DOI: 10.24850/j-tyca-2021-04-04

Articles

Flood frequency analysis with the GEV distribution for *r*-annual events

Análisis de frecuencias de crecientes con la distribución GVE para *r* eventos anuales

Daniel Francisco Campos-Aranda[1]

[1]Retired professor of the Autonomous University of San Luis Potosí, Mexico, campos_aranda@hotmail.com

Correspondence author: Daniel Francisco Campos Aranda, campos_aranda@hotmail.com

Abstract

Floods are the maximum annual flows of a river and their *frequency analysis* are statistical techniques that allow estimating values associated with low probabilities of being exceeded. Such *Predictions* allow the hydrological design and review of hydraulic works. The fundamental stage of frequency analysis is the selection and fit, to the available data or sample, of a probability distribution function to make predictions. The Generalized Extreme Value (GEV) distribution has theoretical bases that make it a probabilistic model suitable for floods, maximum sea levels, extreme wind speeds, etc. Due to its genesis, the GEV processes one data per year, the maximum. This disadvantage has been overcome with the method of the *r* annual maximum events. Such procedure is described and applied to the adjustment of the GEV function, through the maximum likelihood method, using the Complex algorithm of multiple restricted variables, for numerical maximization. Five records of floods with five events per year were integrated from hydrometric stations of the Hydrological Region No. 10 (Sinaloa), Mexico. Also, two records of maximum sea levels with *r* = 5 were processed, taken from specialized literature, one of them is non–stationary since it shows a trend. Results were analyzed and the predictions of the GEV adjustment with *r* annual events, were compared against those of the classical L– and LH–moment methods. Finally, Conclusions are formulated, which highlight the convenience of the described method and point out the advantages of using the Complex algorithm as a numerical technique; due to the above, the systematic application of the described procedure is recommended.

Keywords: GEV distribution, maximum likelihood method, Complex algorithm, decision and dependent variables, Wald–Wolfowitz test, standard error of fit, covariates, predictions.

Resumen

Las *crecientes* son los gastos máximos anuales de un río y su *análisis de frecuencias* son técnicas estadísticas que permiten estimar valores asociados con bajas probabilidades de ser excedidos. Tales *predicciones* permiten el diseño y la revisión hidrológica de las obras hidráulicas. La etapa fundamental del análisis de frecuencias consiste en seleccionar y ajustar, a los datos o muestra disponible, una función de distribución de probabilidades para realizar las predicciones. La distribución general de valores extremos (GVE) tiene bases teóricas que la convierten en un modelo probabilístico adecuado a crecientes, niveles máximos del mar, velocidades extremas de viento, etcétera. Debido a su génesis, la GVE procesa un dato por año, el máximo. Esta desventaja fue superada con el método de los *r* eventos máximos anuales. Tal procedimiento se describe y aplica al ajuste de la función GVE, a través del método de máxima verosimilitud, utilizando al algoritmo Complex de múltiples variables restringidas, para la maximización numérica. Se integraron cinco registros de crecientes con cinco eventos por año, en estaciones hidrométricas de la Región Hidrológica No. 10 (Sinaloa), México. Además, se procesaron dos registros de niveles máximos del mar con *r* = 5, tomados de la literatura especializada, uno de ellos es no estacionario al presentar tendencia. Se analizaron los resultados y se contrastaron las predicciones del ajuste de la GVE con *r* eventos anuales contra las de los métodos clásicos de momentos L y LH. Por último, se formulan las Conclusiones, las cuales destacan la conveniencia del método descrito y señalan las ventajas del uso del algoritmo Complex como técnica numérica; debido a lo anterior, se recomienda la aplicación sistemática del procedimiento descrito.

Palabras clave: distribución GVE, método de máxima verosimilitud, algoritmo Complex, variables de decisión y dependientes, test de Wald–Wolfowitz, error estándar de ajuste, covariables, predicciones.

Received: 18/06/2020

Accepted: 03/09/2020

Introduction

In simple terms, *frequency analysis* is a statistical technique that allows a probabilistic model or probability distribution function (PDF) to be fitted to a sample of *n* data. Its purpose is to make *predictions*, beyond the data interval used to estimate the fitting parameters of the selected PDF. Usually, the intended predictions are greater than the largest of the observed events, and therefore frequency analysis consists of the study of past events in order to establish probabilities for future occurrences. For example, the annual maximum flows of a river are studied to obtain the *Design Floods*, which are maximum flows associated with low probability of being exceeded. To summarize, frequency analysis methods do not predict future with certainty, but generate probabilistic models that explain and make efficient use of the extreme events that have occurred in the past (Khaliq, Ouarda, Ondo, Gachon & Bobée, 2006).

Frequency analysis consists of four steps: (1) verification of the statistical quality of available data or sample. In fact, validity and accuracy of the predictions require data to be independent and show *stationarity*, which implies that they have been generated by a stable random process over time; (2) selection of a PDF; (3) selection of a method for estimating PDF fitting parameters and (4) adoption of results or predictions. The standard error of fit is commonly applied, which is a quantitative measure between observed and estimated data with PDF adjusted (Rao & Hamed, 2000; Stedinger, 2017).

In relation to the topic of selecting a PDF, the extreme value theory is a statistical discipline that develops models that describe the unusual nature of data. Being extreme values scarce by nature, their predictions involve strong extrapolations, and the extreme value theory provides models that allow them, on an asymptotic basis (Coles, 2001).

The Generalized Extreme Value (GEV) distribution has a theoretical basis that makes it essential in the analysis of frequencies of extreme events, such as river floods, maximum sea levels, extreme wind speeds, etc. (Guedes, Soares, & Scotto, 2004; Khaliq *et al.*, 2006; An & Pandey, 2007). However, implicit in its own formulation of maximum one span or block, which is later discussed, GEV distribution, uses only one extreme value per year and therefore two approaches have emerged that seek to incorporate more events of each year (Coles, 2001; Khaliq *et al.*, 2006).

The first one is called *peaks over threshold* or POT analysis, using all events that exceed a threshold value, for example, the lowest value of the observed annual maximum. Naturally, the independence condition between the selected events must be met. The second approach to using more data is known as the *r largest order statistics* and is about including more information of maximum events in each year or block, which meet the independence condition between them. In order to fulfill the above, it has been found that $r \leq 5$ (Ramesh & Davison, 2002). The fundamental advantage of both approaches is to use other maximum sample values, at the right end of the GEV function, to estimate more *accurate or robust* predictions, as more data are available to quantify their fitting parameters (Tawn, 1988; Coles, 2001; Khaliq *et al.*, 2006).

The *objectives* of this study focus on the following four: (1) To expose the theory behind the application of the GEV distribution, with *r* independent annual maximum events; (2) to describe the maximum likelihood method for estimating the three fitting parameters (μ,σ,κ) of the GEV distribution; (3) to integrate five flood records with r = 5, within Hydrological Region No. 10 (Sinaloa), Mexico, in order to carry out a prediction contrast of GEV distribution, between those obtained with the annual *r* events method and those from the sextile, L moment and LH-moment methods. These contrasts are established on the basis of the standard error of fit and (4) to process with annual *r* events method, the record of maximum sea levels in Venice, Italy, which is not stationary, since it presents a linear upward trend.

Operational theory

The GEV distribution

In frequency analysis of extreme hydrological data, such as floods or annual flows, sea levels, wind speeds and maximum rains, *extreme value distributions* play an outstanding role. Its origin is set from a sample of random variables (X_1, X_2, . . . ,X_n), *independent and identically distributed* (*iid*) that have a probability distribution function

F(x). In practice, X_i are values of the observed process measured hourly as sea levels, or daily as river flows.

The statistic defined in Equation (1) is the maximum of the process for *n* units of observation time (*block maximum*) and its probabilistic behavior focuses on development of the extreme value theory.

$$M_n = \max[X_1, X_2, \ldots, X_n] \quad (1)$$

The cumulative probability distribution M_n is defined by condition *iid* and it is (Smith, 1986; Tawn, 1988; Dupuis, 1997; Coles, 2001; An & Pandey, 2007):

$$P(M_n \leq x) = P[X_1 \leq x, \ldots, X_n \leq x] \; ; \; P(M_n \leq x) = P(X_1 \leq x) \cdot \cdots \cdot P(X_n \leq x) \; ; \; P(M_n \leq x) = [F(x)]^n \quad (2)$$

To obtain the behavior of Equation (2) according to *n* tends to infinity and avoid trivial or degenerative solutions, a linear normalization of variable M_n is performed, based on constants $a_n > 0$ and b_n; this is:

$$M_n^* = \frac{M_n - b_n}{a_n} \quad (3)$$

Now, extreme value distribution converges to a nontrivial result, defined as follows:

$$P(M_n^* \leq x) = [F(a_n \cdot x + b_n]^n \rightarrow G(x) \text{ when } n \rightarrow \infty \quad (4)$$

The asymptotic distribution *G*(*x*) must converge in one of the three families called Gumbel, Fréchet and Weibull. These three distributions are included in the Generalized Extreme Value (GEV), whose expression is:

$$G(x) = \exp\left\{-\left[1 + \kappa \frac{(x-\mu)}{\sigma}\right]^{-1/\kappa}\right\} \quad (5)$$

in which, μ, σ and κ are the parameters of location, scale and shape; with $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \kappa < \infty$. In addition, it is defined by the set $\{x: [1+\kappa(x-\mu)/\sigma]>0\}$. If $\kappa > 0$ GEV is type II or Fréchet, with no upper limit (μ-σ/κ < x <∞). When $\kappa < 0$ type III or Weibull is defined, with upper limit (∞ < x < μ-σ/κ). Finally, when $\kappa = 0$, in an asymptotic sense, the distribution type I or Gumbel ($-\infty < x < \infty$), also called double exponential, is reached, whose equation is:

$$G(x) = \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\} \quad (6)$$

Selecting each year maximums (Equation (1)), the sought *predictions* (X_{Tr}) for an exceedance probability q or a return period Tr in years ($Tr = 1/q$) are obtained based on Equation (5), since $G(x) = 1 - q$ (Coles, 2001):

$$X_{Tr} = \mu - \frac{\sigma}{\kappa}\{1 - [-\ln(1-q)]^{-\kappa}\} \text{ for } \kappa \neq 0 \quad (7)$$

$$X_{Tr} = \mu - \sigma \cdot \ln[-\ln(1-q)] \text{ for } \kappa = 0 \tag{8}$$

It should be clarified that during the last century GEV distribution equation (Smith, 1986; Tawn, 1988; Dupuis, 1997; Hoking & Wallis, 1997; Rao & Hamed, 2000) was:

$$G(x) = \exp\left\{-\left[1 - \kappa \frac{(x-\mu)}{\sigma}\right]^{1/\kappa}\right\} \tag{9}$$

For which, the Fréchet model is obtained with $\kappa < 0$. This change of κ by $-\kappa$ in Equation (9), has been started, logically by Coles (2001) and accepted by Ramesh and Davison (2002), Katz, Parlange and Naveau (2002), and by Khaliq *et al.* (2006). It has a slight convenience of signs when handling the likelihood function and its positivity constraints in Equation (13), Equation (14) and Equation (15) below and is therefore retained in this study.

Maximum likelihood method

According to the referred text (Kite, 1977; Rao & Hamed, 2000; Coles, 2001; Kottegoda & Rosso, 2008; Meylan, Favre, & Musy, 2012) the *maximum likelihood principle* is explained in different levels of detail. But, in general, it implies the following: Given a sample (X_1, X_2, . . .

,X_n) of *iid* observations, which follow a distribution F_θ with fitting parameters θ_1, θ_2, . . . ,θ_q. Then, by definition, the probability of obtaining a value X_i, will be:

$$P(x_i \leq X \leq x_i + dx_i) = f_\theta(x_i) \cdot dx_i \qquad (10)$$

with, $f_\theta((x_i)$ is the probability density function. As the data are *iid*, the probability of obtaining *n* values of X_i, will be the joint probability or likelihood function, designated *L*, and expressed as:

$$L(\theta) = f_\theta(x_1) \cdot f_\theta(x_2) \cdots f_\theta(x_n) = \prod_{i=1}^{n} f_\theta(x_i) \qquad (11)$$

The maximum likelihood method is to find a vector $\hat{\theta}$ of parameters that make maximum $L(\theta)$ and therefore the probability of obtaining the sample (X_1, X_2, . . . ,X_n). It is often more convenient to take logarithms and work with the *logarithmic likelihood function* (Coles, 2001), i.e.:

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f_\theta(x_i) \qquad (12)$$

this is acceptable because the logarithmic function is monotonic and then the $l(\theta)$ function reaches its maximum at the same point as the function $L(\theta)$.

Function *l*(θ) of the GEV distribution

The probability density function of the GEV distribution is as follows (Coles, 2001; Ramesh & Davison, 2002):

$$f(x;\mu,\sigma,\kappa)=\frac{1}{\sigma}\left[1+\kappa\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-1-1/\kappa}\cdot\exp\left\{-\left[1+\kappa\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-1/\kappa}\right\} \quad (13)$$

where, $\sigma > 0$, $-\infty < \mu$, $\kappa<\infty$ and the x interval is such $[1 + \kappa(x-\mu) / \sigma] > 0$, indicated by the + sign outside the square brackets. Any fitting parameters combination that violates the previous *positivity* condition implies that at least one of the observed points (x), it is beyond the distribution endpoints and then the likelihood function is zero and the logarithmic likelihood function is equal to $-\infty$ (Coles, 2001). If you have a sample of annual maximum X_1, X_2, . . . ,X_n, which are independent, the logarithmic likelihood function is:

$$l(\mu,\sigma,\kappa)=-n\ln\sigma-\left(1+\frac{1}{\kappa}\right)\sum_{i=1}^{n}\ln\left[1+\kappa\left(\frac{x_i-\mu}{\sigma}\right)\right]_{+}-\sum_{i=1}^{n}\left[1+\kappa\left(\frac{x_i-\mu}{\sigma}\right)\right]_{+}^{-\frac{1}{\kappa}} \quad (14)$$

Function *l*(θ) of GEV for *r* annual events

When other annual maximum values are included in the GEV distribution fitting, to look for more accurate prediction inference; then, each year you have *r* events defined as: $x^1 \geq x^2 \geq, \ldots, \geq x^r$ and the *joint* probability density function allows you to define the logarithmic likelihood function for *n* years of *r* independent maximum annual values each (Coles, 2001; Ramesh & Davison, 2002), which is: $l(\mu, \sigma, \kappa) = -n \cdot r \cdot \ln \sigma - \left(1 + \frac{1}{\kappa}\right) \sum_{i=1}^{n} \sum_{j=1}^{r} \ln \left[1 + \kappa \left(\frac{x_i^j - \mu}{\sigma}\right)\right]_+$

$$- \sum_{i=1}^{n} \left[1 + \kappa \left(\frac{x_i^r - \mu}{\sigma}\right)\right]_+^{-1/\kappa} \quad (15)$$

The fitting parameters constraints defined in Equation (13) are applied to the above expression and $\left[1 + \kappa\left(x^j - \mu\right)/\sigma\right]$ must also be positive for each j = 1,..., *r*. Equation (14) is a special case of Equation (15), which is obtained when *r* = 1. Tawn (1988) presents an equation similar to Equation (15), which considers the number of *r* events to be variable each year.

Function *l*(θ) of GEV for non-stationary series

Non–stationary processes have characteristics that change systematically over time. In climatological and hydrological processes, non–stationarity is observed over periods, due to different climatic patterns effects in several months. It is also present as a trend, mainly caused by regional or global climate change (Coles, 2001).

There is no general extreme value theory for non–stationary processes and thus a pragmatic approach is followed to use the GEV distribution fitted by maximum likelihood, changing over time or with some other *covariate*, its fitting parameters (Katz *et al.*, 2002). For the specific case of a linear trend, the location parameter is caused to vary over time (Coles, 2001):

$$\mu(t) = \beta_0 + \beta_1 \cdot t \qquad (16)$$

The above expression is replaced in Equation (15), in order to maximize that function and obtain the four fitting parameters (($\beta_0, \beta_1, \sigma, \kappa$). Other models based on GEV for non–stationary series, with the *covariate* approach, are described by Coles (2001) in his chapter 6.

The Complex algorithm

The maximization of Equation (15) to obtain the optimal fitting parameters (μ,σ,κ) of the sought GEV, must be addressed *numerically* and for this, the Complex algorithm, with multiple variables (z) restricted or dimensioned, was selected. Its theoretical approach is as follows (Box, 1965):

$$\text{Minimizar } \mathrm{F}(z_1, z_2, \cdots, z_s) \quad (17)$$

Subject to *m* dependent variables (*y*), function of decision variables (*z*):

$$y_1 = \mathrm{F}(z_1, z_2, \cdots, z_s) \;\vdots \qquad \vdots\; y_m = \mathrm{F}(z_1, z_2, \cdots, z_s) \quad (18)$$

Both variables have lower and upper limits of the type ≤, i.e.: $z_{inf} \leq z_i \leq z_{sup}$ and $y_{inf} \leq y_i \leq y_{sup}$. The *Complex algorithm* is a local scanning technique, which is guided solely by what is in its way; its background, a brief description of its operational process and its OPTIM code, can be found in Campos–Aranda (2003). Another description and code of this search method is given in Bunday (1985).

The primary designations in the OPTIM code are NX and NY, which define the number of decision and dependent variables; for the case analyzed, three (*μ*, *σ*, *κ* and *n* (number of record years), because the dependent variables are positivity restrictions, with *j* = 1,. . ,*r* values per year $[1 + \kappa(x^j - \mu)/\sigma]$. MI = 500 is the maximum number of evaluations of the objective function and NQ = 25 is the number of such calculations between printing results. These variables are accessed in data reading subroutine.

An important advantage of the OPTIM code is to allow easy access of the limits (L = lower, U = upper), names and initial values of the variables, in the above mentioned subroutine, by means of the following designations: XL(I), XU(I), XN$(I), X(I), YL(J), YU(J), YN$(J), and Y(J). For the studied case, I varies from 1 to 3, and J from 1 to *n* (number of record years). Then the convergence criteria FA and FR for absolute and relative F deviations are included. The following values were used: 0.0002 and 0.00001 respectively.

The objective function is called F in the OPTIM code and it is accessed at the end of the program, logically corresponding to Equation (15), with name FO$ =" FLMV", of the maximum likelihood logarithmic function. It is assigned with a negative sign to F, because the Complex algorithm minimizes the function (Equation (17)) and it is desired to maximize the FLMV.

Finally, it is clarified that because of the nature of the numerical problem raised, we have another variable that is NR = 1,..., 5, which is equivalent to each annual sample of levels or flows. Data capture is done with two nested cycles, one for *i* varying from 1 *to n* or number of record years processed and the other for *j* ranging from 1 to 5, that is, *r* from Equation (15).

Wald–Wolfowitz test

This non–parametric test has been used by Bobée & Ashkar (1991), Rao & Hamed (2000) and Meylan, Favre & Musy (2012) to test *independence* and *stationarity* in annual maximum flow records (x_i). For the studied case, it was proposed to apply the test to the annual maximum flow record (r = 1), which should be a sample of random values. Wald and Wolfowitz based on the work of Anderson on the serial correlation coefficient developed such a test, whose statistic is:

$$R = \sum_{i=1}^{n-1} x_i \cdot x_{i+1} + x_n \cdot x_1 \qquad (19)$$

When the size (n) of the series or sample (x_i) is not small and its data are independent, R comes from a Normal distribution with mean and variance, given by the following expressions:

$$E[R] = \bar{R} = \frac{S_1^2 - S_2}{n-1} \qquad (20)$$

$$Var[R] = \frac{S_2^2 - S_4}{n-1} + \frac{S_1^4 - 4 \cdot S_1^2 \cdot S_2 + 4 \cdot S_1 \cdot S_3 + S_2^2 - 2 \cdot S_4}{(n-1)(n-2)} - \bar{R}^2 \qquad (21)$$

in which:

$$S_w = \sum_{i=1}^{n} x_i^w \qquad (22)$$

Finally, U is calculated, using the equation:

$$U = \frac{R - \bar{R}}{\sqrt{Var[R]}} \qquad (23)$$

The U value follows a Normal distribution with a zero mean and unit variance, and it can be used to prove the independence of the series data with a significance level α, usually 5%. In a two-tailed test, the standardized normal variable is $Z_{\alpha/2} \cong 1.96$; then, when $|U| < 1.96$ the series will be made up of independent values (random sample).

Standard error of fit

It is the most common indicator (Chai & Draxler, 2014) for the contrast of a probability distribution to real data, established in the mid–1970s (Kite, 1977) and it has been applied in Mexico using the empirical Weibull formula (Benson, 1962). It will now be applied using the Cunnane formula (Equation (25)), which according to Stedinger (2017) leads to non–exceedance probabilities (*p*) approximately unbiased with the distributions used in hydrology. The standard error of fit (*EEA*) expression is:

$$EEA = \left[\frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{(n-np)}\right]^{1/2} \quad (24)$$

X_i are the annual maximum levels and flows ($r = 1$) ordered from low to high, whose number is n and $\hat{X}_i$ the estimated maximum flows under Equation (7) and the probability evaluated with Equation (25); $np = 3$ is the number of GEV fitting parameters.

$$p = \frac{i - 0.40}{n + 0.20} \quad (25)$$

Search for the optimal and robust solution

According to the progress that is made from r = 1 to r = 5, applying Equation (15) and finding its maximum with the Complex algorithm, more and more data is being processed, first n and finally 5n. This modifies the optimal values of the fitting parameters (μ, σ, κ) and also changes the *EEA*, which is calculated *exclusively* with the annual series of levels or annual maximum flows (r = 1).

The above, in order to establish an *objective comparison*, in the use of more information of each year processed (robust solution) and look for a combination of μ, σ and κ (optimal solution) leading to an *EEA*, less than or similar to the minimum obtained with the sextile adjustment methods (Clarke, 1973; Campos–Aranda, 2001), L moments (Hoking & Wallis, 1997; Stedinger, 2017) and LH moments (Wang, 1997a,b; Campos–Aranda, 2016).

Processed data

North Sea level records

Guedes Soares and Scotto (2004) present a graph of five annual maximum levels (meters) in the North Sea, in the 1976–1999 period; therefore, $n = 24$. Their approximate values are shown in Table 1. These authors, to ensure that they use r independent annual values, use the concept of standard storm length or duration (Tawn, 1988), defined in 480 hours (20 days), when sampling a record of maximum levels taken every three hours.

Table 1. Five independent annual maximum levels (meters) in the North Sea (Guedes Soares & Scotto, 2004).

Year	***r* annual values**				
	1	**2**	**3**	**4**	**5**
1976	10.38	9.39	9.21	8.59	8.45
1977	10.60	9.57	9.03	7.80	7.69
1978	11.34	9.06	7.82	7.57	6.78
1979	10.33	10.22	9.42	8.37	7.53
1980	10.54	9.43	9.30	8.50	7.90
1981	11.22	10.13	9.44	9.24	9.17
1982	8.85	8.67	8.34	7.96	7.85
1983	10.53	10.16	9.14	8.78	7.27
1984	9.20	8.97	8.43	8.11	7.14
1985	10.68	9.40	8.89	8.44	7.60
1986	9.92	9.49	8.60	8.47	8.14
1987	9.36	8.22	7.96	7.92	7.18

1988	12.95	10.17	9.08	8.67	8.34
1989	10.32	10.03	9.80	9.14	8.66
1990	10.46	10.00	9.60	9.00	8.47
1991	9.98	9.48	8.98	8.47	8.00
1992	11.47	9.80	8.46	7.98	7.48
1993	12.00	10.98	10.40	9.48	8.98
1994	9.00	8.46	7.98	7.50	7.20
1995	11.48	10.98	10.47	9.98	7.98
1996	10.97	10.00	9.00	8.00	7.50
1997	12.00	9.33	8.47	7.66	7.48
1998	8.52	8.40	8.18	8.04	7.88
1999	8.40	7.86	7.44	7.00	6.87

Flood records integration

Five records were processed and integrated based on information available in the BANDAS system (IMTA, 2002), on CD 1, which is called "Maximum Monthly Flows" and includes the day, time and scale reading for each of the 12 maximum flows for the analyzed year. The five records to be processed from the Hydrological Region No. 10 (Sinaloa), Mexico, are presented from greater to smaller basin area. The divide and main collector of the hydrometric stations: Huites, Santa Cruz, Jaina, Guamuchil and El Bledal are shown in figure 1 of Campos–Aranda (2014), along with 17 other basins in that region.

For its integration, it was accepted that a period of 15 days, assures different meteorological conditions in each flood formation, within the Hydrological Region No. 10 (Sinaloa), Mexico (Schulz, 1976; Campos–Aranda, 2000).

Extreme event verification was not performed, because it is accepted that the hydrometric information contained in the BANDAS system (IMTA, 2002) has been checked and purged against field data or measurements.

In order to cause that the records are integrated with *independent* flows, the following process was followed, each year (i): the annual maximum flow (x_i^1) is selected and reviewed back and forth of that value, if the adjacent monthly flow is at least 15 days apart from the date of this first maximum. Monthly flow that does not meet this time frame is eliminated. The next flow in magnitude is then sought (x_i^2), among the maximum monthly flows available (11, 10 or 9 remaining) and the process described is followed. This is the way to integrate the five independent annual maximum flows (x_i^5).

Record of floods in the Huites hydrometric station

The Huites gauging station in the Fuerte river of the Hydrological Region No. 10 (Sinaloa), Mexico, with code number 10037 and basin area of 26057 km^2, started operations in September 1941 and concluded in December, 1992 (n = 51), when the construction of the Luis Donaldo Colosio Dam began. With the process described, its record of five independent annual maximum flows, as set out in Table 2, was integrated.

Table 2. Five independent annual maximum flows (m^3/s) at the Huites hydrometric station in Hydrological Region No. 10 (Sinaloa), Mexico.

Year	***r* annual values**				
	1	**2**	**3**	**4**	**5**
1942	2531.0	2 037.6	1 868.8	780.3	427.6
1943	14 376.0	3 283.0	1 085.0	416.4	414.2
1944	2 580.0	1 262.5	1 024.8	768.0	474.6
1945	1 499.2	1 250.0	268.4	197.4	24.3
1946	1 164.8	445.0	427.5	151.0	62.2
1947	1 127.3	754.8	718.8	634.0	205.0
1948	3 215.0	799.0	623.2	493.5	118.8
1949	10 000.0	2 297.5	895.3	942.4	826.4
1950	3 229.3	1 384.0	961.0	439.0	438.9
1951	677.0	587.5	322.2	37.6	37.4
1952	1 266.0	895.0	355.8	238.3	223.5
1953	1025.0	885.0	68.3	57.3	39.0
1954	954.8	540.1	481.2	406.6	171.4

1955	4 780.3	1 069.9	662.0	561.8	35.8
1956	695.7	531.7	493.2	278.6	204.8
1957	593.0	489.0	380.0	362.0	156.2
1958	3 010.0	1 045.2	894.0	608.5	307.1
1959	1 908.0	1 831.0	1 345.5	652.0	544.3
1960	15 000.0	1 046.0	985.2	721.4	140.5
1961	1 396.3	905.9	831.6	771.2	682.0
1962	1 620.0	912.0	892.8	501.0	374.5
1963	2 702.0	1 054.0	980.1	969.2	323.0
1964	1 319.1	938.5	216.0	145.2	105.0
1965	1 944.0	1 787.6	491.2	360.0	147.6
1966	2 420.0	892.2	738.0	688.9	340.0
1967	2 505.8	1 310.5	1 192.8	462.5	272.6
1968	1 534.3	1 118.0	1 019.6	703.8	401.6
1969	1 508.0	736.0	504.0	417.6	300.0
1970	1 558.0	1 330.0	970.0	328.7	206.0
1971	2 200.0	1 176.0	592.0	591.2	576.3
1972	2 225.0	2 040.0	1 142.0	1 109.0	732.1
1973	7 960.0	2 256.5	1 546.1	1 380.0	372.5
1974	3 790.0	3 315.0	1 120.0	886.7	587.5
1975	1 095.0	965.5	532.9	213.2	59.2
1976	2 677.0	1 350.2	1 211.0	69.6	50.4
1977	1 135.0	622.0	266.9	153.5	120.0

1978	4 790.0	1 750.0	1 119.2	756.0	475.4
1979	6 860.0	1 001.0	820.0	480.0	302.0
1980	1 496.0	1 475.0	1 197.8	660.2	452.0
1981	4 828.1	2 448.0	2 280.0	2 052.0	1 000.3
1982	2 450.0	1 571.6	1 085.2	624.0	412.8
1983	8 275.0	1 439.0	1 400.0	1 006.7	893.0
1984	5 580.0	1 623.0	1 132.0	624.5	369.2
1985	3 585.0	1 250.0	1 121.4	925.0	307.6
1986	1 348.8	1 329.3	436.8	434.0	193.7
1987	1 429.2	679.2	310.4	78.7	62.9
1988	1 866.3	1 494.3	132.6	82.7	47.2
1989	1 868.5	1 413.9	1 230.3	1 164.7	249.9
1990	11 558.6	3 544.2	970.1	815.6	376.9
1991	2 563.1	2 370.0	1 721.5	1 517.6	1 266.7
1992	2 025.3	1 564.8	1 348.9	787.3	528.5

Flood record in the Santa Cruz hydrometric station

The Santa Cruz gauging station in the San Lorenzo river of the Hydrological Region No. 10 (Sinaloa), Mexico, with code number 10040 and basin area of 8919 km^2, began operations in May 1943 and concluded its continuous record in December 1980 (n = 38). With the process described, its record of five independent annual maximum flows, as set out in Table 3 and Figure 1, was integrated.

Table 3. Five independent annual maximum flows (m^3/s) at the Santa Cruz hydrometric station in the Hydrological Region No. 10 (Sinaloa), Mexico.

Year	***r* annual values**				
	1	**2**	**3**	**4**	**5**
1943	2 102.9	2 067.3	1 166.0	485.0	266.2
1944	2 142.0	1 210.0	173.3	257.0	125.4
1945	1 023.4	623.8	427.4	267.7	66.8
1946	837.6	374.0	265.4	141.1	133.0
1947	1 161.2	734.0	672.8	211.5	134.6
1948	1 062.0	787.8	618.6	606.0	562.6
1949	784.2	503.6	270.4	233.8	207.7
1950	1 086.3	836.7	309.0	302.0	51.4
1951	487.8	421.0	333.4	272.0	146.0
1952	677.0	244.2	161.9	110.0	75.3
1953	807.0	480.6	385.0	122.0	92.0
1954	553.0	424.0	276.8	126.0	14.2
1955	1 252.0	895.6	768.4	426.0	26.8
1956	369.5	329.7	334.1	129.0	38.6
1957	293.0	279.0	190.6	154.0	11.6
1958	1 157.2	576.7	455.0	438.3	337.8

1959	762.2	232.1	48.0	40.4	28.0
1960	1 074.0	679.0	591.5	536.7	364.2
1961	1 280.0	765.0	587.6	521.0	112.5
1962	1 002.0	574.0	284.4	44.5	43.7
1963	3 680.0	707.4	615.1	566.7	277.0
1964	861.0	728.5	400.0	322.3	151.6
1965	888.8	346.0	187.4	175.0	55.9
1966	1 166.4	692.6	103.4	99.6	75.0
1967	950.0	651.2	603.6	252.6	225.8
1968	7 000.0	766.8	458.6	442.0	421.6
1969	484.0	438.0	365.0	323.0	192.0
1970	920.6	780.0	512.0	161.5	49.5
1971	812.0	747.0	443.3	280.6	126.0
1972	3 332.4	1 748.5	610.0	543.0	267.4
1973	898.0	890.0	800.0	90.5	73.3
1974	2 790.0	1 748.0	580.0	256.6	8.7
1975	620.0	410.0	392.5	182.0	23.8
1976	1 495.0	410.0	359.6	330.0	290.0
1977	836.0	696.0	512.9	114.6	78.0
1978	940.0	420.0	315.5	265.6	46.0
1979	3 080.0	523.3	484.0	67.9	22.0
1980	1 550.0	682.9	422.7	411.2	252.5

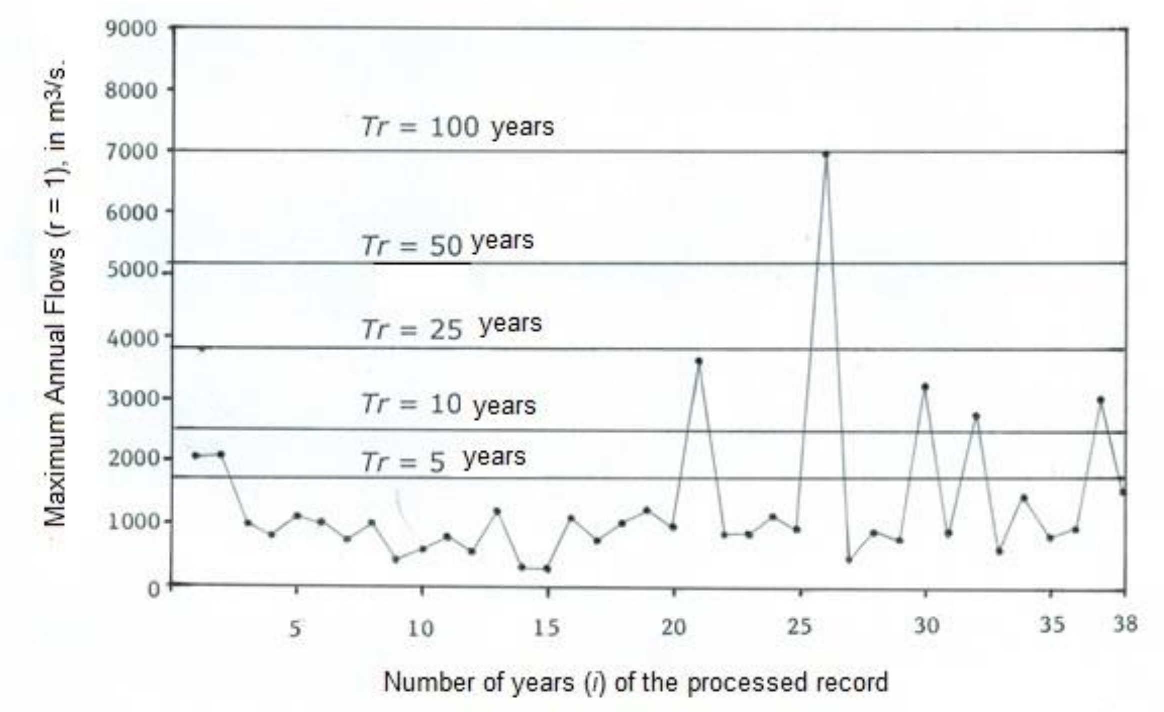

Figure 1. Series of annual maximum flows (r = 1) and stationary predictions straight lines at the Santa Cruz hydrometric station in the Hydrological Region No. 10 (Sinaloa), Mexico.

Record of floods in the Jaina hydrometric station

The Jaina gauging station in the Sinaloa river of the Hydrological Region No. 10 (Sinaloa), Mexico, with code number 10036 and basin area of 8179 km^2, started operations in January 1942 and concluded its continuous record in December 1997 (n = 56). With the process described, its record of five independent annual maximum flows, as set out in Table 4, was integrated.

Table 4. Five independent annual maximum flows (m^3/s) at the Jaina hydrometric station in Hydrological Region No. 10 (Sinaloa), Mexico.

Year	*r* annual values				
	1	2	3	4	5
1942	2 065.2	707.0	219.5	217.0	114.4
1943	6 991.3	1 715.0	1 091.4	501.9	427.0
1944	580.0	514.0	404.9	366.0	253.8
1945	714.2	501.6	312.2	208.2	83.5
1946	746.5	280.0	217.0	81.9	70.3
1947	771.2	452.3	291.5	279.0	175.8
1948	692.5	622.8	438.6	311.0	192.7
1949	2 614.0	914.0	463.0	379.4	283.5
1950	2 336.0	438.3	287.2	135.5	110.9
1951	437.0	369.8	328.2	261.2	31.1
1952	594.2	304.0	163.0	40.6	26.9
1953	545.5	346.3	173.4	19.7	12.3
1954	516.4	327.8	213.1	81.8	58.0
1955	1 600.0	1 143.0	835.0	310.0	19.0
1956	639.0	514.0	212.0	25.1	23.0
1957	362.0	314.0	180.5	42.8	15.3
1958	2 232.0	594.0	389.0	305.6	230.6
1959	615.5	586.0	370.4	359.0	176.5

1960	2 003.0	545.6	394.4	316.1	302.0
1961	795.0	719.0	439.0	351.5	307.8
1962	1 137.0	514.8	297.6	295.2	59.7
1963	1 226.0	718.1	506.8	264.9	100.5
1964	453.7	433.2	233.0	129.8	36.7
1965	649.8	325.2	144.5	42.0	35.6
1966	958.0	381.6	274.8	143.6	109.0
1967	900.0	368.8	327.6	201.9	106.0
1968	1 338.0	734.9	522.0	250.6	215.5
1969	340.0	205.0	193.0	67.1	44.3
1970	356.6	287.8	208.0	107.0	19.0
1971	1 109.0	831.9	467.0	282.0	198.0
1972	932.0	551.0	468.2	412.0	403.1
1973	1 349.0	520.0	411.6	395.0	364.5
1974	680.0	675.0	545.8	418.8	220.0
1975	488.0	164.8	164.3	35.7	17.8
1976	900.0	390.2	313.3	290.5	216.5
1977	790.7	229.4	195.2	188.0	74.4
1978	988.0	521.0	315.2	26.9	26.0
1979	1 620.0	309.0	300.5	239.8	115.6
1980	400.3	321.2	132.5	121.2	103.5
1981	2 831.7	452.1	342.0	233.4	316.9
1982	4 440.4	325.0	105.0	66.5	51.6

1983	178.9	152.8	108.5	89.3	80.2
1984	693.7	586.4	451.2	315.6	220.5
1985	493.9	468.9	152.4	127.2	123.0
1986	416.2	276.7	114.4	88.9	78.7
1987	518.2	86.4	64.2	62.5	61.7
1988	105.0	63.7	36.7	35.8	26.7
1989	227.4	140.1	85.9	75.7	68.4
1990	638.2	412.2	286.9	181.8	86.9
1991	308.9	292.0	138.6	129.1	108.6
1992	371.8	143.7	125.3	114.9	92.1
1993	216.2	90.6	80.6	78.6	77.1
1994	199.0	119.0	91.9	85.2	80.5
1995	173.6	117.4	98.8	85.5	75.5
1996	343.6	336.8	89.9	77.1	57.6
1997	169.1	116.2	86.5	74.1	66.4

Record of floods in the Guamuchil hydrometric station

The Guamuchil gauging station in the Mocorito river of the Hydrological Region No. 10 (Sinaloa), Mexico, with code number 10031 and basin area of 1645 km^2, began operations in October 1938 and concluded in December 1971 (n = 33), when construction of the Eustaquio Buelna Dam began. With the process described, its record of five independent annual maximum flows, as set out in Table 5, was integrated.

Table 5. Five independent annual maximum flows (m^3/s) at the Guamuchil hydrometric station in Hydrological Region No. 10 (Sinaloa), Mexico.

Year	***r* annual values**				
	1	**2**	**3**	**4**	**5**
1939	299.0	299.0	164.4	78.0	12.2
1940	254.5	37.2	34.1	31.9	3.2
1941	65.3	36.1	13.5	3.0	2.1
1942	445.0	298.0	161.0	83.0	4.9
1943	1 550.0	1 236.4	298.0	92.3	71.5
1944	391.8	71.4	22.1	7.9	2.7
1945	916.0	336.0	276.0	228.7	1.7
1946	241.0	48.4	12.7	0.3	0.3
1947	530.0	133.0	26.0	15.5	0.9
1948	648.0	548.0	195.9	40.4	2.8
1949	375.0	145.4	89.2	72.8	34.0
1950	272.3	74.2	69.3	1.4	0.2
1951	422.3	409.7	82.8	13.8	2.3
1952	376.8	26.8	16.7	0.2	0.1

1953	1173.0	261.3	213.0	3.0	1.1
1954	219.0	115.4	101.2	24.9	0.4
1955	3507.0	189.0	11.4	2.3	1.1
1956	165.0	148.2	76.4	0.4	0.2
1957	526.0	342.0	57.4	8.2	0.4
1958	1 014.0	534.0	221.0	168.0	2.3
1959	1 610.0	374.0	372.8	22.4	2.7
1960	137.0	130.0	94.5	69.3	49.4
1961	524.5	302.0	134.1	28.5	5.4
1962	985.0	524.0	168.0	112.5	4.9
1963	459.5	311.2	211.5	90.7	22.7
1964	390.0	202.1	123.2	4.4	1.1
1965	449.0	382.5	3.1	1.4	0.1
1966	793.9	687.8	3.4	3.2	0.6
1967	719.5	325.0	105.7	10.6	6.8
1968	200.0	146.2	132.5	24.8	9.3
1969	312.0	126.6	84.7	30.3	5.3
1970	520.0	295.0	256.0	2.7	0.3
1971	1 045.0	790.0	175.0	52.8	25.0

Record of floods in the El Bledal hydrometric station

The El Bledal gauging station in the same name stream of the Hydrological Region No. 10 (Sinaloa), Mexico, with code number 10027 and basin area of 371 km^2, began operations in September 1937 and concluded in December 1994 (n = 57). With the process described, its record of five independent annual maximum flows, as set out in Table 6, was integrated.

Table 6. Five independent annual maximum flows (m^3/s) at the El Bledal hydrometric station in Hydrological Region No. 10 (Sinaloa), Mexico.

Year	*r* annual values				
	1	2	3	4	5
1938	766.1	120.8	89.8	77.7	2.9
1939	597.3	204.0	16.5	13.0	11.9
1940	170.0	28.0	19.0	8.4	1.3
1941	118.6	17.1	10.1	4.8	3.6
1942	64.6	33.8	6.4	0.3	0.2
1943	157.2	150.5	40.9	32.3	12.3
1944	197.0	46.2	1.0	0.8	0.5
1945	414.0	253.2	144.0	21.4	0.9
1946	291.0	95.3	14.0	7.1	0.3
1947	174.0	8.4	0.9	0.5	0.5
1948	233.0	141.2	106.0	3.5	0.8

1949	155.2	25.1	15.8	1.4	1.0
1950	44.1	40.4	22.8	1.8	0.2
1951	227.4	68.0	29.6	26.7	2.5
1952	169.0	15.4	15.4	0.8	0.2
1953	668.9	334.0	155.0	12.5	0.5
1954	30.7	15.4	8.1	1.9	0.5
1955	152.0	79.8	3.9	3.7	0.3
1956	91.4	51.3	48.1	1.0	0.3
1957	283.0	30.5	28.3	0.3	0.2
1958	433.0	121.0	95.4	30.3	3.0
1959	132.4	59.3	2.4	1.1	0.3
1960	529.0	161.6	20.0	18.6	10.3
1961	122.0	92.3	3.8	1.3	0.5
1962	1 000.0	108.5	56.5	34.7	0.6
1963	335.0	123.0	61.3	0.4	0.2
1964	258.0	189.8	59.0	12.8	0.6
1965	91.5	56.8	52.2	1.3	0.4
1966	121.8	98.6	95.9	1.5	0.2
1967	325.6	115.4	92.7	46.5	8.6
1968	1 576.0	190.4	70.6	6.1	2.9
1969	228.0	42.0	24.4	0.7	0.3
1970	82.0	37.0	0.6	0.2	0.1
1971	276.0	260.0	191.0	4.3	3.1

1972	380.0	112.0	68.6	20.6	0.5
1973	296.0	2.2	0.6	0.4	0.2
1974	256.1	98.0	64.2	46.9	25.0
1975	490.0	73.4	30.6	1.1	0.4
1976	59.6	25.0	25.0	18.4	6.1
1977	123.0	61.3	4.0	0.2	0.1
1978	255.0	21.7	5.6	0.1	0.1
1979	283.0	52.3	10.8	0.2	0.1
1980	226.0	211.1	42.6	1.3	0.2
1981	210.0	118.0	79.8	1.4	0.3
1982	67.0	62.5	18.0	11.2	4.5
1983	50.7	47.6	32.5	13.4	10.8
1984	160.0	153.1	105.2	90.5	70.5
1985	278.0	240.0	95.5	86.5	3.2
1986	723.5	25.6	15.1	13.5	1.9
1987	107.2	92.4	11.2	2.4	0.3
1988	234.3	102.5	80.6	2.8	1.6
1989	424.9	197.0	36.0	31.7	1.1
1990	459.3	332.4	86.2	18.1	2.2
1991	73.8	58.2	35.2	10.3	4.6
1992	181.1	161.9	77.6	12.7	0.9
1993	334.0	127.9	70.4	10.2	0.2
1994	278.2	143.3	101.0	3.3	0.9

Record of levels with trend in Venice

Table 7 and Figure 2 have the first five annual sea maximum levels, out of the 10 that Smith (1986) set out, for Venice, Italy.

Table 7. Record of five annual maximum levels (meters) for 51 years in Venice, Italy (Smith, 1986).

Year	*r* annual values					Year	*r* annual values				
	1	2	3	4	5		1	2	3	4	5
1931	103	99	98	96	94	1957	119	107	100	98	98
1932	78	78	74	73	73	1958	124	114	113	110	108
1933	121	113	106	105	102	1959	118	117	108	107	105
1934	116	113	91	91	91	1960	145	126	123	116	114
1935	115	107	105	101	93	1961	122	108	104	100	100
1936	147	106	93	90	87	1962	114	110	108	107	106
1937	119	107	107	106	105	1963	118	116	114	112	110
1938	114	97	85	83	82	1964	107	104	104	103	102
1939	89	86	82	81	80	1965	110	108	106	102	101
1940	102	101	98	97	96	1966	194	127	126	104	103
1941	99	98	96	95	94	1967	138	118	118	107	100
1942	91	91	87	83	83	1968	144	132	123	114	112

1943	97	88	82	79	78	1969	138	120	116	114	108
1944	106	96	94	90	89	1970	123	122	119	110	105
1945	105	102	98	88	86	1971	122	116	116	109	104
1946	136	104	103	101	100	1972	120	118	113	111	96
1947	126	108	101	99	98	1973	114	111	99	98	97
1948	132	126	119	107	101	1974	96	95	95	93	92
1949	104	102	102	101	93	1975	125	110	109	103	102
1950	117	96	91	89	88	1976	124	122	114	109	108
1951	151	117	114	109	106	1977	120	102	100	98	96
1952	116	104	103	98	91	1978	132	114	110	107	105
1953	107	102	98	98	92	1979	166	140	131	130	122
1954	112	100	95	94	94	1980	134	114	111	109	107
1955	97	96	96	95	94	1981	138	136	130	128	119
1956	95	91	90	85	85	–	–	–	–	–	–

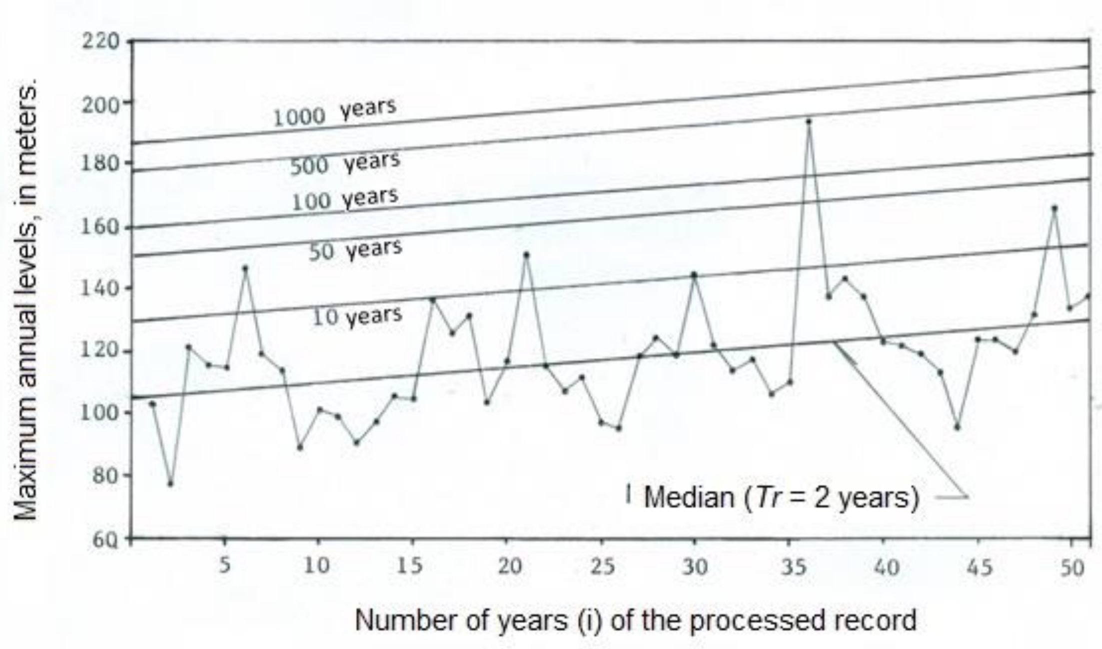

Figure 2. Series of annual maximum levels (*r* = 1) and prediction straight lines with linear trend and covariate time (*t*), in Venice, Italy.

Results

Randomization test

The Wald–Wolfowitz test results, defined in equations (19) to (23), for each of six processed records (Table 1, Table 2, Table 3, Table 4, Table 5 and Table 6, columns 2), are as follows: North Sea U = –0.518; Huites U = –0.113; Santa Cruz U = –0.827; Jaina U = 1.510; Guamuchil U = –1.414; El Bledal U = 0.224, and Venice U = 2.641. Therefore, the first six records of annual maximum values (r = 1) are random, and that of Venice is not.

Fittings to annual series (r = 1)

Table 8 has concentrated the results of the GEV distribution fittings, to each of the annual series of maximum values (*r* = 1), by means of sextile, L–moment and LH–moment methods. These results are the fitting parameters (*μ*, *σ*, *κ*) and the standard error of fit (*EEA*), whose minimum value reached in each series is indicated in parenthesis.

Table 8. Standard error of fit (*EEA*) and optimal parameters of the annual series (*r* = 1), in the six records indicated with the three mentioned methods.

Record:	***EEA***	**Sextiles**		
		μ	***σ***	***κ***
North Sea	0.189	10.048	1.139	-0.3075
Huites	991.2	1 724.345	1 099.256	0.4778
Santa Cruz	305.3	814.822	507.620	0.3495
Jaina	208.5	514.672	390.172	0.4192
Guamuchil	177.2	373.026	260.903	0.3363
El Bledal	36.3	169.232	118.795	0.3062
Venice	3.984	111.438	16.966	-0.1102
Record:	***EEA***	**L Moments**		
		μ	***σ***	***κ***
North Sea	(0.184)	10.034	1.187	-0.3040
Huites	930.0	1755.541	1178.274	0.4459
Santa Cruz	282.2	826.631	402.385	0.4544
Jaina	(190.8)	516.752	381.623	0.4353

Guamuchil	153.4	362.642	250.526	0.3803
El Bledal	33.4	168.181	123.130	0.2982
Venice	3.636	110.994	16.854	-0.0764
Record:	***EEA***	**LH Moments**		
		μ	***σ***	***κ***
North Sea (L1)	0.186	10.038	1.106	-0.2391
Huites (L2)	(830.3)	1 509.511	1 598.497	0.3264
Santa Cruz (L3)	(275.3)	743.090	473.732	0.4044
Jaina (L1)	191.6	514.368	387.161	0.4296
Guamuchil (L3)	(152.7)	361.325	249.637	0.3827
El Bledal (L4)	(32.0)	171.392	116.514	0.3222
Venice (L1)	(3.587)	111.690	14.241	0.0405

Guidelines for Complex algorithm application

At the beginning, we sought to define limits and initial values of fitting variables (μ, σ, κ) and dependents or constraints on positivity, which would work for the five series to be processed (r = 1,..., 5) of each record.

During the first applications of the OPTIM code, several difficulties arose, one of which was associated with the limits of the variables, which are of the ≤ type. Then for the lower limits you should not set zero, but 0.10 for the location (μ) and scale (σ) parameters and 0.01 for the dependent variables.

With regard to upper limits, it was found advisable to use at least twice the calculated values for the GEV parameters μ and σ, of the annual maximum flow series (r = 1), by the method leading to the minimum *EEA* (Table 8). Upper limits, it was never necessary to change them for the OPTIM code to work. A higher value of 100 was adopted for dependent variables.

In relation with the limits of the shape parameter (κ), they were generally used, as lower −0.50 and higher 0.50; or −1 and 1. Rarely, its limit had to be restricted with the aim of obtain results in the order of magnitude of the *EEA* of the first processed record (r = 1).

Finally, with respect to the initial values of the fitting parameters (μ, σ, κ), in general the values near the best GEV fitting worked well (Table 8), between sextile, L and LH moments with the series of annual maximum flows (r = 1). Table 9 shows the initial limits and values used to process the 30 annual maximum values series.

Table 9. Limits and initial values adopted for the fitting parameters (μ, σ, κ) in the application of the Complex algorithm.

Record:	**μ**			**σ**			**κ**		
	low	**upp**	**init**	**Low**	**upp**	**init**	**low**	**upp**	**init**

North Sea	0.10	100	15	0.10	100	5	-0.50	0.50	-0.25
Huites	0.10	4000	1500	0.10	3000	1000	-1.00	1.00	0.45
Santa Cruz	0.10	2000	900	0.10	1000	500	-0.50	0.50	0.35
Jaina	0.10	1000	500	0.10	800	400	-0.50	0.50	0.45
Guamuchil	0.10	800	300	0.10	500	200	-0.50	0.50	0.30
El Bledal	0.10	300	150	0.10	200	100	-0.40	0.40	0.35

The Venice record was processed by replacing Equation (16) in Equation (15), now with four fitting parameters. The lower and upper limits and the initial value were as follows: β_0(0.10,200,100), β_1(0.10,2,0.75), σ(0.10,30,15) and κ(−0.50,0,−0.05).

When processing some series of certain records, an initial data had to be changed so that the Complex algorithm reached a solution close to the optimal one and not to a local minimum. For example, in the North Sea record, for the series of r = 3, −0.05 was used for the initial κ. For El Bledal record, the shape parameter value κ was limited to 0.40.

Results of the Complex algorithm

Table 10 shows the main indicators of the Complex algorithm (aC) application, such as: The maximum likelihood logarithmic function (FLMV), calculated with Equation (15), the standard error of fit (*EEA*), evaluated with Equation (23) and the optimum values of the fitting parameters (μ, σ, κ).

Table 10. Results of the Complex algorithm, with each annual series of *r* maximum flows in the indicated records.

1	2	3	4	5	6	7	8	9
Record:	***r***	**FLMV Initial**	**FLMV final**	**No. Eval.**	***EEA* aC**	**Fitting parameters**		
						μ	**σ**	**κ**
North Sea $EEA_{mín}$ = 0.184	1	79.2	36.9	307	0.184	10.037	1.124	-0.2712
	2	108.8	50.7	176	0.225	10.024	0.995	-0.2001
	3	123.4	60.5	154	0.211	10.007	1.029	-0.2164
	4	158.4	66.7	164	0.196	9.993	1.071	-0.2300
	5	181.3	75.2	198	0.190	10.000	1.118	-0.2753
Huites $EEA_{mín}$ = 830.3	1	455.1	453.3	96	1363.0	1690.25	1081.08	0.5713
	2	829.5	825.9	96	1186.1	1666.73	1054.85	0.5490
	3	1185.8	1182.5	98	971.5	1759.61	1193.36	0.3971
	4	1515.7	1508.9	93	1040.4	1781.62	1196.44	0.3687
	5	1839.9	1825.6	94	900.3	1746.67	1267.03	0.4372
Santa Cruz $EEA_{mín}$ = 275.3	1	300.9	300.6	111	349.5	856.806	457.724	0.3632
	2	552.8	552.1	132	267.3	819.749	463.835	0.4040
	3	793.5	787.3	122	408.0	830.896	496.459	0.3006
	4	1033.3	1009.9	116	364.9	815.842	518.103	0.3090
	5	1295.4	1226.4	128	470.6	860.041	539.006	0.2255
Jaina $EEA_{mín}$ = 190.8	1	436.1	435.6	127	100.4	505.486	372.456	0.4894
	2	788.2	784.3	85	230.9	550.876	377.100	0.4194
	3	1106.2	1098.4	120	216.9	550.285	376.230	0.4266
	4	1403.2	1394.7	115	238.3	547.134	380.289	0.4131

	5	1676.0	1663.4	117	134.9	527.104	381.679	0.4593
Guamuchil	1	243.6	241.9	118	153.7	370.294	254.203	0.3724
$EEA_{mín}$ = 152.7	2	448.2	444.9	145	138.0	375.509	281.706	0.3518
	3	630.5	625.2	103	152.6	338.938	267.647	0.3607
	4	802.1	787.2	213	119.3	292.374	259.547	0.4991
	5	935.1	917.7	158	181.7	240.682	211.943	0.4994
El Bledal	1	374.3	373.5	99	16.6	163.134	118.291	0.3673
$EEA_{mín}$ = 32.0	2	684.1	676.5	110	33.8	144.617	116.827	0.3473
	3	951.4	931.5	167	42.0	131.231	108.428	0.3998
	4	1202.2	1159.1	167	71.2	115.991	99.014	0.4000
	5	1367.1	1326.9	144	140.1	93.270	76.361	0.3999

It is also cited in column 1 of Table 10, the minimum *EEA* obtained (Table 8) with the annual maximum series (r = 1), with one of the three methods applied: sextile, L moments and LH moments. Against this value, the values obtained by the Complex algorithm, shown in column 6 (robust solution), must be compared. The adopted *EEA* values are shown shaded in Table 10, are the minimum values achieved by the Complex algorithm.

The results for the five series of the Venice, Italy record have been concentrated in Table 11; in which it is observed that the best fitting was achieved with r = 2 and an *EEA* value close to the minimum of the LH-moment method of 3.587 meters (Table 8).

Table 11. Results of the Complex algorithm, with each annual series of r maximum levels in the Venice, Italy record, considering linear trend.

1	2	3	4	5	6	7	8	9
r	**FLMV Initial**	**FLMV final**	**No. Eval.**	***EEA* aC**	**Fitting parameters**			
					β_0	β_1	σ	κ
1	224.6	216.1	205	7.444	96.902	0.562	14.490	-0.0276
2	376.0	367.3	188	5.381	101.175	0.476	12.960	-0.0117
3	501.1	494.5	164	6.207	101.419	0.535	12.758	-0.0467
4	616.3	605.6	261	5.983	103.712	0.478	12.539	-0.0459
5	722.2	705.6	244	5.761	104.281	0.456	12.327	-0.0337

Discussion

EEA contrasts

Based on the results in Table 10, the following is deducted: (1) for the North Sea record, that the maximum likelihood method with *r* = 1, equals the minimum *EEA*, achieved with the L–moment method. (2) in the records of the Huites and Santa Cruz stations, the best option is obtained with *r* = 5, in the first one, and *r* = 2, in the second one. In Santa Cruz, an *EEA* less than the minimum (L3) of the LH–moment method is obtained. (3) the following are defined as best options in the records of the stations Jaina, Guamuchil and El Bledal: *r* = 1, *r* = 4 and *r* = 1, leading to an *EEA* less than the minimum obtained with the L–

moment method in Jaina, and the LH-moment method in the other two stations.

Predictions of the Complex algorithm

Selecting, in each of the six records processed, the *r* series that led to the smaller *EEAs*, the predictions obtained and set out in Table 12 are contrasted against those of the method that led to the smaller *EEAs* in Table 8, in order to establish which ones will be adopted, because they are more severe or critical; they appear shaded.

Table 12. Predictions for the indicated return periods, with each annual series of *r* independent maximum flows in the indicated records.

Record:	**Met.**	**Return periods (*Tr*), in years**						
	r	**5**	**10**	**25**	**50**	**100**	**500**	**1000**
North Sea	ml	11.46	11.97	12.46	12.75	12.97	13.35	13.46
(meters)	1	11.42	11.93	12.44	12.74	12.99	13.41	13.54
	2	11.31	11.83	12.38	12.72	13.02	13.56	13.75
	3	11.33	11.84	12.38	12.72	13.01	13.52	13.70
	4	11.35	11.87	12.42	12.75	13.03	13.53	13.70
	5	11.37	11.88	12.35	12.67	12.92	13.33	13.46
Huites	L2	4603	6821	10523	14113	18593	33829	43286

(m^3/s)	1	4256	6642	11562	17380	25997	65651	97674
	2	4123	6355	10869	16112	23758	57976	84967
	3	4206	6099	9457	12906	17428	34194	45434
	4	4178	5976	9089	12213	16228	30607	39953
	5	4432	6600	10581	14805	20501	42685	58214
Santa Cruz (m^3/s)	L3	1720	2482	3842	5247	7098	14024	18704
	1	1770	2450	3624	4796	6297	11637	15087
	2	1776	2521	3851	5225	7035	13801	18370
	3	1772	2428	3499	4516	5762	9870	12349
	4	1804	2500	3644	4738	6086	10575	13308
	5	1822	2440	3387	4232	5214	8174	9817
Jaina (m^3/s)	mL	1324	1975	3168	4431	6133	12745	17364
	1	1330	2034	3386	4882	6975	15670	22108
	2	1338	1962	3091	4271	5843	11832	15945
	3	1341	1972	3120	4328	5945	12162	16465
	4	1337	1959	3077	4240	5783	11615	15593
	5	1351	2032	3307	4684	6569	14116	19526
Guamuchil (m^3/s)	L3	867	1252	1928	2613	3503	6744	8883
	1	881	1266	1934	2607	3474	6592	8628
	2	932	1342	2042	2735	3615	6702	8672
	3	872	1268	1949	2628	3497	6576	8560
	4	872	1371	2339	3418	4938	11329	16109
	5	714	1122	1913	2795	4037	9264	13175
El Bledal (m^3/s)	L4	396	557	823	1081	1402	2487	3158
	1	400	577	884	1191	1586	2997	3912
	2	375	543	830	1112	1470	2719	3512
	3	354	527	834	1151	1566	3112	4151

	4	320	477	758	1047	1427	2840	3790
	5	250	372	589	811	1104	2194	2927
Venice	L1	133.7	145.2	160.3	171.9	183.7	212.3	225.2
(meters)	1	146.9	157.2	169.9	179.2	188.2	208.3	216.7
	2	144.7	154.2	166.1	174.9	183.5	203.1	211.5
	3	147.2	156.0	166.6	174.2	181.5	197.5	204.0
	4	146.2	154.9	165.4	172.9	180.1	195.9	202.3
	5	145.5	154.2	164.9	172.6	180.0	196.6	203.5

It is clarified that the predictions shown for the five series in Venice, Italy, correspond to the end of the historical period, i.e., using the covariate t with a value of 51, in Equation (16), when applied in Equation (7).

It is clarified that due to the extent shown ($33 \leq n \leq 57$) by the five records of floods processed for the Hydrological Region No. 10 (Sinaloa), Mexico, their reliable predictions may span up to the 100-year return period (*Tr*); however, those of *Tr* = 500 and 1 000 years are presented, to observe the degree of dispersion that such robust predictions show at both extreme recurrence intervals.

For the Santa Cruz record, in return periods (*Tr*) greater than 50 years, slightly larger predictions are shown in Table 12 with the LH moments method, than with robust solutions, of the annual maximum *r* events method. The same happens in the Venice record, but only in the return periods (*Tr*) of 500 and 1000 years.

Moreover, for the records of Huites (r = 5), Santa Cruz (r = 2) and Guamuchil (r = 4), more severe or critical predictions were obtained with the fitting of the GEV and the method of *r* annual events, than those achieved with classical methods using a maximum annual value. The above is remarkable in the Guamuchil record.

Finally, in the remaining records (North Sea, Jaina and El Bledal) with r = 1, the maximum likelihood method equals or reduces the *EEA*, achieved by the classical methods of L and LH moments. At the Jaina station, it is feasible to adopt the results of the series with r = 5, if a robust solution is desired, with a slightly larger *EEA*.

Conclusions

Fitting the GEV distribution, using the maximum likelihood method and taking r annual maximum events, is a procedure that allows to use more information on the observed extreme values, and therefore makes more reliable *predictions*, as its fitting parameters are being estimated more accurately. To summarize, it leads to a robust solution, from a statistical point of view.

The seven numerical applications described showed that maximizing the logarithmic likelihood function, using the Complex algorithm, did not present computational difficulties. It also finds optimal solutions that improve the achieved fitting with the classic L– and LH–moment methods, which use only the annual maximum value. This was the case, in four of the five flood records processed (Table 10).

Based on the latest numerical application, relative to the five maximum sea levels in Venice, Italy, the simplicity of processing non–stationary records, which present a linear trend, is observed. This is due to the ease of increasing decision variables, changing their limits and/or initial values and modifying the objective function in the Complex algorithm.

With regard to predictions, which varied from a return period of 5 to one of 1000 years, solutions achieved with the Complex algorithm generally reported larger or more critical values. This was remarkable in the records of Guamuchil, Jaina and El Bledal. The exception was the record of the Santa Cruz station, in the high periods of return, for which slightly lower predictions were obtained (Table 12).

Taking into account that it does not involve great difficulty to integrate records of independent floods with five maximum values per year, and that the implementation of the Complex algorithm is simple, it is recommended to systematically apply the fitting of the GEV distribution, with *r* annual events in the analysis of floods frequencies.

Acknowledgements

The constructive comments of the anonymous referees A and B are appreciated, since they allowed several spelling errors to be corrected, and two explanations regarding the data processed and the predictions made were included.

References

An, Y., & Pandey, M. D. (2007). The *r* largest order statistics model for extreme wind speed estimation. *Journal of Wind Engineering and Industrial Aerodynamics*, 95(3), 165-182.

Benson, M. A. (1962). Plotting positions and economics of engineering planning. *Journal of Hydraulics Division*, 88(6), 57-71.

Bobée, B., & Ashkar, F. (1991). Chapter 1: Data requirements for hydrologic frequency analysis. In: *The Gamma Family and derived distributions applied in Hydrology* (pp. 1-12). Littleton, USA: Water Resources Publications.

Box, M. J. (1965). A new method of constrained optimization and a comparison with other methods. *Computer Journal*, 8(1), 42-52.

Bunday, B. D. (1985). Theme 6.2: The Complex method. In: *Basic Optimisation Methods* (pp. 98-106). London, UK: Edward Arnold publishers, Ltd.

Campos-Aranda, D. F. (2000). Descripción y aplicación del modelo Poisson-Pareto para el análisis de frecuencia de crecientes con series de duración parcial. *Ingeniería. Investigación y Tecnología*, 1(5), 199-207.

Campos-Aranda, D. F. (2001). Contraste de cinco métodos de ajuste de la distribución GVE en 31 registros históricos de eventos máximos anuales. *Ingeniería Hidráulica en México*, 16(2), 77-92.

Campos-Aranda, D. F. (2003). Capítulo 9: Optimización Numérica. En: *Introducción a los métodos numéricos: software en Basic y aplicaciones en hidrología superficial* (pp. 172-211). San Luis Potosí, México: Editorial Universitaria Potosina.

Campos-Aranda (2014). Análisis regional de frecuencia de crecientes en la Región Hidrológica No. 10 (Sinaloa), México. 1: índices de estacionalidad y regiones de influencia. *Agrociencia*, 48(2), 147-158.

Campos-Aranda, D. F. (2016). Ajuste de las distribuciones GVE, LOG y PAG con momentos L de orden mayor. *Ingeniería. Investigación y Tecnología*, 17(1), 131-142.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7, 1247-1250.

Clarke, R. T. (1973). Chapter 5: The estimation of floods with given return period. In: *Mathematical Models in Hydrology* (pp. 130-146) (Irrigation and Drainage Paper 19). Rome, Italy: Food and Agriculture Organization.

Coles, S. (2001). Chapter 2: Basics of Statistical Modeling, chapter 3: Classical Extreme Value Theory and Models, and chapter 6: Extremes of Non-stationary Sequences In: *An introduction to Statistical Modeling of Extreme Values* (pp. 18-44, 45-73, 105-123). London, UK: Springer-Verlag.

Dupuis, D. J. (1997). Extreme value theory based on the *r* largest annual events: a robust approach. *Journal of Hydrology*, 200(1-4), 295-306.

Guedes-Soares, C., & Scotto, M. G. (2004). Application of the *r* largest-order statistics for long-term predictions of significant wave height. *Coastal Engineering*, 51(5-6), 387-394.

Hosking, J. R., & Wallis, J. R. (1997). Appendix: *L*-moments for some specific distributions. In: *Regional Frequency Analysis. An approach based on L-moments* (pp. 191-209). Cambridge, UK: Cambridge University Press.

IMTA, Instituto Mexicano de Tecnología del Agua. (2002). *Banco Nacional de Datos de Aguas Superficiales* (*BANDAS*). 8 CD's. Jiutepec, México: Comisión Nacional del Agua-Secretaría de Medio Ambiente y Recursos Naturales-Instituto Mexicano de Tecnología del Agua.

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8-12), 1287-1304.

Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J. C., Gachon, P., & Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, 329(3-4), 534-552.

Kite, G. W. (1977). Chapter 3: Frequency distributions and chapter 12: Comparison of frequency distributions. In: *Frequency and Risk Analyses in Hydrology* (pp. 27-39, 156-168). Fort Collins, USA: Water Resources Publications.

Kottegoda, N. T., & Rosso, R. (2008). Chapter 3: Random variables and their properties. In: *Applied Statistics for Civil and Environmental Engineers* (pp. 83-164). Oxford, UK: Blackwell Publishing Ltd.

Meylan, P., Favre, A. C., & Musy, A. (2012). Chapter 3: Selecting and checking data series and chapter 5: Estimation of model parameters. In: *Predictive Hydrology. A Frequency Analysis Approach* (pp. 29-70, 91-102). Boca Raton, USA: CRC Press.

Ramesh, N. I., & Davison, A. C. (2002). Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology*, 256(1-2), 106-119.

Rao, A. R., & Hamed, K. H. (2000). Theme 1.8: Tests on hydrologic data and chapter 4: Parameter and quantile estimation. In: *Flood Frequency Analysis*. (pp. 12-21, 73-82). Boca Raton, USA: CRC Press.

Schulz, E. F. (1976). Section 9: Probability Analysis. In: *Problems in Applied Hydrology* (pp. 411-457). Fort Collins, USA: Water Resources Publications.

Smith, R. L. (1986). Extreme value theory based on the *r* largest annual events. *Journal of Hydrology*, 86(1-2), 27-43.

Stedinger, J. R. (2017). Flood frequency analysis. In: Singh, V. P. (ed.). *Handbook of Applied Hydrology*, 2nd ed. (pp. 76.1-76.8). New York, USA: McGraw-Hill Education.

Tawn, J. A. (1988). An extreme-value theory model for dependent observations. *Journal of Hydrology*, 101(1-4), 227-250.

Wang, Q. J. (1997a). Using higher probability weighted moments for flood frequency analysis. *Journal of Hydrology*, 194(1-4), 95-106.

Wang, Q. J. (1997b). LH moments for statistical analysis of extreme events. *Water Resources Research*, 33(12), 2841-2848.