

On accuracy of upper quantiles estimation

I. Markiewicz, W. G. Strupczewski, and K. Kochanek

Institute of Geophysics Polish Academy of Sciences, Warszawa, Poland

Received: 1 July 2010 – Published in Hydrol. Earth Syst. Sci. Discuss.: 21 July 2010

Revised: 29 September 2010 – Accepted: 5 October 2010 – Published: 3 November 2010

Abstract. Flood frequency analysis (FFA) entails the estimation of the upper tail of a probability density function (PDF) of annual peak flows obtained from either the annual maximum series or partial duration series. In hydrological practice, the properties of various methods of upper quantiles estimation are identified with the case of known population distribution function. In reality, the assumed hypothetical model differs from the true one and one cannot assess the magnitude of error caused by model misspecification in respect to any estimated statistics. The opinion about the accuracy of the methods of upper quantiles estimation formed from the case of known population distribution function is upheld. The above-mentioned issue is the subject of the paper. The accuracy of large quantile assessments obtained from the four estimation methods is compared to two-parameter log-normal and log-Gumbel distributions and their three-parameter counterparts, i.e., three-parameter log-normal and GEV distributions. The cases of true and false hypothetical models are considered. The accuracy of flood quantile estimates depends on the sample size, the distribution type (both true and hypothetical), and strongly depends on the estimation method. In particular, the maximum likelihood method loses its advantageous properties in case of model misspecification.

with varying degrees of flood risk, estimating the risk of exploitation of floodplains, as well as for the valuation of the contributions of many branches of the insurance market. FFA provides support for the governing bodies of water resources in decision-making processes and plays a very important role in reducing the flood risk.

The flood frequency analysis boils down to the estimation of the upper tail, i.e., the upper quantiles of the probability density function of the annual (or partial duration) maximum flows, and the distribution function assumed is the statistical hypothesis. The problem of flood frequency modelling refers to the choice of the probability distribution describing the annual peak flows along with the method of estimation parameters and, thus, quantiles of this distribution. This issue is called the distribution and estimation (D/E) procedure. The accuracy of quantile estimate is measured by the mean square error (MSE) and the bias (B). In a classical hydrological approach, the properties of the estimation methods are analysed under the assumption that the hypothetical distribution adopted is true. In the literature, there are several papers concerning an analysis of the accuracy of the estimates of large quantiles for the selected probability distribution (e.g., Landwehr et al., 1980; Kuczera, 1982; Hoshi et al., 1984). The properties of the estimation method observed for some distribution are often automatically generalized to other distributions. In the literature, three estimation methods have been usually compared, including the method of conventional moments (MOM), the method of linear moments (LMM) and the maximum likelihood method (MLM). In this paper, another method is proposed for the comparative analysis; it is the method built on the mean deviation (MDM). Due to the analytical intractability of the mean deviation in statistics, this method was not yet widely applied in the FFA. However, using the simulation techniques can cope with this inconvenience. The application of the MDM to the estimation of the flood quantiles has been proposed in Markiewicz et al. (2006) and Markiewicz and Strupczewski (2009).

1 Introduction

Flood frequency analysis (FFA) provides information about the probable size of flood flows. The estimates of the quantiles of maximum flows obtained in this way have many practical applications. This information is required for designing hydraulic structures, determining the limits of flood zones



Correspondence to: I. Markiewicz
(iwonamar@igf.edu.pl)

Table 1. Dispersion and skewness measures.

| Estimation method | | Dispersion measure | Skewness measure | |
|-------------------|--------------------|--|-------------------|--|
| MOM | Standard deviation | $\sigma = \mu_2^{1/2} = \left[\int_{-\infty}^{+\infty} (x - \mu)^2 dF(x) \right]^{1/2}$ | Third moment | $\mu_3 = \int_{-\infty}^{+\infty} (x - \mu)^3 dF(x)$ |
| LMM | Second L -moment | $\lambda_2 = \int_{-\infty}^{+\infty} 2(x - \mu) F(x) dF(x)$ | Third L -moment | $\lambda_3 = \mu + \int_{-\infty}^{+\infty} 6x \left[F^2(x) - F(x) \right] dF(x)$ |
| MDM | Mean deviation | $\delta_\mu = \int_{-\infty}^{+\infty} x - \mu dF(x)$ | Skewness measure | $\delta_S = \mu - x_{0.5}$ |

As the objective function for the selection of the probability density function to the data should be the best fit of the distribution to empirical data, primarily in the range of the upper quantiles, making allowance for low quality of largest sample data. Moreover, no simple statistical model can reproduce the dataset in its entire range of variability. This would require the use of too many parameters that cannot be estimated reliably and efficiently from a data series which is usually of relatively small size. The probability of the correct identification of density function on the basis of short hydrological samples is very low, even in the ideal case, when a set of alternative distributions contains the true density function (e.g., Mitosek et al., 2006). Therefore, the traditional approach based on the knowledge of theoretical distribution is not acceptable. In papers by Strupczewski et al. (2002a,b) and Weglarczyk et al. (2002), the asymptotic bias of quantile in the case of assuming the wrong distribution has been derived for various estimation methods and for selected pairs of probability functions. If the hypothetical distribution is a true one, then for a given estimation method, the bias of quantile estimate results from a finite random sample on the basis of which we assess the value of a quantile, but when the hypothetical distribution differs from the true one, the total bias of quantile estimator also includes the error resulting from the model.

The aim of the study is to show that the theoretical properties of various estimation methods vary significantly when the choice of a hypothetical distribution is incorrect, which is very likely in the realities of hydrology. The paper is organized as follows. After providing some introduction to the topic, the four estimation methods and the probability distributions analysed in the paper are presented in Sects. 2 and 3, respectively. The next section provides studies on the accuracy of upper quantile estimates for these two- and three-parameter distributions under the assumption of true hypothetical distribution. A similar discussion, for the case of false hypothetical distribution, is presented in Sect. 5. The paper is concluded in the final section.

2 Estimation methods

Several systems of summary statistics describing the properties of a random sample have been developed. Based on different principles they provide, in particular, the measures of location, dispersion, skewness and kurtosis, which consecutively serve for identifying and fitting PDFs. It is convenient to use dimensionless versions of the summary statistic sets in the form of summary statistic ratios. They measure the shape of a distribution independently of its scale of measurement. Among the systems of summary statistics, the most popular are the system of conventional moments and that of linear moments (L -moments). The L -moments create an attractive system because their estimators, in contrast to the classical moments estimators, are not biased and the sampling L -moment ratios have very small biases for moderate and large samples (e.g., Hosking and Wallis, 1997). For both the system of conventional and that of linear moments, the measure of location is expressed by the mean ($\mu \equiv \lambda_1$), and the measures of dispersion and skewness are presented in Table 1.

For the estimation of statistical characteristics, the method of moments (MOM) (e.g., Kendall and Stuart, 1969), and the method of linear moments (LMM) (e.g., Hosking and Wallis, 1997), have been alternatively used. The MDM is an innovative method based on applying the mean deviation δ_μ about the mean value (μ) as a measure of dispersion (see Table 1), with the mean as a measure of location and δ_S as a measure of skewness (Markiewicz et al., 2006; Markiewicz and Strupczewski, 2009). The complement to the estimation methods based on distribution characteristics is the MLM (e.g., Kendall and Stuart, 1973), which is based on the main probability mass. The MLM is sometimes regarded as the most appropriate method because it allows us to obtain the asymptotically most efficient estimators. However, the MLM involves relatively large computational difficulties and that, the maximum likelihood, estimators do not always exist.

Table 2. Probability density functions of log-normal and GEV distributions.

| Distribution | Probability density function (PDF) |
|---|---|
| Log-normal 3 (LN3) $\varepsilon = 0$: log-normal 2 (LN2) | $f(x) = \frac{1}{(x-\varepsilon)b\sqrt{2\pi}} \exp\left[-\frac{(\ln(x-\varepsilon)-m)^2}{2b^2}\right]$ m -scale, $b > 0$ -shape; $\varepsilon < x < \infty$ |
| Generalized extreme values (GEV) $\varepsilon = 0$: log-Gumbel (LG) | $f(x) = \frac{1}{\alpha} \left[-\frac{\kappa}{\alpha}(x-\varepsilon)\right]^{1/\kappa-1} \exp\left\{-\left[-\frac{\kappa}{\alpha}(x-\varepsilon)\right]^{1/\kappa}\right\}$ $\alpha > 0$ -scale, $\kappa < 0$ -shape; $\varepsilon < x < \infty$ |

3 Probability distributions

The true probability distribution, which reflects the time series of extreme flows for a given gauging station, is not known. The study on a distribution form which would describe the observed data series is the subject of many papers, such as Jenkinson (1969) or NERC (1975). The hydrological report of the World Meteorological Organization from 1989 (Cunnane, 1989) shows that the most commonly used and recommended were Gumbel and log-normal distributions. Nowadays, the researchers of hydrological extreme events recommend the use of the heavy-tailed distributions for modelling the annual maximum flows (e.g., FEH, 1999; Rao and Hamed, 2000; Katz et al., 2002). However, as yet, the certificate of a heavy tail of hydrological variables is not sufficiently convincing (e.g., Rowinski et al., 2002; Weglarczyk et al., 2002). The heavy-tailed distributions have conventional moments only in a certain range of shape parameter values and the range decreases with growing moment order. Since the hydrological samples of peak flows are usually of a relatively small size, in order to estimate many parameters reliably and efficiently, both two- and three-parameter distributions are used in FFA, where in the three-parameter distributions the third parameter (ε) serves as lower bound (e.g., Rao and Hamed, 2000). In the paper, to assess the accuracy of the estimates of high quantiles, two two-parameter distributions have been selected, i.e., log-normal 2 (LN2) and log-Gumbel (LG) and their three-parameter counterparts, LN3 and GEV. Density functions of distributions are shown in Table 2. Both two- and three-parameter log-normal distributions represent the classical (albeit borderline) type of distribution, while the LG and GEV are heavy-tailed.

4 True hypothetical distribution

Since the true probability distribution of an observed peak flow series is not known, it would seem that the choice of a hypothetical distribution is the key point to the accurate estimation of high quantiles. However, as discussed in this section, the case where the assumed distribution is consistent with the real one, shows that the ranking of the methods with

respect to the accuracy of large quantile estimate strongly depends on the type of the true distribution and its shape.

The issue is analysed in the example of the quantile $x_{=0.99}$, otherwise known as the quantile 1%. This is likely the most commonly estimated design value for the dimensioning of hydrological structures and it defines the flow value which is exceeded, on average, once every 100 years.

4.1 Simulation experiment

Simulation experiments are performed for two-parameter distributions, LN2 and LG, the variation coefficient C_V ($C_V = \sigma/\mu$) varying from 0.2 to 1.0, with any mean $\mu > 0$. The N -element samples are considered for $N = 20(10)100$. In each case, 20 000 random samples are generated. The value of $x_{F=0.99}$ is calculated using four estimation methods under the right assumption that the population is log-normal and log-Gumbel distributed, respectively. The accuracy of the quantile $x_{F=0.99}$ estimates is expressed by the relative root mean square error (δ RMSE) and the relative bias (δ B):

$$\delta \text{ RMSE}(\hat{x}_{0.99}) = \frac{\sqrt{E(\hat{x}_{0.99} - x_{0.99})^2}}{x_{0.99}}, \quad (1)$$

$$\delta \text{ B}(\hat{x}_{0.99}) = \frac{E(\hat{x}_{0.99} - x_{0.99})}{x_{0.99}}$$

The results of the experiment are presented in Tables 3 and 4 for LN2 and LG distributions, respectively. For the sake of brevity, the selected sample sizes are shown in the tables, i.e., N equals 20, 60 and 100. In all tables, the best values of the relative RMSE and B on each row are bolded. In the asymptotic case, i.e., for $N \rightarrow \infty$, $\delta \text{ RMSE}(\hat{x}_{0.99})$ and $\delta \text{ B}(\hat{x}_{0.99})$ converge to zero. The quantile value in the first column is the true value.

For three-parameter distributions, LN3 and GEV, the mean equals zero, the standard deviation equals one and various values of skewness coefficient C_S ($C_S = \mu_3/\mu_2^{3/2}$) are assumed for the Monte Carlo experiment. The results are shown in Tables 5 and 6. The range of C_S value considered here is conditioned by the existence of skewness coefficient

Table 3. Relative accuracy [%] of $\hat{x}_{0.99}$ for a sample from LN2, assuming LN2 model.

| $T = \text{LN2}, H = \text{LN2}$ | | MOM | | LMM | | MDM | | MLM | |
|--------------------------------------|-----|---------------|------------|---------------|--------------|---------------|---------------|---------------|------------|
| $\mu > 0$ | N | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B |
| $C_V = 0.2$ $x_{0.99} = 1.554\mu$ | 20 | 8.690 | -1.574 | 8.875 | 0.191 | 8.960 | -0.850 | 8.634 | -1.416 |
| | 60 | 5.089 | -0.531 | 5.049 | 0.061 | 5.172 | -0.276 | 4.985 | -0.451 |
| | 100 | 3.963 | -0.327 | 3.901 | 0.023 | 4.014 | -0.179 | 4.070 | -0.322 |
| $C_V = 0.6$ $x_{0.99} = 3.115\mu$ | 20 | 25.80 | -5.500 | 26.51 | 1.189 | 26.15 | -1.130 | 23.98 | -2.087 |
| | 60 | 16.40 | -2.087 | 14.97 | 0.387 | 15.06 | -0.355 | 13.85 | -0.674 |
| | 100 | 13.21 | -1.271 | 11.56 | 0.209 | 11.69 | -0.244 | 10.70 | -0.450 |
| $C_V = 1.0$ $x_{0.99} = 4.905\mu$ | 20 | 38.64 | -11.41 | 43.30 | 2.097 | 42.53 | -0.661 | 37.17 | -0.999 |
| | 60 | 26.38 | -5.319 | 24.33 | 0.723 | 24.24 | -0.131 | 21.01 | -0.306 |
| | 100 | 22.14 | -3.524 | 18.82 | 0.443 | 18.79 | -0.087 | 16.16 | -0.245 |

Table 4. Relative accuracy [%] of $\hat{x}_{0.99}$ for a sample from LG, assuming LG model.

| $T = \text{LG}, H = \text{LG}$ | | MOM | | LMM | | MDM | | MLM | |
|--------------------------------------|-----|---------------|------------|---------------|--------------|---------------|---------------|---------------|--------------|
| $\mu > 0$ | N | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B |
| $C_V = 0.2$ $x_{0.99} = 1.711\mu$ | 20 | 15.86 | -3.409 | 15.30 | 0.387 | 15.35 | -0.612 | 12.67 | -1.403 |
| | 60 | 10.33 | -1.371 | 8.693 | 0.131 | 8.785 | -0.186 | 7.158 | -0.468 |
| | 100 | 8.475 | -0.848 | 6.733 | 0.070 | 6.816 | -0.125 | 5.545 | -0.292 |
| $C_V = 0.6$ $x_{0.99} = 3.183\mu$ | 20 | 32.74 | -14.19 | 40.78 | 1.196 | 41.37 | -0.071 | 30.04 | -0.817 |
| | 60 | 23.52 | -8.713 | 23.69 | 0.489 | 24.05 | 0.182 | 16.62 | -0.314 |
| | 100 | 20.40 | -6.804 | 18.81 | 0.378 | 19.05 | 0.201 | 12.72 | -0.220 |
| $C_V = 1.0$ $x_{0.99} = 4.167\mu$ | 20 | 40.38 | -22.27 | 52.30 | 0.363 | 52.47 | -0.851 | 39.70 | 0.408 |
| | 60 | 28.37 | -16.29 | 32.89 | 0.272 | 33.24 | 0.112 | 21.34 | 0.072 |
| | 100 | 24.66 | -13.40 | 27.20 | 0.367 | 27.35 | 0.320 | 16.24 | 0.000 |

for GEV distribution, which takes values greater than 1.1396 (e.g., Markiewicz et al., 2006, p. 394). Moreover, the maximum likelihood estimation of GEV distribution is not always satisfactory and for some samples it appears that the likelihood function does not have a local maximum (Hosking et al., 1985). In our simulations of the GEV distribution, this non-regularity of the likelihood function causes occasional non-convergence of the modified Powell hybrid algorithm (More et al., 1980; IMSL, 1997) that is used to maximize the log-likelihood. The last column in Table 6 shows the reliability of MLM for the GEV distribution.

4.2 Accuracy of upper quantile estimates for two-parameter distributions

For both distributions, LN2 and LG, for any value of variation coefficient, the method of moments gives the greatest bias, and the higher the C_V value the greater the bias. For small samples of 20 elements from LN2 distribution, the relative bias of quantile $x_{0.99}$ estimated by MOM increases from -1.57% for $C_V = 0.2$ to -11.41% for $C_V = 1.0$

(Table 3), while for LG distribution, these values are, respectively, -3.41% and -22.27% (Table 4). For small values of C_V ($C_V = 0.2$) the difference between $\delta B(\hat{x}_{0.99})$ from the MOM and from the second method in terms of high bias, i.e., the MLM, is not large, but the distance increases with increasing C_V value. For the two distributions, the output of the method of maximum likelihood converges to those of the MDM and LMM, which are the best among the four estimation methods studied; in most cases they give a relative bias lower than 1% in absolute value. A clear negative MOM detachment from other methods is observed for bias of LG. It remains large even for statistically large sample, particularly for a large C_V population value. It is also worth noting that the MDM produces a competitive bias to LMM.

The relative root mean square error of quantile 0.99 estimate is the smallest for MLM both for LN2 and LG distribution except for LN2 with $C_V = 0.2$ and $N = 100$ where it is the largest one. Among the methods built on summary statistics, i.e., MOM, LMM, MDM, the method MOM produces the smallest $\delta \text{RMSE}(\hat{x}_{0.99})$ for small samples ($N = 20$) of LN2 distribution, regardless of the value of C_V and for

Table 5. Relative accuracy [%] of $\hat{x}_{0.99}$ for a sample from LN3, assuming LN3 model.

| $T = \text{LN3}, H = \text{LN3}$ | | MOM | | LMM | | MDM | | MLM | |
|-----------------------------------|-----|---------------|------------|---------------|-------------|---------------|--------------|---------------|------------|
| $\mu = 0, \sigma = 1$ | N | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B |
| $C_S = 2.0$ $x_{0.99} = 3.519$ | 20 | 11.68 | -4.87 | 14.18 | 1.29 | 14.12 | 1.366 | 21.13 | 4.288 |
| | 60 | 7.644 | -2.46 | 7.844 | 0.252 | 8.326 | 0.132 | 7.954 | 0.937 |
| | 100 | 6.313 | -1.74 | 6.011 | 0.198 | 6.628 | 0.144 | 5.847 | 0.517 |
| $C_S = 4.0$ $x_{0.99} = 3.905$ | 20 | 13.40 | -6.05 | 16.64 | 0.653 | 14.37 | -0.54 | 27.08 | 7.338 |
| | 60 | 9.530 | -3.34 | 9.411 | 0.382 | 9.198 | 0.045 | 9.194 | 1.712 |
| | 100 | 7.851 | -2.51 | 7.126 | 0.035 | 7.058 | 0.024 | 6.604 | 0.971 |

Table 6. Relative accuracy [%] of $\hat{x}_{0.99}$ for a sample from GEV, assuming GEV model.

| $T = \text{GEV}, H = \text{GEV}$ | | MOM | | LMM | | MDM | | MLM | | Reliability |
|-----------------------------------|-----|---------------|---------------|---------------|--------------|---------------|--------------|---------------|--------------|-------------|
| $\mu = 0, \sigma = 1$ | N | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B | of MLM |
| $C_S = 2.0$ $x_{0.99} = 3.479$ | 20 | 36.23 | -12.05 | 53.69 | 22.37 | 45.23 | 17.06 | 82.61 | 14.14 | 93.78% |
| | 60 | 25.02 | -7.640 | 27.91 | 7.510 | 30.06 | 11.08 | 32.71 | 3.076 | 95.83% |
| | 100 | 20.73 | -6.009 | 21.25 | 3.857 | 23.91 | 7.212 | 23.71 | 1.725 | 96.01% |
| $C_S = 4.0$ $x_{0.99} = 3.696$ | 20 | 47.68 | -19.47 | 59.07 | 13.52 | 43.33 | 11.86 | 111.1 | 21.54 | 95.17% |
| | 60 | 34.35 | -12.12 | 34.31 | 2.070 | 29.96 | 8.114 | 38.85 | 4.491 | 98.19% |
| | 100 | 28.68 | -9.514 | 27.14 | 0.319 | 24.37 | 4.958 | 28.00 | 2.165 | 98.67% |

log-Gumbel distribution and $C_V \geq 0.6$, regardless of the sample size. While in other cases considered, the MOM is the method that yields the highest root mean square error among the four estimation methods studied.

It is worth noting that for the heavy-tailed distributions of large C_V value (i.e., with large skewness as well this time), the bias of the MOM estimator of the standard deviation, and consequently of the 1% quantile, decreases very slowly with increasing sample size. This is clearly evident in Table 4.

4.3 Accuracy of upper quantile estimates for three-parameter distributions

The strong inferiority of MOM in respect of the relative bias of $x_{0.99}$ quantile assessment, which has been observed for LN2 and LG distributions, does not occur in the case of tree parameter LN3 and GEV. For 20-element samples, the absolute value of $\delta B(\hat{x}_{0.99})$ obtained from MOM is similar to the analogical value obtained from MLM both for LN3 distribution (Table 5) and GEV (Table 6). Then, with increasing sample size, the absolute $\delta B(\hat{x}_{0.99})$ decreases significantly in the case of MLM and slightly in the case of MOM. For LN3 and the analysed range of C_S , the LMM and MDM yield significantly smaller $\delta B(\hat{x}_{0.99})$ than MOM and MLM, regardless of the sample size, while in the case of GEV this regularity is not observed.

The δ RMSE($\hat{x}_{0.99}$) of MLM is worth the special attention. Comparing the two-parameter distributions, the addition of

the location parameter to the distribution characteristics effects in degradation of MLM position in the δ RMSE ranking both for the LN3 and GEV distributions. The MLM losses its first place in all cases except the large samples ($N = 100$) of the LN3. However even then, the superiority of the MLM over the three other methods is very small. For both distributions and small samples ($N = 20$), the ML-estimates of $\hat{x}_{0.99}$ have the highest δ RMSE of all four estimation methods considered and the differences between the quantile assessments obtained from the MLM and other three methods are considerable. For example, in the case of LN3 distribution of $C_S = 2.0$ and $N = 20$, the relative root mean square error of $\hat{x}_{0.99}$ obtained from the MLM is 21.13%, while δ RMSE($\hat{x}_{0.99}$) from the MOM, LMM and MDM are only 11.68%, 14.18% and 14.12%, respectively (Table 5). For the GEV distribution, analogical values of δ RMSE($\hat{x}_{0.99}$) are 82.61%, 36.23%, 53.69%, 45.23% for methods MLM, MOM, LMM and MDM, respectively (Table 6). The first place of MOM in δ RMSE ranking is observed for LN3 and GEV with population $C_S = 2.0$, while if $C_S = 4.0$, the method based on mean deviation is the best with respect to δ RMSE of quantile 0.99 estimate.

5 False hypothetical distribution

Due to the fact that the true probability distribution of the annual maximum flow series is not detectable, the assumption

Table 7. Relative accuracy [%] of $\hat{x}_{0.99}$ for a sample from LN2, assuming LG model.

| $T = \text{LN2}, H = \text{LG}$ | | MOM | | LMM | | MDM | | MLM | |
|--|----------|---------------|---------------|---------------|------------|---------------|--------------|---------------|------------|
| $\mu > 0$ | N | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B |
| $C_V^{(\text{LN})} = 0.2$ $x_{0.99} = 1.554\mu$ | 20 | 12.83 | 8.065 | 20.67 | 16.59 | 19.30 | 14.84 | 40.44 | 33.91 |
| | 60 | 11.10 | 9.399 | 17.83 | 16.39 | 17.23 | 15.68 | 41.20 | 38.61 |
| | 100 | 10.71 | 9.662 | 17.21 | 16.34 | 16.77 | 15.82 | 41.46 | 39.83 |
| | ∞ | 10.08 | 10.08 | 16.31 | 16.31 | 16.09 | 16.09 | 42.09 | 42.09 |
| $C_V^{(\text{LN})} = 0.6$ $x_{0.99} = 3.115\mu$ | 20 | 21.21 | -3.006 | 45.02 | 29.91 | 42.10 | 26.12 | 190.5 | 142.6 |
| | 60 | 13.26 | 0.169 | 35.24 | 29.45 | 33.77 | 27.70 | 174.9 | 156.4 |
| | 100 | 10.68 | 0.924 | 32.95 | 29.33 | 31.75 | 27.97 | 171.9 | 160.2 |
| | ∞ | 2.194 | 2.194 | 29.22 | 29.22 | 28.46 | 28.46 | 167.4 | 167.4 |
| $C_V^{(\text{LN})} = 1.0$ $x_{0.99} = 4.905\mu$ | 20 | 32.26 | -21.69 | 49.25 | 23.13 | 46.96 | 19.43 | 490.3 | 312.7 |
| | 60 | 23.73 | -18.18 | 35.06 | 23.67 | 33.54 | 21.72 | 385.9 | 324.8 |
| | 100 | 21.19 | -17.19 | 31.22 | 23.76 | 29.84 | 22.12 | 366.5 | 329.0 |
| | ∞ | 15.04 | -15.04 | 23.90 | 23.90 | 22.78 | 22.78 | 337.9 | 337.9 |

Table 8. Relative accuracy [%] of $\hat{x}_{0.99}$ for a sample from LG, assuming LN2 model.

| $T = \text{LG}, H = \text{LN2}$ | | MOM | | LMM | | MDM | | MLM | |
|--|----------|---------------|---------------|---------------|------------|---------------|------------|---------------|------------|
| $\mu > 0$ | N | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B | δ RMSE | δ B |
| $C_V^{(\text{LG})} = 0.2$ $x_{0.99} = 1.711\mu$ | 20 | 18.21 | -11.75 | 16.94 | -12.76 | 17.41 | -13.31 | 17.73 | -14.55 |
| | 60 | 13.58 | -10.13 | 14.34 | -12.88 | 14.45 | -12.97 | 15.02 | -13.74 |
| | 100 | 12.20 | -9.767 | 13.80 | -12.89 | 13.82 | -12.90 | 14.44 | -13.56 |
| | ∞ | 9.157 | -9.157 | 12.97 | -12.97 | 12.87 | -12.87 | 13.30 | -13.30 |
| $C_V^{(\text{LG})} = 0.6$ $x_{0.99} = 3.183\mu$ | 20 | 45.48 | -16.71 | 40.40 | -19.84 | 43.01 | -20.19 | 35.11 | -28.46 |
| | 60 | 35.28 | -10.63 | 28.26 | -20.84 | 28.79 | -20.67 | 30.27 | -27.93 |
| | 100 | 31.16 | -8.830 | 25.76 | -21.08 | 25.99 | -20.82 | 29.25 | -27.84 |
| | ∞ | 2.147 | -2.147 | 21.55 | -21.55 | 21.22 | -21.22 | 27.70 | -27.70 |
| $C_V^{(\text{LG})} = 1.0$ $x_{0.99} = 4.167\mu$ | 20 | 61.95 | -17.71 | 63.39 | -19.47 | 56.80 | -20.20 | 42.12 | -33.77 |
| | 60 | 50.09 | -8.599 | 40.71 | -21.20 | 38.16 | -20.72 | 36.42 | -33.72 |
| | 100 | 45.57 | -5.411 | 35.97 | -21.65 | 31.77 | -21.24 | 35.34 | -33.72 |
| | ∞ | 17.70 | 17.70 | 22.65 | -22.65 | 22.15 | -22.15 | 33.70 | -33.70 |

of a false hypothetical distribution seems to be more realistic. The error of the estimate of quantile 0.99 differs significantly for particular options of true and hypothetical distribution assumed, giving an evidence of strong influence of the type distribution, both true and hypothetical, on the accuracy of the estimators of large quantiles.

5.1 Simulation experiment

The Monte Carlo experiment is carried out similarly as in the case of true hypothetical distribution; however, the hypothetical distribution is incorrectly assumed. Therefore, two options for two-parameter distributions are considered, i.e., $T = \text{LN2}, H = \text{LG}$ (Table 7) and $T = \text{LG}, H = \text{LN2}$ (Table 8),

and two options for three-parameter PDFs, i.e., $T = \text{LN3}, H = \text{GEV}$ (Table 9) and $T = \text{GEV}, H = \text{LN3}$ (Table 10). Note that the bias for the asymptotic case ($N \rightarrow \infty$) can be obtained analytically for two-parameter distributions, see Strupczewski et al. (2002a,b) and Weglarczyk et al. (2002), while for three-parameter PDFs analogical values have not been derived yet. For the option of false hypothetical distribution, if the sample converges to infinity, the bias is the total error of quantile estimate, see Tables 7 and 8.

Table 9. Relative accuracy [%] of $\hat{x}_{0.99}$ for a sample from LN3, assuming GEV model.

| $T = \text{LN3}, H = \text{GEV}$ | | MOM | | LMM | | MDM | | MLM | | Reliability of MLM |
|--|-----|-----------------------|------------|-----------------------|--------------|-----------------------|---------------|-----------------------|--------------|--------------------|
| $\mu = 1$ | N | $\delta \text{ RMSE}$ | δB | $\delta \text{ RMSE}$ | δB | $\delta \text{ RMSE}$ | δB | $\delta \text{ RMSE}$ | δB | |
| $C_S^{(\text{LN3})} = 2.0$ $x_{0.99} = 3.519$ | 20 | 36.31 | -12.45 | 53.79 | 21.57 | 45.39 | 16.15 | 45.11 | 10.13 | 88.03% |
| | 60 | 24.89 | -8.018 | 28.71 | 7.946 | 31.38 | 12.56 | 23.34 | 7.663 | 96.41% |
| | 100 | 20.61 | -6.086 | 22.40 | 5.369 | 26.10 | 10.13 | 17.30 | 6.638 | 96.94% |
| $C_S^{(\text{LN3})} = 4.0$ $x_{0.99} = 3.905$ | 20 | 51.18 | -22.86 | 57.41 | 7.624 | 40.11 | -10.80 | * | * | 77.93% |
| | 60 | 35.11 | -15.09 | 35.06 | 3.333 | 28.22 | -3.860 | 65.60 | 35.76 | 92.99% |
| | 100 | 28.67 | -12.36 | 27.29 | 3.257 | 23.75 | -0.470 | 55.85 | 35.48 | 92.18% |

* Values are unreliable due to a low percentage of successful estimation.

Table 10. Relative accuracy [%] of $\hat{x}_{0.99}$ for a sample from GEV, assuming LN3 model.

| $T = \text{GEV}, H = \text{LN3}$ | | MOM | | LMM | | MDM | | MLM | |
|--|-----|-----------------------|------------|-----------------------|---------------|-----------------------|--------------|-----------------------|------------|
| $\mu = 1$ | N | $\delta \text{ RMSE}$ | δB | $\delta \text{ RMSE}$ | δB | $\delta \text{ RMSE}$ | δB | $\delta \text{ RMSE}$ | δB |
| $C_S^{(\text{GEV})} = 2.0$ $x_{0.99} = 3.479$ | 20 | 41.29 | -17.54 | 48.99 | 1.887 | 48.17 | 1.643 | 66.80 | 6.154 |
| | 60 | 28.10 | -8.912 | 27.49 | -2.329 | 28.91 | -3.493 | 25.53 | -4.279 |
| | 100 | 23.04 | -6.317 | 21.25 | -2.673 | 22.67 | -4.055 | 19.44 | -5.439 |
| $C_S^{(\text{GEV})} = 4.0$ $x_{0.99} = 3.696$ | 20 | 51.45 | -21.70 | 61.93 | -0.290 | 55.06 | -4.047 | 75.95 | 5.476 |
| | 60 | 36.79 | -11.01 | 34.95 | -3.349 | 33.48 | -6.472 | 28.89 | -8.278 |
| | 100 | 30.70 | -7.657 | 27.03 | -3.733 | 26.30 | -6.715 | 22.69 | -10.01 |

5.2 Accuracy of upper quantile estimates for two-parameter distributions

In the case of $T = \text{LN2}, H = \text{LG}$, the method of maximum likelihood is ranked the worst among four analysed estimation methods. MLM generates the greatest both relative bias and relative root mean square error (Table 7). The MLM errors are large even for small values of the coefficient of variation and for $C_V = 0.2$, $\delta B(\hat{x}_{0.99})$ varies from 33.91% to 39.83%, depending on the size of the sample, and $\delta \text{ RMSE}(\hat{x}_{0.99})$ is 40.44% to 41.46%, reaching 42% for the asymptotic case. With increasing C_V of LN2 distribution, the errors increase significantly. For $C_V = 1.0$, $\delta B(\hat{x}_{0.99})$ equals 312.7% for 20-element samples and 329.0% for 100-element samples, while $\delta \text{ RMSE}(\hat{x}_{0.99})$ is 490.3% and 366.5%, respectively. For $N \rightarrow \infty$ both errors are 337.9%. Other methods of estimation, built on summary statistics, are much more accurate. In the considered range of C_V , MOM turns out to be the best with the smallest both δB (absolute values are compared) and $\delta \text{ RMSE}$. The second place is occupied by MDM and next is LMM.

For the option $T = \text{LG}, H = \text{LN2}$, as in the previous case, the relative bias of the $x_{0.99}$ estimator is the greatest in case for MLM, but the differences compared to other methods are not so great (Table 8). The lowest $\delta B(\hat{x}_{0.99})$ values are produced by MOM. The method of moments is the only method

for which the value of the relative bias strongly depends on the sample size. Other methods, MDM, LMM and MLM, already for $N = 20$ give values of $\delta B(\hat{x}_{0.99})$ almost consistent with the asymptotic case. The methods MDM and LMM yield almost identical bias and they are classified between the best method MOM and the worst MLM. Regarding the relative root mean square error, for any C_V value the rank of estimation methods strongly depends on sample size. In general, for small C_V values, of the size 0.2, the method MLM is the worst and MOM is the best. With the increase of C_V , $\delta \text{ RMSE}(\hat{x}_{0.99})$ increases for MOM and decreases for MLM.

It is worth noting that in the case of false hypothetical distribution, the absolute total bias does not necessarily decrease with increasing sample size. This is due to the fact that the sampling and model biases may have the opposite signs. While in the case of true hypothetical distribution, the absolute total bias does decrease with the sample size.

5.3 Accuracy of upper quantile estimates for three-parameter distributions

In the case of $T = \text{LN3}, H = \text{GEV}$ and $C_S = 2.0$, the four analysed estimation methods yield comparable values of both $\delta B(\hat{x}_{0.99})$ (in absolute value) and $\delta \text{ RMSE}(\hat{x}_{0.99})$, without clear superiority of one method over the other (Table 9). However, for $C_S = 4.0$, the maximum likelihood method is

strongly inferior to other three methods and the method based on mean deviation ranks very well. For 100-element samples, the relative bias of $\hat{x}_{0.99}$ obtained from MLM is 35.48%, while $\delta B(\hat{x}_{0.99})$ from the methods MOM, LMM and MDM are only -12.36% , 3.257% and -0.470% , respectively. The analogical values of $\delta \text{RMSE}(\hat{x}_{0.99})$ are 55.85%, 28.67%, 27.29%, and 23.75% for MLM, MOM, LMM and MDM, in turn.

For the option $T = \text{GEV}$, $H = \text{LN3}$, the relative bias of the estimate of quantile 0.99 is the largest for MOM, followed by MLM, and then by MDM and LMM (Table 10). The rank of estimation method in respect of δRMSE value strongly depends on sample size. For the considered range of C_S and $N = 20$, the sequence of methods from that which gives the smallest $\delta \text{RMSE}(\hat{x}_{0.99})$ to this which gives the highest $\delta \text{RMSE}(\hat{x}_{0.99})$ is as follows: MOM, MDM, LMM and MLM, while for $N > 60$ the order is opposite.

6 Conclusions

Since the upper quantiles are design values for the dimensioning of hydrological structures, the accuracy of their estimates is a major and extremely important issue for flood frequency analysis. The studies presented in this paper show that the accuracy of the estimates of flood quantiles depends on the sample size, type of distributions, both real and hypothetical, and strongly depends on the method of estimation. Therefore, the properties of estimation methods cannot be generalized in respect to distribution type or sample size, even if the hypothetical distribution is true. However, it is worthy to note that for two-parameter distributions, in the case of model misspecification, the MLM yields the highest bias of upper quantile estimates regardless on the sample size, while the MOM the smallest one. The correct identification of the distribution on the basis of short data series is not possible in hydrological reality. This finding essentially diminishes the practical usefulness of MLM in hydrological extremes analysis because its efficiency may not compensate for the (frequently) huge bias produced by the assumption of a false PDF in the region of high non-exceedance probability quantiles which the user is often interested in. It marks a departure of hydrological extreme value analysis from the classical, statistical theory of extremes whose core is maximum likelihood method. The person making the choice of the distribution and estimation (D/E) procedure, e.g., explorer, hydrologist, designer, should be aware of the impact of the procedure selection on the value of desirable estimate. Presented in this paper a comparative analysis of large quantile estimates obtained by various methods of estimation under the assumption of true or false, but close to the true type of distribution, can be a source of information about the properties of selected D/E procedures. The studies on the estimation methods of flood quantiles when the hypothetical model is

untrue should be continued. Despite a century of research, the problem of flood flows modelling is still open.

Acknowledgements. This work was partly financed by the Polish Ministry of Science and Higher Education under the Grant 0920/B/P01/2009/37 titled “Implementation of new parametric methods for hydrological design in time-variant environment” and COST Action ES0901 titled “European Procedures for Flood Frequency Estimation (FloodFreq)”. The authors would like to express their thankfulness to M. Sc. A. Kozłowski for his fruitful comments.

Edited by: S. Grimaldi

References

- Cunnane, C.: Statistical distributions for flood frequency analysis. Operational Hydrology Report No. 33, World Meteorological Organization, Geneva, 1989.
- FEH: Flood Estimation Handbook, Institute of Hydrology, Wallingford, Oxfordshire, UK, 1999.
- Hoshi, K., Stedinger, J. R., and Burges, S. J.: Estimation of log-normal quantiles: Monte Carlo results and first-order approximations, *J. Hydrol.*, 71, 1–30, 1984.
- Hosking, J. R. M., Wallis, J. R., and Wood, E. F.: Estimation of the generalized extreme value distribution by the method of probability weighted moments, *Technometrics*, 27(3), 251–261, 1985.
- Hosking, J. R. M. and Wallis, J. R.: *Regional Frequency Analysis. An Approach Based on L-Moment*, Cambridge University Press, Cambridge, UK, 1997.
- IMSL: IMSL Fortran Numerical Library, User’s Guide Math Library, Version 7.0, <http://www.vni.com/products/imsl/documentation/fort07/help/math/math.pdf>, Chap. 7, 1337–1340, 2010.
- Jenkinson, A. F.: Statistics of extremes, in: *Estimation of Maximum Floods*, WMO No. 233, TP126, Tech. Note No. 98, Secretariat of the World Meteorological Organization, Geneva, Switzerland, 183–228, 1969.
- Katz, R. W., Parlange, M. B., and Naveau, P.: Statistics of extremes in hydrology, *Adv. Water Resour.*, 25, 1287–1304, 2002.
- Kendall, M. G. and Stuart, A.: *The advanced theory of statistics*, Vol. 1, *Distribution theory*, Charles Griffin & Company Limited, London, 1969.
- Kendall, M. G. and Stuart, A.: *The advanced theory of statistics*, Vol. 2, *Inference and relationship*, Charles Griffin & Company Limited, London, 1973.
- Kuczera, G.: Robust flood frequency models, *Water Resour. Res.*, 18(2), 315–324, 1982.
- Landwehr, J. M., Matalas, N. C., and Wallis, J. R.: Quantile estimation with more or less floodlike distributions, *Water Resour. Res.*, 16(3), 547–555, 1980.
- Markiewicz, I., Strupczewski, W. G., Kochanek, K., and Singh, V. P.: Relations between three dispersion measures used in flood frequency analysis, *Stoch. Env. Res. Risk A*, 20, 391–405, doi:10.1007/s00477-006-0033-x, 2006.
- Markiewicz, I. and Strupczewski, W. G.: Dispersion measures for flood frequency analysis, *Phys. Chem. Earth*, 34, 670–678, doi:10.1016/j.pce.2009.04.003, 2009.

- Mitosek, H. T., Strupczewski, W. G., and Singh, V. P.: Three procedures for selection of annual flood peak distribution, *J. Hydrol.*, 323, 57–73, doi:10.1016/j.hydrol.2005.06.016, 2006.
- More, J., Garbow, B., and Hillstom, K.: User guide for MINPACK-1, Argonne National Labs Report ANL-80-74, Argonne, Illinois, 1980.
- NERC: Flood Studies Report Volume 1: Hydrological Studies, Natural Environment Research Council, London, 1975.
- Rao, A. R. and Hamed, K. H.: Flood Frequency Analysis, CRC Press, Boca Raton, Florida, USA, 2000.
- Rowinski, P. M., Strupczewski, W. G., and Singh, V. P.: A note on the applicability of log-Gumbel and log-logistic probability distributions in hydrological analyses: I. Known pdf, *Hydrolog. Sci. J.*, 47(1), 107–122, 2002.
- Strupczewski, W. G., Singh, V. P., and Weglarczyk, S.: Asymptotic bias of estimation methods caused by the assumption of false probability distribution, *J. Hydrol.*, 258, 122–148, 2002a.
- Strupczewski, W. G., Weglarczyk, S., and Singh, V. P.: Model error in flood frequency estimation, *Acta Geophysica Polonica*, 50(2), 279–319, 2002b.
- Weglarczyk, S., Strupczewski, W. G., and Singh, V. P.: A note on the applicability of log-Gumbel and log-logistic probability distributions in hydrological analyses: II. Assumed pdf, *Hydrolog. Sci. J.*, 47(1), 123–137, 2002.