Stochastic Hydrology and Hydraulics 12 (1998) 267–283

Using the TCEV distribution function with systematic and non-systematic data in a regional flood frequency analysis

F. Francés

Abstract: Due to the social and economic implications, flood frequency analysis must be done with the highest precision. For this reason, the most suitable statistical model must be selected, and the maximum amount of information must be used. Floods in Mediterranean rivers can be produced by two different mechanisms, which forces the use of a non-traditional distribution like the TCEV. The information can be increased by using additional non-systematic data, or with a regional analysis, or both. Through the statistical gain concept, it has been shown that in most cases the use of additional non-systematic information can decrease the quantile estimation error in about 50%. In a regional analysis, the benefit of additional information in one station, is propagated to the rest of the stations with only a small decrease with respect to the at-site equivalent analysis.

Key words: Flood frequency analysis, TCEV, non-systematic information, regional, statistical gain.

1 Introduction

Flood frequency analysis consists basically of obtaining the relationship between flood quantiles and their non-exceedence probability (also referred to as risk or return period). It has been one of the main issues in hydrological survey, as it is the basis for the design of hydraulic structures (e.g., dam spillways, diversion canals, dikes and river channels), urban drainage systems, cross drainage structures (e.g., culverts, bridges and dips), flood risk mapping, etc. Due to the social and economic implications, flood frequency analysis must be done with the highest precision. For example, if the accepted risk for a dam spillway is given, the value of the design flow should be estimated as accurately as possible, since a flood quantile estimate lower than real would increase the flood risks at the dam and downstream; on the other hand, an overestimate would increase the cost of the spillway structure unnecessarily.

Mediterranean rivers present very dramatic floods, caused both by cyclonic and convective rainfalls. The heaviest rainfalls occur mainly during the fall,

F. Francés
Departamento de Ingeniería Hidráulica y Medio Ambiente,
Universidad Politécnica, Apdo. 22012, 46071 Valencia, Spain,
Phone: +34-96 3877 612, Fax: +34-96 3877 618,
e-mail: ffrances@hma.upv.es

This research has been partially sponsored by the CICYT grant with reference HID96–1318.

producing river flows of several orders of magnitude higher than the mean flow. These infrequent but large floods cause huge damage at the floodplain. The 3,000 million US$ damages caused by floods in Spain in 1982 and 1983, and over 1,100 human lives lost in the last 30 years (Berga, 1991), not only justify investment on structural and non-structural measures to prevent floods, but also the high cost of such investments demands a careful analysis of flood behavior.

The most straightforward method for estimating flood risk at a particular point where a gauge station is located is to adjust a probability distribution function to the recorded annual maximum flows. Unfortunately this method may give rise to highly variable flood quantile estimators due to: i) the uncertainty of the statistical model; ii) errors in data recording (especially for the largest floods); iii) data series shorter than the return period; and finally iv) the high variance and skewness of annual maximum flows, which cause great variability in the statistical properties of the recorded samples.

In order to increase flood quantile estimator reliability, great efforts have been made in the last years to find the most precise distribution function, the most accurate and robust estimation method, as well as to increase the amount of used information. For the last case, some of the proposed methods are: i) regional analysis; ii) the use of additional non-systematic at-site information, such as historical data or paleofloods; iii) using the flow peaks above a certain value as a random variable; iv) or the combination of any of the procedures mentioned.

The work presented here is framed within this context. In a first part, this paper presents a statistical model for flood frequency analysis of Mediterranean rivers, which permits to include at-site historical data and/or paleofloods in a regional framework. In a second part, the paper demonstrates the improvements in estimating flood quantile with this method relative to the use of at-site systematic records alone. Finally, an example of the practical application of the model is also shown.

2
The statistical model

2.1
The Two-Component Extreme Value distribution function

The series of annual maximum flows of many rivers in the world are characterized by what is called the "dog leg effect" and the "separation phenomenon", first described by Potter (1958) and Matalas et al. (1975) respectively. However, the existence of "separation phenomenon" has been a controversial issue in the last years (see for example the very interesting discussion between Ourda et al., 1997, and Dawdy and Gupta, 1997).

Mediterranean rivers also present these features due mainly to the existence of two kinds of flood populations (Rossi et al., 1984). One of them will be referred to as "ordinary floods". These floods are generated by frontal-type rainfalls with low or medium convectivity, which is the most frequent type of rainfall and produces the smaller floods. On the other hand, the floods called "extraordinary floods" are less frequent, larger and generated by heavy convective rainfall events occurring mainly in the summer and in the fall.

The traditional distribution functions, like the Gumbel or the logPearson Type III, are not capable of reproducing the existence of these two different kinds of events, thus being inappropriate and, as it will be illustrated in the practical application, even dangerous for the flood frequency analysis of any Mediterranean river. On the other hand, new distribution functions have been developed considering this situation, such as the Two-Component Extreme Value or TCEV (Rossi et al., 1984).

The simplest way of getting the expression of the TCEV is to assume that annual maximum values for ordinary (X_1) and extraordinary (X_2) floods come from independent Gumbel populations. Therefore the annual maximum flow is the highest of these two, and its distribution function is the product of the initial distribution functions:

$$F_X(x) = \exp\left(-\lambda_1 e^{-\theta_1 x} - \lambda_2 e^{-\theta_2 x}\right) \tag{1}$$

where λ_1 and θ_1 are the parameters of the ordinary floods, and λ_2 and θ_2 of the extraordinary floods. The Gumbel distribution function can be considered as a particular case, where the two scale parameters are equal. The probability density function is the derivative of Eq. (1):

$$f_X(x) = F_X(x)\psi(x) \tag{2}$$

where:

$$\psi(x) = \lambda_1\theta_1 e^{-\theta_1 x} + \lambda_2\theta_2 e^{-\theta_2 x} \tag{3}$$

There are two standardized parameters, which can be obtained by the following equations (Beran et al. 1986):

$$\theta = \frac{\theta_2}{\theta_1} \tag{4}$$

$$\lambda = \lambda_2\lambda_1^{-\theta} \tag{5}$$

The standardized parameter θ represents the relationship in order of magnitude between ordinary and extraordinary floods. On the other hand, the probability for an annual maximum flow to be extraordinary is given by:

$$p_2 = P[X_1 < X_2] = \int_{-\infty}^{\infty} F_{X_1}(x) f_{X_2}(x)\, dx \tag{6}$$

and after some manipulations it becomes:

$$p_2 = \int_0^{\infty} \exp\left(-\lambda\xi^{\theta} - \xi\right) d\xi \tag{7}$$

which in the normal range of θ can be approximated by λ (Francés, 1995), i.e., λ approximately represents the extraordinary flood probability.

With respect to the moments, the coefficient of variation is a function of λ_1, λ_2 and θ, but the skewness coefficient is only a function of the standardized parameters (Francés, 1995). Lastly, the TCEV has enough flexibility to reproduce the "separation phenomenon" (Beran et al. 1986).

2.2
Regional analysis

A regional flood frequency analysis consists of standardizing previously the series of data recorded at different gauge stations, so that all the information can be analyzed jointly. By increasing the sample size, errors in parameter estimation are supposed to decrease if the region is statistically homogeneous and space correlation is low, although this second condition is less important (Hosking and Wallis, 1988). Regional analysis models differ in the method of standardizing the parameters, in the distribution function employed and/or in the estimation method. Cunnane (1988) presents an interesting revision of most of them. For the TCEV, the method of standardization employed consists of (Rossi et al., 1984):

$$Y_{i,j} = \theta_{i,1} X_{i,j} - \ln \lambda_{i,1} \tag{8}$$

where $X_{i,j}$ = initial series at the gauge station i, and $\lambda_{i,1}$ and $\theta_{i,1}$ = the parameters corresponding to ordinary at-site floods. This standardization implies that the standardized parameters λ and θ defined by Eqs. (4) and (5) remain constant within the region.

If Eq. (8) is replaced in Eq. (1), it follows that this method of standardization is equivalent to assume that the regional parameters corresponding to ordinary floods (named λ'_1 and θ'_1) must be equal to 1. Arnell and Gabrielle (1988) relax this condition and obtain the parameters by means of an iterative method; whereas Ferrer and Ardiles (1997) and Ferrari (1997) employ a combined method which includes the concept of an "index variable" to improve the estimation of the at-site parameters ($\lambda_{i,1}$ and $\theta_{i,1}$) used in the standardization. In order to simplify the way of obtaining the expressions of the asymptotic variances of the quantile estimator, the method proposed in this paper for regional analysis is the following:

i) For the standardization of the at-site data in Eq. (8) a Gumbel distribution function is adjusted to the ordinary floods at every gauge station, obtaining an estimation of parameters $\lambda_{i,1}$ and $\theta_{i,1}$. Prior to this, it is necessary to select the ordinary floods; one possibility is to use the skewness test for outliers elimination described by Kottegoda (1984).

ii) To adjust the TCEV to the standardized data, thus obtaining the 4 regional parameters λ'_1, θ'_1, λ'_2, θ'_2, without forcing the ordinary flood regional parameters to be 1.

2.3
Non-systematic information

Non-systematic information at a gauge station is any censored information for a period prior to the systematic record. Depending on the source, we can distinguish between historical information and paleofloods. Historical information comes from the marks in buildings, photographs, written records in newspapers or books, verbal communications, etc., and it was first used for flood frequency analysis by Benson (1950). Data from paleofloods are available through evidence on the vegetation of the floodplain (for example, ring abnormalities) and paleolevels (for example, slackwater deposits). Paleohydrological techniques have been described thoroughly by Baker (1987) and Baker et al. (1988).

Non-systematic information can be statistically classified based on the type of censoring (Francés et al., 1994). When there is a given censoring limit X_H, it is called censored information type 1. This is the case when the information source is historical, since there is always a threshold level of perception below which floods are not large enough to be remembered. The value of non-censored floods may or may not be known. We will follow Stedinger and Cohn's classification (1986), naming it "censored information" (CE) when the k floods that exceeded the threshold level of perception during the non-systematic period of length M are known. If their values are unknown, it will be called "binomial censored" (BC). If there is no censoring limit, the information is censored information type 2. In this case the k largest floods during the non-systematic period are known, k being a deterministic variable. As the largest paleoflood tends to remove the evidence left by other large paleofloods, k is usually equal to 1; this information is called "maximum flood" (MF).

2.4 Parameter estimation by the method of Maximum Likelihood

The selection of the method of Maximum Likelihood (ML) for estimating the parameters was not only based on its features (like the existence of asymptotic variance and its capability of analyzing any additional quantified data) but also because of its general ability relative to other estimation methods for both additional non-systematic information and regional analysis.

The ML method with additional non-systematic data has been used by many investigators: with a Gumbel population by Leese (1973), Hosking and Wallis (1986a), and Guo and Cunnane (1991); with the Two-Parameter Extreme Value distributions by Francés et al. (1994); with GEV by Phien and Fang (1989); with the lognormal distribution by Condie and Lee (1982), Cohn and Stedinger (1987) and Kroll and Stedinger (1996) (in this case with censored information alone), and with the logPearson Type 3 by Pilon and Adamowski (1993).

With regard to regional analysis, the ML method has been used by Boes et al. (1989) with the Weibull distribution; in the comparisons of Landwehr et al. (1979) with the Gumbel distribution; Jin and Stedinger (1989) with GEV including non-systematic information; and it is the most widely used method with TCEV (Rossi et al. 1984; Arnell and Gabriele, 1988; Ferrari, 1997; Ferrer and Ardiles, 1997).

The ML method consists of selecting those parameters which maximize the likelihood function $L(.)$, which is any function proportional to the joint probability density function of all the random variables of interest. In the case of using systematic information alone, the log-likelihood function (much easier to maximize) shows the following expression for the TCEV:

$$LL_{\mathrm{SY}}(\underline{\Theta}) = -\lambda_1 \sum_{i=1}^{N} \mathrm{e}^{-\theta_1 x_i} - \lambda_2 \sum_{i=1}^{N} \mathrm{e}^{-\theta_2 x_i} + \sum_{i=1}^{N} \ln\psi(x_i) \tag{9}$$

where $\Theta =$ set of distribution parameters, $x_i =$ systematic recorded data, and $N =$ length of the systematic recording period. When using binomial censored information, the increase in the log-likelihood function is given by:

$$LL_{\mathrm{BC}}(\underline{\Theta}) = (M-k)\left(-\lambda_1 \mathrm{e}^{-\theta_1 X_H} - \lambda_2 \mathrm{e}^{-\theta_2 X_H}\right) + k \ln p \tag{10}$$

where: $k =$ number of floods above the threshold level of perception X_H, $p =$ the exceedence probability of X_H, and $M =$ length of the non-systematic period. When using censored information, a new random variable Z appears, which corresponds to floods exceeding the threshold level of perception. In this case the increase in the log-likelihood function becomes:

$$LL_{\mathrm{CE}}(\underline{\Theta}) = (M-k)\left(-\lambda_1 e^{-\theta_1 X_H} - \lambda_2 e^{-\theta_2 X_H}\right) - \lambda_1 \sum_{j=1}^{k} e^{-\theta_1 z_j} - \lambda_2 \sum_{j=1}^{k} e^{-\theta_2 z_j} + \sum_{j=1}^{k} \ln\psi(z_j) \quad (11)$$

where $z_j =$ known historical floods above the threshold level of perception. If the available information for the non-systematic period is that of the maximum flood W, the increase of the log-likelihood function becomes:

$$LL_{\mathrm{MF}}(\underline{\Theta}) = -M\lambda_1 e^{-\theta_1 w} - M\lambda_2 e^{-\theta_2 w} + \ln\psi(w) \quad (12)$$

When more than one type of information is available simultaneously (e.g., systematic and non-systematic information), the function $LL(.)$ will be obtained by adding the corresponding functions of each type of information. In a regional flood frequency analysis, the only difference is that all systematic and non-systematic information from each gauge station must be firstly standardized following Eq. (8).

3
Flood quantile estimation error

The mean square error is a measurement of the X_T flood quantile estimation reliability. It is defined as:

$$\mathrm{MSE}(\hat{X}_T) = E[(\hat{X}_T - X_T)^2] \quad (13)$$

where $E[.]$ is the expected value operator. This error is very difficult to obtain analytically. Thus, either Monte Carlo simulations (to obtain the sample mean square error), or the asymptotic value are used.

3.1
Sample mean square error

The sample mean square error of the flood quantile estimator, is defined as:

$$\mathrm{SMSE}_i(\hat{X}_T) = \frac{1}{S} \sum_{j=1}^{S} \left(\hat{X}_{Tij} - X_T\right)^2 \quad (14)$$

where $X_T =$ the quantile value of return period T, $\hat{X}_{Tij} =$ quantile estimator in simulation j using information i, and $S =$ total number of simulations. However, the SMSE defined by Eq. (14) is a random variable. It is, thus, an approximated value which will tend to the true MSE value when the number of simulations tends to infinity. To obtain this, it is necessary to generate a great number of synthetic series. In this work, many different scenarios have been simulated, each with 5,000 simulations. The algorithm used to obtain random values with uniform distribution at the interval [0, 1] has been a linear congruential generator (Bratley et al., 1987).

3.2
Asymptotic variance

Another method of solving the problem is by using the asymptotic variance. The asymptotic variance or Cramer-Rao Lower Bound (CRLB) is a lower bound of the variance of all unbiased estimators. The ML method is asymptotically unbiased, thus the CRLB is the lower bound of its MSE when the amount of information tends to infinity. The matrix of the asymptotic variances and covariances of the parameter estimators of a statistical model is given by the inverse of the Fisher information matrix (Kendall and Stuart, 1967). The elements of this matrix are obtained as:

$$\mathbf{I}_i(j,k) = E\left[-\frac{\partial^2 LL_i(\Theta)}{\partial\Theta_j \partial\Theta_k}\right] \tag{15}$$

where $LL_i(.)$ = the log-likelihood function, Θ_j, Θ_k = the parameters used in the model, and i = type of information used: $i = 0$ for systematic information alone (SY), and $i = 1$ for SY plus additional BC information, $i = 2$ for SY plus additional CE information, and $i = 3$ for SY plus additional MF information. Unfortunately, for the TCEV distribution function, this matrix is highly complex and it is not possible to write them in a compact form. In any case, all the expressions can be found in Francés (1995).

From the information matrix, the quantile asymptotic variance or Cramer-Rao Lower Bound is given by:

$$\mathrm{CRLB}i(\hat{X}_T) = \mathrm{V}\mathrm{I}_i^{-1}V' \quad i = 0,1,2,3 \tag{16}$$

where $\mathbf{V}$ is the derivative vector of flood quantile with respect to the parameters. Even when X_T cannot be obtained explicitly, from Eq. (1) we can get its derivatives:

$$\begin{aligned}\mathbf{V} &= \left(\frac{\partial X_T}{\partial\lambda_1}, \frac{\partial X_T}{\partial\theta_1}, \frac{\partial X_T}{\partial\lambda_2}, \frac{\partial X_T}{\partial\theta_2}\right)\\ &= \left(\frac{\mathrm{e}^{-\theta_1 X_T}}{\psi(X_T)}, -\frac{\lambda_1 X_T \mathrm{e}^{-\theta_1 X_T}}{\psi(X_T)}, \frac{\mathrm{e}^{-\theta_2 X_T}}{\psi(X_T)}, -\frac{\lambda_2 X_T \mathrm{e}^{-\theta_2 X_T}}{\psi(X_T)}\right)\end{aligned} \tag{17}$$

3.3
Comparison between quantile asymptotic and sample variances

As SMSE requires Monte Carlo simulations with high computing demand, it is better to use the asymptotic variance CRLB as an approximation to the mean square error, MSE. If asymptotic and sample values are similar, the former will also be close to the true values, and then CRLB could be used instead of MSE. To compare CRLB and SMSE, the asymptotic and sample coefficients of variation of the quantile estimator have been plotted. Their expressions are, respectively:

$$\mathrm{ACV}i(\hat{X}_T) = \frac{\mathrm{CRLB}(\hat{X}_T)}{X_T} \tag{18}$$

$$\mathrm{SCV}i(\hat{X}_T) = \frac{\mathrm{SMSE}_i(\hat{X}_T)}{X_T} \tag{19}$$

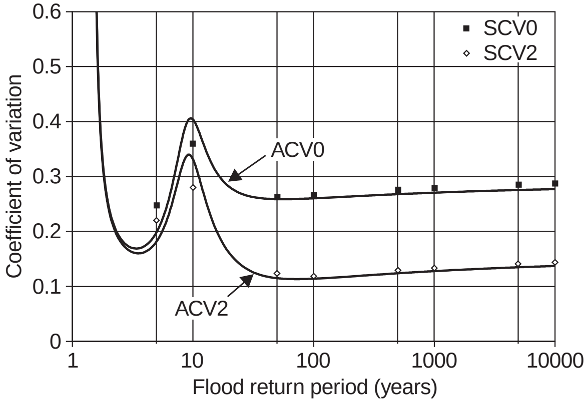

Fig. 1. For the initial scenario, comparison of the asymptotic (ACV) and sample (SCV) coefficient of variation with systematic information alone (0) and with additional censored information (2)

Fig. 1 compares asymptotic and sample coefficients of variation when using systematic information alone ($i = 0$) or additional censored information ($i = 2$). The scenario represented, referred to as "initial scenario", has a systematic period N of 100 years, a non-systematic period M of 500 years, the return period of historical threshold level of perception H is 50 years, and the TCEV parameters are $\lambda_1 = 1$, $\theta_1 = 1$, $\lambda_2 = 0.2$ (by numerical integration of Eq. (7) this is equivalent to an outlier probability $p_2 = 0.18$) and $\theta_2 = 0.1$. As it can be seen in the figure, the differences between ACV and SCV are very small for medium and high quantiles, whereas for low quantiles the difference increases. Similar results are obtained with other types of additional information and a wide range of scenarios. Therefore, the asymptotic variance is a good approximation of the mean square error, which is the same conclusion obtained by Francés et al. (1994) for the Two-Parameter Extreme Value distributions. However, this result is contradictory with the results obtained by Phien and Fang (1989), who concluded that for the GEV distribution the CRLB is much lower than sample variance, suggesting the use of the observed information matrix rather than using the Fisher information matrix.

4
Asymptotic statistical gains with additional non-systematic information

The concept of statistical gain can be used as a way of measuring reliability in flood quantile estimation when using any type of additional information besides systematic records (Francés et al., 1991). Asymptotic statistical gain is defined as:

$$\mathrm{ASG}i = 1 - \frac{\mathrm{CRLB0}}{\mathrm{CRLB}i} \quad i = 1, 2, 3 \tag{20}$$

The analytical expressions have not been obtained for the TCEV; however it has been numerically proved that asymptotic statistical gains with additional censored information type 1 (BC and CE) are exclusively a function of: i) the ratio between non-systematic and systematic period lengths; ii) the return period H of the historical threshold level of perception; iii) the return period T of the quantile of interest; and iv) the standardized parameters λ and θ, defined in Eqs. (4) and (5). On the other hand, asymptotic statistical gain using additional censored information type 2 (MF) is exclusively a function of: i) the length N of the systematic records; ii) the length M of the historical period; iii) the return period

T of the quantile of interest; and iv) the standardized parameters λ and θ. The influence of these factors on the statistical gain is described qualitatively below.

4.1
Sensitivity to the standardized parameters

As it can be seen in Fig. 2, in the usual range of the standardized parameters the statistical gain with additional binomial censored information (ASG1) increases as the number of extraordinary floods increases, i.e., as λ increases. On the contrary, it decreases as both flood populations become closer in magnitude (i.e., as θ increases). With regard to statistical gain using additional censored information (Fig. 3) as the differences are small, for practical purposes it can be considered as independent of the 4 TCEV parameters. Fig. 4 shows the asymptotic statistical gain using the additional information of maximum flood. This gain decreases as the number of extraordinary floods increases, and the influence of the standardized parameter θ is lower.

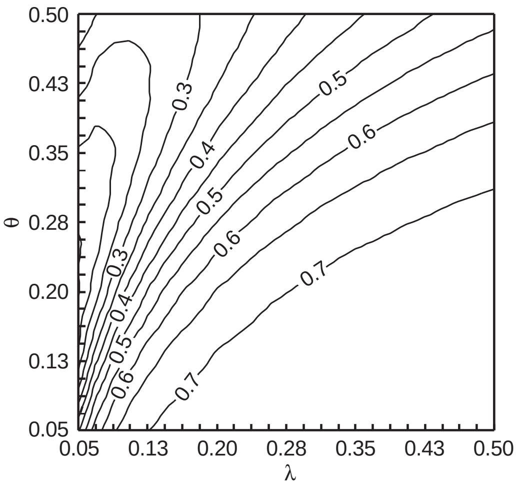

Fig. 2. Asymptotic statistical gain with additional binomial censored information (ASG1) as a function of the standardized parameters, for the 500 years quantile and initial scenario

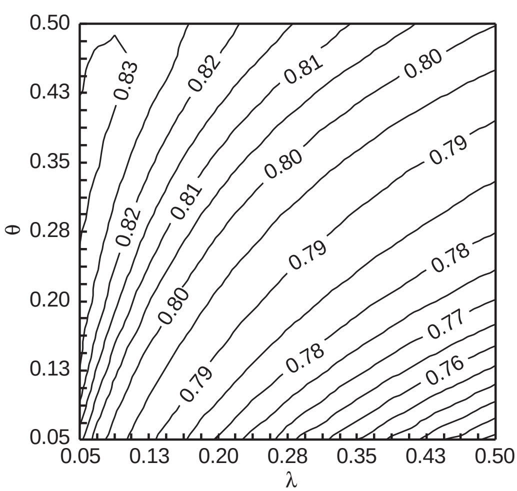

Fig. 3. Asymptotic statistical gain with additional censored information (ASG2) as a function of the standardized parameters, for the 500 years quantile and initial scenario

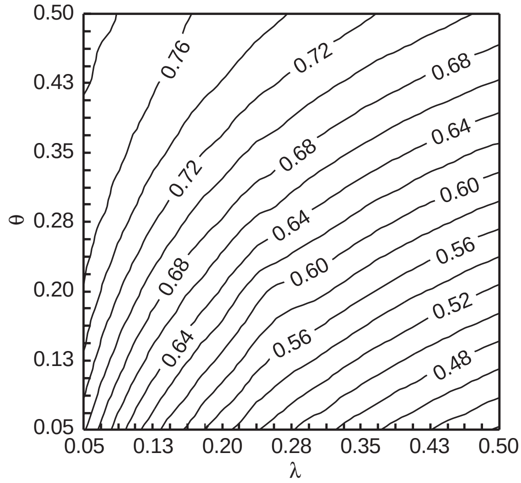

Fig. 4. Asymptotic statistical gain with additional maximum flood information (ASG3) as a function of the standardized parameters, for the 500 years quantile and initial scenario

4.2
Sensitivity to the flood return period

The influence of the flood return period on the ASG can be seen in Fig. 5 for the initial scenario. The statistical gain is minimum for a low return period, and it is only appreciable for medium or high return periods. With censored information type 1, maximum asymptotic statistical gain is reached with ASG1 for quantile values equal to the threshold level of perception H, or slightly higher for ASG2, decreasing slowly from this maximum value. Notice the small difference between ASG1 and ASG2 in the initial scenario for all the range of quantiles analyzed. The quantile which maximizes its statistical gain with censored information type 2 is close to the non-systematic length M, although in practice, for medium and high quantiles the statistical gain is constant.

4.3
Sensitivity to the threshold level of perception return period

The historical threshold level of perception is one of the factors affecting asymptotic statistical gains with censored information type 1 (ASG1 and ASG2). Contrarily, obviously it does not affect statistical gain with censored information type 2 (ASG3). As Fig. 6 shows, only with threshold levels of perception higher than 500 year return period, the maximum flood information is more valuable

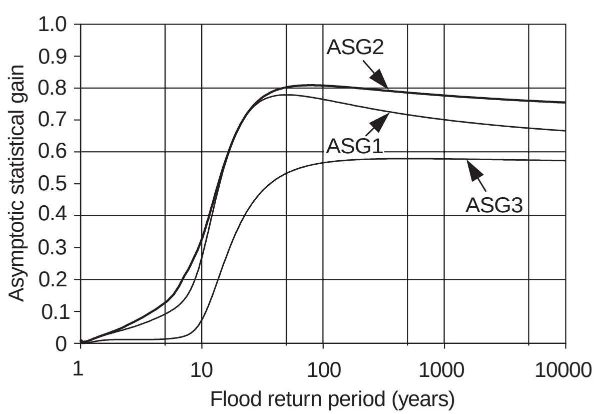

Fig. 5. Sensitivity of the asymptotic statistical gain to the flood return period, for the initial scenario and different types of additional information

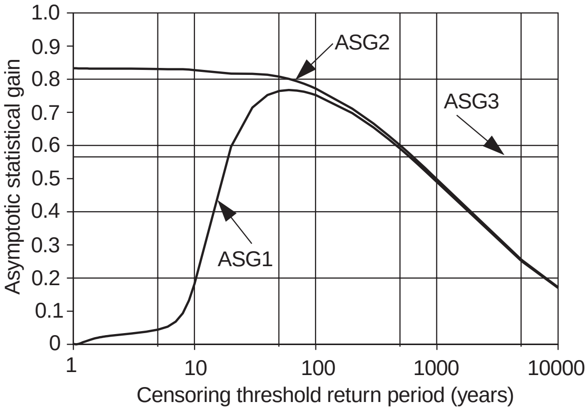

Fig. 6. Sensitivity of the asymptotic statistical gain to the threshold level of perception return period, for the initial scenario and different types of additional information

than the censored information type 1, but this is not the usual case. The statistical gain with binomial censored information is lower than with censored information. This difference is negligible for medium and high threshold levels of perception, whereas for low threshold levels of perception, ASG1 is very small.

4.4
Sensitivity to the length ratio *r* between systematic and non-systematic periods

The three statistical gains increase with ratio r, as can be seen in Fig. 7; but if r is lower than 1 (in the figure, $M < 100$ years) the influence of additional information is negligible. With censored information type 1, statistical gains present an asymptotic behavior. Furthermore, with censored information ASG2 tends to 1 as r increases; i.e., if historical length is infinite, quantile estimator variance with CE information is null. This asymptotic behavior of ASG1 and ASG2 makes it statistically little beneficial to increase the length of the non-systematic information period. Contrarily, ASG3 increases more slowly, showing a maximum value for very high M.

5
The application of the TCEV model in regional analysis

Once the at-site standardized parameters and the four regional TCEV parameters have been estimated, the at-site flood quantile is obtained from Eq. (8) by the following expression:

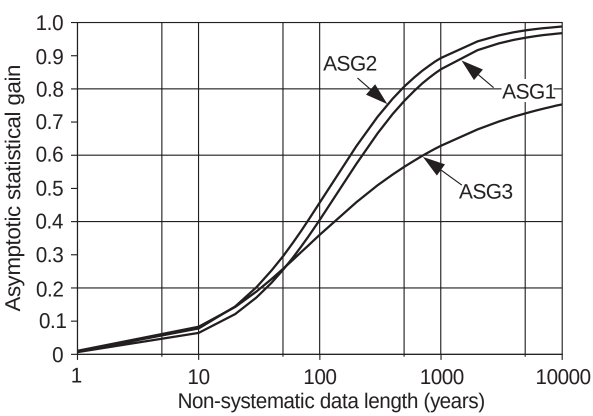

Fig. 7. Sensitivity of the asymptotic statistical gain to the length of the non-systematic period, for the initial scenario and different types of additional information.

$$X_{i,T} = \frac{Y_{i,T} + \ln \lambda_{i,1}}{\theta_{i,1}} \tag{21}$$

By taking the first elements of a Taylor series in Eq. (21), the quantile estimator variance can be approximated by

$$\mathrm{Var}(X_{i,T}) \approx \left(\frac{\partial X_{i,T}}{\partial Y_{i,T}}, \frac{\partial X_{i,T}}{\partial \lambda_{i,1}}, \frac{\partial X_{i,T}}{\partial \theta_{i,1}}\right) \mathbf{S} \left(\frac{\partial' X_{i,T}}{\partial Y_{i,T}}, \frac{\partial X_{i,T}}{\partial \lambda_{i,1}}, \frac{\partial X_{i,T}}{\partial \theta_{i,1}}\right) \tag{22}$$

where $\mathbf{S}$ is the variance and covariance matrix of the random variables $Y_{i,T}$, $\lambda_{i,1}$ and $\theta_{i,1}$. With the hypothesis of independence between the regional flood quantile and the at-site standardized parameters, and if the variances and covariances are approximated by the corresponding CRLB, we obtain:

$$\begin{aligned}\mathrm{Var}(X_{i,T}) \approx{}& \frac{1}{\theta_{i,1}^2}\mathrm{CRLB}(Y_{i,T}) + \frac{1}{\lambda_{i,1}^2\theta_{i,1}^2}\mathrm{CRLB}(\lambda_{i,1}) - 2\frac{Y_{i,T} + \ln\lambda_{\mathrm{i},1}}{\lambda_{i,1}\theta_{i,1}^3}\mathrm{CRLB}(\lambda_{i,1}\theta_{i,1}) \\ &+ \frac{(Y_{i,T} + \ln\lambda_{i,1})^2}{\theta_{i,1}^4}\mathrm{CLRB}(\theta_{i,1})\end{aligned} \tag{23}$$

where CRLB ($Y_{i,T}$) is the asymptotic variance of the quantile of a TCEV distribution given by Eq. (16). On the other hand, CRLB ($\lambda_{i,1}$), CRLB ($\theta_{i,1}$) and CRLB ($\lambda_{i,1}$, $\theta_{i,1}$) are the asymptotic variances and covariances of the parameters of the Gumbel distribution; i.e., the elements of the inverse matrix of the Fisher information matrix for the Gumbel distribution using systematic information alone, which, in this case, can be obtained in a compact form (Leese, 1973; Francés, 1995):

$$\mathbf{I}_0(1,1) = -E\left[\frac{\partial^2 LL_{\mathrm{SY}}(\Theta)}{\partial \lambda^2}\right] = \frac{N}{\lambda^2} \tag{24}$$

$$\mathbf{I}_0(2,2) = -E\left[\frac{\partial^2 LL_{\mathrm{SY}}(\Theta)}{\partial \theta^2}\right] = \frac{N[1 + \Gamma''(2) + \ln^2\lambda - 2\Gamma'(2)\ln \lambda]}{\theta^2} \tag{25}$$

$$\mathbf{I}_0(1,2) = \mathbf{I}_0(2,1) = -E\left[\frac{\partial^2 LL_{\mathrm{SY}}(\Theta)}{\partial \theta \partial \lambda}\right] = \frac{N[\Gamma'(2) - \ln \lambda]}{\lambda\theta} \tag{26}$$

From Eq. (23) it seems clear that any improvement in regional quantile estimation automatically affects at-site quantile estimation, with the same properties as if the at-site analysis had been performed with an equivalent number of years. The only difference lies in the fact that regional statistical gain must be lower than the equivalent at-site value (Fig. 8), since in the first case there are two more parameters to analyze for each gauge station, which obviously decrease the quantile estimator reliability. The effect of a fixed amount of additional at-site information decreases as the number of gauge stations in the region increases; likewise, for at-site analysis, as systematic period length increases, statistical gain decreases. Fig. 8 shows that the statistical gain obtained by additional at-site information of type BC falls below 10% only with more than 20 stations, 100 years long each.

These results confirm those obtained by Jin and Stedinger (1989); however, Hosking and Wallis (1986a, 1986b) were skeptic about the advantages of the use

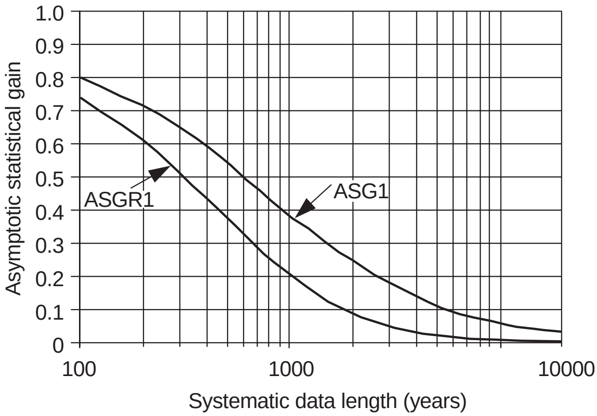

Fig. 8. ASG comparison with at-site (1) and regional analysis (R1), as a function of the total number of systematic data, for the initial scenario and with additional binomial censored information

of non-systematic information for regional analysis. Yet, it should be taken into account that Hosking and Wallis consider as additional information only one maximum flood (which usually gives rise to the lowest statistical gain, as it has been demonstrated in previous paragraphs) and they do not consider any other sources of information. In addition, these authors add recording errors to that single historical flood or paleoflood, when, in practice, the errors produced in determining a historical flood and a large systematic flood are of the same magnitude; furthermore, if the non-systematic information values present significant errors, they can be analyzed as binomial censored information.

6
Application to the Jucar and Turia rivers

The reason for applying the model to a practical case is to judge its validity from two points of view: applicability and profitability. The example presented here is the joint analysis of the floods in the Jucar and Turia rivers, located on the Spanish Mediterranean coast, by using the historical information available.

The area of the Jucar river catchment is 22,000 km^2, although the top half of the catchment does not contribute to the floods produced in the lower catchment area. The systematic recording period is 42 years. As for its historical information, it is known that from 1388, 70 flood events occurred causing damages in the villages located at the river floodplain. The Hydrological Research Center of the Public Works Ministry of Spain (Centro de Estudios Hidrográficos, 1983) quantified the 6 most important floods which occurred since the 17th century; if the censoring limit is placed at 6,200 m^3/s, the values of the 5 floods exceeding it are known with an approximation similar to that of the great floods recorded during the systematic period. In this case, the historical information of the Jucar river used is of the CE type with $M = 154$ years.

The Turia river, located to the north of the Jucar river, has a catchment area of 6,300 km^2 and flows in the city of Valencia. The total length of the systematic record is 41 years. According to Carmona (1990), 22 floods affecting the city of Valencia occurred from 1321 to 1977. At the end of the 16th century the river bed began to be channelized, with no other important structural changes from the beginning of the 18th century to the beginning of the systematic recording period. During this period the four flood events affecting the city had to exceed the artificial river channel capacity, which can be evaluated in 2,300 m^3/s. However, the exact values of these historical floods are not accurately known. Therefore, in this case the historical information has been used as BC information, with $M = 235$ years.

The coefficient of spatial correlation between these two rivers is –0.0384, hence they can be considered as independent series. On the other hand, the joint analysis of the systematic and historical data of these two rivers was possible not only due to their geographical proximity but also due to their similar hydro-morphological features. Because of the small size of the region (only 2 gauge stations) the analysis of their statistical homogeneity was done once the estimation of the model was completed; then the plotting positions and the adjusted function for both the standardized and at-site parameters were compared. The regional parameters are: $\lambda'_1 = 0.9638$, $\theta'_1 = 1.0434$, $\lambda'_2 = 0.0906$ and $\theta'_2 = 0.0329$, thus nearly fulfilling the regionalization hypothesis (regional parameters of the ordinary floods equal to 1) in only one iteration. Figs. 9 and 10

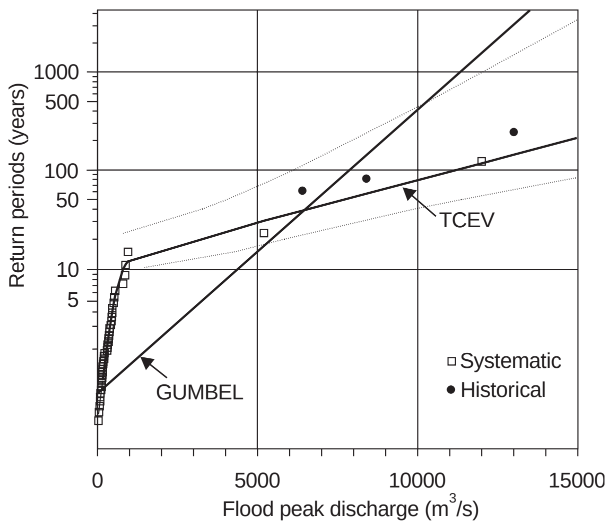

Fig. 9. Results for the Jucar river comparing the plotting positions, an at-site Gumbel, and a regional TCEV with additional censored information (dotted lines are the 95% TCEV confidence limits)

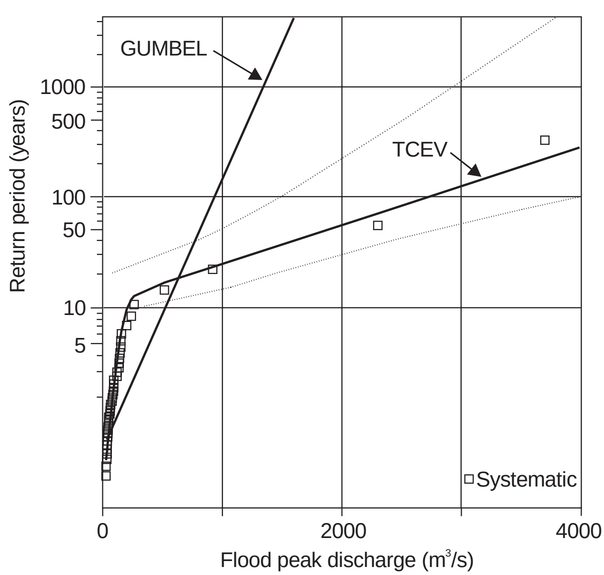

Fig. 10. Results for the Turia river comparing the plotting positions, an at-site Gumbel, and a regional TCEV with additional binomial censored information (dotted lines are the 95% TCEV confidence limits)

Table 1. 95% confidence limits of the 1,000 years quantile with different information levels for the case study

	X_{1000} (m^3/s)	at-site SY	at-site SY + historical	regional SY + historical
Jucar river	22.711	±24.700	±12.700	±10.700
Turia river	2.222	± 6.130	± 3.250	± 2.650

illustrate the TCEV adjustment for historical information and regional analysis; the 95% confidence limits assuming a normal distribution of the estimation error with a variance equal to the corresponding CRLB value; the plotting positions by using Hazen expression (Hirsch, 1987); and the adjustment of the Gumbel function using systematic record alone. It is clear that the TCEV model is much more appropriate than the Gumbel distribution, basically due to the fact that it is capable of reproducing the "dog leg effect" presents in these rivers.

The use of additional historical information and joint analysis has decreased considerably estimation error values in the statistical model. Table 1 shows how error values decrease as the amount of information used in estimating the 1,000 year flood quantile increases. It can be noted that the use of historical information reduces the estimation error by somewhat less than 50% and, in addition, if regional analysis is carried out, another 10% decrease can be further added.

7 Conclusions

This paper has presented an approach which permits to increase the information used in estimating river flood quantiles by means of the use of non-systematic information in a regional analysis framework. The TCEV has been the distribution function employed which, as it has been proved in the case study illustrated in this paper, fits well the statistical features of the Mediterranean rivers. On the contrary, traditional distribution functions, like the Gumbel function, can give rise to disastrous results, underestimating the quantile estimates for medium and high return periods, and overestimating those for low return periods.

The improvements provided by using additional information and regional analysis compared to using at-site systematic records alone have been numerically measured through the concept of statistical gain. To do that it has firstly been checked that the asymptotic variance (CRLB) obtained analytically is a good approximation to the mean square error of the quantile estimator.

For a particular quantile, the value of the non-systematic censored information type 1 is a function of the length ratio between the non-systematic and the systematic periods, of its return period, of the threshold level of perception return period, and of the standardized parameters. In the case of non-systematic censored information type 2, its value is a function of the lengths of the systematic and non-systematic periods, of its return period, and of the standardized parameters. Francés et al. (1994) showed for the Two-Parameter Extreme Value distributions the statistical gain is not a function of their parameters. These results do not contradict each other, because the TCEV standardized parameters represent the frequency of extraordinary events and the relationship of the magnitudes between ordinary and extraordinary floods, which are properties derived from the assumption of the existence of two different flood populations.

On the basis of the sensitivity analysis, two very important observations on the use of additional non-systematic information with a TCEV population should be pointed out, also applicable at least to the Two-Parameter Extreme Value distributions:

i) If the censoring threshold has a medium or high return period, the differences between ASG1 and ASG2 are small; in this case it is advisable to use the additional information as binomial censored information. In the case study, the censoring threshold for the Turia river had a return period of 70 years, whereas for the Jucar river it was 35 years. Therefore, for the former, the historical information has been used as binomial censored information without evaluating the magnitude of the floods, and for the latter as censored information.

ii) For practical purposes, it may be useless to compare statistical gains type 1 and type 2 as they may have different sources. However, if there is enough historical information available, it may be non profitable statistically to increase the length of the historical period or to add paleoflood information.

Finally, with the regional statistical model presented here, the use of additional information from any gauge station means greater reliability in flood quantile estimation in the whole region, with a statistical gain similar to that obtained in an equivalent at-site analysis.

References

Arnell NW, Gabriele S (1988) The Performance of the Two-Component Extreme Value Distribution in Regional Flood Frequency Analysis. Water Res. Res., 24: 879–887

Baker VR (1987) Paleoflood hydrology and extraordinary flood events. Jour. of Hydrol., 96: 79–99

Baker VR, Kochel RC, Patton PC (eds.) (1988) Flood geomorphology. Wiley, New York

Benson MA (1950) Use of Historical Data in Flood Frequency Analysis. Eos. Trans. AGU, 31 (3): 419–424

Beran M, Hosking JRM, Arnell N (1986) Comment on "Two-Component Extreme Value Distribution for Flood Frequency Analysis". Water Res. Res., 22: 263–266

Berga L (1991) Flood Forecasting in Spain. Procc. of XXIV IAHR International Congress, Madrid, volume A: 79–88

Boes DC, Heo J, Salas JD (1989) Regional Flood Quantile Estimation for a Weibull Model. Water Res. Res., 25: 979–990

Bratley P, Fox BL, Schrage LE (1987) A Guide to Simulation. Springer-Verlag

Carmona P (1990) La Formación de la Plana Aluvial de Valencia. Institució Valenciana d'Estudis i Investigació, Valencia, (in spanish)

Centro de Estudios Hidrográficos (1983) Actualización de los Estudios de Desagüe de la Autopista Tarragona-Alicante, Tomo IV: Cruce con el río Júcar. Report for the Spanish Ministry of Public Works, Madrid, (in spanish)

Cohn TA, Stedinger JR (1987) Use of Historical Information in a Maximum Likelihood Framework. Jour. of Hydrol., 96: 215–233

Condie R, Lee KH (1982) Flood Frequency Analysis with Historic Information. Jour. of Hydrol., 58: 47–61

Cunnane C (1988) Methods and merits of regional flood frequency analysis. Jour. of Hydrol., 100: 269–290

Dawdy DR, Gupta VK (1997) Reply. Water Res. Res., 33(1): 273–275

Ferrari E (1997) Regional Rainfall and Flood Frequency Analysis in Italy. In Friend Amhy Annual Report no. 4, Developments in Hydrology of Mountainous Areas, UNESCO IHP–IV, Technical Documents in Hydrology, no. 8: 45–52

Ferrer J, Ardiles L (1997) Frequency analysis of daily rainfall annual maximum series in Spain. In Friend Amhy Annual Report no. 4, Developments in Hydrology of Mountainous Areas, UNESCO IHP–IV, Technical Documents in Hydrology, no. 8: 53–59

Francés F (1995) Utilización de la Información Histórica en el Análisis Regional de las Avenidas. Centro Internacional de Métodos Numéricos en Ingeniería, no. 27, Barcelona, (in spanish)

Francés F, Salas JD, Boes DC (1991) Flood Frequency Analysis by Using Historical Information. Procc. of XXIV IAHR International Congress, Madrid, volume A: 11–21

Francés F, Salas JD, Boes DC (1994) Flood frequency analysis with systematic and historical or paleoflood data based on the two-paramater general extreme value models. Water Res. Res., 30(6): 1653–1664.
Hirsch RM (1987). Probability Plotting Position Formulas for Flood Records with Historical Information. Jour. of Hydrol., 96: 185–199
Hosking JRM, Wallis JR (1986a) Palaeoflood Hydrology and Flood Frequency Analysis. Water Res. Res., 22(4): 543–550
Hosking JRM, Wallis JR (1986b) The Value of Historical Data in Flood Frequency Analysis. Water Res. Res., 22(11): 1606–1612
Hosking JRM, Wallis JR (1988) The effect of intersite dependence on regional flood frequency analysis. Water Res. Res., 24(4): 588–600
Guo SL, Cunnane C (1991) Evaluation of the usefulness of historical and paleological floods in quantile estimation. Jour. of Hydrol., 129: 245–262

Jin M, Stedinger JR (1989) Flood Frequency Analysis with Regional and Historical Information. Water Res. Res., 25(5): 925–936
Kendall MG, Stuart A (1967) The Advanced Theory of Statistics. Vol. II, 2nd Edition, Hafner Publ. Co., New York
Kottegoda NT (1984) Investigation of Outliers in Annual Maximun Flow Series. Journal of Hydrology, 72: 105–137
Kroll CN, Stedinger JR (1996) Estimation of moments using censored data. Water Res. Res., 32(4): 1005–1012
Landwehr JM, Matalas NC, Wallis JR (1979) Probability Weighted Moments Compared with some Traditional Techniques in Estimating Gumbel Parameters and Quantiles. Water Res. Res., 15: 1055–1064
Leese MN (1973) Use of Censored Data in the Estimation of Gumbel Distribution Parameters for Annual Maximum Flood Series. Water Res. Res., 9: 1534–1542
Matalas NC, Slack JR, Wallis JR (1975). Regional Skew in Search of a Parent. Water Res. Res., 11(6): 815–826
Ourda TBM, Bobée B, Rasmussen PF, Bernier J (1997) Comment on "Multiscaling and skew separation in regional floods" by David R. Dawdy and **Vijay K. Gupta**. Water Res. Res., 33 (1): 271–272
Phien HN, Fang TE (1989) Maximum likelihood estimation of the parameters and quantiles of the general extreme-value distribution from censored samples. Jour. of Hydrol., 105: 139–155
Pilon PJ, Adamowski K (1993) Asymptotic variance of flood quantile in log Pearson type III distribution with historical information. Jour. of Hydrol., 143: 481–503
Potter WD (1958) Upper and Lower Frequency Curves for Peak Rates of Runnof. EOS. Trans. AGU, 39: 100–105
Rossi F, Fiorentino M, Versace P (1984) Two-Component Extreme Value Distribution for Flood Frequency Analysis. Water Res. Res., 20: 847–856
Stedinger JR, Cohn TA (1986) Flood Frequency Analysis with Historical and Paleoflood Information. Water Res. Res., 22(5): 785–793