Hydrol. Earth Syst. Sci. Discuss., 7, 5413–5440, 2010
www.hydrol-earth-syst-sci-discuss.net/7/5413/2010/
doi:10.5194/hessd-7-5413-2010

Hydrology and Earth System Sciences Discussions

Estimation of high return period flood quantiles using additional non-systematic information with upper bounded statistical models

B. A. Botero[1] and F. Francés[2]

[1]Facultad de Ingeniería y Arquitectura, Universidad Nacional de Colombia Sede Manizales, Colombia
[2]Instituto de Ingeniería del Agua y Medio Ambiente, Universidad Politécnica de Valencia, Spain

Received: 19 June 2010 – Accepted: 28 June 2010 – Published: 5 August 2010

Correspondence to: B. A. Botero (baboteroh@unal.edu.co)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This paper proposes the estimation of high return period quantiles using upper bounded distribution functions with Systematic and additional Non-Systematic information. The aim of the developed methodology is to reduce the estimation uncertainty of these quantiles, assuming the upper bound parameter of these distribution functions as a statistical estimator of the Probable Maximum Flood (PMF). Three upper bounded distribution functions, firstly used in Hydrology in the 90's (referred to in this work as TDF, LN4 and EV4), were applied at the Jucar River in Spain. Different methods to estimate the upper limit of these distribution functions have been merged with the Maximum Likelihood (ML) method. Results show that it is possible to obtain a statistical estimate of the PMF value and to establish its associated uncertainty. The behaviour for high return period quantiles is different for the three evaluated distributions and, for the case study, the EV4 gave better descriptive results. With enough information, the associated estimation uncertainty is considered acceptable, even for the PMF estimate. From the robustness analysis, EV4 distribution function appears to be more robust than the GEV and TCEV unbounded distribution functions in a typical Mediterranean river and Non-Systematic information availability scenario.

1 Introduction

Flood frequency analysis is one of the most common methods to estimate the design flood for hydraulic structures and for flood hazard/risk mitigation programs. In Europe, the national legislation for flood risk assessment is based on flood frequency analysis to estimate discharges associated with different return periods, from 50 to 500 years (Benito et al., 2004). In some projects the focus is on extreme floods, which have been defined according to different authors as floods with an annual probability of occurrence of about 10^{-3} to 10^{-7} (Jarret and Tomlinson, 2000), 10^{-3} or lesser (Naghettini et al., 1996) and in other cases, as floods with return periods greater than 500 years (England et al., 2003). Traditionally, extreme flood estimates have been associated with large

HESSD

7, 5413–5440, 2010

dam projects or with the location of nuclear and other high vulnerable facilities, in which the release of hazardous materials to the environment is in consideration (Stevens, 1992). For some of these projects, the design criteria commonly include the Probable Maximum Flood (PMF) estimation. The PMF is the biggest flood physically possible at a specific catchment (Smith and Ward, 1998). It has a physical meaning and it provides an upper limit of the interval within which the decision maker must operate and design. The PMF is the flood generated by the Probable Maximum Precipitation (PMP) with the worst but reasonable hydrological conditions in the studied basin. The PMP is defined by the World Meteorological Organization as a precipitation upper limit for a given region, duration and time of the year (WMO, 1986).

Related to high return period quantiles estimation, flood frequency analysis has a well known drawback, as pointed out by Klemes (1993) and more recently by Merz and Blöschl (2008): the lack of available information about large events in a relatively short data series recorded systematically at a flow gauge station (from now, Systematic information). This fact involves the extrapolation of very high return period quantiles from data records which rarely exceed a hundred of years, producing quantile estimates with a high level of uncertainty.

In the last decades, as a way to solve this problem, many authors as Stedinger and Cohn (1986), Hosking and Wallis (1986), Stedinger and Baker (1987), Jin and Stedinger (1989), Stevens (1992), Pilon and Adamowski (1993), Frances et al. (1994), Cohn et al. (1997), Jarret and Tomlinson (2000), Martins and Stedinger (2001), O'Conell et al. (2002), England et al. (2003), Naulet et al. (2005), Reis and Stedinger (2005) and Merz and Blöschl (2008), have included historic and palaeoflood information (from now, Non-Systematic information) in flood frequency analysis with very good results.

Probability distribution functions with 2, 3 or 4 parameters have been used in extreme floods frequency analysis, with the common characteristic of having a positive skewness coefficient (γ_x) and no upper bound. The use of parametric distribution functions allows the increment of the return period of the requested quantile as much

HESSD
7, 5413–5440, 2010

as it is required (obviously increasing at the same time its uncertainty). However, as the return period increases, with this kind of parametric distribution functions, the estimated quantiles increase too with no limit. Though, the question to pose at this point is: would it be possible to have a flood with such a high magnitude, as large as it could be obtained with these unbounded distribution functions, for a certain catchment with specific area and geomorphologic characteristics? The straight answer is no, this is not possible. There must be a limiting flood discharge which is the biggest physically possible flood for the specific climatic and hydrologic characteristics of the catchment, which indeed corresponds with the PMF definition. Not considering the existence of this upper limit must introduce an additional significant model error in the high return period estimated quantiles. Moreover, in our opinion, this additional error could produce in most cases the underestimation of the high return period quantiles, which is one of the most frequent causes of dam failure (ASCE, 1988).

In accordance with reality, some distribution functions incorporate an additional parameter, which is actually the upper limit to the random variable. This class of functions has been applied to the extreme frequency analysis of annual maximum daily precipitation by Elíasson (1994 and 1997), Takara and Loebis (1996) and Takara and Tosa (1999) and in frequency analysis of annual maximum flood by Takara and Tosa (1999). All these authors concluded that upper bounded distribution functions fit properly to extreme data and improve the quantile estimates.

In this paper we propose the use of upper bounded distribution functions, in order to better estimate high return period quantiles. The upper limit of these distribution functions can be fixed a priori or not. Following the classification of Merz and Blöschl (2008) for additional information, in the first case the PMF value can be considered as a causal information expansion. This was the option followed previously by Elíasson (1994 and 1997), Takara and Loebis (1996) and Takara and Tosa (1999). In the last case, the PMF can be estimated as one of the parameters of the statistical model, using in this paper additional Non-Systematic information, called temporal information expansion, in terms of Merz and Blöschl (2008) to obtain enough estimate reliability.

HESSD
7, 5413–5440, 2010

B. A. Botero and F. Francés

2 Upper bounded distribution functions

The three upper bounded distributions functions applied in the case study were chosen because they had been previously successfully applied to hydrological extremes series. Other distribution functions commonly used in Hydrology which have an upper bound are the GEV and the Generalized Pareto. The former has an upper bound when its shape parameter is bigger than 0, which occurs for $\gamma_x < 2.0$. The latter presents an upper bound when the shape parameter is also positive, but then, γ_x becomes less than the Gumbel's constant skewness coefficient, which is equal to 1.14. Our aim is to analyse rivers with high skewness coefficient (γ_x clearly bigger than 2), like those with a Mediterranean regime, which is the reason to not include in this paper the GEV and Generalized Pareto distribution functions.

Following paragraphs presents a short description concerning the three selected distributions. More behavioural and statistical details can be found in Botero (2006).

2.1 The Extreme Value with Four Parameters distribution function (EV4)

This probability distribution function was firstly proposed by Kanda (1981), who empirically derived it from the EV distribution function family. The EV4 cumulative distribution function (cdf) is given by

$$F_X(x) = \exp\left[-\left\{\frac{g-x}{v(x-a)}\right\}^k\right] k>0; \quad v>0; \quad a \le x \le g \tag{1}$$

where g and a are respectively the upper and lower bounds of the random variable, and v and k are parameters which characterize the scale and shape of the distribution.

Takara and Tosa (1999) applied this distribution to annual maximum daily precipitation and annual maximum daily discharge at Ohtsu (Japan). The authors made a sensitivity analysis fixing the upper and lower bounds to a priori values obtained empirically. They concluded in a comparison with other distribution functions, the EV4 is the most appropriated to datasets with sample skewness coefficient higher than about 2.

2.2 The Slade-type Four Parameter LogNormal distribution function (LN4)

Proposed by Slade (1936) and named in this manner by Takara and Loebis (1996), the LN4 can be obtained if a Slade-type random variable transformation is applied to a Two Parameters LogNormal distributed random variable. This transformation is given by

$$y = \ln\left(\frac{x-a}{g-x}\right) \quad a \le x \le g \tag{2}$$

Where g and a are respectively the upper and lower bounds. The resulting LN4 pdf is defined as following

$$f_X(x) = \frac{g-a}{(x-a)(g-x)\sigma_y\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left\{\frac{y-\mu_y}{\sigma_y}\right\}^2\right] \tag{3}$$

where μ_y and σ_y are the well known LogNormal cdf parameters. From the LN4 application to annual maximum daily precipitation data, Takara and Loebis (1996) concluded that if the four parameters of the distribution are estimated, the variability of quantile estimates is higher than if one or both limits are fixed previously in a known value. In addition, they suggested the use of the PMF as the upper bound when dealing with floods. In a posterior paper, Takara and Tosa (1999) conclude that the LN4 distribution function fits well to many hydrological datasets with sample skewness coefficient less than about 1.5.

2.3 The Transformed Extreme Value type distribution function (TDF):

This distribution was proposed by Elíasson (1994) as a statistical model for frequency analysis of extreme precipitation. The author suggested that bounded data fitted by an unbounded distribution as the EV1 (also called Gumbel), must deviate from the distribution at high return periods. In Elíasson (1997) is defined a Transformed Distribution Function (TDF) derived from a Base Distribution Function (BDF) selected by the author, which corresponds with the EV1. In his work, Elíasson (1997) fits the resulting TDF to standardized annual maximum daily precipitation from Iceland and Washington State (USA) with very good results fixing a prior estimate of the PMP. The final expression of the TDF cdf is given by

$$F_X(x) = \exp\left[-\exp\left(\frac{-x}{\alpha} + \frac{\alpha k^*}{(g-x)} - b\right)\right] \quad \alpha > 0; \quad k^* < 0 \text{ and } x \le g \tag{4}$$

where g is the upper bound, α is a scale parameter, b is a location parameter, and k^* is a negative constant.

3 Data classification

As it was mentioned above, data series recorded systematically at a flow gauge station located in a river section will be called Systematic information. In opposition, the Non-Systematic information is that information not recorded systematically. If there is not a gauge station, all river flow information can be considered as Non-Systematic. The sources for this information can be historical or from palaeofloods studies. The former are associated with past human registered observations and the latest, are floods identified using physical or botanical indicators irrespective of any direct human observation (Stedinger and Baker, 1987), but not necessarily, previous to human registers. In practice, Non-Systematic information is always censored type I, in such a way we have some information concerning a flood at a given time during the Non-Systematic period because this flood was bigger than a threshold level of perception X_H (Stedinger and Cohn, 1986; Frances et al., 1994), where H is the threshold return period. The value of the peak flow for the floods above X_H can be known or not. Concerning the floods below X_H, always it is not known their exact values, but at least it is known they

HESSD
7, 5413–5440, 2010

were smaller than X_H. The threshold level of perception can be, for example, the position of the cave where flood sediments are deposited (for palaeoflood information) or the minimum discharge which produces damages in a city (for historical information). It can change with time and, in some cases, there can be an upper and lower thresholds for the same flood (following the palaeoflood example if there are two caves at different positions, the lower one with sediments of a particular flood and the upper one without any trace of this flood). On the other hand, usually the Systematic data is completely known, but some times the uncertainty in the data forces to treat them also as censored. In this framework, as was presented by Naulet (2002), any piece (Systematic or Non-Systematic) of flood data can be classified in (see Fig. 1):

EX type: the flood peak value is known. It will correspond with most of the Systematic data and with some Non-Systematic floods with enough information to reconstruct the peak discharge.

LB type: in this case, it is only known that the flood was bigger than a lower bound L, which is the known X_H

UB type: for this type of data it is known that at the time of the flood, an upper bound (U) with known magnitude was not exceeded. Again, this U corresponds with the X_H.

DB type: the flood peak is unknown and the only information about this flood is that it has a double bound. This means that the flood was between an interval with known values for the upper (U) and lower (L) bounds.

When dealing with Non-Systematic information, what is commonly available is a combination of the different types of data described above. It was stressed by Frances et al. (1994), that from the statistical point of view there is no difference concerning the source of the Non-Systematic information, historical or palaeoflood, and their treatment must be completely similar. More over, with this new data classification there is not any difference also between Systematic and Non-Systematic information (as it is shown in next section), with the additional advantage that always the data can have a time assigned, which will be crucial for future non-stationary flood frequency models.

HESSD
7, 5413–5440, 2010

4 Parameter estimation methodology

In this study, the parameters set for each distribution function is estimated based on the Maximum Likelihood (ML) estimation method. This methodology has been selected on the basis of its statistical features for large samples, and also because of its ability to incorporate easily in the estimation process any type of additional data. For annual maxima analysis, depending on the type of data, each year i contributes to the likelihood function through one of these general expressions

$$\mathrm{L}_{\mathrm{SY}}(\underline{\Theta};x_i) = f_X(x_i;\underline{\Theta}) \tag{5}$$

$$\mathrm{L}_{\mathrm{LB,i}}(\underline{\Theta};\mathrm{L}_i) = 1 - F_X(\mathrm{L}_i;\underline{\Theta}) \tag{6}$$

$$\mathrm{L}_{\mathrm{UB,i}}(\underline{\Theta};\mathrm{U}_i) = F_X(\mathrm{U}_i;\underline{\Theta}) \tag{7}$$

$$\mathrm{L}_{\mathrm{DB,i}}(\underline{\Theta};\mathrm{L}_i,\mathrm{U}_i) = F_X(\mathrm{U}_i;\underline{\Theta}) - F_X(\mathrm{L}_i;\underline{\Theta}) \tag{8}$$

where the independent and identically distributed random variable X is described for all years through its probability density function $f_X(\cdot)$ or its cumulative distribution function $F_X(\cdot)$; $\underline{\Theta}$ represents the parameters set; x_i is the magnitude of the flood presented in the i-year; U_i is the upper bound which is not exceeded in the i-year; and L_i is the lower bound exceeded in the i-year.

The ML parameter set estimates are obtained by maximizing the logarithm of the likelihood function over the parameters space.

4.1 Upper limit estimation

Dealing with upper bounded distribution functions, it must be carefully undertaken the estimation of the upper limit parameter (g). The first possibility is to prefix g in a specific value previously computed (called G): i.e., the parameter g is not estimated jointly with the other parameters. This was the approach used by Takara and Tosa (1999), Elíasson (1994) and Takara and Loebis (1996). In this case, G is the PMF estimate by

HESSD

7, 5413–5440, 2010

Estimation of flood quantiles with upper bounded models

B. A. Botero and F. Francés

traditional means. The PMF can be estimated by different procedures, being the most advisable the rainfall-runoff modelling, where rainfall input is the PMP estimated by the WMO maximization and transposition procedure (WMO, 1986).

A second possibility is to estimate the whole parameters set, including g using the ML method. However, with EV4 and TDF distribution functions and when the available Non-Systematic information is a mix of EX and UB data (also called Censored Information or CE by Stedinger and Cohn, 1986), the estimated $\hat{g}$ is equal to the maximum observed value of the data set (which obviously it is not the PMF) and thus, other methods to estimate g must be introduced. This follows from the fact that the likelihood function for these two distributions and this type of information decreases monotonically for $g \to \infty$ (Kijko and Sellevoll, 1989).

Kijko (2004) proposed what he called a "Generic Equation" (GE) for the estimation of the maximum earthquake magnitude, which corresponds with the magnitude of the largest possible earthquake, conceptually equivalent to the PMF for floods. This author developed this equation based on the limit of a random variable estimator proposed by Cooke (1979). The GE is valid for any cdf and it is given by

$$g = x_{\max} + \int_{-\infty}^{g} \left[F_X\left(x; \underline{\Theta}\right)\right]^n dx \tag{9}$$

where $x_{\max}$ is the maximum observed value of the Systematic and Non-Systematic data, and n is the number of observed values (i.e., the $x_{\max}$ order). It must be notice that it is not possible to apply the Generic Equation when there are LB or DB data in the available information, because with this kind of data the $x_{\max}$ order cannot be known.

HESSD

7, 5413–5440, 2010

4.2 Proposed estimation methodologies

In this paper and in accordance with the selected method to estimate g, the whole parameters set estimation method will be referred as following

ML-C: when the whole parameters set of the distribution function is estimated by the ML method, including g as another free parameter in the maximization process:

$$\max \mathrm{L}(\underline{\Theta}) \tag{10}$$

ML-GE: it will be referred to the Maximum Likelihood-Generic Equation estimation method. This method consists on the use of the Generic Equation presented above (Eq. 9) to estimate g and the ML method for the rest of parameters:

$$\left.\begin{array}{c} \max \mathrm{L}\left(\underline{\Theta'}\right) \\ g = x_{\max} + \int\limits_{-\infty}^{g} \left[F_X\left(x;\underline{\Theta'},g\right)\right]^n dx \end{array}\right\} \tag{11}$$

where $\underline{\Theta}'$ is the parameters set excluding g. The expresion in Eq. (11) must be solved iteratively, fixing the g when maximizing the likelihood function and fixing the rest of parameters when obtaining a new g value with the Generic Equation. This procedure is repeated until the estimated g converges with the proper tolerance

ML-PG: finally, in this case, g is fixed at the value previously calculated (G) as the best approximation for the true unknown PMF, and the other parameters are estimated by ML method:

$$\left.\begin{array}{c} g = G \\ \max \mathrm{L}(\underline{\Theta'}) \end{array}\right\} \tag{12}$$

HESSD
7, 5413–5440, 2010

5 The Jucar River case study

The statistical models described above have been applied to the available data of the Jucar River at "Huerto-Mulet" flow gauge station. This river is located in a semi-arid climate region of Eastern Spain and has a long period with type CE Non-Systematic information. The point of interest is close to the river sea mouth and has a basin of 21 500 m^2, but because meteorological reasons, actually only one third of the basin is contributing to the floods in this area. The basin mean annual precipitation is 450 mm, but it must be underlined that the Jucar River presents the typical high variability (or torrentiallity) of Mediterranean rivers, with frequent observed daily precipitation events with more than 100 mm during the Fall season. These extreme events are generated by strong Convective Mesoscale Systems positioned in the Western Mediterranean Sea (Rigo and Llasat, 2007). The available Systematic record at the flow gauge station starts at 1946. The main sample statistical characteristics of the instantaneous annual maximum floods are: mean=713 $m^3 s^{-1}$, coefficient of variation=2.74 and skewness coefficient=5.26.

The Spanish Centro de Estudios Hidrográficos (Frances, 1998) quantified the peak flow of 4 Non-Systematic floods who reached the level of an ancient convent, located within the floodplain, in 1778, 1805, 1814 and 1864. The discharge required to flood the convent is 6200 $m^3 s^{-1}$, which can be considered as the threshold level of perception for this source of information. To avoid a bias estimation in the number of known floods, the used historical period is 153 years long, since 1792 (the average between 1778 and 1805) to 1945. During this period, the Non-Systematic information can be classified as CE with 3 EX data type. At 1982 the threshold was also exceeded with a flood of 12 000 $m^3 s^{-1}$ and it was almost exceeded in 1987 with a flood of 5200 $m^3 s^{-1}$. The stationarity of this long information period (more than 200 years) was tested and proved with a test of stationarity described by Lang et al. (1999 and 2004), using as random variable the cumulative number of floods over the threshold of perception.

HESSD
7, 5413–5440, 2010

The models applied to the flood frequency analysis of the Jucar River are combinations of the three bounded distribution functions presented in Sect. 2 and the three parameter estimation methods shown in Sect. 4. The lower bound for the EV4 and LN4 has been fixed to zero, hence reducing to three the number of parameters to be estimated. Frances (1998) studied this case study, using the unbounded distribution function TCEV.

In order to apply the ML-PG parameter estimation method, a previous PMF value to the Jucar River catchment must be calculated. Cifres and Abad (1992), using for the PMP the maximization and transposition procedure (WMO, 1986), computed the PMF in 25 000 $m^3 s^{-1}$ for the Tous dam, which is located upstream our point of interest. Assuming the same specific discharge (overestimating hypothesis), taking into account the catchment area increment and considering only the meteorologically active basin, the PMF can be extrapolated to 33 900 $m^3 s^{-1}$ at Huerto-Mulet flow gauge station. With a high probability, it will be an overestimation of the PMF. With any better estimation, due to the scope of this study, this value will be used for G.

Usually, to test the model performance (distribution and estimation method) from a descriptive point of view (Cunnane, 1986), the fitted cdf and the plotting positions are compared graphically. In this work, the probability plotting positions with Systematic and CE Non-Systematic information were calculated with the E-formula proposed by Hirsh and Stedinger (1987). Figure 2 shows the plotting positions for the Jucar data and the fits of the applied distribution functions by each parameter estimation method. A very interesting first conclusion about the three upper bounded distribution functions is their completely different behaviour approaching the upper limit: the EV4 do it faster than the LN4 and the TDF is the slowest. This different behaviour can be generalized for the usual parameter range of the three functions.

Concerning the Jucar sample data, Fig. 2 shows the characteristic "dog-leg" effect in torrential regime rivers. It is clear the TDF distribution can not reproduce the shape of the plotting positions. The reason for this unsuitability could be that the TDF is based on a Gumbel distribution function, which, according to Frances (1998), is not

HESSD
7, 5413–5440, 2010

appropriated for Mediterranean rivers. On the other hand, the EV4 distribution function with all the parameter estimation methods is the cdf that better reproduces the shape of the plotting positions. The triangle in Fig. 2 represents the non-exceedence probability for the threshold of perception considering the complete sample: only the EV4 (for the three estimation methods) can approach to it. In fact, the sample skewness coefficient is in the range recommended by Takara and Tosa (1999) for the EV4. This descriptive skill of the EV4 distribution function, when the "dog leg" effect is present, makes the EV4 the recommended distribution function to the Jucar River annual maximum floods.

Concerning the estimated PMF ($\hat{g}$), for the EV4/ML-C and TDF/ML-C models is equal to 13 000 $m^3 s^{-1}$, which is the maximum observed data (the 1864 flood). This poor result was expected from the likelihood function properties in these two cases, as it was explained earlier in this paper. The $\hat{g}$ estimated by models TDF/ML-GE (93 100 $m^3 s^{-1}$) and LN4/ML-C (99 300 $m^3 s^{-1}$) are approximately three times G (the PMF deterministic value), which are unreasonable values, whereas the $\hat{g}$ estimated with EV4/ML-GE is 18 100 $m^3 s^{-1}$, almost half of G value and about 40% above the maximum observed value in two centuries, which is more reasonable.

6 Uncertainty analysis for the EV4 model

Once the EV4 distribution function has been selected for the Jucar River case study, an uncertainty analysis has been made in order to establish the reliability of the quantiles and PMF, estimated with this distribution function and using the different parameter estimation methods (ML-C, ML-GE and ML-PG).

The uncertainty analysis has been made by Monte Carlo simulations with two skewness coefficient scenarios. One population has a high skewness coefficient (γ_x) of γ_x= 5.77 (scenario 1), which corresponds with the EV4's skewness coefficient calculated with the parameters estimated for the Jucar River. The other population has a γ_x=2.39 (scenario 2), which is lower than the first one, but it can still be considered large and possible in Mediterranean rivers.

HESSD
7, 5413–5440, 2010

Estimation of flood quantiles with upper bounded models

B. A. Botero and F. Francés

The length of the generated series was 450 years, with a Non-Systematic period of M=400 years with a X_H with return period equal to 50 years and a Systematic period of N=50 years, which can be typical characteristics when dealing with historical information in European cities.

The parameter estimation methods compared here are those exposed in Sect. 4, but with a variation in the ML-PG method introducing some random and systematic errors in the G prefixed value. It has been assumed that the error in G value is normal with a coefficient of variation (CV_G) of 0.3 and mean equal to 10% bigger than the PMF (i.e., a systematic additional positive error of a 10% of the theoretical PMF). The aim of this modification in the ML-PG method has been to analyze how the PMF uncertainty is propagated to the quantile estimation uncertainty. Figure 3 shows the uncertainty of $\hat{q}_T$ and $\hat{g}$, reflecting how it varies with the quantile return period (T). The uncertainty is measured with the next error index:

$$E(\%) = 100 \frac{\sqrt{\frac{1}{S}\sum_{i=1}^{S}\left(\hat{\theta}_i - \theta\right)^2}}{\theta} \tag{13}$$

where S is the number of generated samples; $\hat{\theta}_i$ is the estimated quantile or upper limit; and θ is the theoretical quantile or upper limit value of the distribution function.

From Fig. 3 left, which corresponds with scenario 1, it can be seen that the three estimation methods have an error between 15% and 25%, from 50 to 500 years quantiles. Below 200 years, the quantile errors are similar, but for quantiles larger than $\hat{q}_{500}$, the parameter estimation method with less error is the ML-GE, which gives the less sensitive error to the quantile return period. ML-PG is the method whit higher error, which results in an error of 30% for the $\hat{q}_{10\,000}$, in opposition to the ML-GE with only 18%. Obviously, the errors for the ML-PG method can be reduced if the error in the a priori G value is reduced, either, its coefficient of variation or its bias.

On the other hand, results for scenario 2 (Fig. 3 right) show that for ML-C and ML-GE methods, the errors for $\hat{q}_T$ and $\hat{g}$ are limited to about 10%. The quantiles errors with ML-PG method are strongly controlled by the G error, even for quantile return periods

HESSD
7, 5413–5440, 2010

smaller than 1000 years. In both scenarios ML-PG gives $\hat{q}_{10\,000}$ and $\hat{g}$, errors close to 30%, which corresponds with the coefficient of variation of the introduced error. It is clear that if the CV_G were zero, this method would be the best, but it deteriorates as CV_G increases.

A second Monte Carlo simulation was performed in order to analyze how is the influence of the X_H, characterized by its return period H, in the $\hat{q}_T$ and $\hat{g}$ uncertainty. The return period H of the generated series was established at 25, 50, 100 and 250 years. Figure 4 shows the quantile and $\hat{g}$ estimation errors with the ML-GE method. For both scenarios, it can be observed a minimum error, located near to the $\hat{q}_T$ with a return period equal to H. After this point, the estimation error is slightly higher, but remains in a similar order. It means that the Non-Systematic information contributes to reduce the error in the flood quantiles estimation only for those quantiles of equal or larger return period than the threshold of perception.

7 Robustness analysis

Robustness analysis has been also made by Monte Carlo simulations. Series have been generated coming from 3 different populations: 1) EV4 population, to analyse the robustness respect to itself; 2) TCEV population, in order to analyse robustness respect to an unbounded distribution with 4 parameters, which have shown good results in Mediterranean rivers; and 3) GEV population, to analyse robustness with respect to an unbounded distribution commonly used when dealing with flood frequency analysis.

When an EV4 population is assumed (Fig. 5), the quantiles estimated with EV4 and TCEV distributions give similar errors, whereas the GEV $\hat{q}_T$ error is larger than 80% for T >200 years in both scenarios. In other words, for our case study (with high skewness coefficient) the TCEV can be used even when there is a flood upper bound.

Assuming a GEV population (Fig. 6), the three distribution functions give similar $\hat{q}_T$ errors, being the maximum difference of an absolute increase of only about 20%, compared with the GEV $\hat{q}_T$ error. For low skewness coefficient (Fig. 8 right), the EV4

HESSD
7, 5413–5440, 2010
Estimation of flood quantiles with upper bounded models
B. A. Botero and F. Francés

gives almost the same errors that the GEV quantile estimations for all return periods, and the TCEV only for very large ones (T >1000 years).

Finally, for samples with a TCEV population (Fig. 7), the quantiles estimated with EV4 give similar errors to the TCEV, thought the $\hat{q}_T$ error increment in scenario 1 (Fig. 7 left) with EV4 is slightly larger than in scenario 2 (Fig. 7 right). In both scenarios, the GEV gives large error increments, increasing with T.

8 Conclusions

Dealing with PMP or PMF, a general assumption in the hydrological community is that these upper bounds can be only estimated deterministically. However, it has been shown in this paper that, using statistical analysis with upper bounded distribution functions and introducing additional Non-Systematic information, it is possible to estimate high return period quantiles and the PMF with relative low errors.

Based on the ML method, three different methodologies to estimate the parameters of the bounded distribution functions can be implemented, depending on the available flood information. If there is a good deterministic a priori estimation of the PMF, the ML-GP method is advised. Otherwise, ML-C can be used considering the upper limit as one more parameter. But in some combinations of distribution function and information (for example, the EV4 and TDF with CE additional Non-Systematic information), the upper limit estimated by ML-C method is equal to the maximum observation. In these situations, only the ML-GE method can be used.

Between the LN4, TDF and EV4 upper bounded distribution functions, the latest is the distribution function which better represents the shape of the empirical distribution function of the Jucar River, which has a high skewness coefficient and, consequently, in accordance with the results obtained by Takara and Loebis (1996) and Takara and Tosa (1999). Combining this distribution function with the three proposed parameter estimation methods, the Jucar series has been fitted and it has been possible to estimate statistically a PMF value for this river. The resulting estimate is not out the possible range of the Jucar River PMF at its sea mouth.

The uncertainty analysis shows that the EV4/ML-GE statistical model is the most adequate among those presented in this paper, for the estimation of high return period quantiles and PMF when dealing with CE type Non-Systematic information.

For the case study, the $\hat{q}_{10\,000}$ and $\hat{g}$ (PMF estimate) estimation error using the EV4/ML-GE model is approximately about 20%. Nevertheless, this model shows a slight sensitivity to the sample skewness, giving a reduction for the $\hat{q}_T$ and $\hat{g}$ estimation error as the skewness coefficient is reduced. It must be pointed out the $\hat{g}$ error with all methods is close to the $\hat{q}_{10\,000}$ error. Actually, if we acknowledge the 10 000 years return period quantile estimation error is admissible, we must admit also the $\hat{g}$ estimation error using the statistical approach given by ML-C or ML-GE.

The $\hat{q}_{10\,000}$ error obtained with the model EV4/ML-PG is approximately the error associated with the deterministic estimation of the PMF. It means that if it is available a previous value of the PMF, with its associated uncertainty, it is possible to estimate high return period quantiles with equal or less uncertainty using the EV4/ML-PG model.

For the information scenario used in the robustness analysis, it can be concluded the EV4/ML-GE model can satisfactory fit samples coming from unbounded populations (GEV) or unbounded mixed populations (TCEV), showing a higher robustness than the TCEV. On the contrary, from the robustness point of view, the GEV distribution function only can be accepted for the estimation of low return period quantiles.

Acknowledgements. This research was funded by the European SPHERE project (EVG1-CT-1999-00010), a Doctoral research fellowship of the Universidad Politécnica de Valencia and the Spanish Research Project FLOODMED (CGL2008-06474-C02-02/BTE). Also we are very grateful to the comments of A. Kijko, Director of the Benfield Natural Hazard Centre at University of Pretoria, who gave us bright ideas in the precise moment of our research work.

References

American Society of Civil Engineers (ASCE): Evaluation procedures for hydrologic safety of dams. Spillway Design Flood Selection, Surface Water Hydrology Committee, Report. 8726-26520, ASCE, New York, USA, 1988.

HESSD

7, 5413–5440, 2010

Benito, G., Lang, M., Barriendos, M., Llasat, M. C., Frances, F., Ouarda, T., Thorndycraft, V. R., Enzel, Y., Bardossy, A., Coeur, D., and Bobée, B.: Use of Systematic, Palaeoflood and Historical Data for the improvement of flood risk estimation. Review of scientific methods, Nat. Hazards, 31, 623–643, 2004.
Botero, B. A.: Estimación de Crecidas de alto período de retorno mediante funciones de distribución con límite superior e información No Sistemática, Ph.D. dissertation, Universidad Politécnica de Valencia, Spain, 223 pp., 2006.
Cifres, E. and Abad, P.: Assessing the PMF for its use in the Algar Dam Project, in: Proceedings of the International Symposium on Dams and Extreme Floods, ICOLD–CIGB, Granada, Spain, 16 September 1992, 91–100, 1992.
Cohn, T. A., Lane, W. L., and Baier, W. G.: An Algorithm for Computing Moment-Based Flood Quantile Estimates when Historical Flood Information is available, Water Resour. Res., 339, 2089–2096, 1997.
Cooke, P.: Statistical Inference for Bounds of Random Variables, Biometrika, 66, 367–374, 1979.
Cunnane, C.: Review of Statistical Models for Flood Frequency Estimation, in: Proceedings of the International Symposium on Flood Frequency and Risk Analysis, Baton Rouge, USA, 14–17 May 1986, 49–96, 1986.
Elíasson, J.: Statistical Estimates of PMP Values, Nord. Hydrol., 25, 301–312, 1994.
Elíasson, J.: A Statistical Model for Extreme Precipitation, Water Resour. Res., 333, 449–455, 1997.
England, J. F., Jarrett, R. D., and Salas, J. D.: Data-Based Comparisons of Moments Estimators Using Historical and Paleoflood Data, J. Hydrol., 278, 172–196, 2003.
Frances, F.: Using the TCEV Distribution Function with Systematic and Non – Systematic Data in a Regional Flood Frequency Analysis, Stoch. Hydrol. Hydraul., 12, 267–283, 1998.
Frances, F., Salas, J. D., and Boes, D. C.: Flood Frequency Analysis with Systematic and Historical or Paleofood Data Based on the Two Parameter General Extreme Value Models, Water Resour. Res., 306, 1653–1664, 1994.
Hirsch, R. M. and Stedinger, J. R.: Plotting Positions for Historical Floods and Their Precision, Water Resour. Res., 234, 715–727, 1987.
Hosking, J. R. M. and Wallis, J. R.: Paleofood Hydrology and Flood Frequency Analysis, Water Resour. Res., 224, 543–550, 1986.

HESSD
7, 5413–5440, 2010

Jarrett, R. D. and Tomlinson, E. M.: Regional Interdisciplinary Paleoflood Approach to Assess Extreme Flood Potential, Water Resour. Res., 3610, 2957–2984, 2000.
Jin, M. and Stedinger, J. R.: Flood Frequency Analysis with Regional and Historical Information, Water Resour. Res., 255, 925–936, 1989.
Kanda, J.: A New Value Distribution with lower and upper Limits for earthquake motions and wind speeds, Theor. Appl., 31, 351–360, 1981.
Kijko, A.: Estimation of the Maximum Earthquake Magnitude-Mmax, Pure Appl. Geophys., 161, 1–27, 2004.
Kijko, A. and Sellevoll, M. A.: Estimation of earthquake hazard parameters form incomplete data files, part I, Utilization of extreme and complete catalogues with different threshold magnitudes, B. Seismol. Soc. Am., 79, 645–654, 1989.
Klemeš, V.: Probability of extreme hydrometeorological events:A different approach, in: Extreme Hydrological Events: Precipitation, Floods and Droughts, edited by: Kundzewicz, Z. W., Rosbjerg, D., Simonovic, S. P., and Takeuchi, K., IAHS Publications, 213, 167–176, 1993.
Lang, M., Ouarda, T. B. M. J., and Bobée, B.: Towards Operational Guidelines for Over-Threshold Modelling, J. Hydrol., 225, 103–117, 1999.
Lang, M., Renard, B., Dindar, L., Lemaitre, F., and Bois, P.: Use of Statistical Test Based on Poisson Process for Detection of Changes in Peak-Over-Threshold Series, in: Hydrology: Science Practice for the 21st Century, Proceedings of the London Conference, London, UK, 12–16 July 2004, 1, 158–164, 2004.
Martins, E. S. and Stedinger, J. R.: Historical Information in Generalized Maximum Likelihood Framework with Partial Duration and Annual Maximum Series, Water Resour. Res., 3710, 2559–2567, 2001.
Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, Water Resour. Res., 44, W08432, doi:10.1029/2007WR006744, 2008.
Naghettini, M., Potter, K. W., and Illangasekare, T.: Estimating the upper tail of flood-peak frequency distributions using hydrometeorological information, Water Resour. Res., 326, 1729–1740, 1996.
Naulet, R.: Utilisation de L' information Des Crues Historiques Pour Une Meilleure Prédétermination Du Risque D' inondation, Ph.D. thesis, Université Joseph Fourier, Université du Québec Institut national de la recherche scientifique, France, 133 pp., 2002.
Naulet, R., Lang, M., Ouarda, T. B. M. J., Coeur, D., Bobee, B., Recking, A., and Moussay, D.:

HESSD
7, 5413–5440, 2010

Estimation of flood quantiles with upper bounded models

B. A. Botero and F. Francés

Flood Frequency Analysis on the Ardeche River Using French Documentary Sources from the Last Two Centuries, J. Hydrol., 313, 58–78, 2005.
O'Connell, D. R. H., Ostenaa, D. A., Levish, D. R., and Klinger, R. E.: Bayesian Flood Frequency Analysis with Paleohydrologic Bound Data, Water Resour. Res., 385, 16.1–16.14, 2002.
Pilon, P. J. and Adamowski, K.: Asymptotic Variance of Flood Quantile in Log Pearson Type III Distribution with Historical Information, J. Hydrol., 143, 481–503, 1993.
Reis, D. S. and Stedinger, J. R.: Bayesian MCMC Flood Frequency Analysis with Historical Information, J. Hydrol., 313, 97–116, 2005.
Rigo, T. and Llasat, M. C.: Analysis of mesoscale convective systems in Catalonia using meteorological radar for the period 1996–2000, Atmos. Res., 83, 458–472, 2007.
Slade, J.: An Asymmetric Probability Function, Transactions of American Society of Civil Engineers, 62, 35–104, 1936.
Smith, K. and Ward, R.: Floods. Physical Processes and Human Impacts, John Wiley and Sons, Baffins Lane, Chichester, UK, 1998.
Stedinger, J. R. and Cohn, T. A.: Flood Frequency Analysis with Historical and Paleoflood Information, Water Resour. Res., 22(8), 785–793, 1986.
Stedinger, J. R., and Baker, V. R.: Surface Water Hydrology: Historical and Paleofood Information, Reviews of Geophysics, 252, 119–124, 1987.
Stevens, E. W.: Uncertainty of Extreme Flood Estimates Incorporating Historical Data, Water Resour. Bull., 286, 1057–1068, 1992.
Takara, K. and Loebis, J.: Frequency Analysis Introducing Probable Maximum Hydrologic Events: Preliminary Studies in Japan and in Indonesia, in: Proceedings of International Symposium on Comparative Research on Hydrology and Water Resources in Southeast Asia and the Pacific, Yogyakarta, Indonesia, 18–22 November 1996, Indonesian National Committee for International Hydrology Programme, 67–76, 1996.
Takara, K. and Tosa, K.: Storm and Flood Frequency Analysis Using PMP/PMF Estimates, in: Proceedings of International Symposium on Floods and Droughts, Nanjing, China, 18–20 October 1999, 7–17, 1999.
World Meteorological Organization (WMO): Manual for Estimation of Probable Maximum Precipitation, Operational Hydrology Report no. 1, 190 pp., 1986.

HESSD
7, 5413–5440, 2010

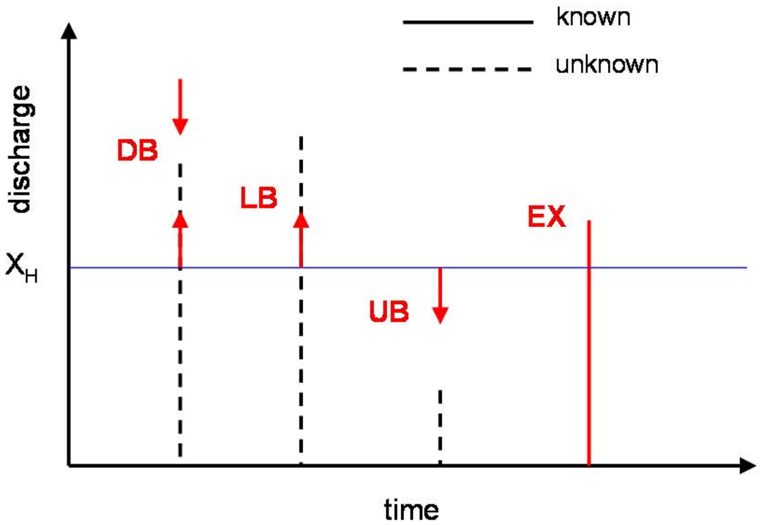

Fig. 1. Proposed flood data classification (after Naulet, 2002).

HESSD
7, 5413–5440, 2010

Estimation of flood quantiles with upper bounded models

B. A. Botero and
F. Francés

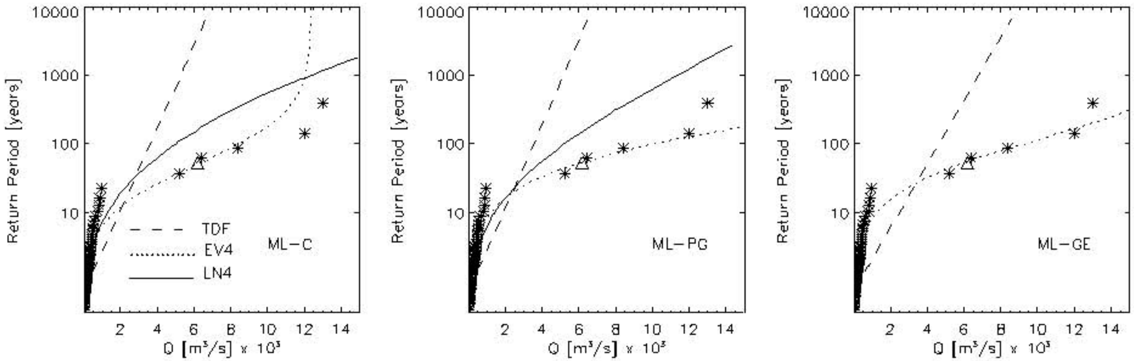

Fig. 2. Applied models to the Jucar River. Stars are the sample plotting positions given by the E-formula. Triangle represents the X_H=6200 m^3 s^{-1} plotting position. Figures correspond to each estimation method: ML-C, ML-PG and ML-GE.

HESSD
7, 5413–5440, 2010

Estimation of flood quantiles with upper bounded models

B. A. Botero and F. Francés

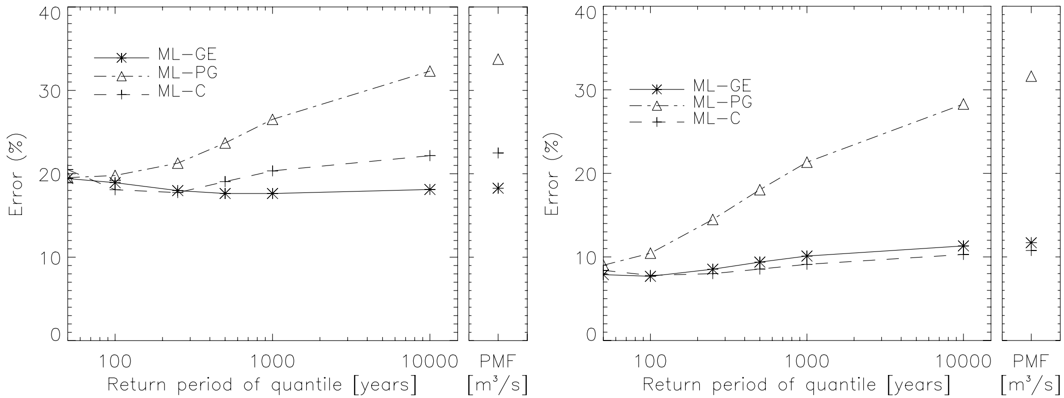

Fig. 3. Estimation error (in %) of $\hat{q}_T$ and $\hat{g}$ (PMF estimate). H= 50, M=400 and N=50 years. Left: scenario 1, right: scenario 2.

HESSD
7, 5413–5440, 2010

Estimation of flood quantiles with upper bounded models

B. A. Botero and F. Francés

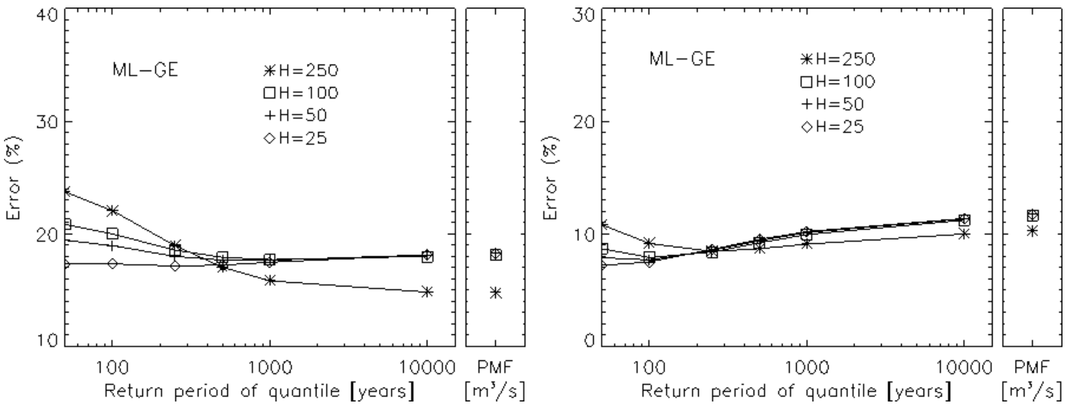

Fig. 4. Estimation error (in %) of $\hat{q}_T$ and $\hat{g}$ (PMF estimate) for different H. M=400 and N=50 years. Left: scenario 1, right: scenario 2.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

7, 5413–5440, 2010

Estimation of flood quantiles with upper bounded models

B. A. Botero and F. Francés

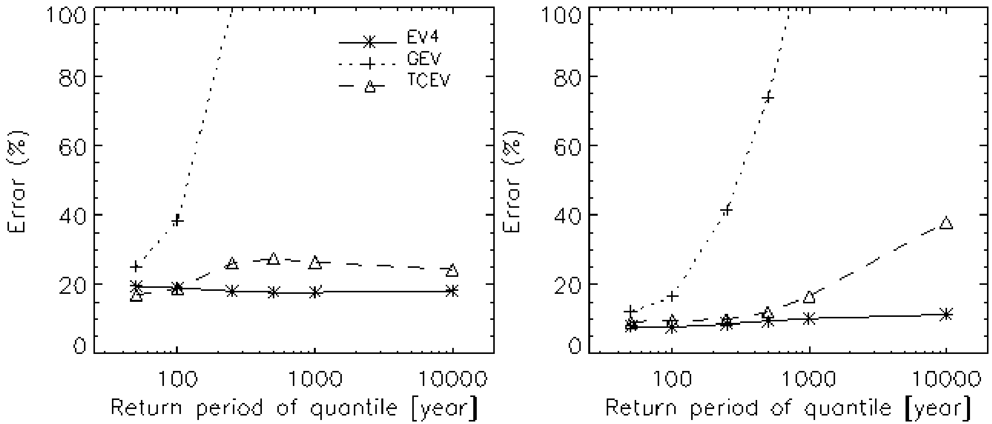

Fig. 5. Estimation error (in %) of $\hat{q}_T$ with ML-GE method for EV4, GEV and TCEV distribution functions, assuming an EV4 population. H= 50, M=400 and N=50 years. Left: scenario 1, right: scenario 2.

HESSD
7, 5413–5440, 2010

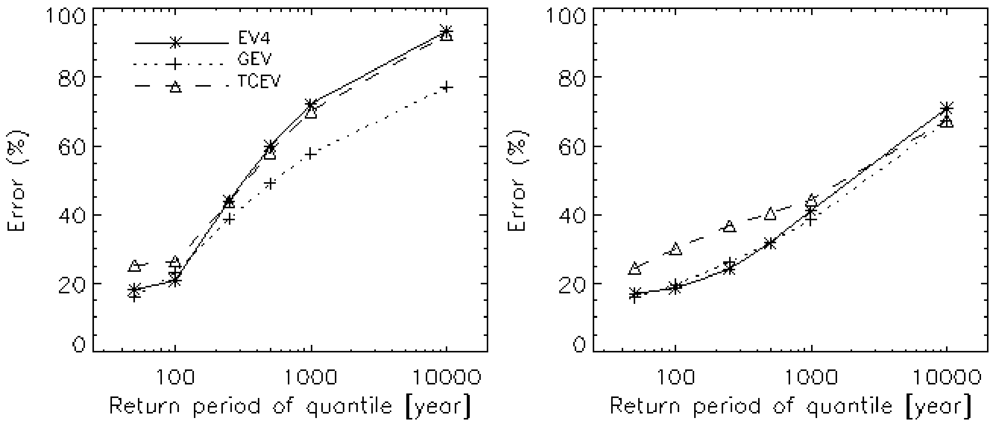

Fig. 6. Estimation error (in %) of $\hat{q}_T$ with ML-GE method for EV4, GEV and TCEV distribution functions, assuming a GEV population. H=50, M=400 and N=50 years. Left: scenario 1, right: scenario 2.

HESSD
7, 5413–5440, 2010

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |
Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

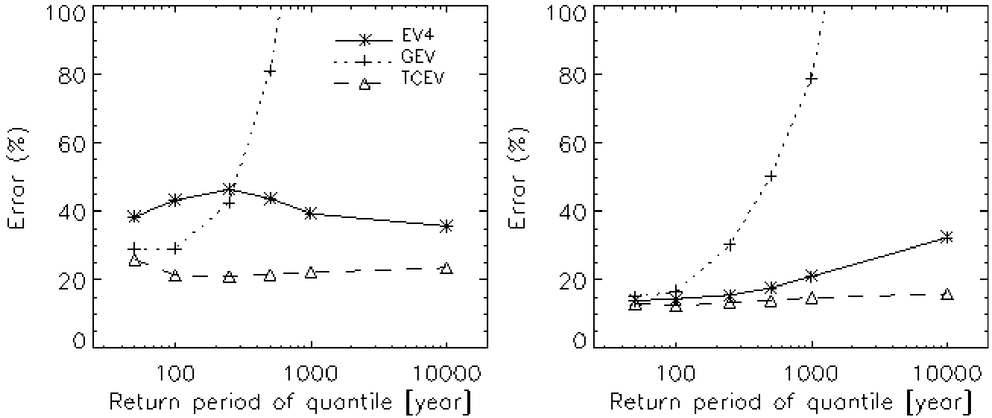

Fig. 7. Estimation error (in %) of $\hat{q}_T$ with ML-GE method for EV4, GEV and TCEV distribution functions, assuming a TCEV population. H=50, M=400 and N=50 years. Left: scenario 1, right: scenario 2.

HESSD
7, 5413–5440, 2010