DOI: 10.24850/j-tyca-13-06-11

Articles

# Application of the bivariate GEV distribution in the joint flood frequency analysis

# Aplicación de la distribución GVE bivariada en el análisis de frecuencias conjunto de crecientes

Daniel Francisco Campos-Aranda[1], ORCID: https://orcid.org/0000-0002-9876-3967

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

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

**Abstract**

The floods in our country every year cause damage and endanger the reservoirs. Therefore, its hydrological dimensioning is based on the *hydrograph of the design flood*, and its most straightforward estimation is based on the joint frequency analysis of the annual peak flow and volume. In this study, the *bivariate general extreme values distribution* (*GVEb*) was adjusted to the record of the 55 annual floods at the La Cuña hydrometric station on the Río Verde of Hydrological Region No. 12-3, Mexico. This study encompasses the following nine stages: (1) selection and testing of annual records; (2) verification of the randomness of the records; (3) estimation of the joint empirical probabilities; (4) adjustment of the *GVEb* function through the maximum likelihood method; (5) validation of the *GVEb* function; (6) ratification of GVE marginal functions; (7) verification of probability constraints; (8) estimation of hybrid univariate design events, and (9) estimation of joint design events and selection of the critical subgroup. In stage 1, a simple test is applied based on the shape parameter of the marginal GVE. Stage 2 is carried out based on the Wald–Wolfowitz Test. In stage 4, the Complex algorithm is used. Stages 5 and 6 use the Kolmogorov-Smirnov Test. In stage 9, the graphs of the joint return period of type AND are used. Finally, conclusions are formulated, which highlight the maximization approach adopted and the advantages of the bivariate joint frequency analysis through the *GVEb*.

**Keywords**: Types of design floods, *GVEb* distribution, conditional distributions, joint empirical probabilities, complex algorithm, *GVEb* function validation, hybrid univariate return periods, joint return periods.

## Resumen

Las crecientes que ocurren en nuestro país cada año generan daños y ponen en peligro a los embalses, cuyo dimensionamiento hidrológico está basado en el *hidrograma de la creciente de diseño.* La estimación más simple de tal hidrograma se basa en el análisis de frecuencias conjunto del gasto pico y volumen anuales. En este estudio se ajustó la *distribución general de valores extremos bivariada* (*GVEb*), al registro de 55 crecientes anuales en la estación hidrométrica La Cuña, sobre el Río Verde de la Región Hidrológica No. 12-3, México. Este proceso abarca nueve etapas: (1) selección y prueba de los registros anuales; (2) verificación de su aleatoriedad; (3) estimación de las probabilidades empíricas conjuntas; (4) ajuste de la función *GVEb* a través del método de máxima verosimilitud; (5) validación de la función *GVEb*; (6) ratificación de las marginales GVE; (7) verificación de las restricciones de probabilidad; (8) estimación de eventos de diseño univariados híbridos, y (9) estimación de eventos de diseño conjuntos y selección del subgrupo crítico. En la etapa 1 se aplica un test simple de la GVE. La etapa 2 se realiza con base en el Test de Wald-Wolfowitz. En la etapa 4 se emplea el algoritmo Complex. Las etapas 5 y 6 utilizan el Test de Kolmogorov–Smirnov. En la etapa 9 se usan las gráficas del periodo de retorno conjunto de tipo AND. Por último, se formulan las conclusiones, las cuales destacan el enfoque de maximización adoptado y las ventajas de aplicar la *GVEb*.

**Palabras clave**: tipos de crecientes de diseño, distribución *GVEb*, distribuciones condicionales, probabilidades empíricas conjuntas, algoritmo Complex, validación de la función *GVEb*, periodos de retorno univariados híbridos, periodos de retorno conjuntos.

Received: 03/02/2021

Accepted: 16/08/2021

# Introduction

## Types of hydraulic works

In our country, due to its geographical location, severe *floods* occur annually in various regions, which generate social, environmental, and material damage. Such floods risk all *Hydraulic Works* for use and protection (Aldama, 2000). The most crucial water use works are *reservoirs*, classified from a simple hydrological point of view into: large and small. Several large reservoirs operate on a multi-year basis and have various uses. Due to their size, their rupture or collapse can create a catastrophe of such magnitude that cannot be allowed to happen. On the other hand, thousands of small reservoirs operate on an annual basis with a single purpose, commonly irrigation. Their rupture or collapse causes great destruction at the local level and must be avoided.

The main hydraulic protection works are the *levee* or *retaining walls*, which prevent the river from overflowing in its flood plains, allowing the agricultural or recreational use of these areas. *Bridges*, although they are road and rail crossing works, their correct hydrological dimensioning prevents their destruction or superficial or lateral damage to the road or rail embankment due to their reduced size.

Hydraulic protection works include all *urban drainage structures*, *flood control dams,* and *peak break dams*. The former structures store the designed flood temporarily and release it at little flow; the latter reduces the flood due to its transit in their reservoir or storage.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

# Types of design floods

The process of estimating the possible impact of extreme hydrological events on a hydraulic work, the selection of its dimensions, and the operating policy for its correct operation are known as *Hydrological Design* (Ramírez-Orozco & Aldama, 2000). The concept associated with hydraulic works, to avoid damage that can lead to their rupture or collapse, is called *Hydrological Safety* (Aldama, Ramírez, Aparicio, Mejía-Zermeño & Ortega-Gil, 2006). Logically, there are other concepts associated with Structural safety and Hydraulic operation.

In the case of levees or retaining walls, bridges, rectifications, canalizations, and some urban drainage works, their hydrological dimensioning is based exclusively on the maximum flow or flow associated with a low probability of being exceeded or *Design Flood*.

On the other hand, for all reservoirs, control dams, peak breakers, and urban ponds, their hydrological dimensioning requires the design flood *hydrograph* since this determines their operation and/or performance. The flood hydrograph is the chart that defines the evolution over time (abscissa) of the flow (ordinate). It has four basic characteristics or variables: peak flow, volume (area under the hydrograph), time to peak flow, and total duration.

# Univariate design floods

Hydrological flood estimation methods are divided into two large groups: (1) *hydrometeorological* and (2) *probabilistic*. The former groups are deterministic, applied by sub-basins, and rely on the rainfall data from the pluviographic and pluviometric stations to define the design storms, which are transformed into flows through the mathematical modeling of the rainfall-runoff process. The estimated partial hydrographs are translated and integrated into the requested hydrograph (Teegavarapu, 2012; Mujumdar & Nagesh-Kumar, 2012).

Probabilistic methods are more reliable and accurate but require data from the maximum annual flows observed at the project site. When only the data of the maximum annual flows are processed, the statistical technique is known as *Flood Frequency Analysis* (FFA). It consists of representing the available record of peak flow by a *probability distribution function* (PDF) and, with base in such a model, making the desired inferences or *predictions* associated with a certain probability of exceedance (*Floods Design*), and the reciprocal is the return period in years (Kite, 1977; Stedinger, Vogel, & Foufoula-Georgiou, 1993; Hosking & Wallis, 1997; Rao & Hamed, 2000; Khaliq, Ouarda, Ondo, Gachon, & Bobée, 2006; Meylan, Favre, & Musy, 2012; Stedinger, 2017; Teegavarapu, Salas, & Stedinger, 2019).

## Bivariate design floods

As stated above, a flood is a *multivariate* extreme hydrological event. In this regard, it has been shown that reservoirs are not sensitive to the value of time at peak flow (*Qp*) and that the total duration is related to the volume (*V*); so these two variables (*Qp* and *V*) that are correlated are sufficient to define, approximately, the hydrograph of the design flood (Goel, Seth, & Chandra, 1998; Yue, Ouarda, Bobée, Legendre, & Bruneau, 1999; Aldama, 2000; Yue & Rasmussen, 2002; Aldama *et al.*, 2006).

The *bivariate* analysis of floods is the *simplest* multivariate approach, yet it still involves several mathematical complications; for example: (1) a bivariate PDF should be used; (2) its validation requires the estimation of the bivariate empirical probabilities; (3)

now there are joint and conditional probabilities; (4) a joint return period must be defined, for which there are *infinite* pairs of *Qp* values and *V* which meet them and (5) the critical or design events must be selected from the mentioned pairs (Aldama, 2000; Ramírez-Orozco & Aldama, 2000; Escalante-Sandoval y Reyes-Chávez, 2002; Volpi & Fiori, 2012; Requena, Mediero- & Garrote, 2013).

## Objectives

The *overall objective* of this study consisted of detailing the nine stages that make up the joint frequency analysis of floods, which processes annual peak flow and volume. The *specific objectives* can be summarized in the following six: (1) selection and testing of the annual maximum flows and volume records; (2) fitting with L moments of the GEV distribution; (3) description and fitting of the bivariate General Extreme Values distribution (*GEVb*), using the maximum likelihood method. The validation of the *GEVb* to the bivariate data and ratifying its marginal data are included. (4) a description of the theory of univariate and joint return periods; (5) a selection of the critical events in the AND type joint return period graph and (6) the described theory applied to the 55-year record of the La Cuña hydrometric station, of the Rio Verde in Hydrological Region No. 12–3 (Rio Santiago), Mexico.

## Operative theory

### General Extreme Values distribution

The GVE (by its acronym in Spanish) or *General Extreme Values distribution* includes three families of probabilistic models called: Gumbel, Fréchet, and Weibull, which are straight lines and two types of curves in the Gumbel–Powell extreme probability paper (Chow, 1964). The expression of its PDF is (Stedinger *et al.*, 1993; Escalante-Sandoval & Raynal-Villaseñor, 1994; Hosking & Wallis, 1997; Rao & Hamed, 2000):

$$F(x) = \exp\left\{-\left[1 - k\frac{(x-u)}{\alpha}\right]_{+}^{1/k}\right\} \qquad (1)$$

in which *u*, *α,* and *k* are the location, scale, and shape parameters, with $-\infty < u,k < \infty$ and $\alpha > 0$. The GEV distribution is defined by the set {xi: $[1-k(x_i-u)/\alpha]>0$}, which is indicated by the + sign outside the rectangular parenthesis.

Coles (2001) indicates that any combination of fit parameters that violates the previous condition of positivity implies that at least one of the observed points ($x_i$) is beyond the endpoints of the distribution. Then the likelihood function is zero, and the logarithmic likelihood function equals −∞.

If $k < 0$, the GEV has type II or Fréchet without an upper limit ($u + \alpha / k < x < \infty$) and concavity upwards. When $k > 0$, type III or Weibull is defined, with an upper limit ($\infty < x < u + \alpha / k$) and downward concavity. Finally, when $k = 0$, we arrive, in an asymptotic sense, at type I or Gumbel distribution, also called double exponential, are straight lines on the extreme probability paper ($-\infty<x<\infty$).

By representing the maximum values in each year with the GEV distribution, the *predictions* sought ($X_{Tr}$) for an exceedance probability *q* or a return period *Tr* in years ($Tr = 1/q$) are obtained based on Equation (1) since $F(x) = 1–q$ (Hosking & Wallis, 1997; Rao & Hamed, 2000):

$$X_{Tr} = u + \frac{\alpha}{k}\{1 - [-\ln(1-q)]^k\} \text{ para } k \neq 0 \qquad (2)$$

## Bivariate GEV probabilistic Model

Emil Julius Gumbel proposed in the early sixties the so-called *Logistic model*, which accepts as *marginal* distributions those of extreme values of the type: Gumbel, GEV, Mixed Gumbel, and TCEV (Escalante-Sandoval & Raynal-Villaseñor, 1994; Ramírez-Orozco & Aldama, 2000; Escalante-Sandoval & Reyes-Chávez, 2004: Yue & Wang, 2004 and Escalante-Sandoval, 2005), and is expressed as:

$$F(x, y, \theta) = \exp\left\{-[(-\ln F_X(x))^m + (-\ln F_Y(y))^m]^{1/m}\right\} \, (m \geq 1) \qquad (3)$$

in which $\theta$ represents the seven fit parameters, $m$ is the association parameter, which describes the dependence between the two random variables $X$, $Y$. When the marginals have Gumbel distribution, their expression is:

$$m = \frac{1}{\sqrt{1-\rho}}\ (0 \leq \rho \leq 1) \quad (4)$$

where $\rho$ is the Pearson correlation coefficient, with the following equation:

$$\rho = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \cdot \sigma_Y} \quad (5)$$

in which, $(\mu_X, \sigma_X)$ and $(\mu_Y, \sigma_Y)$ are the *population* mean and standard deviation of the random variables $X$ and $Y$. If the marginal distributions $F_X(x)$ and $F_Y(y)$ are GEV functions, their expressions are:

$$F(x) = \exp\left\{-\left[1 - k_1\left(\frac{x-u_1}{\alpha_1}\right)\right]_+^{1/k_1}\right\} \quad (6)$$

$$F(y) = \exp\left\{-\left[1 - k_2\frac{(y-u_2)}{\alpha_2}\right]_+^{1/k_2}\right\} \quad (7)$$

Substituting equations (6) and (7) in (3), we obtain the bivariate PDF with marginal GEV (Escalante-Sandoval & Raynal-Villaseñor, 1994):

$$F(x,y,\theta) = \exp\left\{-\left[\left(1-k_1\left(\frac{x-u_1}{\alpha_1}\right)\right)_+^{m/k_1} + \left(1-k_2\left(\frac{y-u_2}{\alpha_2}\right)\right)_+^{m/k_2}\right]^{1/m}\right\} \qquad (8)$$

with $\theta = u_1, \alpha_1, k_1, u_2, \alpha_2, k_2, m$, where the fitting parameters are obtained with the maximum statistical likelihood. The univariate and joint probabilities of the logistic model must meet the following restriction (Escalante-Sandoval & Raynal-Villaseñor, 1994; Escalante-Sandoval, 2005):

$$F_X(x) \cdot F_Y(y) < F(x,y) < min[F_X(x), F_Y(y)] \qquad (9)$$

## Fitting of the GEV distribution with L moments

The L moments ($\lambda_s$) are linear combinations of the weighted probability moments ($\beta_r$) developed by Greenwood, Landwehr, Matalas, and Wallis (1979), which are statistical parameters of the ordered data. In a sample of size *n*, with its elements $x_i$ arranged in ascending order ($x_1 \leq x_2 \leq \cdots \leq x_n$) the unbiased estimators of $\beta_r$ are obtained with the following general expression:

$$\beta_r = \frac{1}{n}\sum_{j=r+1}^{n} \frac{(j-1)\cdot(j-2)\cdots(j-r)}{(n-1)\cdot(n-2)\cdots(n-r)} x_j \tag{10}$$

The L moments are an efficient and robust system for estimating the fit parameters of various PDFs used in AFC, with the calculation equations (Stedinger *et al.*, 1993; Hosking & Wallis, 1997; Rao & Hamed, 2000; Stedinger, 2017):

$$\lambda_1 = \beta_0 \tag{11}$$

$$\lambda_2 = 2 \cdot \beta_1 - \beta_0 \tag{12}$$

$$\lambda_3 = 6 \cdot \beta_2 - 6 \cdot \beta_1 + \beta_0 \tag{13}$$

The quotient ($\tau$) of moments L of similarity with that of asymmetry is:

$$\tau_3 = \lambda_3/\lambda_2 \quad (14)$$

The equations that allow estimating the three fit parameters of the GEV distribution are:

$$k \cong 7.8590 \cdot c + 2.9554 \cdot c^2 \quad (15)$$

being:

$$c = \frac{2}{3+t_3} - 0.63093 \quad (16)$$

$$\alpha = \frac{\lambda_2 \cdot k}{(1-2^{-k}) \cdot \Gamma(1+k)} \quad (17)$$

$$u = \lambda_1 - \frac{\alpha}{k}[1 - \Gamma(1+k)] \quad (18)$$

For the estimation of the Gamma function Γ(ω), the Stirling formula (Davis, 1972) was used:

$$\Gamma(\omega) \cong e^{-\omega} \cdot \omega^{\omega-\frac{1}{2}} \cdot (2\pi)^{1/2} \cdot F1 \quad (19)$$

being:

$$F1 = \left(1 + \frac{1}{12 \cdot \omega} + \frac{1}{288 \cdot \omega^2} - \frac{139}{51840 \cdot \omega^3} - \frac{571}{2488320 \cdot \omega^4} + \cdots\right)$$

## Maximum likelihood method

The *principle of maximum likelihood* is explained with different levels of detail according to the text consulted (Kite, 1977; Metcalfe, 1997; Rao & Hamed, 2000; Coles, 2001; Kottegoda & Rosso, 2008; Meylan *et al.*, 2012). But in general, it defines the following: given a random sample ($X_1$, $X_2$, . . . , $X_n$) of independent and identically distributed observations (*iid*) that follow a PDF named $F_\theta$ with fitting parameters $\theta_1$, $\theta_2$,. . . , $\theta_q$. Then, by definition, the probability of obtaining a value $X_i$ will be:

$$P(x_i \leq X \leq x_i + dx_i) = f_\theta(x_i) \cdot dx_i \quad (20)$$

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

$$L(\theta) = f_\theta(x_1) \cdot f_\theta(x_2) \cdots f_\theta(x_n) = \prod_{i=1}^{n} f_\theta(x_i) \quad (21)$$

The maximum likelihood method consists in finding a vector $\hat{\theta}$ of parameters that make $L(\theta)$ a maximum 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), that is:

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f_\theta(x_i) \quad (22)$$

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

## Fitting of the GEV with maximum likelihood

The following description focuses on the iterative process that maximizes the logarithmic likelihood function (Rao & Hamed, 2000):

$$l(\theta) = -n \cdot \ln \alpha - (1-k)\sum_{i=1}^{n} v_i - \sum_{i=1}^{n} e^{-v_i} \quad (23)$$

being:

$$v_i = -\frac{1}{k}\ln\left[1 - k\left(\frac{x_i - u}{\alpha}\right)\right] \quad (24)$$

The process begins by evaluating *P*, *Q,* and *R*, with the expressions:

$$P = n - \sum_{i=1}^{n} e^{-v_i} \quad (25)$$

$$Q = \sum_{i=1}^{n} e^{kv_i - k} - (1-k)\sum_{i=1}^{n} e^{kv_i} \quad (26)$$

$$R = n - \sum_{i=1}^{n} v_i + \sum_{i=1}^{n} v_i e^{-v_i} \quad (27)$$

The calculation of the increments in each iteration (*j*) is performed with the following equations:

$$\delta u_j = -(\frac{\alpha_j}{n})\left\{b \cdot Q_j + h \cdot \left(\frac{P_j+Q_j}{k_j}\right) + \frac{f}{k_j} \cdot \left[R_j - \left(\frac{P_j+Q_j}{k_j}\right)\right]\right\} \quad (28)$$

$$\delta \alpha_j = -(\frac{\alpha_j}{n})\left\{h \cdot Q_j + a \cdot \left(\frac{P_j+Q_j}{k_j}\right) + \frac{g}{k_j} \cdot \left[R_j - \left(\frac{P_j+Q_j}{k_j}\right)\right]\right\} \quad (29)$$

$$\delta k_j = -(\frac{1}{n})\left\{f \cdot Q_j + g \cdot \left(\frac{P_j+Q_j}{k_j}\right) + \frac{c}{k_j} \cdot \left[R_j - \left(\frac{P_j+Q_j}{k_j}\right)\right]\right\} \quad (30)$$

In the previous expressions, *a*, *b*, *c*, *f*, *g,* and *h* are the coefficients of the variance-covariance matrix, a function of the shape parameter (*k*), with the following general expression and the values shown in Table 1 (Campos-Aranda, 2003):

$$coef. = C0 + C1 \cdot k + C2 \cdot k^2 + C3 \cdot k^3 + C4 \cdot k^4 \quad (31)$$

**Table 1**. Fitting parameters of Equation (31).

| ***coef.*** | **Numerical values of:** | | | | |
|---|---|---|---|---|---|
| | **C0** | **C1** | **C2** | **C3** | **C4** |
| *a* | 0.6524995 | −0.5580180 | 1.0877760 | −0.0617186 | −0.1209662 |
| *b* | 1.2474540 | −0.1987501 | −0.2179270 | 0.0626601 | 0.1053422 |
| *c* | 0.4737703 | −0.7670770 | 0.2818301 | −0.1587989 | 0.1703716 |
| *f* | 0.2577352 | −0.1293403 | −0.3234050 | 0.0312508 | 0.1636221 |
| *g* | 0.1441769 | 0.4209507 | −0.4192687 | −0.0467502 | −0.0986341 |
| *h* | 0.3358007 | −1.1933840 | −0.1146433 | −0.0658476 | 0.0381460 |

The process begins by adopting the fit parameters obtained with the L moments method as initial values: $u_j$, $a_j$, and $k_j$, to estimate their values in the next iteration with the equations:

$$u_{j+1} = u_j + \delta u_j \quad (32)$$

$$\alpha_{j+1} = \alpha_j + \delta \alpha_j \quad (33)$$

$$k_{j+1} = k_j + \delta k_j \qquad (34)$$

In each iteration, new values of the six coefficients *a*–*h* (Equation (31)) are calculated since $k_j$ changes. The new values $u_{j+1}$, $a_{j+1}$, and $k_{j+1}$, are taken to Equations (24) to (30), and the process concludes when such increases are less than the adopted tolerance; for example, 0.00001.

## $I(\theta)$ of the bivariate GEV

The logarithmic likelihood function of the *GEVb* probabilistic model, for the simple case of random variables *X*, *Y* with equal record amplitude, has been exposed by Escalante-Sandoval and Raynal-Villaseñor (1994) and is the following:

$$l(\theta) = \sum_{i=1}^{n} \left\{ -(\ln\alpha_1 + \ln\alpha_2) + \ln\left[1 - k_1\left(\frac{x_i - u_1}{\alpha_1}\right)\right]_+^{\frac{m}{k_1}-1} + \ln\left[1 - k_2\left(\frac{y_i - u_2}{\alpha_2}\right)\right]_+^{\frac{m}{k_2}-1} + \ln\left[\left(1 - k_1\left(\frac{x_i - u_1}{\alpha_1}\right)\right)_+^{m/k_1} + \left(1 - k_2\left(\frac{y_i - u_2}{\alpha_2}\right)\right)_+^{m/k_2}\right]^{\frac{1}{m}-2} + \ln\left[(m-1) + \left(\left(1 - k_1\left(\frac{x_i - u_1}{\alpha_1}\right)\right)_+^{m/k_1} + \left(1 - k_2\left(\frac{y_i - u_2}{\alpha_2}\right)\right)_+^{m/k_2}\right)^{1/m}\right] - \left[\left(1 - k_1\left(\frac{x_i - u_1}{\alpha_1}\right)\right)_+^{m/k_1} + \left(1 - k_2\left(\frac{y_i - u_2}{\alpha_2}\right)\right)_+^{m/k_2}\right]^{1/m} \right\} \quad (35)$$

The previous expression is one of the five parts that make up the $l(\theta)$ of the GEV trivariate model, with different amplitudes in each *X*, *Y*, and *Z* record, developed by Escalante-Sandoval and Raynal-Villaseñor (1994).

## The Complex algorithm

The maximization of Equation (35) to obtain the seven optimal fit parameters $(\theta = u_1, \alpha_1, k_1, u_2, \alpha_2, k_2, m)$ of the *GEVb* model must be approached numerically given the complexity of such equation and its partial derivatives concerning *θ,* and for this, the Complex algorithm was selected, with multiple restricted or bounded variables ($z$). Its theoretical approach is as follows (Box, 1965):

$$\text{Minimize F}(z_1, z_2, \cdots, z_s) \tag{36}$$

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

$$\begin{gathered} y_1 = \text{F}(z_1, z_2, \cdots, z_s) \\ \vdots \vdots \\ y_w = \text{F}(z_1, z_2, \cdots, z_s) \end{gathered} \tag{37}$$

Both variables have lower and upper limits of the type ≤, that is: $z_{inf} \leq z_i \leq z_{sup}$ y $y_{inf} \leq y_i \leq y_{sup}$. The *Complex algorithm* is a local exploration technique guided exclusively by what it finds in its path; its antecedents, a brief description of its operative process, and its OPTIM code in the Basic language can be consulted in Campos-Aranda (2003). In Bunday (1985), there is another description and code of this search method.

The main designations in the OPTIM code are NX and NY, which define the number of decision and dependent variables; for the case analyzed, seven $(u_1, \alpha_1, k_1, u_2, \alpha_2, k_2, m)$ and 2*n* (twice the number of years of the record) because the dependent variables are the positivity restrictions: $[1 - k_1(x_i - u_1)/\alpha_1]$ y $[1 - k_2(y_i - u_2)/\alpha_2]$. MI = 500 is the maximum number of evaluations of the objective function (FO), and NQ = 25 is the number of such calculations between printing results. These variables are defined in the data reading subroutine.

An important advantage of the OPTIM code lies in allowing easy access to the limits (L=lower, U=upper), names, and initial values of the variables, in the cited 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 case studied, I ranges from 1 to 7 and J from 1 to 2*n* (double the number of years of the record). Then, the FA and FR convergence criteria are included for the absolute and relative deviations of the FO. The following values were used: 0.0002 and 0.00001, respectively.

The objective function is called FO in the OPTIM code and is defined at the end of the program, it logically corresponds to Equation (35), named FO$="FLV", of the logarithmic likelihood function. FO is assigned with a negative sign because the Complex algorithm minimizes the function (Equation (36)), and it is desired to maximize the FLV.

## Univariate return periods

The classic concept of probability of an event is defined as the quotient of the number of favorable cases (*ncf*) to such event, between the number of possible cases (*ncp*) to said event; therefore, it varies from zero to one. Due to the annual management of the variable $X$, the exceedance probability $F_X'(x)$ corresponds to the reciprocal of the return period ($T_X$) in years; since in each year, we have *ncf* = 1 and *ncp* = $T_X$, that is (Yue & Rasmussen, 2002; Shiau, 2003):

$$T_X = \frac{1}{F_X'(x)} = \frac{1}{1-F_X(x)} \quad (38)$$

$$T_Y = \frac{1}{F'_Y(y)} = \frac{1}{1-F_Y(y)} \quad (39)$$

In the previous expressions, $F_X(x)$ and $F_Y(y)$ are the probability of non–exceedance that is estimated with Equation (6) and Equation (7).

## Joint return periods

The first *joint return period* of the event ($X$, $Y$) is defined under the OR condition and Equation (38) as follows (Goel *et al.*, 1998; Yue, 2000b; Shiau, 2003; Requena *et al.*, 2013; Vogel & Castellarin, 2017):

$$T(x,y) = \frac{1}{1-F(x,y)} \text{ siendo } F(x,y) = P(X \leq x \text{ ó } Y \leq y) \quad (40)$$

In the previous expression, $F(x,y)$ is the *joint* non–exceedance probability that is estimated with Equation (8) after estimating its adjustment parameters $(u_1, \alpha_1, k_1, u_2, \alpha_2, k_2, m)$ with the method's maximum likelihood. This event represents the case in which the limits *x* or *y*, or both, can be exceeded (*X* > *x*, or *Y* > *y*; or *X* > *x* and *Y* > *y*).

The second joint return period of the event (*X*, *Y*) is associated with the case in which both limits are exceeded (*X* > *x*, *Y* > *y*) or the AND condition; its equation is (Goel *et al.*, 1998; Aldama, 2000; Ramírez-Orozco & Aldama, 2000; Yue, 2000b and Shiau, 2003; Requena *et al.*, 2013; Vogel & Castellarin, 2017):

$$T'(x,y) = \frac{1}{F'(x,y)} = \frac{1}{1 + F(x,y) - F_X(x) - F_Y(y)} \tag{41}$$

Aldama (2000) obtains the expression *F*'(*x*,*y*) of the joint probability of exceedance using logical and simple reasoning of probabilities applied in the Cartesian plane. Instead, Yue and Rasmussen (2002) resort to the Cartesian plane to numerically define a bivariate event (*X*, *Y*), which can occur in any of the four quadrants, and then obtain the equation of the joint probability of exceedance (denominator of the Equation (41)).

Yue and Rasmussen (2002) also establish the following relationships between univariate return periods and joint return periods:

$$T(x,y) \leq \text{mínimo}\ (T_X, T_Y) \qquad (42)$$

$$T'(x,y) \geq \text{máximo}\ (T_x, T_Y) \qquad (43)$$

Various authors (Yue, 2000b; Yue & Rasmussen, 2002; Shiau, 2003; Vogel & Castellarin, 2017) have shown the two joint return periods' graphs and discussed their differences. But Volpi and Fiori (2012) established a procedure to select their *critical events*.

Figure 1 shows the graph of the joint return period *T'*(*Q*, *V*), built with the data from the numerical application described later.

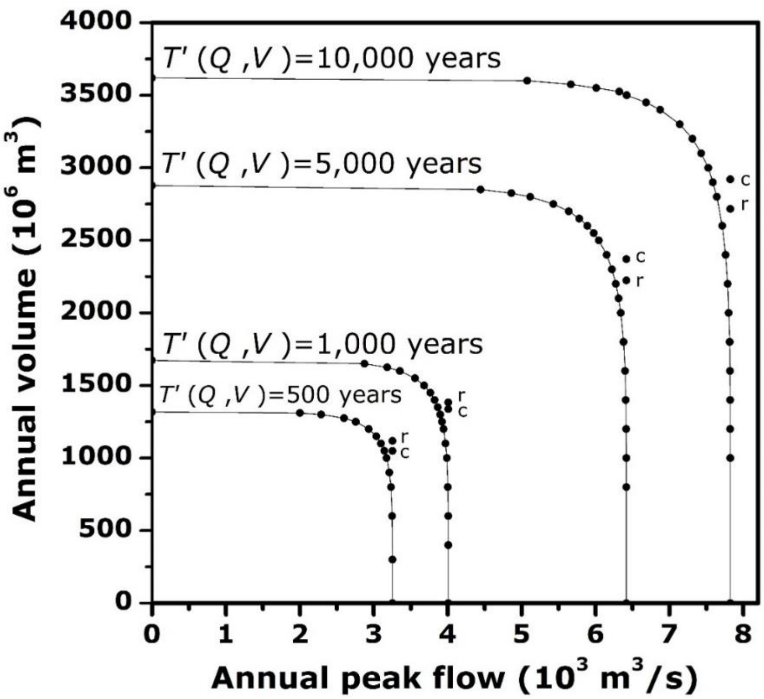


**Figure 1**. Graphs of the four design joint return periods *T′*(*Q*, *V*) estimated with the *GVEb* distribution for the annual floods recorded in the La Cuña hydrometric station, Mexico.

According to Yue *et al.* (1999), Yue (2000b), and Yue and Rasmussen (2002), there is a third type of *joint return period*, which is applied in hydrological practice and is defined for an event *X* given that $Y \leq y$ or for an event *Y* given that $X \leq x$ y are therefore designated *conditionals*. For such events, their conditional probability distributions are defined with these equations:

$$F(X|Y \leq y) = \frac{F(x,y)}{F(y)} \quad (44)$$

$$F(Y|X \leq x) = \frac{F(x,y)}{F(x)} \quad (45)$$

Substituting the previous expressions in Equation (38) and Equation (39), the formulas for the *conditional joint return period* are obtained:

$$T_{X|Y} = \frac{1}{1-F(X|Y\leq y)} \quad (46)$$

$$T_{Y|X} = \frac{1}{1-F(Y|X\leq x)} \quad (47)$$

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

## Critical or design events of the *T′*(*x*,*y*)

Volpi and Fiori (2012) highlight that the joint return period graph of the AND type, shown in Figure 1, presents a severe inconsistency as it contains univariate critical thresholds in a bivariate context. Due to the above, such a graph is considered to be made up of two portions: the *simple* (*naive part*) and the *correct one* (*proper part*). The straight parts are the tails or lines asymptotes to the curved part. The probability of occurrence of an event or pair of *Qp* and *V* is variable in the curved part and decreases along the straight part, although all the values define the same period of joint return. In summary, the pairs of values of the asymptotes have a low probability of occurrence and therefore should not be included in the search analyses for critical or severe floods (*Qp* and *V*). For practical purposes, the end points of the curved part can be defined according to their empirical distribution or close to the beginning of the asymptotic lines.

# Wald-Wolfowitz test

This non–parametric test has been used by Bobée and Ashkar (1991), Rao and Hamed (2000), and Meylan *et al.* (2012) to test *independence and stationarity* in records of maximum annual expenses ($X_i$). According to the type of hydrometric information processed, it was proposed to apply the test to the records of peak flows and annual volumes, which should be samples of random values. The test statistic is:

$$R = \sum_{i=1}^{n-1} X_i \cdot X_{i+1} + X_n \cdot X_1 \quad (48)$$

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:

$$E[R] = \bar{R} = \frac{S_1^2 - S_2}{n-1} \quad (49)$$

$$Var[R] = \frac{S_2^2 - S_4}{n-1} + \frac{S_1^4 - 4 \cdot S_1^2 \cdot S_2 + 4 \cdot S_1 \cdot S_3 + S_2^2 - 2 \cdot S_4}{(n-1)(n-2)} - \bar{R}^2 \quad (50)$$

in which:

$$S_k = \sum_{i=1}^{n} X_i^k \quad (51)$$

Finally, $U$ is calculated with the equation:

$$U = \frac{R - \bar{R}}{\sqrt{VAR[R]}} \quad (52)$$

The value of $U$ follows a Normal distribution (0.1) and can be used to test the independence of the series data with a significance level $\alpha$, commonly 5 %. In a two-tailed test, the standardized normal variable is $Z\alpha/2 \cong 1.96$; when the absolute value of $U$ is less than 1.96, the series will be made up of independent values (random sample).

## Selection and testing of records to be processed

The book by Aldama, Ramírez, Aparicio, Mejía-Zermeño, and Ortega-Gil (2006) is a benchmark in the bivariate frequency analysis of floods. It presents 16 joint records of maximum annual flow and volume of entry into important dams in Mexico. Page 43 indicates that they were obtained from existing records of maximum annual flows and maximum annual hydrographs. This implies that the joint records comprised the maximum annual events, not belonging to each flood. On the other hand, page 44 indicates that the continuous hydrometric records of a graphic or tabular nature made it possible to identify the flow and volume data of the avenues.

In the author's opinion, the second clarification of the book by Aldama *et al.* (2006) is correct, as Yue (1999) and Yue *et al.* (1999), who established that the maximum flows and their volume must come from the hydrograph of the annual flood. In addition, and in general, both the maximum peak flow and its respective runoff volume include the base flow (Shiau, Wang, & Tsai, 2006).

On the other hand, to avoid subjectivity in the search for *Qp* and *V* records that *accept* the GEV distribution as a marginal function, it is proposed to apply the powerful test developed by Stedinger *et al.* (1993), which establishes that when the absolute value of the *Z* statistic exceeds the Normal standard deviation of 1.645, it follows that the shape parameter $k$ is statistically different from zero. Therefore the GEV distribution must be adjusted and not the Gumbel. The statistic is:

$$Z = k\sqrt{n/0.5633} \quad (53)$$

To search for records that accept the GEV distribution as marginal, such a probabilistic model is fitted with the method of L moments (equations (10) to (19)), and Equation (53) is applied if the absolute value of $Z$ exceeds 1.645, the GEV distribution can represent such data.

## Estimation of empirical probabilities

The empirical univariate and bivariate probabilities of non–exceedance were estimated based on the Cunnane formula, which according to

Stedinger (2017), leads to approximately unbiased probabilities of non–exceedance ($p$) for many PDFs used in AFC; its expression is:

$$p = \frac{i-0.40}{n+0.20} \quad (54)$$

where $i$ is the number of the data when ordered from lowest to highest, and $n$ is the total number or years of the records of peak quantity and annual volume.

In the case of bivariate probabilities, we worked on the two–dimensional plane, with the flows in the rows and the volumes in the columns (Yue *et al.*, 1999; Yue, 2000b; Yue & Rasmussen, 2002; Yue & Wang, 2004). The numerical process begins by saving the historical records of maximum flow ($Q$) and annual volume ($V$) in files $Qh$ and $Vh$; they were also ordered in progressive order of magnitude in $Qo$ and $Vo$ files. They were then processed in annual pairs, and each $Qh$ value was compared against $Qo$, and the times the second was less than or equal were counted, and $NQ$ was designated. The same was done with $Vh$ against $Vo$ to obtain $NV$. This is equivalent to changing the original data, of each pair of historical annual values, by its order number or *rank*.

Afterward, each historical pair of ranges is compared against all the others, and the times in which both ranges (AND condition) are lower are counted, and such quantity is called *NQV*; that is, the number of occurrences of minor *q* and *v* combinations in the two–dimensional plane. Finally, the Cunnane graphical position formula is applied; for the bivariate case, it is the following:

$$F_e(x,y) = P(Q \leq q, V \leq v) = \frac{NQV_i - 0.40}{n + 0.20} \quad (55)$$

## Validation of the *GEVb* distribution

Yue (2000a) indicates that the relationship between the empirical and theoretical joint probabilities of the peak flow and the volume allows the validity of the proposed joint distribution. The simplest way to represent them consists of taking the empirical non–exceedance probability to the abscissa axis and the theoretical one on the ordinate axis; logically, each data pair defines a point that coincides with or moves away from the line at 45º. Inspection of the graph described and the value of the correlation coefficient, in these cases, more significant than 0.98, confirm the validity of the joint probabilistic model used.

Yue (2000b) and Yue and Rasmussen (2002) apply the Kolmogorov–Smirnov test with a significance level ($\alpha$) of 5 % to accept or reject the *maximum absolute difference* (*dif*) between the joint probabilities. To evaluate the statistics ($D_n$) of the test, which is a function of the number of data ($n$), the expression reported by Meylan *et al.* (2012) for $\alpha$ = 5 % is as follows:

$$D_n = \frac{1.358}{\sqrt{n}} \tag{56}$$

If *dif* is less than $D_n$, the joint probabilistic model or bivariate GVE distribution is accepted.

## Data records to be processed

The La Cuña gauging station, with code 12054 and a basin area of 19097 $km^2$, is located on the Río Verde in Hydrological Region No. 12–3 (Río Santiago), Mexico. Gómez, Aparicio, and Patiño (2010) present the record of maximum flow in $m^3/s$ and annual volumes in millions of $m^3$ ($Mm^3$)

from the period 1947 to 2004, with 55 pairs of data (*n*), since the years 1983, 1984 and 1985 do not have volume information. Such data are shown in Table 2.

**Table 2**. Peak flows and annual volumes at the La Cuña hydrometric station of the Hydrological Region No. 12–3 (Río Santiago), Mexico (Gómez *et al.*, 2010).

| **Año** | ***Qp*** **($m^3/s$)** | ***V*** **($Mm^3$)** | **Año** | ***Qp*** **($m^3/s$)** | ***V*** **($Mm^3$)** |
|---|---|---|---|---|---|
| 1947 | 784.0 | 146.80 | 1975 | 622.1 | 249.07 |
| 1948 | 736.8 | 155.12 | 1976 | 1374.0 | 527.96 |
| 1949 | 510.0 | 111.40 | 1977 | 439.7 | 111.77 |
| 1950 | 461.0 | 94.06 | 1978 | 280.2 | 66.23 |
| 1951 | 411.0 | 111.55 | 1979 | 267.2 | 45.80 |
| 1952 | 326.0 | 70.82 | 1980 | 287.3 | 99.60 |
| 1953 | 349.8 | 144.75 | 1981 | 280.7 | 28.70 |
| 1954 | 130.4 | 23.22 | 1982 | 156.5 | 35.37 |
| 1955 | 690.0 | 203.31 | 1986 | 698.2 | 193.51 |
| 1956 | 266.0 | 106.76 | 1987 | 184.7 | 55.39 |

| | | | | | |
|---|---|---|---|---|---|
| 1957 | 199.0 | 45.92 | 1988 | 595.2 | 242.21 |
| 1958 | 690.0 | 188.71 | 1989 | 110.2 | 42.49 |
| 1959 | 340.6 | 47.91 | 1990 | 523.9 | 248.07 |
| 1960 | 249.6 | 91.58 | 1991 | 1636.3 | 443.30 |
| 1961 | 350.0 | 130.68 | 1992 | 1168.0 | 172.49 |
| 1962 | 317.0 | 51.27 | 1993 | 295.0 | 96.50 |
| 1963 | 732.6 | 127.90 | 1994 | 212.8 | 53.55 |
| 1964 | 265.1 | 82.75 | 1995 | 367.4 | 114.61 |
| 1965 | 743.6 | 295.34 | 1996 | 144.6 | 57.43 |
| 1966 | 463.9 | 202.90 | 1997 | 78.4 | 16.55 |
| 1967 | 1 474.9 | 598.38 | 1998 | 261.9 | 66.17 |
| 1968 | 323.0 | 118.25 | 1999 | 196.3 | 41.15 |
| 1969 | 160.4 | 32.22 | 2000 | 46.8 | 18.62 |
| 1970 | 763.8 | 187.75 | 2001 | 313.8 | 75.78 |
| 1971 | 578.0 | 166.61 | 2002 | 319.6 | 153.79 |
| 1972 | 191.8 | 26.39 | 2003 | 621.1 | 326.28 |
| 1973 | 2440.0 | 920.30 | 2004 | 824.5 | 384.45 |
| 1974 | 238.4 | 66.66 | Mediana | 340.6 | 111.40 |

# Sequence of bivariate frequency analysis

The joint frequency analysis process is considered to be made up of two parts, the first is called "Fitting *GEVb*" and the second "Estimation of Design Events". Once the joint record of maximum flows and annual volumes was integrated, its randomness was tested with the Wald–Wolfowitz test. If both series proved not to have deterministic components, whether they come from a GEV distribution is verified. If so, such a model is adjusted with the L moments method. Next, the empirical bivariate non–exceedance probabilities are estimated, against which the values of the theoretical probability, estimated with the *GEVb* model, fitting through the Complex algorithm to maximize the maximum likelihood function. The *GEVb* is validated, and the marginal GEVs are ratified using the Kolmogorov–Smirnov test. This first part ends by verifying the joint probability constraints.

The second part estimates the univariate predictions with the GEV distribution and then the hybrid, regression, and conditional univariate design events. Next, the joint design events of type AND are estimated, with the help of the *GEVb* model, and their critical values are selected, located in the curved part of the graphs of the joint return period.

# Results and discussion

## Verification of the randomness

Equations (48) to (52) were applied to the records of peak flow and annual volume in Table 2 to test their independence and stationarity. Both series were found to be *random*, with the following values for the $U$ statistic: 0.284 and 0.213.

# Acceptance of marginal GEV

To the records of annual peak flow and volume in Table 2, GEV distributions were fitting with the methods of L moments (equations (10) to (19)) and maximum likelihood (equations (24) to (34)). Table 3 shows the basic statistics and the values of the fit parameters of each GEV distribution.

**Table 3**. Statistical and fit parameters of the GEV distribution in the annual peak flow and volume records of the La Cuña hydrometric station, Mexico.

| **Data (ma*)** | **Moments L** | | | **Fit parameters** | | |
|---|---|---|---|---|---|---|
| | **$\lambda_1$** | **$\lambda_2$** | **$\tau_3$** | **location** | **scale** | **shape** |
| *Qp* (mL) | 499.8746 | 205.4937 | 0.38713 | 296.28 | 201.70 | −0.3131 |
| *Qp* (mv) | | | | 302.40 | 205.04 | −0.3068 |
| *V* (mL) | 154.8391 | 74.48538 | 0.43326 | 79.51 | 66.03 | −0.3734 |
| *V* (mv) | | | | 85.67 | 69.71 | −0.3543 |

*fitting method: (mL) moments L and (mv) maximum likelihood.

The application of Equation (53) to the results of the maximum likelihood method, with $n$ = 55, leads to an absolute value of $Z$ of 3.03 for the peak flow record and 3.50 for the volume record; Therefore, the records in Table 2 can be represented by the GEV distribution.

## ***Initial*** **search for optimal parameters of the** ***GEVb*** **distribution**

Initially, we sought to maximize Equation (35) using the *Complex algorithm,* and for this, initial values were assigned to the first six decision variables with a similar magnitude to the fitting parameters of each GEV marginal function (Table 3). The lower and upper limits of $u_1$, $a_1$, $u_2$, and $a_2$ were arbitrarily varied. The limits adopted for the shape parameter ($k$) were −0.15 and 0.75. For the association parameter ($m$), its limits were roughly defined with Equation (4), varying from 1.0 to 3.8; this last value corresponds to a $\rho$=0.9302 or correlation calculated for the records in Table 2.

With this approach, it was found that the optimal adjustment parameters ($u$,$a$,$k$) were very different from the initial ones: In general, the final association parameter $m$ was the maximum value adopted, and the optimal $k$–shape parameters their lower limits assigned; this would lead to joint predictions lower than the univariate ones. In addition, it was observed that the maximization process of Equation (35) depended notoriously on the initial values and their limits in the seven decision variables, not allowing them to achieve consistent results or to obtain the global minimum.

## ***Combined* search for optimal parameters of the *GEVb* distribution**

In this new approach to maximizing Equation (35), the Complex algorithm was applied under the following three rules: (1) *unique* initial values were assigned to the six equal fit parameters ($u_1$, $a_1$, $k_1$, $u_2$, $a_2$, $k_2$) to those obtained with the method of maximum likelihood in the fitting of the marginal distributions GEV (Table 3); (2) their lower and upper limits were obtained by multiplying the initial values by 0.90 and 1.10; therefore, they did not vary either and (3) five intervals were defined for the association parameter (*m*) and the value medium, such limits were: 1.0–1.5, 1.5–2.0, 2.0–2.5, 2.5–3.0, and 3.0–3.5. Table 4 shows the main results of the five described applications of the Complex algorithm, except its last column, whose estimate is detailed below.

**Table 4**. Optimal results of the Complex algorithm during the maximization of Equation (35), with the *combined search*, at the La Cuña hydrometric station, Mexico.

| FO initial | FO final | No. Iter. | Fit parameters: | | | *m*, init. | $r_{xy}$ |
|---|---|---|---|---|---|---|---|
| | | | $u_1$, $u_2$ | $a_1$, $a_2$ | $k_1$, $k_2$ | *m* final | |
| 630.1 | 612.6 | 187 | 278.269 | 185.005 | −0.2700 | 1.25 | 0.9918 |
| | | | 77.014 | 63.000 | −0.3192 | 1.50 | |
| 618.4 | 601.5 | 284 | 270.387 | 185.027 | −0.2700 | 1.75 | 0.9942 |
| | | | 80.555 | 63.000 | −0.3129 | 2.00 | |
| 607.8 | 591.6 | 221 | 274.302 | 185.026 | −0.2700 | 2.25 | 0.9928 |
| | | | 90.005 | 63.000 | −0.3181 | 2.50 | |
| 598.5 | 582.9 | 267 | 270.000 | 185.052 | −0.2700 | 2.75 | 0.9954 |
| | | | 79.120 | 63.000 | −0.3194 | 3.00 | |
| 590.3 | 575.2 | 232 | 270.464 | 185.008 | −0.2700 | 3.25 | 0.9920 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | 92.831 | 63.010 | −0.3176 | 3.50 | |

Based on the results in columns 4 to 7 of Table 4, the probabilities of theoretical joint non–exceedance $F(x,y)$ were estimated with Equation (8) using the data $x_i$, $y_i$ of Table 2. On the other hand, applying Equation (55) and its described numerical procedure, the so-called empirical bivariate non-exceedance probabilities $F_e(x,y)$ were calculated, against which the values of the estimated theoretical $F(x,y)$ were contrasted.

The best correspondence or similarity between both joint probabilities $F_e(x,y)$ and $F(x,y)$, was obtained for the fourth application or run of the Complex algorithm, with a value of the correlation coefficient ($r_{xy}$) of 0.9954 and maximum differences positive and negative of 0.0482 and −0.0969, which are indicated by shading in Table 5.

**Table 5**. Joint non–exceedance probabilities and their differences for the annual floods of the La Cuña hydrometric station, Mexico.

| **No.** | **$F_e(x,y)$ empiric** | **$F(x,y)$ theoretic** | **Differences** | **No.** | **$F_e(x,y)$ empiric** | **$F(x,y)$ theoretic** | **Differences** |
|---|---|---|---|---|---|---|---|
| 1 | 0.6449 | 0.6695 | −0.0246 | 29 | 0.7174 | 0.7909 | −0.0735 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2 | 0.6630 | 0.6919 | −0.0289 | 30 | 0.9348 | 0.9670 | −0.0322 |
| 3 | 0.5000 | 0.5213 | −0.0213 | 31 | 0.5000 | 0.5036 | −0.0036 |
| 4 | 0.4275 | 0.4367 | −0.0092 | 32 | 0.2645 | 0.2470 | 0.0175 |
| 5 | 0.4819 | 0.4895 | −0.0076 | 33 | 0.1558 | 0.1513 | 0.0045 |
| 6 | 0.3370 | 0.2888 | 0.0482 | 34 | 0.3551 | 0.3504 | 0.0047 |
| 7 | 0.5000 | 0.4905 | 0.0095 | 35 | 0.0833 | 0.0770 | 0.0063 |
| 8 | 0.0471 | 0.0375 | 0.0096 | 36 | 0.0652 | 0.0727 | −0.0075 |
| 9 | 0.6993 | 0.7809 | −0.0816 | 37 | 0.6812 | 0.7688 | −0.0876 |
| 10 | 0.3007 | 03303 | −0.0296 | 38 | 0.1196 | 0.1387 | −0.0191 |
| 11 | 0.1558 | 0.1247 | 0.0311 | 39 | 0.6812 | 0.7738 | −0.0926 |
| 12 | 0.6630 | 0.7599 | −0.0969 | 40 | 0.0471 | 0.0517 | −0.0046 |
| 13 | 0.2101 | 0.1734 | 0.0367 | 41 | 0.6630 | 0.7253 | −0.0623 |
| 14 | 0.2464 | 0.2860 | −0.0396 | 42 | 0.9348 | 0.9618 | −0.0270 |
| 15 | 0.5000 | 0.4768 | 0.0232 | 43 | 0.7174 | 0.7426 | −0.0252 |
| 16 | 0.2101 | 0.1883 | 0.0218 | 44 | 0.3551 | 0.3535 | 0.0016 |
| 17 | 0.5906 | 0.6036 | −0.0130 | 45 | 0.1739 | 0.1574 | 0.0165 |
| 18 | 0.2645 | 0.2882 | −0.0237 | 46 | 0.4819 | 0.4674 | 0.0145 |
| 19 | 0.8261 | 0.8546 | −0.0285 | 47 | 0.0833 | 0.0986 | −0.0153 |
| 20 | 0.6268 | 0.6583 | −0.0315 | 48 | 0.0109 | 0.0150 | −0.0041 |
| 21 | 0.9529 | 0.9740 | −0.0211 | 49 | 0.2283 | 0.2351 | −0.0068 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 22 | 0.4457 | 0.4263 | 0.1964 | 50 | 0.1196 | 0.1082 | 0.0114 |
| 23 | 0.0652 | 0.0670 | −0.0018 | 51 | 0.0109 | 0.0082 | 0.0027 |
| 24 | 0.6993 | 0.7671 | −0.0678 | 52 | 0.3188 | 0.3048 | 0.0140 |
| 25 | 0.6449 | 0.6953 | −0.0504 | 53 | 0.4457 | 0.4500 | −0.0043 |
| 26 | 0.0652 | 0.0596 | 0.0056 | 54 | 0.7174 | 0.8029 | −0.0855 |
| 27 | 0.9891 | 0.9933 | −0.0042 | 55 | 0.8986 | 0.8911 | 0.0075 |
| 28 | 0.2283 | 0.2180 | 0.0103 | – | – | – | – |

## Validation of the probabilistic model

In Figure 2, both probabilities of joint non–exceedance (empirical and theoretical from Table 5) have been drawn, observing a predominance of negative differences, that is, of points above the 45º line. The value of the Kolmogorov–Smirnov test statistic is 0.1831 (Equation (56)), therefore, the bivariate GEV distribution is accepted as a joint probabilistic model of the data in Table 2, since $dif$ = 0.0969 < $D_n$ = 0.1831.

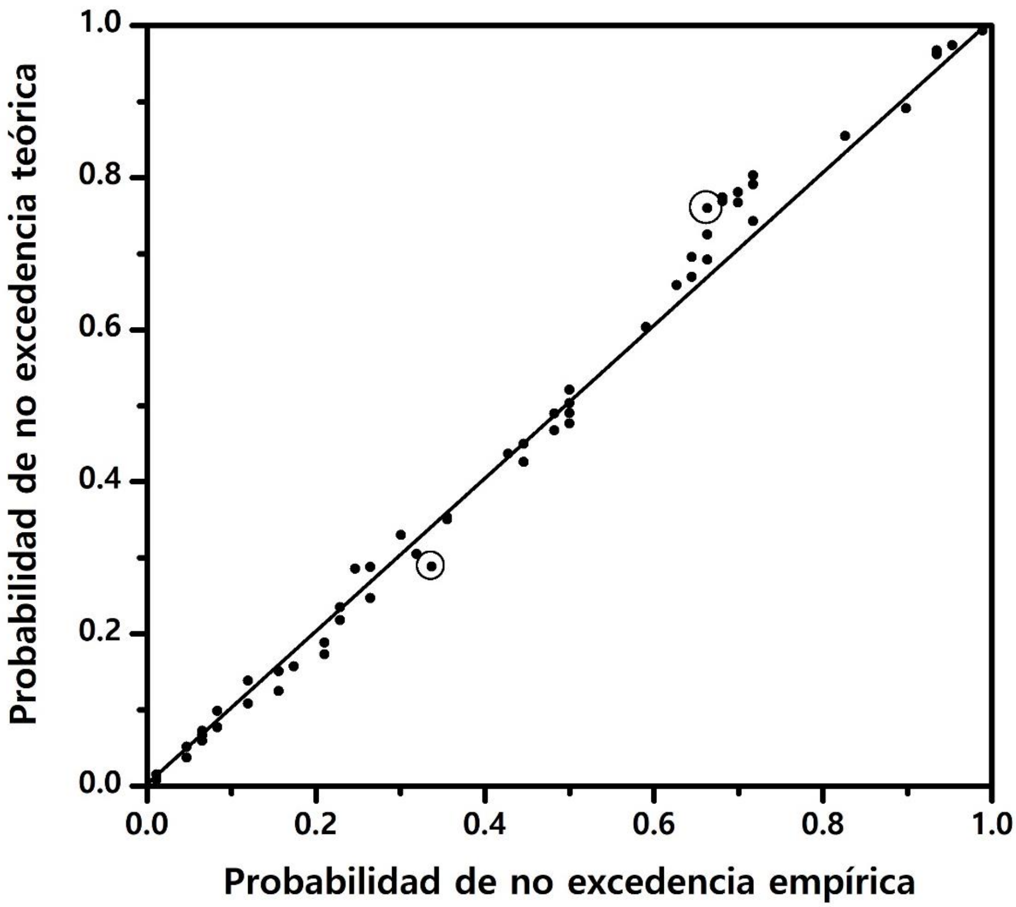


**Figure 2**. Graphic contrast of estimated joint probabilities with the *GVEb* distribution for the annual floods of the La Cuña hydrometric station, Mexico.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

## Ratification of the marginals

This graphical and ratification contrast based on the Kolmogorov–Smirnov Test is necessary due to the change in the univariate fitting parameters ($u$,$a$,$k$) (Table 3) during the application of the Complex algorithm. That is, it allows defining whether the accepted 10 % change in these parameters is acceptable or not.

First, the peak flows and volumes in Table 2 were ordered from lowest to highest. Then, their *theoretical* non–exceedance probabilities were calculated with equations (6) and (7), using the optimal fit parameters of Table 5, for the fourth run of the Complex algorithm. Both series' empirical non–exceedance probabilities (*Qp* and *V*) were estimated with Equation (54).

Figure 3 and Figure 4 show the graphical contrast of probabilities for each ordered series. A much better fit is observed in the volume register. The maximum absolute differences between empirical and theoretical probabilities of peak flows and volumes were 0.0954 and 0.0457; the first occurred in ordered data number 38 and the second in 16, as shown in Table 6 of partial results. As both differences are less than $D_n$ = 0.1831 obtained with Equation (56), it is accepted that the *Qp* and *V* records of Table 2 can be represented by the GEV distribution, with the fitting parameters of Table 4, in the fourth run of the Complex algorithm.

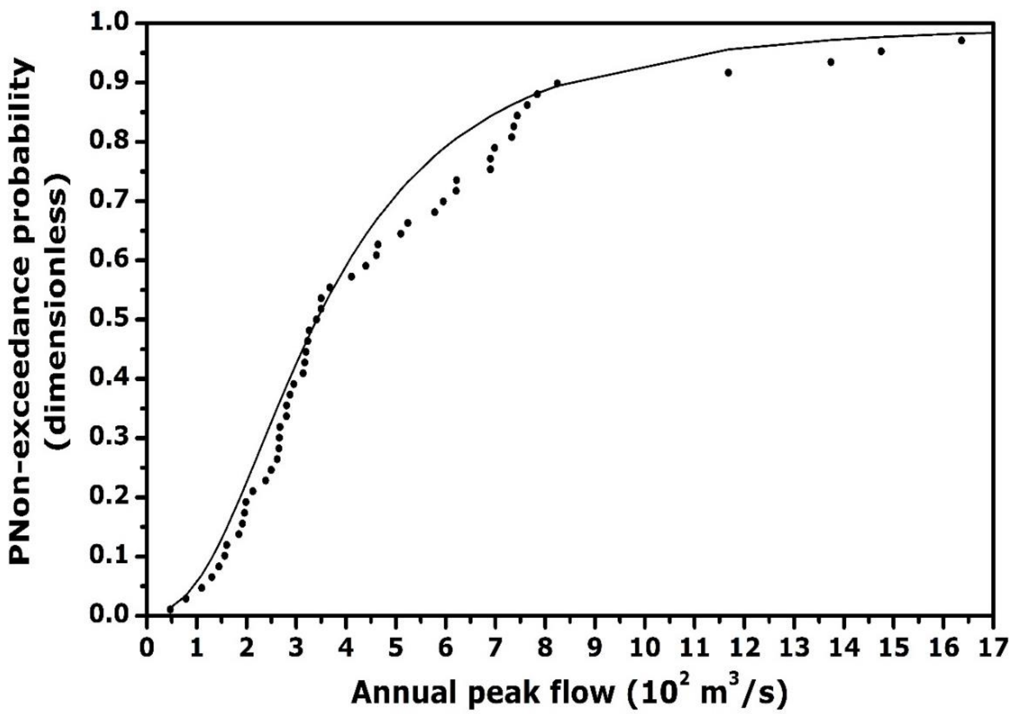


**Figure 3**. *GEVb* marginal distribution of the annual peak flow of the floods of the La Cuña hydrometric station, Mexico.

Tecnología y Ciencias del Agua

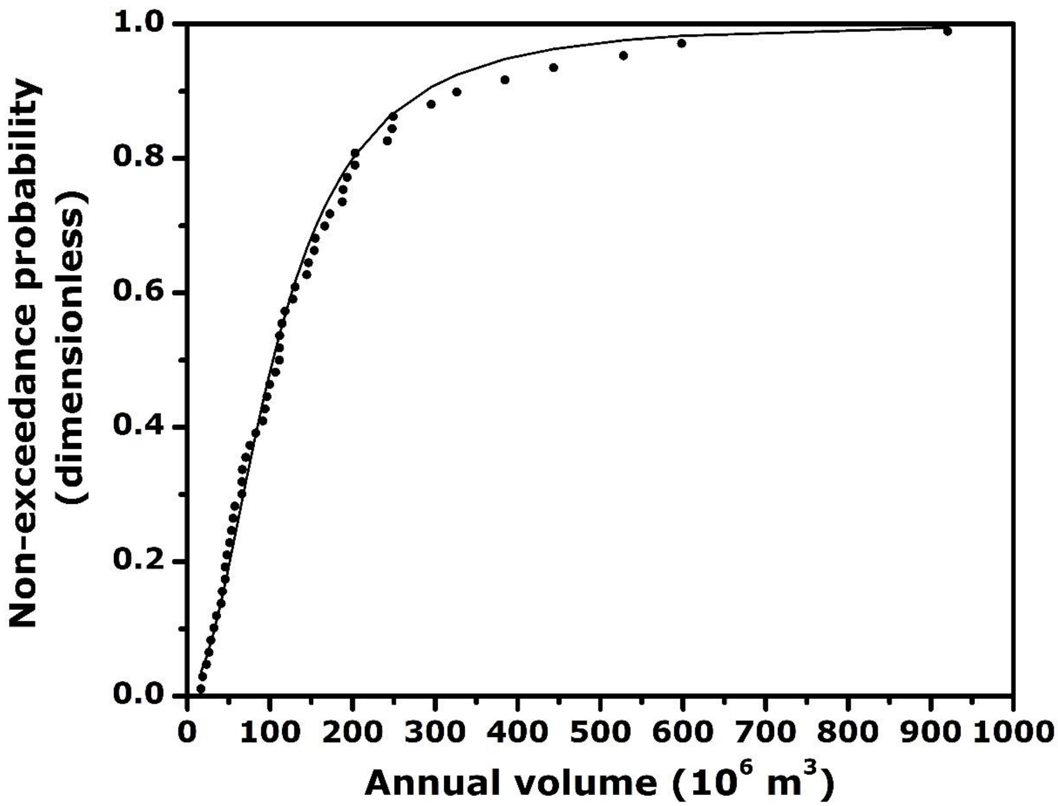


**Figure 4**. *GEVb* marginal distribution of the annual volume of floods at the La Cuña hydrometric station, Mexico.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

**Table 6**. Empirical ($F_e$) and theoretical probabilities in the *GEVb* marginal distributions and their differences in the floods of the La Cuña hydrometric station, Mexico.

| No. | *Qp* | *V* | $F_e$ | *F*(*Qp*) | *F*(*V*) | Dif *Qp* | Dif *V* |
|---|---|---|---|---|---|---|---|
| 1 | 46.8 | 16.55 | 0.0109 | 0.0135 | 0.0368 | −0.0027 | −0.0259 |
| 8 | 184.7 | 41.15 | 0.1377 | 0.1948 | 0.1418 | −0.0571 | −0.0042 |
| 16 | 265.1 | 57.43 | 0.2826 | 0.3581 | 0.2369 | −0.0755 | 0.0457 |
| 27 | 326.0 | 106.76 | 0.4819 | 0.4735 | 0.5152 | 0.0084 | −0.0333 |
| 38 | 578.0 | 155.12 | 0.6812 | 0.7765 | 0.6974 | −0.0954 | −0.0162 |
| 45 | 732.6 | 203.31 | 0.8080 | 0.8624 | 0.8051 | −0.0544 | 0.0028 |
| 55 | 2440.0 | 920.30 | 0.9891 | 0.9949 | 0.9945 | −0.0058 | −0.0054 |

## Verification of the probability restrictions

Before proceeding to estimate the joint design return periods $T'(Q, V)$, it is convenient to verify Equation (9), which establishes the probability restrictions. This is shown in Table 7 for a reduced number of pairs of historical data.

**Table 7**. Verification of the probability restriction of the annual peak flow and volume of the floods of the La Cuña hydrometric station, Mexico.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| **No.** | ***Qp*** | ***V*** | ***F*(*x*)** | ***F*(*y*)** | ***F*(*x*)·*F*(*y*)** | ***F*(*x*,*y*)** |
| 1 | 784.0 | 146.80 | 0.8817 | 0.6723 | 0.5928 | 0.6695 |
| 10 | 266.0 | 106.76 | 0.3599 | 0.5152 | 0.1854 | 0.3303 |
| 20 | 463.9 | 202.90 | 0.6720 | 0.8044 | 0.5406 | 0.6583 |
| 30 | 1374.0 | 527.96 | 0.9718 | 0.9759 | 0.9484 | 0.9670 |
| 40 | 110.2 | 42.49 | 0.0690 | 0.1492 | 0.0103 | 0.0517 |
| 50 | 196.3 | 41.15 | 0.2178 | 0.1418 | 0.0309 | 0.1082 |
| 55 | 824.5 | 384.45 | 0.8947 | 0.9479 | 0.8481 | 0.8911 |

It is observed in Table 7 above and in its complete version that the value of column 7 is always greater than that of 6 and less than the smallest of columns 4 or 5. Therefore, the restriction is met imposed by Equation (9).

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

## Univariate predictions

Applying in Equation (2) the six optimal fit parameters of the fourth run of the Complex algorithm, shown in Table 4, the final predictions in Table 8 were obtained.

**Table 8**. Final univariate predictions were calculated with the *GEVb* distribution at the La Cuña hydrometric station in Mexico.

| **Data** | **Return period, in years** | | | | | | |
|---|---|---|---|---|---|---|---|
| | **25** | **50** | **100** | **500** | **1000** | **5000** | **10000** |
| *Qp* | 1210 | 1550 | 1958 | 3254 | 4009 | 6418 | 7823 |
| *V* | 430 | 568 | 739 | 1317 | 1673 | 2877 | 3619 |

## Design events for hydraulic works

Assuming that in the vicinity of waters below the La Cuña hydrometric station and on the Rio Verde, dikes will be built to protect flood plains and a bridge for crossing, it will be necessary to estimate design events with joint return periods of 500 and 1000 years. In addition, a review of the hydrological safety of the project reservoir will be carried out at the site of the gauging station, with joint return periods of 5,000 and 10 000 years. Therefore, estimating peak flows and annual volumes with the four joint return periods T'(Q,V) cited is necessary.

## Design events obtained with regression

The scatter diagram of the 55 original data pairs (Table 2) showed a cloud of points with a linear trend with a linear correlation coefficient ($r_{xy}$) of 0.9302. The linear regression equation that represents it is the following (Campos-Aranda, 2003):

$$V = -20.0532 + 0.34987 \cdot Qp \qquad (57)$$

Table 8 shows the following predictions for the peak flow ($Qp$) and the four joint design return periods: 3 254, 4 009, 6 418, and 7 823 $m^3/s$, respectively. Based on Equation (57), the following annual volumes are defined: 1 118, 1 383, 2 225, and 2 717 $Mm^3$ for the joint design floods.

Like Serinaldi and Grimaldi (2011), a similarity is found between the volumes estimated with regression and their predictions in Table 8, which are: 1 317, 1 673, 2 877 and 3 619 $Mm^3$; that is, on the order of 18 to 33 % higher.

## Conditional design events type $T(V|Q)$

They are defined by equations (45) and (47), whose application uses expressions 6 and 8. For the four defined joint design return periods, the following four peak flows are obtained from Table 8: 3 254, 4 009, 6 418, and 7 823 m$^3$/s. Adopting such flows as *conditioning* values ($X \leq x$), we proceeded by trial and error of the volume ($y$) to estimate, with Equation (47), the conditional return period that must equal that of the peak flow. The estimated volumes were: 1049, 1338, 2317, and 2920 Mm$^3$.

## Graphs of the joint return period $T'(Q, V)$

The joint return periods of type AND are estimated based on Equation (41). Once the four design joint return periods are defined, peak flows and volumes are arbitrarily selected to obtain their joint and marginal non–exceedance probabilities. The first is estimated with equations (6) and (7), and the second with Equation (8). Table 9 shows the pairs of peak flow and annual volume used to define the four graphs in Figure 1 relative to the design joint return periods of the numerical application described.

**Table 9**. Pairs of peak flow and annual volume were used to define the joint return period graphs (Figure 1) in the floods at the La Cuña hydrometric station in Mexico.

| ***T′*(*Q*, *V*) 500 years** | | ***T′*(*Q*, *V*) 1000 years** | | ***T′*(*Q*, *V*) 5000 years** | | ***T′*(*Q*, *V*) 10000 years** | |
|---|---|---|---|---|---|---|---|
| ***Vol.* Mm$^3$** | ***Qp* m$^3$/s** | ***Vol.* Mm$^3$** | ***Qp* m$^3$/s** | ***Vol.* Mm$^3$** | ***Qp* m$^3$/s** | ***Vol.* Mm$^3$** | ***Qp* m$^3$/s** |
| 0 | 3 254 | 0 | 4 009 | 0 | 6 418 | 0 | 7 823 |
| 300 | 3 254 | 400 | 4 009 | 800 | 6 418 | 1 000 | 7 823 |
| 600 | 3 249 | 600 | 4 008 | 1 000 | 6 417 | 1 200 | 7 822 |
| 800 | 3 233 | 800 | 4 003 | 1 200 | 6 415 | 1 400 | 7 821 |
| 900 | 3 212 | 1 000 | 3 988 | 1 400 | 6 409 | 1 600 | 7 820 |
| 1000 | 3 173 | 1 100 | 3 971 | 1 600 | 6 399 | 1 800 | 7 814 |
| 1050 | 3 142 | 1 200 | 3 944 | 1 800 | 6 379 | 2 000 | 7 802 |
| 1100 | 3 098 | 1 250 | 3 925 | 2 000 | 6 342 | 2 200 | 7 787 |
| 1150 | 3 034 | 1 300 | 3 900 | 2 100 | 6 313 | 2 400 | 7 759 |
| 1 200 | 2 933 | 1 350 | 3 867 | 2 200 | 6 275 | 2 600 | 7 714 |
| 1 250 | 2 756 | 1 400 | 3 823 | 2 300 | 6 222 | 2 800 | 7 641 |
| 1 275 | 2 598 | 1 450 | 3 765 | 2 400 | 6 149 | 2 900 | 7 588 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 300 | 2 289 | 1 500 | 3 682 | 2 500 | 6 046 | 3 000 | 7 524 |
| 1 310 | 2 000 | 1 550 | 3 559 | 2 550 | 5 976 | 3 100 | 7 429 |
| 1 317 | 0 | 1 600 | 3 354 | 2 600 | 5 890 | 3 200 | 7 310 |
| | | 1 625 | 3 181 | 2 650 | 5 781 | 3 300 | 7 140 |
| | | 1 650 | 2 875 | 2 700 | 5 637 | 3 400 | 6 874 |
| | | 1 673 | 0 | 2 750 | 5 431 | 3 450 | 6 685 |
| | | | | 2 800 | 5 115 | 3 500 | 6 422 |
| | | | | 2 825 | 4 863 | 3 525 | 6 322 |
| | | | | 2 850 | 4 445 | 3 550 | 6 009 |
| | | | | 2 877 | 0 | 3 575 | 5 667 |
| | | | | | | 3 600 | 5 075 |
| | | | | | | 3 619 | 0 |

On the other hand, Table 10 shows the peak flows and annual volumes defined with the hybrid univariate return periods. Such pairs of *Qp* and *V* have also been drawn in Figure 1 and are indicated by the letters "r" and "c".

**Table 10**. Design events obtained with the hybrid univariate return periods in the La Cuña hydrometric station, Mexico floods.

| Design events | Design joint return period, in years | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 500 | | 1000 | | 5000 | | 10000 | |
| | ***Qp*** | ***V*** | ***Qp*** | ***V*** | ***Qp*** | ***V*** | ***Qp*** | ***V*** |
| con regresión (r) | 3 254 | 1 118 | 4 009 | 1 383 | 6 418 | 2 224 | 7 823 | 2 717 |
| Condicional (c) | 3 254 | 1 049 | 4 009 | 1 338 | 6 418 | 2 317 | 7 823 | 2 920 |

In Figure 1 or Table 9, infinite pairs of *Qp* and *V* can be selected, which satisfy the joint design return period and define as a *subgroup of critical pairs* those that are within the curved portion of each graph of *T'*(*Q*, *V*), outside the asymptotic lines (Volpi & Fiori, 2012).

The combinations of peak flow and volume with the same joint return period define floods or *hydrographs* that will produce different effects in the reservoir that is designed or revised, adopting for safety the one that generates the most critical, severe, or unfavorable conditions. The foregoing incorporates the hydrological design, the spillway's physical characteristics, and reservoir storage or basin in the project or under review.

To form each design hydrograph, there are theoretical and empirical methods (Aldama, 2000; Aldama *et al.*, 2006; Ramírez–Orozco & Aldama, 2000; Serinaldi & Grimaldi, 2011). Campos-Aranda (2008) has exposed an empirical process that defines slender and flattened Gamma-type hydrographs.

## Other bivariate distributions

The first bivariate frequency analyses used the bivariate Normal distribution and were applied to annual floods (Yue, 1999) and daily rainstorms (Yue, 2000a). Subsequently, the Log-normal bivariate version was applied (Yue, 2000c), and the use of the so-called *Logistic model* (Equation (3)) was also generalized, which accepts Gumbel marginal distributions and has an explicit solution (Yue *et al.*, 1999; Yue, 2000b). The logistics model gained popularity for accepting as *equal* marginal functions those of extreme values: GEV, mixed Gumbel, TCEV, and mixed GEV. Currently, it is known as the Gumbel–Hougaard *copula function* and accepts *different* marginal functions (Shiau *et al.*, 2006), with the possibility of processing marginal distributions with one and/or two populations.

## Conclusions

A combined approach of maximizing the logarithmic likelihood function of the *bivariate GEV distribution* (*GEVb*) was proposed and applied, which tries to preserve the univariate fits of the marginal functions, leaving the association parameter ($m$) variable, to find the best correspondence or similarity between the observed non–exceedance probabilities and the bivariate theoretical ones. In this study, a 10 % variation was allowed in the values of the adjustment parameters ($u_1, a_1, k_1, u_2, a_2, k_2$) of the marginal functions.

The application of the *GEVb* was presented in the joint frequency analysis of the 55 peak flows and annual volumes of the floods registered in the La Cuña hydrometric station, Mexico. The adjustment of the *GEVb* allows the calculation of the univariate, joint and conditional probabilities. This part of the process concludes with contrasting the marginal functions and validating the *GEVb*, after numerical estimation of the univariate and bivariate empirical probabilities.

Then the calculation of the hybrid univariate and joint return periods was addressed. The former used a dominant design variable, for example, peak flow, and defined the associated volume by regression or estimated it by conditional probability. Figure 1 of the AND type joint return period shows the unique pairs of estimated peak flow and volume.

In the AFC with the *GEVb*, dozens of *critical hydrographs* are defined that will produce different effects in the designed or revised reservoir, adopting the one that generates the most severe or adverse conditions for safety. In this way, the physical characteristics of the reservoir under study or review are incorporated into the hydrological design.

The *GEVb* is suitable for jointly processing records of peak flow and flood volumes, which present dispersed extreme values (*outliers*), but which are not made up of mixed populations. Therefore, the *GEVb* allows to process floods of medium and large basins, of areas or regions with unique meteorological mechanisms for the formation of floods.

**Acknowledgments**

We are grateful for the corrections suggested by the anonymous referees A and C, which made it possible to amend drafting errors and include two explanations regarding the processed data. In addition, they helped to place the work in the bivariate context, of its antecedents and of its future analyzes.

## References

Aldama, A. A. (2000). Hidrología de avenidas. Conferencia Enzo Levi 1998. *Ingeniería Hidráulica en México*, 15(3), 5-46.

Aldama, A. A., Ramírez, A. I., Aparicio, J., Mejía-Zermeño, R., & Ortega-Gil, G. E. (2006). *Seguridad hidrológica de las presas en México.* Jiutepec, México: Instituto Mexicano de Tecnología del Agua.

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 optimization* (pp. 98-106). London, UK: Edward Arnold Publishers, Ltd.

Campos-Aranda, D. F. (2003). Capítulo 5: Ajuste de curvas y Capítulo 9: Optimización numérica. En: *Introducción a los métodos numéricos: software en Basic y aplicaciones en hidrología superficial* (pp. 93-127, 72-211). San Luis Potosí, México: Editorial Universitaria Potosina.

Campos-Aranda, D. F. (2008). Procedimiento para revisión (sin hidrometría) de la seguridad hidrológica de presas pequeñas para riego. *Agrociencia*, 42(5), 551-563.

Coles, S. (2001). Chapter 2: Basics of statistical modeling and chapter 3: Classical extreme value theory and models. In: *An introduction to statistical modeling of extreme values* (pp. 18-44, 45-73). London, UKA: Springer-Verlag.

Chow, V. T. (1964). Statistical and probability analysis of hydrologic data. Section 8–I: Frequency Analysis. In: Chow, V. T. (ed.). *Handbook of Applied Hydrology* (pp. 8.1-8.42). New York, USA: McGraw-Hill Book Co.

Davis, P. J. (1972). Gamma function and related functions. Chapter 6 (pp 253–296). In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of mathematical functions*. New York, USA: Dover Publications.

Escalante-Sandoval, C. A., & Raynal-Villaseñor, J. A. (1994). A trivariate extreme value distribution applied to flood frequency analysis.

*Journal of Research of the National Institute of Standards and Technology*, 99(4), 369-375.

Escalante-Sandoval, C. A., & Reyes-Chávez, L. (2002). Capítulo 9: Análisis conjunto de eventos hidrológicos. En: *Técnicas estadísticas en hidrología* (pp. 203-246). México, DF, México: Facultad de Ingeniería, Universidad Nacional Autónoma de México.

Escalante-Sandoval, C. A., & Reyes-Chávez, L. (2004). Análisis bivariado de gastos máximos anuales con distribuciones marginales TCEV. *XVIII Congreso Nacional de Hidráulica* (pp. 523-529), 10-12 de noviembre, San Luis Potosí, México.

Escalante-Sandoval, C. A. (2005). Análisis de eficiencia de la distribución Bi-Gumbel. *Ingeniería. Investigación y Tecnología*, 6(1), 13-17. DOI: 10.22201/fi.25940732e.2005.06.n1.002

Goel, N. K., Seth, S. M., & Chandra, S. (1998). Multivariate modeling of flood flows. *Journal of Hydraulic Engineering*, 124(2), 146-155. DOI: 10.1061/(ASCE)0733–9429(1998)124:2(146)

Gómez, J. F., Aparicio, M., & Patiño, C. (2010). Capítulo 6: Análisis de frecuencias bivariado para la estimación de avenidas de diseño. En: *Manual de análisis de frecuencias en hidrología* (pp. 106-127). Jiutepec, México: Instituto Mexicano de Tecnología del Agua.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5), 1049-1054. DOI: 10.1029/WR015i005p01049

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.

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. DOI: 10.1016/j.jhydrol.2006.03.004

Kite, G. W. (1977). Chapter 3: Frequency distributions. In: *Frequency and risk analyses in Hydrology* (pp. 27-39). 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.

Metcalfe, A. V. (1997). Chapter 2: Probability distributions and Monte Carlo simulations. *Statistics in Civil Engineering* (pp. 7-38). London, UK: Arnold Publishers.

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. 9-102, 29-70). Boca Raton, Florida, USA: CRC Press.

Mujumdar, P. P., & Nagesh-Kumar, D. N. (2012). *Floods in a changing climate. Hydrologic Modeling*. Cambridge, UK. International Hydrology Series (UNESCO) and Cambridge University Press.

Ramírez-Orozco, A. I., & Aldama, A. A. (2000). Capítulo 1: Teoría estadística y análisis de frecuencias conjunto. En: *Análisis de frecuencias conjunto para estimación de avenidas de diseño* (pp. 25-58). Avances en Hidráulica No. 7. México, DF, México: Asociación Mexicana de Hidráulica, Instituto Mexicano de Tecnología del Agua.

Rao, A. R., & Hamed, K. H. (2000). Theme 1.8: Tests on hydrologic data and Chapter 7: Extreme Value Distributions (pp. 12-21, 207-257). In: *Flood frequency analysis*. Boca Raton, USA: CRC Press.

Requena, A. I., Mediero, L., & Garrote, L. (2013). A bivariate return period based on copulas for hydrologic dam design: Accounting for reservoir routing in risk estimation. *Hydrology and Earth System Sciences*, 17(8), 3023-3038. DOI: 10.5194/hess–17–3023–2013

Serinaldi, F., & Grimaldi, S. (2011). Synthetic design hydrographs based on distribution functions with finite support. *Journal of Hydrologic Engineering*, 16(5), 434-446. DOI: 10.1061/(ASCE)HE.1943–5584.0000339

Shiau, J. T. (2003). Return period of bivariate distributed extreme hydrological events. *Stochastic Environmental Research and Risk Assessment*, 17(1-2), 42-57. DOI: 10.1007/s00477–003–0125–9

Shiau, J. T., Wang, H. Y., & Tsai, C. T. (2006). Bivariate frequency analysis of floods using copulas. *Journal of the American Water Resources Association*, 42(6), 1549-1564. DOI: 10.1111/j.1752–1688–2006.tb06020.x

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.

Stedinger, J. R., Vogel, R. M., & Foufoula–Georgiou, E. (1993). Chapter 18: Frequency Analysis of Extreme Events. In: Maidment, D. R. (ed.). *Handbook of Hydrology* (pp. 18.1–18.66). New York, USA: McGraw-Hill, Inc.

Teegavarapu, R. S. V. (2012). *Floods in a changing climate. Extreme precipitation*. Cambridge, UK: International Hydrology Series (UNESCO) and Cambridge University Press.

Teegavarapu, R. S. V., Salas, J. D., & Stedinger, J. R. (2019). *Statistical analysis of hydrologic variables*. Reston, USA: American Society of Civil Engineers.

Vogel, R. M., & Castellarin, A. (2017). Risk, reliability, and return periods and hydrologic design. In: Singh, V. P. (ed.). *Handbook of applied* *hydrology* (2nd ed.) (pp. 78.1-78.10). New York, USA: McGraw-Hill Education.

Volpi, E., & Fiori, A. (2012). Design event selection in bivariate hydrological frequency analysis. *Hydrological Sciences Journal*, 57(8), 1506-1515. DOI: 10.1080/02626667.2012.726357

Yue, S. (1999). Applying bivariate Normal distribution to flood frequency analysis. *Water International*, 24(3), 248-254.

Yue, S., Ouarda, T. B. M. J., Bobée, B., Legendre, P., & Bruneau, P. (1999). The Gumbel mixed model for flood frequency analysis. *Journal of Hydrology*, 226(1-2), 88-100. DOI: 10.1016/S0022–1694(99)00168–7

Yue, S. (2000a). Joint probability distribution of annual maximum storm peaks and amounts as represented by daily rainfalls. *Hydrological Sciences Journal*, 45(2), 315-326. DOI: 10.1080/02626660009492327

Yue, S. (2000b). The Gumbel mixed model applied to storm frequency analysis. *Water Resources Management*, 14(5), 377-389.

Yue, S. (2000c). The bivariate log-normal distribution to model a multivariate flood episode. *Hydrological Processes*, 14(14), 2575-2588.

Yue, S., & Rasmussen, P. (2002). Bivariate frequency analysis: Discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14), 2881-2898. DOI: 10.1002/HYP.1185

Yue, S., & Wang, C. Y. (2004). A comparison of two bivariate extreme value distributions. *Stochastic Environmental Research and Risk Assessment*, 18(2), 61-66. DOI: 10.1007/s00477-003-0124-x