allowing bias of estimates one can get the gain in variance and in mean square error of large quantiles even if ML method is used.

Key words: floods, log-Gumbel distribution, estimation methods, bias, Monte Carlo simulation, truncation, censoring, non-exceedance probability threshold.

1. INTRODUCTION

The main interest in flood frequency analysis (FFA) is the estimation of large quantiles from a given time series of annual flow maxima (AM) or of peaks over threshold (POT). Hence, the reliability of the largest element of a series is a problem of special concern. Due to technical difficulties of flow discharge measurements during floods and short duration of flood events, as well as due to the difficulties in getting reliable results from physical models for complex geometries of river channels, the scarcity of information for high water levels is a norm. As a result, the upper limb of a rating curve, which is used for the conversion of water gauge heights into flow discharges, may be severely biased. Therefore, one deals with error-corrupted data, while the accuracy of measured data decreases with increasing flood magnitudes. The highest values in annual maximum flow series may be considered as poor quality data and this inaccuracy undermines their information value. Quite often the largest element in a series considerably deviates from other elements and, hence, one suspects that it corresponds to a much longer return period than the period of a time series. However, there is a lack of an objective effective method for determining whether the largest element of hydrological size sample is an outlier and whether its removal is justified. If the largest sample elements are very low quality data, their values may be considered as unavailable and then the parameters of a distribution are estimated from a censored data set. This study also encompasses the case when the largest sample element was erroneously excluded from the analysis and a sample deprived of its largest element was considered as a complete sample from a given distribution. Whereas, the view that heavy floods are generated by mechanisms different from those generating other smaller floods is in conformity with the statistical meaning of outliers. Then, one deals with double censoring of data. For the sake of short time series, the problem with this view is that it can be hardly implemented without using paleoflood data.

The problem posed in this study is to evaluate the effect of omission of the largest sample element on the accuracy of large quantile estimates for the known distribution function. Common sense should tell us that this will probably reduce the accuracy of large quantile estimation. As stated in the statistical literature (e.g. Kendall and Stuart 1973), censoring always results in the loss of estimation efficiency. Although the maximum likelihood (ML) and the least squares method have normally been employed, this property has been without any proof ascribed to any estimation method. It may not be the case for heavy tailed distributed data if the Probability Weighted Moments (PWMs) or the method of moments (MOMs) is employed and estimation accuracy is compared with the one got by the same method for a complete sample. Taking the log-Gumbel distribution, it is shown here by simulation experiments that right singly censoring need not decrease accuracy of large quantile estimates expressed by the mean square error (*MSE*). Although a censored distribution is employed for parameter estimation, the same holds for censored samples estimates. To be precise, according to Rao (1958) truncation need not result in the loss of estimation efficiency. Swamy (1962) shows that the truncation of normal distribution always reduces efficiency when both mean and variance are estimated. Note that the efficiency can serve as the measure of accuracy for "unbiased" estimators while we will deal also with biased estimators. Then the root mean square error *RMSE* and bias *B* are commonly used measures of the performance of an estimator. As shown in the paper, allowing for biased estimates one can get the gain in variance and *RMSE* of large quantile estimates even if ML method is employed.

Presently the method of *L*-moments (LMM) (Hosking and Wallis 1997), being a modification of the method of Probability Weighted Moments (PWMs) (Greenwood *et al*. 1979), dominates in FFA because of its computational simplicity and satisfactory performance for hydrological sample sizes. Hosking *et al*. (1985) and Hosking and Wallis (1987) found that with small and moderate samples the method of *L*-moments is often more efficient than the maximum likelihood, particularly for estimating quantiles in the upper tail of the distribution. Hosking (1995) extended the theory of *L*-moments to the analysis of upper bound censored samples. His concept of the "A"-type PWMs with Type II censoring is employed in this study and extended for the two other estimation methods, i.e., MOMs and MLM. Its performance with reference to a large quantile of the LG distribution is the subject of our investigation. The censored distribution (the term coined by Hosking 1995) is applied for a fixed proportion m/n of the sample size n and a given non-exceedance probability threshold, F_T. Hence, the upper threshold T of a variable is a random value, as it varies from sample to sample. Replacement of T by F_T creates possibility of controlling the estimator property. For the sake of brevity, the only largest element of a sample is removed from the sample, i.e., $m = n - 1$.

There are several reasons for selecting the LG distribution for this study. First, it was found in our previous study (Strupczewski *et al.* 2007) that for two-parameter heavy tailed and log-normal distributions the removal of the largest sample element need not result in a decrease in the accuracy of large quantile estimates, while it produces a negative bias. Two-parameter models are recommended by Cunnane (1989) for at-site FFA. Second, nowadays there is a growing consensus that hydrological extremes are heavy-tail distributed (e.g. Katz *et. al*. 2002). Among two-parameter heavy-tailed distributions, the log-logistic (LL), the log-Gumbel (LG), i.e. the two-parameter General Extreme Value (GEV) distribution, and Pareto are most popular for modeling floods, precipitation and sea level extremes. Third, a large dispersion of largest sample elements makes heavy-tail distributed data interesting for this study. Fourth, being a two-parameter distribution bounded-at-zero, LG has explicit forms of both the cumulative density function (CDF) and the inverse thereof. Hence, all the three estimation methods can be easily applied.

The performance of large quantile estimates, $\hat{x}_F$, is assessed by Monte Carlo simulation experiments and compared with that of a complete sample by the same method. Three estimation methods, i.e. PWMs, MOMs and MLM, are employed for the purpose. Therefore, using a truncated distribution, an informative value of the largest sample element in respect to upper quantile estimates is assessed.

The papers is organized as follows. Followed by the notation (Section 2), the next section provides a short review of hydrological literature in respect to censored data. Application of three estimation methods to the distribution censored by a non-exceedance probability threshold is described in Section 4 and exemplified by the LG distribution in Section 5. In Section 6 performance measures that are applied to large quantile estimators are introduced and four ways (variants) of the non-exceedance probability threshold F_T are formulated. Design of simulation experiments employed to assess the informative value of the largest element of LG samples is described in Section 7. Section 8 presents, discusses and compares the experimental results got by the three estimation methods for the estimation of quantile $x_{0.99}$ for complete samples (subsection 8.1), for the censored data with four F_T variants (subsection 8.2), and for quantile $x_{0.80}$ (subsection 8.3), and it is completed by a comparative assessment of the various procedures applied (subsection 8.4). The last part, Section 9, summarizes and concludes the paper.

2. NOTATION

$B(\cdot)$	bias (eq. 28)
$B_\delta(\cdot)$	relative bias (eq. 31)
β_r	r-th theoretical probability weighted moment (eq. 3)
b_r	r-th sampling probability weighted moment (eq. 5)
CDF	cumulative density function
CS	complete sample
C_V	coefficient of variation
$E(\cdot)$	expected value
EM	estimation method
$f(\cdot)$	probability density function
$F(\cdot)$	cumulative density function
FFA	flood frequency analysis
F_T	non-exceedance probability threshold
$g(\cdot)$	PDF of censored distribution
$G(\cdot)$	CDF of censored distribution
LG	log-Gumbel distribution
L_T	likelihood function of censored distribution (eq. 10)
m	censored sample size
MLM	maximum likelihood method
MOMs	method of moments
m_r	r-th theoretical moment about the origin (eq. 8)

$\hat{m}_r$	r-th sampling moment about the origin (eq. 9)
$MSE(\cdot)$	mean square error
n	total sample size
PDF	probability density function
PWMs	probability weighted moments
$SD(\cdot)$	standard deviation
$SD_\delta(\cdot)$	relative standard deviation (eq. 31)
$R^v(\cdot)$	*RMSE* ratio (eq. 32)
$RMSE(\cdot)$	root mean square error (eq. 29)
$RMSE_\delta(\cdot)$	relative *RMSE* (eq. 31)
T	censoring threshold
V	variant for selecting the F_T value (V=A, B, C, D)
var $(\cdot)$	variance
$x(F)$	quantile function
$\hat{x}(F)$ or $\hat{x}_F$	F-th quantile estimator
$x_{i:n}$	the i-th element of an ordered sample of size n
α	parameter of LG distribution
θ	parameters of a distribution function
ξ	parameter of LG distribution

3. REVIEW OF HYDROLOGICAL LITERATURE

The data set in which largest elements are not available can be in fact considered as a censored data set and appropriate estimation techniques can then be applied. Because of its simplicity, the quantile method has been used in practice to deal with such cases, despite the fact that the method is recognized as inefficient. Kaczmarek (1977, p. 205) was one of the first who employed MLM to censored-on-the-right annual peak flow data, where the data were assumed to come from the log-normal population. He stated “Since moments from the sample are undeterminable in this case, to solve the above problem only the maximum likelihood and the quantile method can be used”. Phien and Fang (1989) used MLM to GEV samples censored on the right, left and both sides (censoring of Type I). It was found that censoring may reduce the bias of the parameter estimators but does not necessarily increase the variances.

Presently, censoring and truncation can be tackled by two other methods, namely, method of moments (MOMs) (Jawitz 2004, Strupczewski *et al.* 2006) and method of probability weighted moments (PWMs) (Wang 1990, 1996; Hosking 1995). Introducing partial PWMs, Wang (1990, 1996) extended the definition of PWMs to singly and doubly censored samples. Computing sampling partial PWMs, the unavailable values of sample elements are replaced by the zeros. While integrating to get the population partial PWMs, the censoring threshold T is expressed by the threshold non-exceedance probability F_T. Note that in ML estimation from censored sample, the censoring threshold T is used both in Type I and Type II censoring (e.g. Kendall and Stuart 1973). In fact, for a specified T, the exact value of F_T from a sample is not known. Given that m of the n events in the sample do not exceed the threshold T, the threshold non-exceedance probability is estimated by Wang (1990, 1996) and Hosking (1995) as

$$F_T = m/n \; . \tag{1}$$

As opposed to PWMs, the partial PWMs cannot be used as measures of the shape of the distribution; for example, the second L-moment as a measure of dispersion is always positive, while the second partial L-moment can be also negative. Employing Monte Carlo simulation for generating samples from the GEV distribution, Wang (1990, 1996) showed that lower bound censoring at a moderate level does not unduly reduce the efficiency of large quantile estimation.

Hosking (1995) introduced two ways of extending the definitions of PWMs to upper bounded censored samples. His "A"-type PWMs is based on the concept of conditional probability. The distribution is censored by the threshold T or F_T. In Hosking's "B"-type PWMs, unobserved values are replaced by the censoring threshold T, i.e. PWMs are calculated from the "completed sample". For estimating the parameters, Hosking prefers it to his "A"-type PWMs, and shows that for the zero-bounded exponential distribution it is equivalent to the ML under Type II censoring. Using censored GEV and the inverse Gumbel samples from a simulation study, Hosking's results indicate that PWMs-based estimators of upper quantiles perform well and that they can be competitive to computationally more complex methods, such as maximum likelihood.

Recognizing the robustness of estimates to largest sample elements as a desirable property of an estimation method in FFA, Vogel and Fennessey (1993) and Hosking and Wallis (1997) deleted the largest observation from several records and compared relative changes of L-moment ratios of order $r = 2$, 3 and 4 with those of ordinary moment ratios and found them to be less affected by extreme observations. Following this way and using the data simulated by Monte Carlo experiments, Strupczewski *et al.* (2007) investigated, by employing three estimation methods, the effect of removing the largest sample element on upper quantile estimators.

4. CENSORED DISTRIBUTION APPLIED TO CENSORED DATA

Consider a continuous positively defined variable x with probability density function (PDF) $f(x|\theta)$ and cumulative density function (CDF) $F(x|\theta)$ where θ denotes parameters. The distribution is to be estimated from only sample values that do not exceed a threshold T with $F_T = F(T)$. The smallest m sample values of the n-element sample $(x_{1:n} \le x_{2:n} \le \ldots \le x_{m:n} \le x_{m+1:n} \le \ldots \le x_{n:n})$ are observed.

Conditional on the value of m, the uncensored values constitute a random sample of size m from a censored distribution with PDF $g(y|\theta) = f(x|\theta)/F_T$, CDF $G(y|\theta) = \mathrm{F}(x|\theta)/F_T$, $0 < G < 1$ and quantile function $y(G|\theta) = x(\mathrm{F} = F_T \cdot G|\theta)$. Here m is fixed and the largest $(n - m)$ values are censored above the threshold nonexceedance probability F_T, which corresponds to the random value $\hat{T} \geq x_{m:n}$. The θ parameters are to be estimated from the uncensored sample of m observed values of the censored distribution. Note that for a fixed proportion m/n, one gets asymptotically $\lim_{n\to\infty} \hat{x}\left(F_T = m/n\right) = \lim_{n\to\infty} \hat{y}\left(G = 1\right) = T$.

PWMs estimates

The probability weighted moments (PWMs) of a truncated distribution can be written as

$$\beta_r = \int_0^1 y\left(G\right) G^r\left(y\right) \mathrm{d}G = \frac{1}{F_T^{r+1}} \int_0^{F_T} x\left(F\right) F^r \,\mathrm{d}F \tag{2}$$

and in particular

$$\beta_0 = \int_0^1 y\left(G\right) \mathrm{d}G = \frac{1}{F_T} \int_0^{F_T} x\left(F\right) \mathrm{d}F \,, \tag{3}$$

$$\beta_1 = \int_0^1 y\left(G\right) G \,\mathrm{d}G = \frac{1}{F_T^2} \int_0^{F_T} x\left(F\right) F \,\mathrm{d}F \,. \tag{4}$$

Sampling *PWM*s of a truncated distribution are

$$b_r = \frac{1}{m} \sum_{i=r+1}^{m} \frac{\left(i-1\right)\left(i-2\right)..\left(i-r\right)}{\left(m-1\right)\left(m-2\right)..\left(m-r\right)} x_{i:n} \tag{5}$$

and in particular

$$b_0 = \frac{1}{m} \sum_{i=1}^{m} x_{i:n} \,, \tag{6}$$

$$b_1 = \frac{1}{m} \sum_{i=2}^{m} \frac{\left(i-1\right)}{\left(m-1\right)} x_{i:n} \,. \tag{7}$$

Equating sample and population PWMs, i.e., $b_j = \beta_j$ for $j = 0, 1, \ldots, k$, where $(k + 1)$ stands for the number of parameters, while F_T is a given value, one gets the PWMs-estimate of the θ parameters.

MOMs estimates

Application of the method of moments (MOMs) for censored samples is analogous to that of PWMs. The theoretical r-th moment about the origin can be expressed as

$$m_r = \int_0^1 y^r\left(G\right) \mathrm{d}G = \frac{1}{F_T} \int_0^{F_T} x^r\left(F\right) \mathrm{d}F \,. \tag{8}$$

The sample r-th moment about the origin can be written as

$$\hat{m}_r = \frac{1}{m}\sum_{i=1}^{m} x_{i:n}^r . \tag{9}$$

Solving the system of equations $\hat{m}_r = m_r$ for $r = 1,\ldots, l$, where $l = k + 1$ stands for the number of parameters, while F_T is a given value, one gets the MOMs-estimate of the θ parameters.

Maximum likelihood (ML) estimates

If the smallest m observations are available, then the likelihood function (LF) can be expressed as

$$L_T = \prod_{i=1}^{m} g\left(x_{i;n}|\theta\right) = \frac{\prod_{i=1}^{m} f\left(x_{i:n}|\theta\right)}{\left\{\int_0^T f\left(x|\theta\right)\mathrm{d}x\right\}^m} = \prod_{i=1}^{m} f\left(x_{i:n}|\theta\right)/F^m\left(T\right), \tag{10}$$

where for m fixed in advance $T = x_{m:n}$.

To copy Hosking's PWMs "A"- type parameter estimation technique, we expressed in the equations of the maximum likelihood conditions, $\partial \ln L_T/\partial\theta = 0$, T by F_T using the inverse CDF: $T = \phi\,(F_T\,|\,\theta)$. Substituting the ML-estimates of the θ parameters into the quantile equation of the original distribution we get

$$\hat{x}_F = \phi\left(F\middle|\hat{\theta}\right), \quad 0 \le F \le 1 . \tag{11}$$

Putting $F = F_T$ into eq. (11) one can find and compare the threshold estimate $\hat{T}$ with the largest value of the uncensored sample of m observed values, i.e. with $x_{m:n}$.

5. LOG-GUMBEL DISTRIBUTION TRUNCATED ON THE RIGHT

The log-Gumbel distribution (notation after Rowiński *et al*. 2002) can be written as

$$f\left(x\right) = \frac{\xi}{\alpha} x^{-1/\alpha-1} \exp\left(-\xi x^{-1/\alpha}\right), \qquad \xi,\ \alpha > 0 , \tag{12}$$

$$F\left(x\right) = \exp\left(-\xi x^{-1/\alpha}\right), \tag{13}$$

$$x\left(F\right) = \left(-\frac{\ln F}{\xi}\right)^{-\alpha} . \tag{14}$$

PWMs estimates of parameters

PWMs of the log-Gumbel censored distribution can be expressed as

$$\beta_r = \frac{\alpha}{F_T^{r+1}}(r+1)^{\alpha-1}\xi^a\,\Gamma\left[1-\alpha,\ -(r+1)\ln F_T\right], \tag{15}$$

where $\Gamma(a,z) = \int_z^{+\infty} t^{a-1}e^{-t}dt$ is the incomplete gamma function. In particular,

$$\beta_0 = \frac{1}{F_T}\xi^\alpha\,\Gamma(1-\alpha,\ -\ln F_T), \tag{16}$$

$$\beta_1 = \frac{1}{F_T^2}2^{\alpha-1}\xi^\alpha\,\Gamma(1-\alpha,\ -2\ln F_T). \tag{17}$$

Parameter α is estimated from

$$\frac{\beta_1}{\beta_0} = \frac{2^{\alpha-1}\Gamma(1-\alpha,\ -2\ln F_T)}{F_T\,\Gamma(1-\alpha,\ -\ln F_T)} = \frac{b_1}{b_0} \tag{18}$$

then

$$\xi = \left[\frac{b_0\cdot F_T}{\Gamma(1-\alpha,\ -\ln F_T)}\right]^{\frac{1}{\alpha}}. \tag{19}$$

Conventional moments estimates

The r-th moment about the origin of the censored LG distribution can be written as

$$m_r = \frac{1}{F_T}\xi^{r\alpha}\,\Gamma(1-r\alpha,\ -\ln F_T). \tag{20}$$

In particular,

$$m_1 = \frac{1}{F_T}\xi^\alpha\,\Gamma(1-\alpha,\ -\ln F_T), \tag{21}$$

$$m_2 = \frac{1}{F_T}\xi^{2\alpha}\,\Gamma(1-2\alpha,\ -\ln F_T). \tag{22}$$

The α parameter is set by solving eq. (23) for a given F_T as

$$\frac{\hat{m}_2}{\hat{m}_1^2} = F_T\frac{\Gamma(1-2\alpha,\ -\ln F_T)}{\Gamma^2(1-\alpha,\ -\ln F_T)} \tag{23}$$

and ξ from

$$\xi = \left[\frac{\hat{m}_1}{\Gamma(1-\alpha,\ -\ln F_T)}\right]^{\frac{1}{\alpha}}. \tag{24}$$

ML estimates

Log of L-function (eq. 10) takes the form:

$$\ln L_T = m\ln\xi - m\ln\alpha - \left(1/\alpha + 1\right)\sum_{i=1}^{m}\ln x_{i:n} - \xi\sum_{i=1}^{m} x_{i:n}^{-1/\alpha} + m\xi T^{-1/\alpha} , \tag{25}$$

$$\frac{\partial \ln L_T}{\partial\alpha} = -\frac{m}{\alpha} - \frac{\xi}{\alpha^2}\sum_{i=1}^{m} x_{i:n}^{-1/\alpha}\ln x_{i:n} + \frac{1}{\alpha^2}\sum_{i=1}^{m}\ln x_{i:n} + m\frac{\xi}{\alpha^2}T^{-1/\alpha}\ln T = 0 , \tag{26}$$

$$\frac{\partial \ln L_T}{\partial\xi} = \frac{m}{\xi} - \sum_{i=1}^{m} x_{i:n}^{-1/\alpha} + mT^{-1/\alpha} = 0 . \tag{27}$$

Putting $x = T$ into eq. (14) one gets

$$T = \left(-\frac{\ln F_T}{\xi}\right)^{-\alpha} , \tag{28}$$

hence

$$\ln T = -\alpha\left[\ln\left(-\ln F_T\right) - \ln\xi\right]. \tag{29}$$

Substituting eqs. (28) and (29) into eqs. (26) and (27) one gets

$$\frac{\partial \ln L_T}{\partial\alpha} = -\frac{m}{\alpha} - \frac{\xi}{\alpha^2}\sum_{i=1}^{m} x_{i:n}^{-1/\alpha}\ln x_{i:n} + \frac{1}{\alpha^2}\sum_{i=1}^{m}\ln x_{i:n} + \frac{m}{\alpha}\ln F_T \ln\left(-\frac{1}{\xi}\ln F_T\right) = 0 , \tag{26a}$$

$$\frac{\partial \ln L_T}{\partial\xi} = \frac{m}{\xi} - \sum_{i=1}^{m} x_{i:n}^{-1/\alpha} - \frac{m}{\xi}\ln F_T = 0 . \tag{27a}$$

From eq. (27a) one obtains

$$\hat{\xi} = \frac{m\left(1 - \ln F_T\right)}{\sum_{i=1}^{m} x_{i:n}^{-1/\hat{\alpha}}} \tag{27b}$$

and from eq. (26a) one gets

$$\hat{\alpha} = \frac{1}{m}\frac{\hat{\xi}\sum_{i=1}^{m} x_{i:n}^{-1/\hat{\alpha}}\ln x_{i:n} - \sum_{i=1}^{m}\ln x_{i:n}}{\ln F_T \ln\left(-\frac{1}{\hat{\xi}}\ln F_T\right) - 1} . \tag{26b}$$

Solving eq. (26b) for α by an iterative method one gets the α estimate and then $\hat{\xi}$ from eq. (27b).

6. VARIANTS OF UPPER QUANTILE ESTIMATION

Because the primary interest in FFA is the estimation of large quantiles, this paper concentrates on the performance of estimator $\hat{x}(F=0.99)$ which is then compared to the performance of $\hat{x}(F=0.80)$. Common measures of the performance of an estimator $\hat{x}_F$ are its bias (B) and root mean square error (*RMSE*) defined by

$$B(\hat{x}_F)=E(\hat{x}_F-x_F), \tag{28}$$

$$RMSE(\hat{x}_F)=\left[B^2(\hat{x}_F)+SD^2(\hat{x}_F)\right]^{1/2}=\left[E(\hat{x}_F-x_F)^2\right]^{1/2}, \tag{29}$$

where

$$SD(\hat{x}_F)=\left[\mathrm{var}(\hat{x}_F)\right]^{1/2}=\left\{E\left[\hat{x}_F-E(\hat{x}_F)\right]^2\right\}^{1/2}. \tag{30}$$

It is convenient to express bias, *RMSE* and standard deviation as ratios with respect to the quantile itself, i.e.,

$$B_\delta(\hat{x}_F)=B(\hat{x}_F)/x_F, \tag{28a}$$

$$RMSE_\delta(\hat{x}_F)=RMSE(\hat{x}_F)/x_F \tag{29a}$$

and

$$SD_\delta(\hat{x}_F)=SD(\hat{x}_F)/x_F=\left\{\left[RMSE_\delta(\hat{x}_F)\right]^2-\left[B_\delta(\hat{x}_F)\right]^2\right\}^{1/2}. \tag{30a}$$

Thus, the performance measure quantities used in the paper are expressed as the percentage of the underlying population quantiles (x_F).

The ratio of *RMSE*s

$$R^v(\hat{x}_{0.99})=RMSE_\delta^v(\hat{x}_{0.99})/RMSE_\delta(\hat{x}_{0.99}), \tag{31}$$

where v stands for A, B, C and D Variants, and serves for every estimation method as a measure of censoring on the accuracy of $\hat{x}_{0.99}$.

For the three estimation methods, the performance of upper quantile estimators obtained from the censored LG distribution is compared with that of a complete sample obtained by the same method. This comparison gives the answer to the question "how for a LG sample and a given estimation method the lacking of its largest element value affects the accuracy of upper quantile estimators measured by *RMSE* and bias".

So far, the probability bound F_T for censoring has been considered as a given value. In practice, F_T is generally unknown and it has to be estimated. Taking into account difficulties in getting minimum variance unbiased estimator for truncated distribution (e.g. Kendall and Stuart 1973, 32.14-23), practitioners are satisfied with methods that minimise *RMSE* or *B*, depending on their needs and a share of bias in *RMSE*.

Therefore, the first task in this study is to find F_T which for given ($m = n - 1$, n) and fixed values of LG parameters, minimises *RMSE* of x_F estimator:

$$\min_{F_T} RMSE_T\left(\hat{x}_F\right) = \min_{F_T}\left[E\left(\hat{x}_F - x_F\right)^2\right]^{1/2} . \tag{32}$$

This is represented here as the Variant A. However, one has to bear in mind that the *RMSE* is a good measure of performance only when bias has limited impact on it, i.e. the B^2 is a small fraction of *MSE*. Hence, the Variant B is introduced with the aim of getting an unbiased estimator of $\hat{x}_F$:

$$\min_{F_T}\left|B\left(\hat{x}_F\right)\right| = \min_{F_T}\left|E\left(\hat{x}_F - x_F\right)\right| = 0 . \tag{33}$$

Estimation of F_T by a fraction of observable sample elements (eq. 1) is considered here as the Variant C. Note that this $\hat{F}_T$ may not satisfy the condition $\hat{T} = x\left(\hat{F}_T, \hat{\alpha}, \hat{\xi}\right) \geq x_{m:n}$. In fact, if n is reasonably large and moreover $x_{m:n}$ is located within the main probability mass of the range of x, then $\hat{T} \approx x_{m:n}$ and $\hat{F}_T \approx m/n$. This hardly happens in hydrology, where one deals with short samples and the threshold T is located far in the right tail of PDF. Perceiving some theoretical disadvantages using this simple approach for parameter estimation, Hosking (1995) stated that finite sample corrections may, in some circumstances, be worthwhile. From the unbiased estimator of the mean of the exponential distribution based on the “A” type PWMs (Hosking 1995, eq. 29.3.6) one can find that F_T exceeds the m/n value, asymptotically converging to it. For example, taking $m/n = 0.9$, the probability bound F_T equals 0.9128, 0.9089, 0.9055 and 0.9028 for n equal to 20, 30, 50 and 100, respectively. The ratio m/n can be considered as a conservative estimate of F_T. Recognized in hydrology under the name “California Department of Public Works Formula” it gives the greatest F_T value of all commonly used plotting position formulas (Cunnane 1981, 1989; Rao and Hamed 2000). To tackle the case when the largest sample element is erroneously treated as an outlier and then removed from the sample, we put $F_T = 1$ (the Variant D). It means that the ($n - 1$) element sample, i.e., a sample deprived of its largest element ($x_{n:n}$), is considered as the complete sample of the LG distribution.

Therefore, four variants for the F_T value along with three estimation methods make all together twelve “variant/associated estimation method” procedures, denoted as the V/EM. Their performance is investigated and compared with those from three estimation procedures employed for the complete sample, which are denoted as CS/EM.

For convenience, the original LG parameters ξ, α are replaced by the mean m_1 and the coefficient of variation C_V. This facilitates comparison of results got for various two-parameter models. Putting $F_T = 1$ into eq. (20) one gets

$$m_1 = \xi^{\alpha}\Gamma\left(1-\alpha\right) \tag{34}$$

and from the C_V definition

$$C_V = \sqrt{\frac{\Gamma(1-2\alpha)}{\Gamma^2(1-\alpha)} - 1}\,. \tag{35}$$

The value of the threshold non-exceedance probability F_T is subject to the objective function (32) in Variant A and (33) in Variant B and it will be assessed by the Monte-Carlo simulation experiment. Hence, n is not explicitly used in the parameter estimation. For a given objective function the F_T value is an implicit function of (m, n) and the population C_V (or α) value. For the C and D Variants, there is a fixed value of F_T, which equals m/n and one, respectively.

7. EXPERIMENTAL DESIGN

The Monte Carlo experiments were carried out for each 'point' in the sample space [C_V, n estimation method, variant] with the mean value equal to ten. The sample space consisted of 630 points, covering six values of C_V (0.3, 0.6, 0.8, 1.0, 1.5, 2.0), seven values of n (n = 20, 30, 50, 70, 100, 1000, 10000), three estimation methods, the complete sample (CS) and four variants of F_T, i.e., A, B, C and D, applied to censored samples. Note that usually $C_V < 1$ in annual flow maxima series and their size n is less than one hundred. For each point in the sample subspace [C_V , n], 20,000 sequences of length n were generated. Without loss of generality, the LG distribution mean used to generate the samples was $m_1 = 10$. To obtain the value of quantile $x_{F\,=\,0.99}$ for the LG population with the mean equal to ten and a given value of the coefficient of variation C_V , eqs. (14), (34) and (35) were employed.

For each of the 20,000 sequences associated with a point in the sample subspace [C_V , n], and three estimation methods, the distribution parameters and quantile $x_{0.99}$, denoted as $\hat{x}_{0.99}$, were determined. Then, bias (eq. 28) and *RMSE* (eq. 29) of $\hat{x}_{F=0.99}$ were assessed by averaging over all 20,000 complete samples. They are displayed in Tables 1–3 (columns 4–5) as the percentage of the underlying population quantiles ($x_{0.99}$) for each estimation method separately. They serve as the reference values for the evaluation of the performance of the censored distribution estimates of $x_{0.99}$.

Then the largest sample element ($x_{n:n}$) from each sequence of length n was removed, i.e., $m = n - 1$ and the twelve V/EM procedures were applied to estimate parameters for the censored data. For each of the 20,000 censored samples of a point in the subspace [C_V , n], a given F_T value and $m = n - 1$, the distribution parameters (α, ξ) were estimated by: eqs. (18) and (19) for PWMs; eqs. (23) and (24) for MOMs; and eqs. (26b) and (27b) for MLM. Then the quantile estimate $\hat{x}_{0.99}$ was computed by eq. (14). Its bias (eq. 28) and *RMSE* (eq. 29) were assessed by averaging over all 20,000 censored samples. The results expressed as the percentage of the underlying population quantiles ($x_{0.99}$) are shown in Tables 1–3 (columns 6–14). Note that for the C and D Variants, there is a fixed value of F_T equal to $(n - 1)/n$ and 1, respectively, while for the A and B Variants the F_T value has to be found for every point in the sample sub-space $[C_V, n]$ and every estimation method according to the respective objective function. The search for F_T starts from $F_T = 1$, i.e. from the value of the D Variant, and is continued until the optimum F_T has been found.

8. EXPERIMENTAL RESULTS

The performance of the LG estimator $\hat{x}_{0.99}$ got by the A, B, C and D Variants with the three estimation methods is related to that got from the complete sample by the same estimation method. Therefore, the assessed measures of performance of the CS estimator $\hat{x}_{0.99}$ (columns 4–5 of Tables 1–3) are briefly discussed first, followed by those got from the A, B, C and D Variants (columns 6–14 of Tables 1–3). The value of the coefficient of variation C_V and the corresponding value of the L-variation coefficient τ are displayed in column 1, while the corresponding value of the population quantile $x_{0.99}$ in the next column. For the sake of brevity, the *SD* values are not presented in the tables as they can be calculated by a reader from $RMSE_\delta$ and B_δ values by eq. (30a).

8.1 Performance of estimators of $\hat{x}_{0.99}$ from complete samples

For all points in the sample subspace $[C_V, n]$, the MLM estimators of $x_{0.99}$ (Table 3) are superior to the estimators of the other two methods (Tables 1 and 2) both in respect to the bias (col. 5) and the root mean square error (col. 4). The PWMs estimators of $x_{0.99}$ are less accurate than those of the method of moments but for small C_V values (i.e., represented by $C_V = 0.3$ in Tables 1 and 2) and large hydrological samples for $C_V > 0.6$. However, MOMs produces much larger absolute bias $\left|\mathrm{B}_\delta\left(\hat{x}_{0.99}\right)\right|$ than does either PWMs or MLM, and the difference between them grows rapidly with the C_V value. Therefore, in spite of the competitiveness of $RMSE_\delta\left(\hat{x}_{0.99}\right)$ got from MOMs, both a large bias and its share in $RMSE_\delta\left(\hat{x}_{0.99}\right)$ make MOMs unsatisfactory for the LG distribution.

Moreover, contrary to MLM and PWMs, the MOMs bias remains considerably high for large samples, e.g., for $C_V = 2.0$ and $n = 10{,}000$ it equals –10.8% of the $x_{0.99}$ value and its share in the MSE value is as high as 80%. It is much lower for the $x_{0.99}$ estimate got from the censored distribution (Variants A, B and C) while the *SD* values remain of similar magnitude, which results in lower *RMSE* values. Such a large bias is characteristic of samples derived from two-parameter heavy tailed distributions with a large C_V value. It cannot be solely explained by the algebraic bound of C_V (Katsnelson and Kotz 1957). Note that the algebraic bound depends on the sample size but not on the distribution and its population value of C_V. For a set of n non-negative values x_i, not all equal, the coefficient of variation C_V cannot exceed $(n-1)^{1/2}$, attaining this value if and only if all but one of the x_i's are zero. Hence, for $n = 10{,}000$ one gets the upper bound $\hat{C}_V \leq 100$, while the largest population C_V considered equals two.

Table 1

Measures of the performance of PWMs estimator of $x_{0.99}$ for complete and censored data of the LG distribution in percentage of $x_{0.99}$

C_V (τ)	$x_{0.99}$	n	Complete data		Incomplete data ($m = n - 1$)								
					Variant A			Variant B		Variant C		Variant D	
			$RMSE_\delta$	B_δ	F_T^A	$RMSE_\delta^A$	B_δ^A	F_T^B	$RMSE_\delta^B$	$RMSE_\delta^C$	B_δ^C	$RMSE_\delta^D$	B_δ^D
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
0.3 (0.14)	21.00	20	22.56	0.70	.9874	18.67	–6.57	.9736	20.11	26.91	10.94	20.24	–14.40
		30	18.19	0.48	.9903	15.54	–4.87	.9810	16.35	20.03	7.14	17.03	–11.64
		50	14.02	0.18	.9927	12.43	–3.02	.9877	12.85	14.58	4.20	13.67	–8.80
		70	11.67	0.04	.9943	10.57	–2.35	.9907	10.83	11.85	2.93	11.67	–7.28
		10^2	9.74	0.00	.9958	8.96	–1.83	.9933	9.12	9.75	2.08	9.87	–5.85
		10^3	3.03	–0.15	.9993	2.97	–0.28	.9991	3.01	2.99	0.12	3.18	–1.36
		10^4	0.97	–0.16	.9998	0.95	0.00	.9998	1.00	0.99	0.03	1.05	–0.39
0.6 (0.24)	31.83	20	41.04	1.18	.9937	30.34	–15.93	.9738	36.23	52.75	18.32	31.42	–23.68
		30	33.50	0.92	.9940	25.50	–11.59	.9809	28.65	36.87	11.81	26.96	–19.74
		50	25.87	0.41	.9950	20.63	–7.65	.9876	22.21	26.02	7.02	22.14	–15.47
		70	21.51	0.12	.9955	17.65	–5.54	.9905	18.64	20.88	4.91	19.21	–13.14
		10^2	17.93	0.03	.9965	15.12	–4.18	.9931	15.75	17.16	3.55	16.50	–10.85
		10^3	5.51	–0.31	.9993	5.18	–0.47	.9991	5.30	5.25	0.26	5.72	–3.02
		10^4	1.79	–0.34	.9999	1.73	–0.05	.9999	1.77	1.80	0.09	1.92	–0.91
0.8 (0.28)	37.44	20	48.49	0.89	.9964	35.09	–21.40	.9744	44.88	65.46	22.53	35.79	–27.67
		30	40.39	0.79	.9955	29.70	–15.30	.9812	34.80	46.59	14.55	30.97	–23.35
		50	31.70	0.33	.9958	24.16	–10.07	.9878	26.62	31.96	8.66	25.69	–18.58
		70	26.56	0.04	.9962	20.74	–7.58	.9906	22.30	25.42	6.07	22.45	–15.95
		10^2	22.22	–0.04	.9970	17.88	–5.80	.9933	18.88	20.86	4.43	19.44	–13.32
		10^3	6.85	–0.42	.9993	6.24	–0.60	.9991	6.30	6.35	0.38	7.00	–4.01
		10^4	2.24	–0.46	.9999	2.12	–0.06	.9999	2.11	2.23	0.14	2.42	–1.25
1.0 (0.32)	41.67	20	52.84	0.34	.9974	38.26	–24.76	.9746	48.92	75.89	25.90	38.70	–30.45
		30	44.74	0.45	.9966	32.59	–18.38	.9815	39.59	53.90	16.72	33.69	–25.91
		50	35.72	0.12	.9965	26.62	–12.25	.9878	30.01	36.66	10.00	28.14	–20.83
		70	30.20	–0.14	.9967	22.93	–9.27	.9907	25.03	28.93	7.01	24.71	–18.02
		10^2	25.44	–0.18	.9973	19.86	–7.04	.9932	21.27	23.71	5.14	21.51	–15.17
		10^3	7.92	–0.53	.9994	7.03	–1.10	.9991	7.25	7.19	0.48	7.98	–4.81
		10^4	2.59	–0.57	.9999	2.44	–0.16	.9999	2.46	2.57	0.18	2.83	–1.59
1.5 (0.35)	48.05	20	57.42	–1.05	.9975	42.23	–27.85	.9752	56.60	92.46	31.36	42.59	–34.38
		30	49.72	–0.54	.9975	36.57	–22.30	.9820	44.77	63.68	20.06	37.39	–29.58
		50	40.78	–0.55	.9972	30.08	–15.27	.9881	35.02	44.29	12.18	31.54	–24.13
		70	35.09	–0.67	.9971	26.04	–11.41	.9909	29.07	34.47	8.54	27.90	–21.08
		10^2	29.98	–0.60	.9977	22.72	–9.03	.9934	24.76	28.24	6.32	24.48	–17.94
		10^3	9.69	–0.73	.9994	8.23	–1.14	.9992	8.34	8.48	0.67	9.51	–6.15
		10^4	3.19	–0.77	.9999	2.91	–0.16	.9999	3.10	3.10	0.25	3.53	–2.17
2.0 (0.38)	51.17	20	58.90	–1.96	.9986	44.07	–31.36	.9755	59.98	99.46	33.78	44.33	–36.21
		30	51.49	–1.24	.9965	38.28	–21.56	.9822	47.72	68.50	21.76	39.07	–31.32
		50	42.77	–1.05	.9974	31.70	–16.57	.9884	37.42	47.65	13.24	33.12	–25.72
		70	37.11	–1.07	.9974	27.51	–12.78	.9910	31.09	37.35	9.35	29.39	–22.56
		10^2	31.97	–0.92	.9979	24.10	–10.12	.9936	26.48	30.62	6.93	25.88	–19.31
		10^3	10.63	–0.86	.9994	8.83	–1.22	.9992	8.97	9.14	0.78	10.28	–6.85
		10^4	3.51	–0.89	.9999	3.16	0.00	.9999	3.17	3.39	0.30	3.88	–2.49

Table 2

Measures of the performance of MOMs estimator of $x_{0.99}$ for complete and censored data of the LG distribution in percentage of $x_{0.99}$

C_V (τ)	$x_{0.99}$	n	Complete data		Incomplete data ($m = n - 1$)								
			$RMSE_\delta$	B_δ	Variant A			Variant B		Variant C		Variant D	
					F_T^A	$RMSE_\delta^A$	B_δ^A	F_T^B	$RMSE_\delta^B$	$RMSE_\delta^C$	B_δ^C	$RMSE_\delta^D$	B_δ^D
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
0.3 (0.14)	21.00	20	22.15	–4.48	.9840	19.78	–8.39	.9672	21.83	27.11	8.83	23.20	–19.57
		30	19.43	–3.43	.9880	16.81	–6.22	.9773	18.08	21.30	6.19	20.04	–16.59
		50	16.27	–2.56	.9920	13.79	–4.40	.9850	14.73	16.33	4.02	16.61	–13.36
		70	14.25	–2.12	.9940	11.91	–3.66	.9894	12.42	13.53	2.94	14.57	–11.57
		10^2	12.41	–1.67	.9950	10.34	–2.26	.9924	10.72	11.46	2.26	12.63	–3.27
		10^3	4.53	–0.54	.9993	3.83	–0.42	.9992	3.93	3.89	0.27	4.75	–3.27
		10^4	1.54	–0.39	.9999	1.39	–0.07	.9999	1.43	1.48	0.11	1.73	–1.12
0.6 (0.24)	31.83	20	32.42	–12.94	.9850	30.92	–15.35	.9645	33.91	39.06	9.96	36.73	–33.16
		30	28.84	–11.00	.9890	27.05	–11.93	.9766	29.53	33.71	8.30	32.67	–29.24
		50	24.85	–9.11	.9930	23.04	–9.16	.9853	24.95	27.97	6.32	28.11	–24.77
		70	22.34	–8.03	.9950	20.33	–8.15	.9894	21.84	24.32	5.13	25.33	–22.21
		10^2	20.02	–6.91	.9960	17.99	–5.97	.9923	19.31	21.04	4.23	22.59	–19.58
		10^3	9.65	–2.80	.9994	7.65	–1.30	.9991	8.09	8.01	0.85	10.48	–8.73
		10^4	4.07	–1.73	.9999	3.13	–0.26	.9999	3.54	3.56	0.45	4.81	–3.99
0.8 (0.28)	37.44	20	36.45	–17.72	.9710	32.50	–8.87	.9574	34.07	35.87	4.84	42.15	–38.90
		30	32.39	–15.59	.9840	29.09	–10.13	.9725	30.38	31.91	4.68	37.94	–34.80
		50	27.96	–13.45	.9900	25.28	–7.52	.9831	26.63	27.76	4.23	33.12	–30.07
		70	25.23	–12.17	.9930	22.78	–6.76	.9885	23.65	24.87	3.85	30.18	–27.32
		10^2	22.70	–10.84	.9956	20.36	–6.79	.9922	21.14	22.30	3.62	27.23	–24.47
		10^3	11.72	–5.30	.9995	9.56	–2.46	.9991	10.10	10.11	1.18	13.87	–12.25
		10^4	5.82	–3.32	.9999	4.20	–0.51	.9999	4.63	4.94	0.72	7.16	–6.37
1.0 (0.32)	41.67	20	39.54	–21.23	.9560	31.70	–4.58	.9477	32.26	32.03	–1.39	45.79	–42.79
		30	35.10	–19.04	.9720	28.53	–4.23	.9653	28.77	28.73	–0.63	41.54	–38.64
		50	30.36	–16.82	.9850	24.92	–4.88	.9808	25.09	25.30	0.16	36.62	–33.81
		70	27.46	–15.46	.9890	22.76	–3.36	.9866	23.00	23.17	0.79	33.61	–30.99
		10^2	24.77	–14.04	.9920	20.64	–1.95	.9902	21.08	21.14	1.32	30.57	–28.04
		10^3	13.28	–7.77	.9993	10.33	–1.27	.9992	10.95	10.60	1.01	16.59	–15.12
		10^4	7.37	–5.09	.9999	5.05	–0.91	.9999	5.80	5.84	0.90	9.29	–8.58
1.5 (0.35)	48.05	20	44.53	–26.19	.9230	27.93	–2.33	.9165	28.14	29.56	–12.05	50.71	–48.05
		30	39.50	–24.00	.9480	24.55	–1.62	.9449	24.63	26.12	–10.33	46.47	–43.91
		50	34.27	–21.77	.9650	20.93	0.34	.9653	20.94	22.75	–8.54	41.51	–39.05
		70	31.14	–20.39	.9710	18.87	2.96	.9769	19.05	20.64	–7.03	38.47	–36.18
		10^2	28.25	–18.93	.9825	17.21	1.33	.9841	17.31	18.57	–5.72	35.36	–33.16
		10^3	16.16	–12.14	.9987	10.02	1.00	.9988	10.06	10.06	–0.84	20.88	–19.67
		10^4	10.32	–8.70	.9999	5.87	–0.10	.9999	5.91	6.18	0.57	13.04	–12.47
2.0 (0.38)	51.17	20	47.19	–28.44	.8900	25.73	0.50	.8929	25.77	30.24	–17.38	52.91	–50.41
		30	41.83	–26.28	.9240	21.90	–0.67	.9212	21.95	26.78	–15.43	48.70	–46.30
		50	36.34	–24.09	.9450	18.22	1.71	.9507	18.31	23.35	–13.35	43.76	–41.45
		70	33.10	–22.73	.9560	15.77	2.87	.9633	16.09	21.05	–11.63	40.73	–38.59
		10^2	30.12	–21.28	.9679	14.15	2.98	.9748	14.46	18.83	–10.05	37.61	–35.56
		10^3	17.87	–14.46	.9973	7.93	2.57	.9983	8.23	9.45	–3.30	23.04	–21.95
		10^4	12.11	–10.81	.9998	5.13	1.24	.9999	5.39	5.56	–0.48	15.08	–14.58

Table 3

Measures of the performance of MLM estimator of $x_{0.99}$ for complete and censored data of the LG distribution in percentage of $x_{0.99}$

C_V (τ)	$x_{0.99}$	n	Complete data		Incomplete data ($m = n - 1$)								
			$RMSE_\delta$	B_δ	Variant A			Variant B		Variant C		Variant D	
					F_T^A	$RMSE_\delta^A$	B_δ^A	F_T^B	$RMSE_\delta^B$	$RMSE_\delta^C$	B_δ^C	$RMSE_\delta^D$	B_δ^D
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
0.3 (0.14)	21.00	20	17.52	–1.62	.9780	17.25	–5.92	.9476	18.39	18.23	–0.50	18.33	–11.22
		30	14.29	–1.05	.9830	14.20	–3.80	.9657	14.76	14.71	–0.29	14.98	–8.23
		50	11.12	–0.69	.9890	11.08	–2.37	.9799	11.31	11.31	–0.21	11.56	–5.58
		70	9.36	–0.55	.9916	9.35	–1.71	.9850	9.51	9.48	–0.20	9.73	–4.33
		10^2	7.78	–0.39	.9938	7.78	–1.18	.9896	7.86	7.85	–0.14	8.05	–0.45
		10^3	2.48	–0.02	.9993	2.48	–0.12	.9991	2.48	2.48	0.00	2.50	–0.45
		10^4	0.79	–0.01	.9999	0.79	0.00	.9998	0.79	0.79	0.00	0.79	–0.06
0.6 (0.24)	31.83	20	29.29	–1.11	.9880	27.49	–11.62	.9528	30.50	30.96	0.70	28.00	–16.41
		30	23.57	–0.69	.9910	22.63	–8.14	.9687	24.22	24.50	0.58	23.08	–12.17
		50	18.19	–0.50	.9930	17.74	–4.88	.9802	18.56	18.59	0.30	18.09	–8.34
		70	15.23	–0.45	.9944	15.00	–3.56	.9861	15.44	15.48	0.13	15.31	–6.51
		10^2	12.63	–0.32	.9956	12.50	–2.46	.9903	12.74	12.77	0.09	12.74	–4.92
		10^3	4.01	–0.01	.9994	4.00	–0.23	.9992	4.00	4.01	0.04	4.02	–0.69
		10^4	1.28	–0.01	.9999	1.28	0.00	.9998	1.28	1.28	0.00	1.28	–0.10
0.8 (0.28)	37.44	20	34.70	–0.55	.9930	31.54	–14.80	.9547	35.84	36.87	1.63	31.82	–18.24
		30	27.68	–0.30	.9930	26.05	–9.86	.9694	28.38	28.89	1.19	26.34	–13.58
		50	21.25	–0.28	.9950	20.47	–6.37	.9815	21.52	21.76	0.66	20.73	–9.36
		70	17.75	–0.31	.9955	17.33	–4.51	.9864	17.97	18.06	0.37	17.58	–7.32
		10^2	14.69	–0.23	.9964	14.45	–3.16	.9904	14.82	14.87	0.25	14.66	–5.54
		10^3	4.65	0.01	.9995	4.63	–0.33	.9990	4.65	4.65	0.06	4.66	–0.79
		10^4	1.48	–0.01	.9999	1.48	0.00	.9998	1.48	1.48	0.00	1.48	–0.11
1.0 (0.32)	41.67	20	38.66	–0.03	.9950	34.27	–16.60	.9552	39.84	41.24	2.42	34.42	–19.40
		30	30.63	0.04	.9950	28.37	–11.45	.9705	31.20	32.06	1.71	28.56	–14.48
		50	23.40	–0.08	.9960	22.34	–7.33	.9811	23.80	24.01	0.95	22.53	–10.01
		70	19.51	–0.18	.9962	18.92	–5.20	.9869	19.69	19.87	0.56	19.13	–7.85
		10^2	16.12	–0.14	.9969	15.79	–3.66	.9906	16.24	16.33	0.39	15.97	–5.95
		10^3	5.09	0.02	.9995	5.07	–0.35	.9990	5.09	5.09	0.07	5.09	–0.85
		10^4	1.62	0.00	.9999	1.62	0.00	1.0000	1.62	1.62	0.00	1.62	–0.12
1.5 (0.35)	48.05	20	44.56	0.88	.9990	37.99	–20.10	.9572	45.32	47.78	3.71	38.01	–20.86
		30	34.91	0.64	.9970	31.55	–13.48	.9713	35.38	36.68	2.55	31.62	–15.64
		50	26.49	0.27	.9970	24.91	–8.53	.9824	26.67	27.23	1.44	25.02	–10.85
		70	22.00	0.05	.9970	21.13	–6.13	.9878	22.05	22.45	0.89	21.27	–8.54
		10^2	18.15	0.01	.9980	17.65	–4.75	.9900	18.38	18.39	0.61	17.78	–6.49
		10^3	5.70	0.04	.9996	5.68	–0.47	.9992	5.69	5.71	0.10	5.70	–0.93
		10^4	1.81	0.00	.9999	1.81	0.00	1.0000	1.82	1.81	0.00	1.82	–0.13
2.0 (0.38)	51.17	20	47.47	1.37	1.0000	39.67	–21.49	.9583	47.71	51.02	4.40	39.67	–21.49
		30	36.97	0.97	.9980	33.00	–14.55	.9727	37.05	38.92	3.00	33.03	–16.14
		50	27.95	0.45	.9980	26.09	–9.51	.9839	27.80	28.76	1.69	26.17	–11.22
		70	23.18	0.18	.9976	22.15	–6.76	.9875	23.28	23.67	1.06	22.26	–8.84
		10^2	19.09	0.10	.9978	18.50	–4.74	.9912	19.14	19.36	0.72	18.62	–6.73
		10^3	5.99	0.05	.9996	5.96	–0.49	.9992	5.98	5.99	0.11	5.98	–0.97
		10^4	1.90	0.00	.9999	1.90	0.00	1.0000	1.90	1.90	0.00	1.91	–0.14

8.2 Performance of estimators of $\hat{x}_{0.99}$ from censored samples

Variant A

The objective was to find for each point in the sample subspace $[C_V, n]$ and for each of the three estimation methods (Tables 1–3), the F_T value (column 6) minimizing the *RMSE* (eq. 32) and to compare the performance of this estimator $\hat{x}_{0.99}$ with that got from the complete sample. In general, the results of the three methods are consistent. For each point in the sample subspace $[C_V, n]$ and for each estimation method a gain in accuracy in relation to the complete sample is observed, i.e., $R^{\mathrm{A}}\left(\hat{x}_{0.99}\right)<1$ (eq. 31). A gain is the largest for PWMs and the smallest for MLM. Taking, as an example, $n = 70$ and $C_V = 1$ as the value rarely exceeded by the sample $\hat{C}_V$ in FFA, one can read from Tables 1–3 that $R^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ amounts to 0.759, 0.829 and 0.970, for PWMs, MOMs and MLM, respectively. Note that MLM estimates of complete data (Table 3, col. 4) are superior to two other methods' estimates of incomplete data (Tables 1 and 2, col. 7) for any sample size.

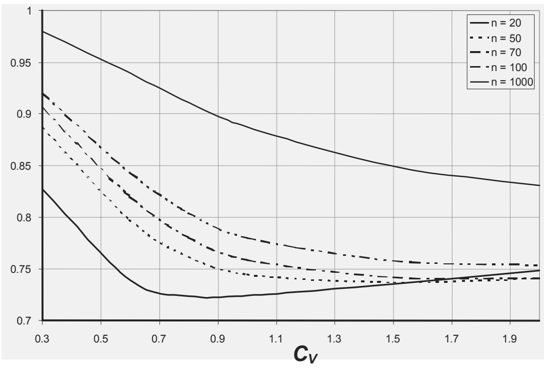

Fig. 1. $R^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ *versus* C_V for selected sample sizes n of PWMs method.

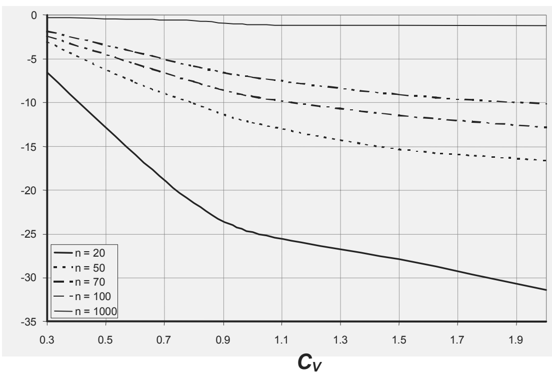

Fig. 2. $B_{\delta}^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ *versus* C_V for selected sample sizes n of PWMs method.

For each estimation method, the standard deviation $SD_\delta^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ is much smaller than that got for the complete samples $SD_\delta\left(\hat{x}_{0.99}\right)$. $R^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ (see eq. 31) decreases with C_V for MLM and for PWMs up to $C_V = 1$ only (Fig. 1) tending to one for infinitely large samples. For these two methods (Tables 1 and 3) the bias $\left|B_\delta^{\mathrm{A}}\left(\hat{x}_{0.99}\right)\right|$ (col. 8) is much greater than that of a complete sample $\left|B_\delta\left(\hat{x}_{0.99}\right)\right|$ (col. 5) and is growing with C_V as shown in Fig. 2 for PWMs. For MOMs, a quite different situation is observed both in respect to $R^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ and $\left|B_\delta^{\mathrm{A}}\left(\hat{x}_{0.99}\right)\right|$. In general a decrease of $R^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ with both C_V and n is observed**.** For $C_V \le 0.6$, $B_\delta\left(\hat{x}_{0.99}\right)$ (col. 5) and $B_\delta^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ (col. 8) do not differ much. Then $\left|B_\delta^{\mathrm{A}}\left(\hat{x}_{0.99}\right)\right|$ sharply decreases with growing C_V, becoming much smaller than $\left|B_\delta\left(\hat{x}_{0.99}\right)\right|$ and it tends to zero with growing n. Therefore, a good performance of MOMs in Variant A**,** observed for large C_V values and expressed by the low values of $R^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$, is mainly due to the reduction of bias of a complete sample's estimator $x_{0.99}$ (Table 2, col. 5).

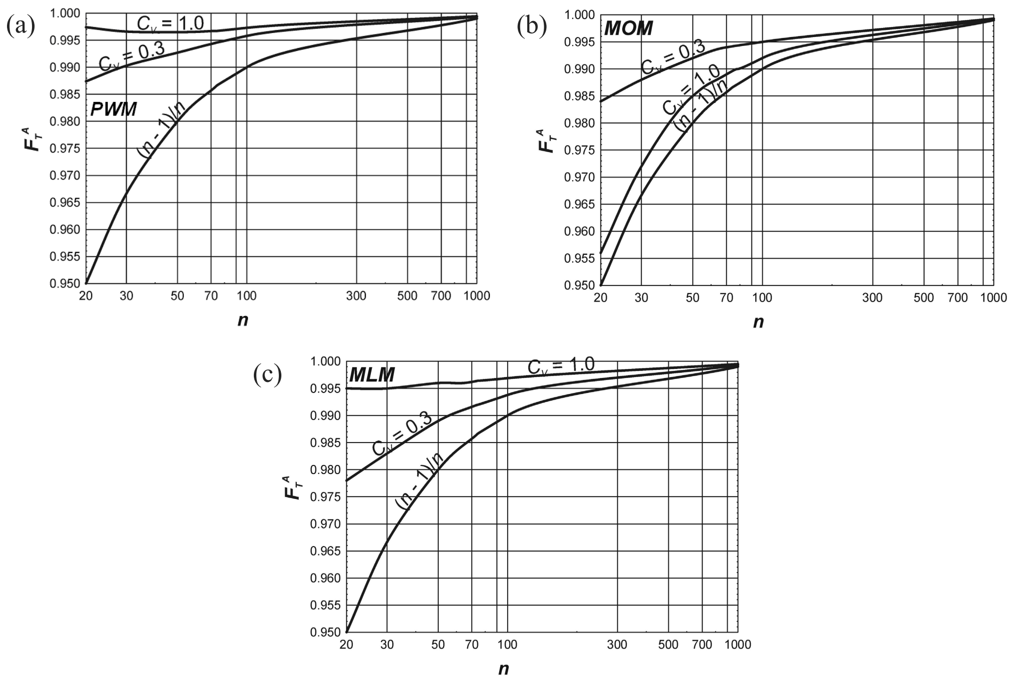

Fig. 3. The $F_T^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ values *versus* n for various C_V values and three estimation methods: (a) PWMs, (b) MOMs, (c) MLM.

For PWMs and MLM, the F_T^{A} values do not change much with C_V. Such a low sensitivity of F_T^{A} to C_V (and to F of the upper quantile – see Section 8.3) is convenient for practical application, when the population C_V value is not known and an interest may concern upper tail estimation but a single large quantile. For MOMs, $F_T^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ slightly varies with C_V up to $C_V = 0.6$, then its decrease with C_V is observed. Comparing the $F_T^{\mathrm{A}}(x_{0.99})$ values in col. 6 with F_T^{C} values (see eq. 1), it is seen that they are greater than $(n-1)/n$ for all points in the sample subspace $[C_V, n]$ and all three estimation methods (Fig. 3), but for MOMs with $C_V \geq 1.5$.

Variant B

For each point in the sample subspace $[C_V, n]$ and for both MOMs and PWMs but not for MLM, $R^{\mathrm{B}}\left(\hat{x}_{0.99}\right) < 1$, pointing to a gain in the efficiency of $\hat{x}_{0.99}$ in relation to the respective complete sample estimator got by the same method. Although MLM gives $R^{\mathrm{B}}\left(\hat{x}_{0.99}\right) > 1$, its unbiased estimator of $x_{0.99}$ is more efficient than that of the other two methods (Tables 1 and 2, col. 10). As a rule, the $R^{\mathrm{B}}\left(\hat{x}_{0.99}\right)$ values are larger than $R^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ for any of the estimation methods. This is a price for getting unbiased

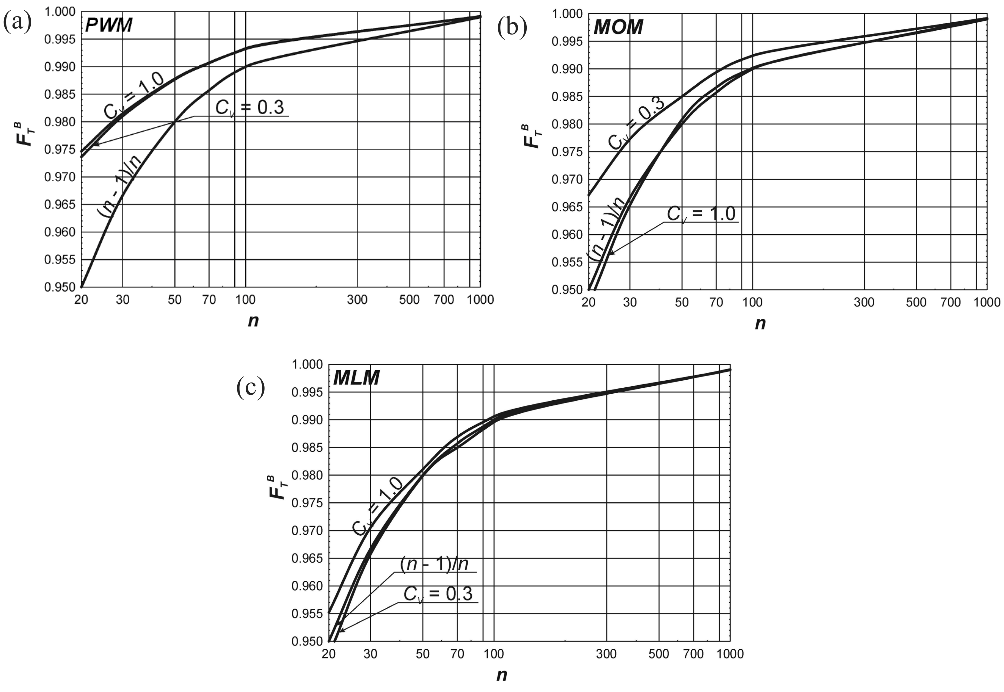

Fig. 4. The $F_T^{\mathrm{B}}\left(\hat{x}_{0.99}\right)$ values *versus* n for various C_V values and three estimation methods: (a) PWMs, (b) MOMs, (c) MLM.

$\hat{x}_{0.99}$. Similarly, the standard deviation of $\hat{x}_{0.99}$ [$SD_{\delta}^{\mathrm{B}}(\hat{x}_{0.99})$] is greater than $SD_{\delta}^{\mathrm{A}}(\hat{x}_{0.99})$ for all three estimation methods, being in general smaller than that of the complete samples for MOMs and PWMs and a bit greater for MLM. Comparing $R^{\mathrm{B}}(\hat{x}_{0.99})$ with $R^{\mathrm{A}}(\hat{x}_{0.99})$ and $F_T^{\mathrm{B}}(\hat{x}_{0.99})$ with $F_T^{\mathrm{A}}(\hat{x}_{0.99})$, one can see that $RMSE_{\delta}(\hat{x}_{0.99})$ is sensitive to the F_T value. Note that $F_T^{\mathrm{B}}(\hat{x}_{0.99}) < F_T^{\mathrm{A}}(\hat{x}_{0.99})$ for all three estimation methods. Moreover note that the ML-probability threshold $F_T^{\mathrm{B}}(\hat{x}_{0.99})$ is only a little greater than that of eq. (1) (Fig. 4c).

Variant C

In the Variant C, the plotting position formula $F_T = m/n$ with $m = n - 1$ is used. The best and only acceptable results are obtained for MLM. The $RMSE_{\delta}^{\mathrm{C}}(\hat{x}_{0.99})$ values (Table 3, col. 11) are here much lower than those of two other methods and only a little greater than that of complete sample (Table 3, col. 4). Moreover, $\left|B_{\delta}^{\mathrm{C}}(\hat{x}_{0.99})\right|$ of MLM (col. 12) being the smallest of the three methods is of similar size as one from the complete sample (col. 5), i.e., as $\left|B_{\delta}(\hat{x}_{0.99})\right|$, and for $C_V \le 0.6$, it is even smaller than $\left|B_{\delta}(\hat{x}_{0.99})\right|$. Because of a large bias and large variance of $\hat{x}_{0.99}$, setting $F_T = (n-1)/n$ is improper for PWMs and MOMs.

Variant D

Removal of the largest sample element $x_{n:n}$ and putting $F_T = 1$ means that the resulting sample is considered as a complete sample of a given distribution. It would be substantiated if $x_{n:n}$ were an outlier. However, one may erroneously remove it from a sample treating the censored $(n-1)$ element sample as the complete sample of the same distribution. It is important for practice to learn consequences of the wrong decision in respect to both the *RMSE* and the bias. As shown by Strupczewski *et al*. (2007) for the log-Gumbel, log-logistic, Pareto and log-normal thirty-element samples and three C_V values, such an error may be beneficial for the accuracy of large quantile estimates measured by *RMSE* but it causes their underestimation. Note that the assumption $F_T = 1$ leads to a contradiction, as then the censoring threshold T is infinite, i.e., $T(F_T = 1) = \infty$, while the removed element $x_{n:n} < \infty$.

The accuracy of $\hat{x}_{.99}$ got by PWMs and MLM exceeds the accuracy of Variants B and C, but only a little lower than one of Variant A. The $R^{\mathrm{D}}(\hat{x}_{0.99})$ values (col. 13/col. 4 in Tables 1 and 3) are less than one at least for hydrological sample sizes and for all C_V values, but for MLM with $C_V = 0.3$ it is slightly greater than one. For a given n and, $C_V \ge 0.6$, $R^{\mathrm{D}}(\hat{x}_{0.99})$ is quite stable, i.e. it does not change much with C_V. Application of MOMs to the incomplete sample does not improve the accuracy of the $\hat{x}_{0.99}$ estimator, as $R^{\mathrm{D}}(\hat{x}_{0.99})$ (Table 2) is greater than one for all points in the sample subspace

$[C_V, n]$. This is so only for $C_V \leq 0.6$ and small samples ($n \leq 30$), whereas for MOMs in the Variant D it shows superiority to Variant C.

As one can expect, the deletion of $x_{n:n}$ will cause underestimation of $\hat{x}_{0.99}$ – unwelcome property for design of hydraulic structures. For every estimation method, $\left|B_\delta^{\mathrm{D}}\left(\hat{x}_{0.99}\right)\right|$ is greater than for any other variants including the complete sample. However, a large $\left|B_\delta^{\mathrm{D}}\left(\hat{x}_{0.99}\right)\right|$ value is compensated for by the smallest $SD_\delta^{\mathrm{D}}\left(\hat{x}_{0.99}\right)$ of all variants including $SD_\delta\left(\hat{x}_{0.99}\right)$. The experimental results of Variant D provide an answer to the question of what to do with the uncertain largest element of the log-Gumbel distributed sample. If it is wrongly identified as an error corrupted value and removed, one would get the performance of $\hat{x}_{0.99}$ as in the Variant D. In the opposite case, the performance of $\hat{x}_{0.99}$ will correspond to that of a complete ($n - 1$) element sample. Referring to comparative assessment of robustness of PWMs and MOMs $x_{0.99}$ estimators to largest sample element, the differences of absolute value of biases (cols. 5 and 14 of Tables 1 and 2) point to MOMs as to a slightly more robust method.

8.3 Performance of the estimator $\hat{x}_{0.80}$

So far the performance of the estimator $\hat{x}_{0.99}$ has been analyzed. One can expect that the largest sample element would have a lesser impact on the "censored distribution" estimate of a smaller quantile than $\hat{x}_{0.99}$. The accuracy of "censored distribution" quantile $\hat{x}_{0.80}$ in relation to the complete sample is investigated here. It is also interesting to learn the extent to which the values of F_T^{A} and F_T^{B} depend on the quantile of interest, i.e., to compare $F_T^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ with $F_T^{\mathrm{A}}\left(\hat{x}_{0.80}\right)$ and $F_T^{\mathrm{B}}\left(\hat{x}_{0.99}\right)$ with $F_T^{\mathrm{B}}\left(\hat{x}_{0.80}\right)$. For the sake of brevity the results are shown for three C_V values only, namely 0.3, 0.6 and 1.0, and PWMs (Table 4).

Applying Variants A and B for the $x_{0.80}$ estimation, one can see (Table 4, cols. 8 and 12) that the "censored distribution" estimates are still superior to those of the complete sample, i.e., $R^{\mathrm{A}}\left(\hat{x}_{0.80}\right) < 1$ and $R^{\mathrm{B}}\left(\hat{x}_{0.80}\right) < 1$ for all C_V and n values considered. As expected, their accuracy is lower than that of the $\hat{x}_{0.99}$ quantile, i.e., $R^{\mathrm{A}}\left(\hat{x}_{0.80}\right) > R^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ and $R^{\mathrm{B}}\left(\hat{x}_{0.80}\right) > R^{\mathrm{B}}\left(\hat{x}_{0.99}\right)$. For the Variant C of the "censored distribution," one can see a similar pattern of $R^{\mathrm{C}}\left(\hat{x}_{0.80}\right)$ and $R^{\mathrm{C}}\left(\hat{x}_{0.99}\right)$ in respect to C_V and n, and that $R^{\mathrm{C}}\left(\hat{x}_{0.80}\right) > R^{\mathrm{C}}\left(\hat{x}_{0.99}\right)$. The $R^{\mathrm{C}}\left(\hat{x}_{0.80}\right)$ values fall below one for large samples only. The $R^{\mathrm{D}}\left(\hat{x}_{0.80}\right)$ and $R^{\mathrm{D}}\left(\hat{x}_{0.99}\right)$ values differ considerably. While $R^{\mathrm{D}}\left(\hat{x}_{0.99}\right)$ is less than one for small and medium size samples, $R^{\mathrm{D}}\left(\hat{x}_{0.80}\right)$ is always greater than one. Therefore, Variant D applied for PWMs procedure results in decrease of *RMSE* for large quantiles only. $F_T^{\mathrm{A}}\left(\hat{x}_F\right)$ and $F_T^{\mathrm{B}}\left(\hat{x}_F\right)$ values are fairly robust for the probability F of estimated upper quantile, as $F_T^{\mathrm{A}}\left(\hat{x}_{0.80}\right)$ and $F_T^{\mathrm{B}}\left(\hat{x}_{0.80}\right)$ are only slightly lower than $F_T^{\mathrm{A}}\left(\hat{x}_{0.99}\right)$ and $F_T^{\mathrm{B}}\left(\hat{x}_{0.99}\right)$, respectively. All of them are greater than $F_T^{\;C} = (n-1)/n$. Note that both $F_T^{\mathrm{A}}\left(\hat{x}_{0.80}\right)$ and $F_T^{\mathrm{B}}\left(\hat{x}_{0.80}\right)$ can be regarded as independent on the C_V.

Table 4

Measures of the performance of PWMs estimator of $x_{0.80}$ for complete and censored data of the LG distribution in percentage of $x_{0.80}$

C_V (τ)	$x_{0.80}$	n	Complete data		Incomplete data ($m = n - 1$)								
					Variant A			Variant B		Variant C		Variant D	
			$RMSE_\delta$	B_δ	F_T^A	$RMSE_\delta^A$	B_δ^A	F_T^B	$RMSE_\delta^B$	$RMSE_\delta^C$	B_δ^C	$RMSE_\delta^D$	B_δ^D
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
0.3 (0.14)	11.55	20	8.34	–0.29	.9839	7.73	–1.48	.9778	7.87	11.31	6.67	9.00	–6.06
		30	6.92	–0.21	.9881	6.44	–1.07	.9839	6.53	8.43	4.34	7.42	–4.76
		50	5.42	–0.18	.9918	5.11	–0.67	.9895	5.16	6.05	2.54	5.81	–3.49
		70	4.58	–0.16	.9937	4.34	–0.52	.9921	4.37	4.91	1.78	4.91	–2.83
		10^2	3.84	–0.12	.9953	3.67	–0.37	.9942	3.69	4.01	1.25	4.10	–2.23
		10^3	1.22	–0.08	.9992	1.21	–0.03	.9992	1.21	1.22	0.14	1.28	–0.50
		10^4	0.39	–0.08	.9998	0.38	–0.01	.9998	0.39	0.39	0.02	0.41	–0.15
0.6 (0.24)	12.12	20	13.38	–1.02	.9837	12.61	–3.35	.9737	13.09	17.24	7.91	14.39	–10.21
		30	11.28	–0.73	.9879	10.52	–2.46	.9813	10.81	13.08	5.15	12.00	–8.18
		50	8.99	–0.56	.9917	8.38	–1.63	.9877	8.54	9.60	3.01	9.51	–6.16
		70	7.66	–0.47	.9935	7.13	–1.21	.9907	7.24	7.87	2.11	8.10	–5.09
		10^2	6.47	–0.35	.9952	6.05	–0.91	.9934	6.11	6.51	1.50	6.81	–4.09
		10^3	2.09	–0.17	.9992	2.02	–0.12	.9991	2.04	2.05	0.21	2.22	–1.05
		10^4	0.69	–0.15	.9998	0.66	0.01	.9998	0.67	0.67	0.04	0.73	–0.33
1.0 (0.32)	12.24	20	16.55	–2.03	.9844	16.01	–5.02	.9722	16.80	21.77	8.92	18.11	–13.31
		30	14.12	–1.50	.9881	13.41	–3.60	.9802	13.93	16.65	5.82	15.22	–10.82
		50	11.43	–1.12	.9919	10.71	–2.45	.9872	10.99	12.29	3.43	12.17	–8.29
		70	9.84	–0.92	.9935	9.14	–1.77	.9904	9.32	10.10	2.40	10.43	–6.94
		10^2	8.39	–0.70	.9952	7.77	–1.34	.9931	7.89	8.38	1.74	8.83	–5.67
		10^3	2.83	–0.28	.9992	2.64	–0.19	.9991	2.67	2.70	0.27	2.98	–1.62
		10^4	0.94	–0.23	.9999	0.89	–0.03	.9998	0.90	0.90	0.07	1.03	–0.54

8.4 Comparative assessment of estimation methods

ML estimates of large quantile are superior to the estimates of two other methods both for the complete samples of log-Gumbel distribution and all variants applied for LG truncated distribution (Table 5). A loss in efficiency due to censoring is observed (Variant B in Table 3). However allowing estimates to be biased, a reduction of *RMSE* can be obtained (Variants A and D in Table 3). The *RMSE* value slightly changes with the probability threshold F_T while its two terms, i.e., B^2 and SD^2, are increasing and decreasing function of F_T, respectively.

As concerns the two other methods, an unbiased estimator (Variant B) is more efficient than the one from a complete sample got by the same method. The asterisk in

Table 5 denotes the variants giving lower value of *RMSE* and |*Bias*| than those of complete samples. Analyzing the bias only (Table 5), the PWMs estimators are superior to the MOMs estimators for the Variants A and C, but still worse than the MLM estimators. Note that PWMs estimators $x_{0.99}$ of complete samples have a little less bias than the ML estimators and a larger *RMSE*.

Table 5

Ranking of the estimation methods for $\hat{x}_{0.99}$ of LG samples with $C_V < 1$

Criterion	Estimation method	CS	Variant				Σ
			A	B	C	D	
Minimum of *RMSE*	PWMs	2	2*	3*	3	2*	12
	MOMs	3	3*	2*	2	3	13
	MLM	1	1*	1	1	1*	5
Minimum of |*Bias*|	PWMs	1	2	UB	3	2	8
	MOMs	3	3	UB	2	3	11
	MLM	2	1	UB	1*	1	5

Note: Numbers 1, 2, 3 denote the ranks while 1 is the highest rank; the asterisk indicates the lower value of the performance measure than that from complete samples (CS) by the same method; UB – unbiased estimator of $\hat{x}_{0.99}$; Σ is the total of ranks.

Since the accuracy of $\hat{x}_{0.99}$ is highly sensitive to the threshold non-exceedance probability F_T, the key point in the application of any of the three methods is setting the F_T value. The estimate $\hat{F}_T = (n-1)/n$ of the threshold non-exceedance probability is not acceptable for PWMs and MOMs and hydrological sample sizes, giving too small a value. From Tables 1–3 one can see that $R(\hat{x}_{.99}) < 1$ comes for any $F_T(\hat{x}_{0.99})$ value from at least the intervals $(F_T^{\mathrm{B}}, 1)$, $(F_T^{\mathrm{B}}, F_T^{\mathrm{A}})$, and $(F_T^{\mathrm{A}}, 1)$, for PWMs, MOMs and MLM, respectively. It is an important statement for application, as in practice, while dealing with a single sample the population C_V is not available

9. CONCLUSIONS

Employing the simulation experiments for the log-Gumbel (LG) distributed data, the effect of the removal of the largest element of a sample on the accuracy of large quantile estimates has been assessed. Three estimation methods are employed for the LG and censored LG distributions, i.e., Probability Weighted Moments (PWMs), conventional Moments (MOMs) and Maximum Likelihood (ML), and four variants for selecting the threshold non-exceedance probability F_T of the censored distribution. Experimental results are analyzed in respect to two cases of practical interest: (i) when the largest element is erroneously deleted from a sample (Variant D) and an incomplete sample is considered as a complete LG sample (i.e., $F_T = 1$); (ii) when it is intentionally removed (Variants A, B, C and D) as error corrupted value. The following conclusions are drawn from the study:

- As it has been expected, censoring results in the loss of efficiency of ML-estimates. However, allowing for bias of estimates, reduction of *MSE* is obtained (Variants A and D for $C_V \geq 0.6$)
- ML-estimators of large quantiles are superior to the estimators of the two other methods for both complete and censored data.
- For PWMs method and MOMs, censoring results in a gain in relative efficiency of upper quantile estimate (Variant B) and in considerable reduction of *RMSE* (Variant A and Variant D for PWMs only).
- For each estimation method, *RMSE* of a large quantile is highly sensitive to the F_T value.
- For PWMs method, low sensitivity of F_T^{A} and F_T^{B} to both the coefficient of variation and the cumulative probability F of upper quantile x_F has been observed – a convenient property for application.
- As shown for PWMs method, the relative accuracy of $\hat{x}_F$ got from censored distribution increases with the increase in F.
- The formula $F_T = m/n$, where $m = n - 1$, is not acceptable for both PWMs method and MOMs and hydrological sample sizes, because it yields too small a value and worst results of all the four variants in respect to both bias and MSE.
- Large bias is observed when MOMs is applied to complete data, which cannot be solely explained by the algebraic bound of C_V.
- Taking into account the fragmentary results obtained for $F_T^{\mathrm{D}} = 1$ (Strupczewski *et al.* 2005, 2007), similar results can be expected for other two-parameter heavy tail distributions, including log-normal which is on the border between PDFs having moments and heavy-tailed PDFs.

Therefore, using an appropriate procedure the removal of the largest element in a sample originated from a given two-parameter heavy tailed distributions like the log-Gumbel, may lead to increase in relative accuracy (often at the cost of bias) if the large-quantile estimators. The competitiveness of PWMs estimators of upper quantiles to those of MLM stated in the literature (e.g., Hosking and Wallis 1997) has not been confirmed by the presented results either for complete or censored samples of the log-Gumbel distribution. However, in reality the true distribution is never known, while the interest is focused on upper quantiles estimation. As shown by Strupczewski *et al.* (2002a, b) *L*-moments estimates of upper quantiles are much more robust to distributional choice than those of ML-method. Therefore, finding the largest element of annual peak flow series as uncertain value and aiming to get estimator less sensitive to the model error, the “A”-type PWMs with type II censoring is advocated for large quantiles estimation.

Acknowledgments. This work was partly supported by the Polish Ministry of Science and Informatics under the Grant 2 P04D 057 29 entitled “Enhancement of statistical methods and techniques of flood events modelling”.

References

Cunnane, C., 1981, *Unbiased plotting positions – A review*, J. Hydrol. **38**, 205-222.

Cunnane, C., 1989, *Statistical Distributions for Flood Frequency Analysis*. World Meteorological Organization Operational Hydrology, Report 33, WMO – No. 718, Geneva, Switzerland.

Greenwood, J.A., J.M. Landwehr, N.C. Matalas and J.R.Wallis, 1979, *Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form*, Water Resour. Res. **15**, 1049-1054.

Hosking, J.R.M., 1995, *The use of L-moments in the analysis of censored data*. **In:** N. Balakrishnan (ed.), “Recent Advances in Life-Testing and Reliability”, CRC Press, Boca Raton, FL, 545-564.

Hosking, J.R.M., and J.R.Wallis, 1987, *Parameter and quantile estimation for the Generalized Pareto distribution*, Technometrics **29**, 339-349.

Hosking, J.R.M., and J.R.Wallis, 1997, *Regional Frequency Analysis*, Cambridge Univ. Press, Cambridge, 224 pp.

Hosking J.R.M., J.R. Wallis and E.F. Wood, 1985, *Estimation of the generalized extreme-value distribution by the method of probability-weighted moments*, Technometrics **27**, 251-261.

Jawitz, J.W., 2004, *Moments of truncated continuous univariate distributions*, Adv. Water Resour. **27**, 269-281.

Kaczmarek, Z., 1977, *Statistical Methods in Hydrology and Meteorology*, Publ. U.S. Dept. of Commerce, National Techn. Inf. Service, Springfield, VI.

Katsnelson, J., and S. Kotz, 1957, *On the upper limits of some measures of variability*, Archiv. f. Meteor. Geophys. u. Bioklimat. (B), 8, 103-107.

Katz, R.W., M.B. Parlange and P. Naveau, 2002, *Statistics of extremes in hydrology*, Adv. Water Resour. **25**, 1287-1304.

Kendall, M.G., and A. Stuart, 1973, *The Advanced Theory of Statistics. Vol. 2 Inference and Relationships*, Charles Griffin, London, 543 pp.

Phien, H.N., and T.S.E. Fang, 1989, *Maximum likelihood estimation of the parameters and quantiles of the General Extreme-Value distribution from censored samples*, J. Hydrol. **105**, 139-155.

Rao, B.R., 1958, *On the relative efficiencies of ban estimates based on double truncated and censored sample*, Proc. Nat. Inst. Sci. India A, **24**, 376.

Rao, A.R., and K.H. Hamed, 2000, *Flood Frequency Analysis*, CRS Press LLC, Boca Raton, FL, 350 pp.

Rowiński, P.M., W.G. Strupczewski and V.P. Singh, 2002, *A note on the applicability of log-Gumbel and log-logistic probability distributions in hydrological analyses*: *I. Known pdf*, Hydrol. Sc. J. **47**, 1, 107-122.

Strupczewski, W.G., V.P. Singh and S. Węglarczyk, 2002a, *Asymptotic bias of estimation methods caused by the assumption of false probability distribution*, J. Hydrol. **258**, 1-4, 122-148.

Strupczewski, W.G., S. Węglarczyk and V.P. Singh, 2002b, *Model error in flood frequency estimation* Acta Geophys. Pol. **50**, 2, 279-319.

Strupczewski, W.G., K. Kochanek, S. Węglarczyk and V.P. Singh, 2005, *On robustness of large quantile estimates of log-Gumbel and log-logistic distributions to largest element of observation series: Monte Carlo vs. first order approximation*, SERRA **19**, 280-291, DOI 10.1007/s00477- 05-0232-x.

Strupczewski, W.G., V.P. Singh, S. Węglarczyk, K. Kochanek and H.T Mitosek, 2006, *Complementary aspects of linear flood routing modelling and flood frequency analysis*, Hydrol. Process. **20**, 3535-3554.

Strupczewski, W.G., K. Kochanek, S. Węglarczyk and V.P. Singh, 2007, *On robustness of large quantile estimates to largest elements of the observation series*, Hydrol. Process. **20**, 1328-1344, DOI:10.1002/hyp.6342.

Swamy, P.S., 1962, *On the joint efficiency of the estimates of the parameters of normal populations based on singly and double truncated samples*, J. Amer. Statist. Ass. **57,** 46.

Vogel, R.M., and N.M. Fennessey, 1993, *L-moment diagrams should replace product moment diagrams*, Water Resour. Res. **29**, 1745-1752.

Wang, Q.J., 1990, *Estimation of the GEV distribution from censored samples by the method of partial probability weighted moments*, J. Hydrol. **120**, 103-114.

Wang, Q.J., 1996, *Using partial probability weighted moments to fit the extreme value distributions to censored samples*, Water Resour. Res. **32,** 6, 1767-1771.

Received 8 February 2007
Accepted 4 June 2007