DOI: 10.24850/j-tyca-14-03-01

Articles

Joint frequency analysis of peak flow and volumes of floods with Gumbel marginals

Análisis de frecuencias conjunto de gastos pico y volúmenes de crecientes con marginales Gumbel

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

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

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

Abstract

For two decades, the estimation of *Design Floods* of reservoirs has been addressed with the simplest multivariate approach, the *bivariate*. This has been accepted because it was proven that the reservoirs are not time sensitive to maximum flow, moreover, that such flow and volume are correlated with each other and the latter, with the total duration of the flood hydrograph. In this study, the bivariate Gumbel distribution or *Logistics model* was adjusted to the 61 annual data of peak flow and volume of floods entering the Adolfo Ruiz Cortines (*Mocúzari*) dam in the Río Mayo of the state of Sonora, Mexico. This process comprehends the following *eight* stages: (1) selection and testing of records to be processed; (2) verification of the randomness of the annual records; (3) acceptance of Gumbel marginal functions; (4) estimation of the joint empirical probabilities; (5) validation of the Logistic model; (6) verification of probability constraints; (7) estimation of design events, peak flow and volume, hybrid univariates, and (8) estimation of joint design events. In stage 1, first a subjective selection is made and then it is verified with the PPCC Test. Stage 2 is carried out based on the Wald-Wolfowitz Test. Stages 3 and 5 use the Kolmogórov-Smirnov Test. In stage 7, design flows are defined, and volumes are obtained by regression and conditional probability. In contrast, in stage 8, several peak flow and volume events are obtained, belonging to the *subgroup of critical pairs*, in the graphs of the joint return period $T'(Q,V)$. Towards the last part of this work, conclusions are formulated, which highlight the advantages of the bivariate joint frequency analysis and the simplicity of application and testing of the Logistic model.

Keywords: Design floods, bivariate Gumbel distributions, conditional distributions of the Logistic model, joint empirical probabilities, validation of the Logistic model, hybrid univariate return periods, joint return periods.

Resumen

Desde hace dos décadas, la estimación de las *crecientes de diseño* de los embalses se aborda con el enfoque multivariado más simple: el *bivariado*. Lo anterior se aceptó pues se demostró que los embalses no son sensibles al tiempo al gasto máximo, y que tal gasto y volumen están correlacionados entre ellos y este último con la duración total del hidrograma de la creciente. En este estudio se ajustó la distribución Gumbel bivariada o *modelo logístico* a los 61 datos anuales de gasto pico y volumen de las crecientes de entrada a la presa Adolfo Ruiz Cortines (*Mocúzari*) en el río Mayo del estado de Sonora, México. Este proceso abarca las *ocho* etapas siguientes: (1) selección y prueba de los registros por procesar; (2) verificación de la aleatoriedad de los registros anuales; (3) aceptación de las funciones marginales Gumbel; (4) estimación de las probabilidades empíricas conjuntas; (5) validación del modelo Logístico; (6) verificación de las restricciones de probabilidad; (7) estimación de eventos de diseño, gasto pico y volumen, univariados híbridos, y (8) estimación de eventos de diseño conjuntos. En la etapa 1 primero se hace una selección subjetiva y después se verifica con el Test PPCC. La etapa 2 se realiza con base en el Test de Wald-Wolfowitz. Las etapas 3 y 5 utilizan el Test de Kolmogórov-Smirnov. En la etapa 7 se definen gastos de diseño, y se obtienen volúmenes por regresión y probabilidad condicional. En contraste, en la etapa 8 se obtienen diversos eventos de gasto pico y volumen que pertenecen al *subgrupo de parejas críticas* en las gráficas del periodo de retorno conjunto $T'(Q,V)$. Por último, se formulan las Conclusiones, las cuales destacan las ventajas del análisis de frecuencias conjunto bivariado y la sencillez de aplicación y prueba del modelo Logístico.

Palabras clave: crecientes de diseño, distribuciones Gumbel bivariadas, distribuciones condicionales del modelo Logístico, probabilidades empíricas conjuntas, validación del modelo Logístico, periodos de retorno univariados híbridos, periodos de retorno conjuntos.

Received: 07/01/2021

Accepted: 21/10/2021

Introduction

Generalities

Preponderantly, the central-southern portion of the Mexican Republic is located under the area of hurricane impact that originate in the Atlantic and Pacific oceans. On the other hand, its northern portion is affected by cold fronts. Both meteorological phenomena generate rains of great magnitude, which produce *floods* or *maximum avenues*, which inundate extensive regions and endanger hydraulic works. Given this situation that repeats every year in various areas of the country, the need to study the floods becomes evident, in order to formulate measures to protect and mitigate the social, environmental and economic damage that such floods produce (Aldama, 2000; Aldama, Ramírez, Aparicio, Mejía-Zermeño, & Ortega-Gil, 2006; Domínguez & Arganis, 2012).

In relation to the hydrological dimensioning of the dikes or protection walls, bridges, rectifications and canalizations, it is necessary to define the *Design Floods* (CD, by its acronym in Spanish), or maximum flows associated with low probabilities of being exceeded. The estimation of the CDs is carried out through the *frequency analysis of floods* (AFC, by its acronym in Spanish), a process that involves representing the annual data of the floods by a probabilistic model or *probability distribution function* (FDP, by its acronym in Spanish), which is used to obtain the searched *predictions* (Kite, 1977; Bobée & Ashkar, 1991; Hosking & Wallis, 1997; Rao & Hamed, 2000; Meylan, Favre, & Musy, 2012; Stedinger, 2017).

For two decades, reservoir design floods have been analyzed as *multivariate events*, since the characteristics that define their hydrograph are correlated. The simplest approach, the *bivariate* one, is justified by the low sensitivity of the storages to the peak flow period that can be reached (Aldama, 2000) and by the correlation that such maximum flow has with the volume and the volume with the duration total hydrograph (Goel, Seth, & Chandra, 1998; Yue, 1999; Yue, Ouarda, Bobée, Legendre, & Bruneau, 1999; Yue, 2000c; Yue & Rasmussen, 2002).

Ramírez-Orozco and Aldama (2000), Escalante-Sandoval and Reyes-Chávez (2002), Volpi and Fiori (2012), Requena, Mediero and Garrote (2013), and Vogel and Castellarin (2017) highlight that the bivariate AFC leads to an infinite number of peak flow and volume combinations for an adopted *joint* exceedance probability. The foregoing implies that for the same *joint return period* there are many floods or hydrographs that would produce different effects on the designed or revised reservoir; logically adopting the one that generates the most critical or severe conditions in its storage and spillway.

Severe storms are another hydrological event that have been analyzed with a bivariate approach, whose maximum daily intensity is correlated with the amount of rain that occurred during the days of such event (Yue, 2000a; Yue, 2000b).

Objectives

The main *objective* of the study was to present the fit of the simplest bivariate probabilistic model, due to its explicit nature, with Gumbel-type marginals, called the *Logistic model*. The foregoing was developed in the following six *objectives*: (1) for such a bivariate distribution, the equations of its joint FDP and conditional probability are cited; (2) the theory behind the topic of univariate return periods with regression and conditionals, and joints is exposed; (3) describes how univariate and bivariate empirical probabilities are estimated; (4) shows how the marginal functions are verified and validates the Logistic model; (5) it concludes with the selection and testing of the records to be processed and (6) the exposed theory is applied to the 61 peak flow rates and annual volumes of the floods entering the Adolfo Ruiz Cortines Dam (*Mocúzari*), on the Mayo River. in the state of Sonora, Mexico.

Operative theory

Description of joint frequency analysis

The analysis begins by selecting random records with stationarity, annual peak flow (*Qp*) and volume (*V*) of floods, which do not show scattered values (*outliers*), nor come from different types of floods. Next, it is verified that both records can be probabilistically represented by Gumbel distributions applying the PPCC test.

Afterwards, the Gumbel distribution is fitted to each *Qp* and *V record* and its acceptance is ratified, graphically and by means of the Kolmogorov-Smirnov (KS) test. Afterwards, the estimation of the correlation between the *Qp* and *V* variables is done, in order to fit the logistic model and estimate the theoretical joint probabilities; which are contrasted with the empirical ones; again, graphically and with the KS test.

Finally, the estimation of the hybrid univariate return periods is addressed, and afterwards, the AND-type sets. Making various calculations of the latter, the graph that includes a curve for each design return period is formed. In the curved part of each graph, the critical design events are obtained.

Fitting of bivariate Gumbel distributions

There are two bivariate distributions of *explicit* extreme values, the first one is named Mixed Gumbel (GM) and the second, Logistic Gumbel (GL), both were proposed by Emil Julius Gumbel in the early sixties. The FDPs of both bivariate distributions are (Yue *et al.*, 1999; Ramirez-Orozco & Aldama, 2000; Yue & Rasmussen, 2002; Yue & Wang, 2004):

$$F_{GM}(x,y) = F_X(x) \cdot F_Y(y) \cdot \exp\left\{-\theta\left[\frac{1}{lnF_X(x)} + \frac{1}{lnF_Y(y)}\right]^{-1}\right\} \ (0 \leq \theta \leq 1) \tag{1}$$

$$F_{GL}(x,y) = \exp\left\{-\left[\left(-lnF_X(x)\right)^m + \left(-lnF_Y(y)\right)^m\right]^{1/m}\right\} \ (m \geq 1) \tag{2}$$

where, $F_X(x)$ and $F_Y(y)$ are the marginal FDPs of the random variables X and Y, whose expressions are:

$$F_X(x) = \exp\left[-\exp\left(\frac{-x-u_x}{\alpha_x}\right)\right] \qquad (x \geq 0) \tag{3}$$

$$F_Y(y) = \exp\left[-\exp\left(\frac{-y-u_y}{\alpha_y}\right)\right] \qquad (y \geq 0) \tag{4}$$

where *u* and *α* are the location and scale parameters of each Gumbel distribution, also designated *double exponential*; Its expressions according to the method of moments, based on the mean (*M*) and unbiased standard deviation (*S*) of the sample are (Kite, 1977; Rao & Hamed, 2000):

$$\alpha = 0.7797 \cdot S \quad (5)$$

$$u = M - 0.5772 \cdot \alpha \quad (6)$$

Equation (5) and Equation (6) are applicable in *X* and *Y* registers that have equal amplitude (*n*). To process records of different sizes, the maximum likelihood adjustment method, exposed by Escalante-Sandoval (2005), must be applied to obtain the five parameters of the logistic model (Equation (2)).

The inverse solution of equations (3) and (4), allows the estimation of predictions (x_p,y_p) associated with a non-exceedance probability $p = F_X(x)$ or $p=F_Y(y)$, are the following:

$$x_p = u_x + \alpha_x \cdot \{-\ln[-\ln(p)]\} \quad (7)$$

$$y_p = u_y + \alpha_y \cdot \{-\ln[-\ln(p)]\} \quad (8)$$

In both bivariate models θ and m are the *association parameters*, which describe the dependence between the two random variables, their expressions are:

$$\theta = 2 \cdot \left[1 - \cos\left(\pi \cdot \sqrt{\frac{\rho}{6}}\right)\right] \qquad (0 \leq \rho \leq 2/3) \qquad (9)$$

$$m = \frac{1}{\sqrt{1-\rho}} \qquad (0 \leq \rho \leq 1) \qquad (10)$$

where ρ is the Pearson correlation coefficient, with the following equation:

$$\rho = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \cdot \sigma_Y} \qquad (11)$$

where (μ_X, σ_X) and (μ_Y, σ_Y) are the *population* mean and standard deviation of the random variables *X* and *Y*.

Yue and Wang (2004) indicate that when the coefficient ρ ranges between zero and 2/3, both bivariate models (equations (1) and (2)) lead to identical joint probabilities and when ρ fluctuates from 2/3 to 1, only the logistic model (GL) must be used. Therefore, from now on only this probabilistic model will be used.

In this sense, Escalante-Sandoval (2005) uses the logistic model at the regional level to process floods and concludes that its predictions are less biased than the univariate estimates. Furthermore, he indicates that the univariate and joint probabilities of the logistic model must satisfy the following restriction:

$$F_X(x) \cdot F_Y(y) < F_{GL}(x,y) < min[F_X(x), F_Y(y)] \quad (12)$$

Conditional distributions of the logistic model

They are defined in an orthodox statistical way, for the variable *X* given that *Y* = *y* and similarly, for *Y* given that *X* = *x*, their expressions are (Yue & Rasmussen, 2002):

$$F(x|Y = y) = F(x,y) \cdot [EMX + EMY]^{(1-m)/m} \cdot \exp(MY + EY) \quad (13)$$

$$F(y|X = x) = F(x,y) \cdot [EMX + EMY]^{(1-m)/m} \cdot \exp(MX + EX) \quad (14)$$

being:

$$EMX = \exp[-m\,(x - u_x)/\alpha_x] \quad (15)$$

$$EMY = \exp\left[-m\left(y - u_y\right)/\alpha_y\right] \quad (16)$$

$$MY = (1 - m)\left(y - u_y\right)/\alpha_y \quad (17)$$

$$MX = (1 - m)\,(x - u_x)/\alpha_x \qquad (18)$$

$$EY = \exp[-(y - u_y)/\alpha_y] \qquad (19)$$

$$EX = \exp[-(x - u_x)/\alpha_x] \qquad (20)$$

Yue (2000b), and Yue and Rasmussen (2002) have pointed out, for hydrological practice, that the condition of X given $Y \leq y$ and Y given $X \leq x$ may be of greater interest. Such mixed model conditional distributions have been exposed by Yue *et al* (1999) and Yue (2000b) and for the logistic model by Yue and Rasmussen (2002). On the other hand, Shiau (2003) has developed the *conditionals* of the logistic model of X given $Y \geq y$ and of Y given $X \geq x$.

Hybrid univariate return periods

Univariate return periods

The classic concept of probability of an event is defined as the ratio 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 handling of the variable X, the probability of exceedance $F'_X(x)$ 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):

$$T_X = \frac{1}{F'_X(x)} = \frac{1}{1 - F_X(x)} \quad (21)$$

In the previous expression, $F_X(x)$ is the probability of non-exceedance that is estimated with Equation (3), or Equation (4) for T_Y.

When flood records are integrated with the POT (peaks-over-threshold) technique, the so-called partial duration series are formed that include more data than the number of years of record and then, according to Shiau (2003), the numerator of the previous equation must be replaced by $E(L) < 1$ or average value of the time between floods.

In the bivariate context, such as the floods one and with a strict approach, the univariate return periods are only applicable when there is a single random variable that is dominant or significant in the design criterion; or the dependence between the two variables is low or scarce (Shiau, 2003).

Conditional return periods

They are obtained by applying Equation (21), after substituting $F_X(\underline{x})$ for $F(x|Y=y)$ or for $F(y|X=x)$, that is:

$$T(x|y) = \frac{1}{1-F(x \vee Y=y)} \quad (22)$$

$$T(y|x) = \frac{1}{1-F(y \vee X=x)} \quad (23)$$

Design events obtained with regression

When it is possible to determine the dominant variable X or the one with the greatest influence (*driving variable*) in the hydrological design, then it is possible to select a return period T_X and, based on Equation (21), obtain the design value x_d, which is obtained with the Marginal FDP of X (Equation (7)). Then, based on a regression equation between the variables X and Y, x_d is applied as an independent variable and the desired y_d value is obtained. This approach does not estimate a joint return period, but provides an estimate (x_d,y_d) that has statistical support (Serinaldi & Grimaldi, 2011; Gräler *et al.*, 2013).

Design events of the conditional bivariate distribution

In this other approach, the value of x_d or prediction of the univariate return period of the dominant or prominent variable *X*, is taken to the *conditional* bivariate distribution defined as $F(y|X = x_d)$, to obtain the value of y_d sought; which generates the same return period of the variable *X*. Again, this approach does not lead to a bivariate design event that has a joint return period, but such an approach has been suggested and applied by several authors (Gräler *et al*., 2013).

Joint return periods

The first *joint return period* of the event *(X,Y)* is defined under the OR condition and Equation (21) as follows (Goel *et al*., 1998; Yue, 2000b; Shiau, 2003):

$$T(x, y) = \frac{1}{1-F(x,y)} \quad \text{being } F(x, y) = P(X \leq x \text{ ó } Y \leq y) \tag{24}$$

In the previous expression, *F*(*x*, y) is the probability of joint non-exceedance that is estimated with Equation (2). 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 AND condition, its equation is (Goel *et al*., 1998; Yue, 2000b; Shiau, 2003):

$$T'(x,y) = \frac{1}{F'(x,y)} = \frac{1}{1+F(x,y)-F_X(x)-F_Y(y)} \quad (25)$$

Aldama (2000) obtains the expression *F'*(*x*,y) of the joint probability of exceedance through a logical and simple reasoning of probabilities applied in the Cartesian graph. Instead, Yue and Rasmussen (2002) use the Cartesian graph to numerically define a bivariate event (*X*, *Y*), which can occur in any of the four quadrants, and then obtain the equation for the joint probability of exceedance (denominator of the Equation (25)).

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

$$T(x,y) \leq \text{minimum}(T_X, T_Y) \quad (26)$$

$$T'(x,y) \geq \text{maximum}(T_x, T_Y) \quad (27)$$

Various authors (Yue, 2000b; Yue & Rasmussen, 2002; Shiau, 2003; Yue & Wang, 2004; Requena *et al*., 2013; Vogel & Castellarin, 2017) have shown the graphs of the two joint return periods and have discussed their differences. Figure 1 shows the graph of the joint return period $T'(Q,V)$, built with the data from the numerical application, which is described later.

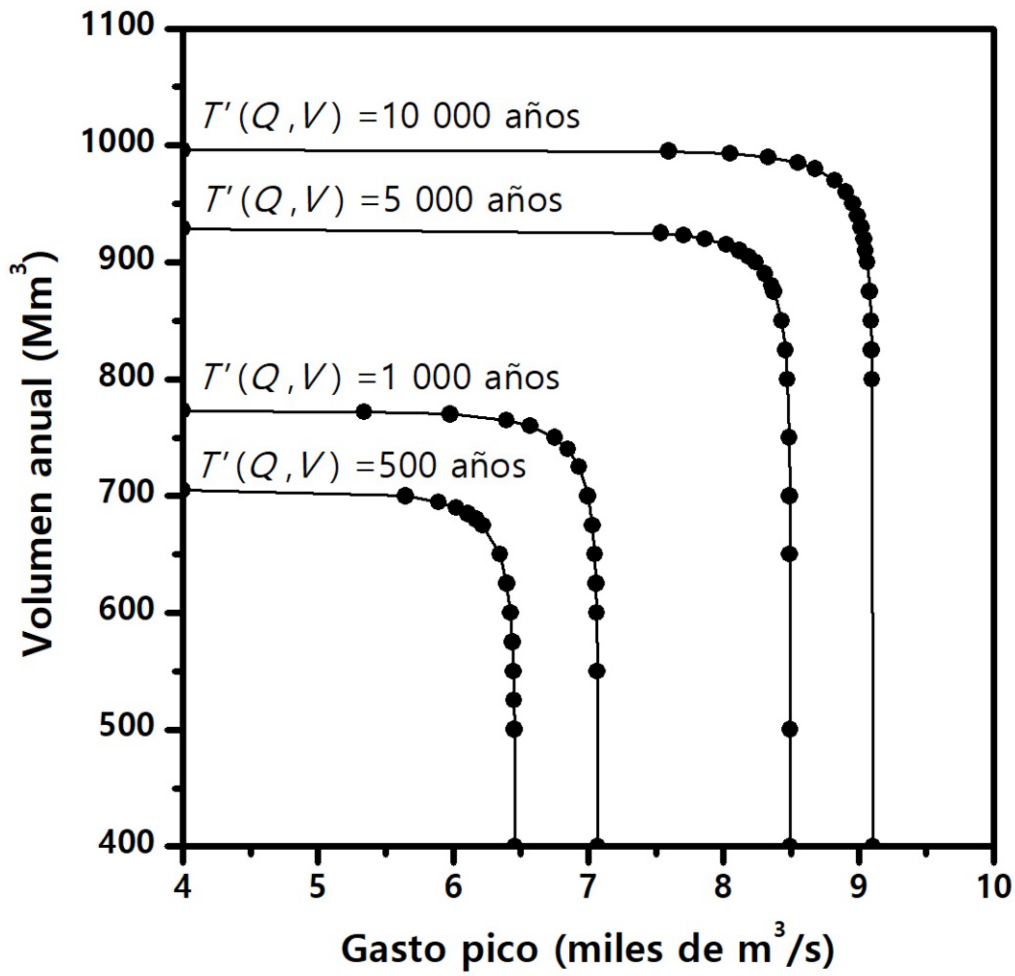

Figure 1. Graphs of the four joint return periods $T'(Q,V)$ of the design, of the peak discharge and annual flood volumes entering the Adolfo Ruiz Cortines dam (*Mocúzari*), Mexico. Abscissa axis: Peak Flow (thousands of m^3/s); ordinate axis: Annual volumen (Mm^3)"

Recently, Salas, Obeysekera and Vogel (2018) have designated the return period as the *waiting time* between the occurrence of a flood that exceeds the design flood. This concept, applied to the bivariate case, implies the estimation of the joint return period based on $F'(x, y)$; that is, with Equation (25).

Estimation of empirical probabilities

The univariate and bivariate empirical probabilities were estimated based on the Gringorten formula (Equation (28)), which according to Yue *et al*. (1999), Yue (2000b), Yue and Rasmussen (2002), and Yue and Wang (2004), leads to unbiased non-exceedance probabilities (*p*) for the Gumbel distribution or extreme values I. Such a formula is:

$$p = \frac{i - 0.44}{n + 0.12} \tag{28}$$

in which, *i* is the number of each data, when they are sorted progressively and *n* is the total number of them, or amplitude of the processed register.

For the estimation of the bivariate empirical probabilities, the same principle applied to Equation (28) was followed, but the work was done on the two-dimensional planes, with the data ordered progressively; peak flows (*Qp*) in the rows and volumes (*V*) in the columns. The plane formed is a square of *n* by *n* cells, with *n* cells on its main diagonal, when the order number of the row is equal to that of the column. Afterwards, each pair of annual data (*Qp* and *V*) is located in the two-dimensional plane and the box defined by the intersection of the row and column is identified with the number *i* that corresponds to the drawn historical year.

When the *n* pairs of data are drawn, year 1 is sought and a rectangular or square area of smaller *Qp* and *V* values is defined, where the count of numbered cells within is NM_1 or smaller combinations of *Qp* and *V*. Once the n values of NM_i are calculated, the Gringorten graphical position formula (Stedinger, 2017) is applied to calculate the joint or bivariate empirical probability:

$$F(x,y) = P(Q \leq q, V \leq v) = \frac{NM_i - 0.44}{n + 0.12} \quad (29)$$

Validation of the logistic model

Yue (2000a) indicates that the relationship between the theoretical and empirical joint probabilities of peak flow and volume allow defining the validity of the proposed joint distribution. The simplest way to represent them is to take the empirical probability of non-exceedance on the abscissa axis and the theoretical probability on the ordinate axis; Logically, each pair of data defines a point that coincides with or moves away from the line at 45º. The inspection of the graph described and the value of the correlation coefficient, in these cases, greater than 0.98, ratify the validity of the joint probabilistic model used.

Yue (2000b), and Yue and Rasmussen (2002) apply the Kolmogorov-Smirnov test with a significance level (α) of 5 %, to accept or reject the maximum difference (*dif*) between the joint probabilities. To evaluate the statistic (D_n) of the test, which is a function of the number of data (n), the expression proposed by Meylan *et al*. (2012), for α = 5 % it is:

$$D_n = \frac{1.358}{\sqrt{n}} \tag{30}$$

If *dif* is less than D_n, the joint probabilistic model or bivariate Gumbel distribution (logistic model) is accepted.

Selection of the record to process

Aldama *et al*. (2006) expose 16 records of peak flow and annual volume of the floods entering to 15 reservoirs and one in project. The General Extreme Values (GVE) distribution was fitted to such records using the L-moments method (Hosking & Wallis, 1997). The pair of records of *Qp* and *V* with the smallest shape parameters (k) was sought, which implies that they can fit a Gumbel distribution (k=0). For the entrances to the La

Boquilla dam on the Conchos River in the state of Chihuahua, -0.200 and -0.125 were found as the lowest values of *k* for *Qp* and *V*, respectively.

As such records were already processed by the author with the bivariate Normal distribution, the following were adopted with *k* values of -0.259 and -0.138 for the 61 floods at the entrance to the Adolfo Ruiz Cortines dam (*Mocúzari*) on the Mayo River in the state of Sonora, Mexico. Its basin area is 10719 km^2 and its *Qp* and *V* data are presented in Table 1, at the end of which its statistical parameters are shown: mean (*M*), standard deviation (*S*) and linear correlation coefficient (*ρ*).

Table 1. Peak flows and annual volumes of floods entering the Adolfo Ruiz Cortines dam (*Mocúzari*), Mexico (Aldama *et al*., 2006).

No.	Year	*Qp* (m^3/s)	*V* (Mm^3)	No.	Year	*Qp* (m^3/s)	*V* (Mm^3)
1	1941	602.00	64.55	33	1973	1 421.40	152.92
2	1942	774.90	47.69	34	1974	1 262.10	119.09
3	1943	880.00	24.43	35	1975	666.60	74.89
4	1944	673.60	23.39	36	1976	950.50	96.05
5	1945	392.25	16.30	37	1977	1 028.40	113.97
6	1946	1 159.25	87.77	38	1978	2 302.50	262.73
7	1947	319.35	10.92	39	1979	1 751.70	191.82
8	1948	1 212.00	72.99	40	1980	1 217.30	159.71
9	1949	6 390.00	636.47	41	1981	1 546.70	188.50
10	1950	437.0	37.99	42	1982	1 418.80	165.50
11	1951	400.00	4.64	43	1983	2 794.70	383.90

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

No.	Year	*Qp* (m^3/s)	*V* (Mm^3)	No.	Year	*Qp* (m^3/s)	*V* (Mm^3)
12	1952	573.78	45.21	44	1984	2 344.30	278.50
13	1953	368.00	10.40	45	1985	1 473.30	231.43
14	1954	273.00	65.93	46	1986	1 046.10	309.86
15	1955	1 350.80	117.71	47	1987	578.50	52.18
16	1956	469.00	58.13	48	1988	1 443.90	175.22
17	1957	1 772.30	146.97	49	1989	635.10	54.10
18	1958	1 009.60	111.71	50	1990	4 158.90	419.64
19	1959	1 661.00	182.99	51	1991	1 501.30	220.22
20	1960	3 955.00	381.46	52	1992	1 914.00	121.30
21	1961	1 517.40	238.58	53	1993	923.30	93.18
22	1962	893.20	96.62	54	1994	1 896.50	160.84
23	1963	1 534.40	216.07	55	1995	3 275.60	317.98
24	1964	1 154.20	335.24	56	1997	1 535.30	174.98
25	1965	1 755.80	157.57	57	1998	2 797.00	144.98
26	1966	1 755.80	147.12	58	1999	813.10	72.00
27	1967	1 782.20	256.19	59	2000	2 006.80	277.60
28	1968	808.30	116.01	60	2001	801.00	70.29
29	1969	813.10	72.00	61	2002	692.50	68.91
30	1970	967.00	94.66	*M*		1 477.266	158.709
31	1971	4 935.40	456.21	*S*		1 132.950	124.374
32	1972	1 326.40	195.06	ρ		0.8942	

Regarding the origin of such data, Aldama *et al*. (2006) indicate that the continuous hydrometric records, of a graphical or tabular nature, made it possible to identify the flow rate and volume data of the floods; Therefore, it follows that both values come from a flood. In addition, and in general, both the peak flow rate and its respective runoff volume include the base flow rate.

Acceptance test of the Gumbel marginals

A first test of the selected record based on the smallest value of the shape parameter (k) of the GVE distribution, consists of plotting such data on the extreme probability or Gumbel-Powell paper (Chow, 1964), to verify that they define roughly a straight line. This is a graphic test that is not very objective.

To avoid subjectivity in the search for Qp and V records that accept the Gumbel distribution, it is proposed to apply a simple test developed by Stedinger, Vogel and Foufoula-Georgiou (1993), which was established by generating random values with $k = 0$ with a GVE distribution fitted with the L moments method; it was observed that the estimator of k from such a sample has zero mean and variance equal to $0.5633/n$. With this, the authors define that when the absolute value of the Z statistic exceeds the standard Normal deviation of 1.645, it follows that k is statistically different from zero and therefore the GVE distribution must be adjusted and not the Gumbel. The equation is:

$$Z = k\sqrt{n/0.5633} \tag{31}$$

Then, to search for records that accept the Gumbel distribution, the GVE distribution is fitted to them using the method of L moments (Hosking & Wallis, 1997) and Equation (31) is applied, if the absolute value of Z does not exceed 1.645, the data can be represented by the Gumbel distribution.

The second test that was applied for the acceptance of the Gumbel marginals is the correlation coefficient of the probability graph, designated *Test PPCC* for its acronym in English (Vogel, 1986). It begins by calculating the linear correlation coefficients (Equation (11)) between the data *ordered* from lowest to highest (x_i,y_i) and their estimates with equations (7) and (8), when using as probability (p) the one estimated with the Gringorten formula (Equation (28)). Vogel (1986) calls such estimates *median* order predictions (M_i) and his statistic for peak flow (x_i) and volume (y_i) will be:

$$r_x = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(M_i-\overline{M})}{\left[\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(M_i-\overline{M})^2\right]^{1/2}} \tag{32}$$

$$r_y = \frac{\sum_{i=1}^{n}(y_i-\overline{x})(M_i-\overline{M})}{\left[\sum_{i=1}^{n}(y_i-\overline{y})^2\sum_{i=1}^{n}(M_i-\overline{M})^2\right]^{1/2}} \tag{33}$$

Vogel (1986) exposes the Critical Points equal to 1000(1-r_c) in Table 2, depending on the amplitude in years of the tested record (*n*) and the significance level of the test. From such values, the critical correlation coefficient (r_c) is obtained. If the values r_x and r_y of the records being tested exceed r_c, then, the Gumbel distribution is accepted.

Table 2. Critical Points (CP) equal to 1000(1-r_c) of the Vogel's *PPCC test* (Vogel, 1986).

n	**Significance level**			***n***	**Significance level**		
	1 %	**5 %**	**10 %**		**1 %**	**5 %**	**10 %**
30	80.9	47.4	37.8	70	49.4	28.0	21.3
40	71.4	40.6	31.1	80	47.5	25.3	19.6
50	61.1	35.4	27.1	90	44.6	23.6	18.1
60	53.3	31.5	24.0	100	40.4	22.1	16.9

For the level of significance in the 5 % test, a polynomial equation of order three (Campos-Aranda, 2003) was adjusted, with the eight data from Table 2, the following expression was obtained:

$$PC = \mathrm{C0} - \mathrm{C1} \cdot n + \mathrm{C2} \cdot n^2 - \mathrm{C3} \cdot n^3 \quad (34)$$

being :

C0 = 76.20922   C1 = 1.226496

$C2 = 9.610048 \cdot 10^{-3}$ $C3 = 2.757887 \cdot 10^{-5}$

Wald-Wolfowitz test

This nonparametric 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 flows (X_i). According to the type of hydrometric information processed, it was proposed to apply the test to peak flow records and annual volumes, which should be samples of random values. Wald and Wolfowitz (1943) based on the work of Anderson (1942) on the serial correlation coefficient developed such a test, the statistic is:

$$R = \sum_{i=1}^{n-1} x_i \cdot x_{i+1} + x_n \cdot x_1 \quad (35)$$

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:

$$E[R] = \overline{R} = \frac{S_1^2 - S_2}{n-1} \quad (36)$$

$$\mathrm{Var}[R] = \frac{S_2^2 - S_4}{n-1} + \frac{S_1^4 - 4 \cdot S_1^2 \cdot S_2 + 4 \cdot S_1 \cdot S_3 + S_2^2 - 2 \cdot S_4}{(n-1)(n-2)} - \overline{R}^2 \tag{37}$$

in which:

$$S_k = \sum_{i=1}^{n} x_i^k \tag{38}$$

Finally, U is calculated, with the equation:

$$U = \frac{R - \overline{R}}{\sqrt{\mathrm{Var}[R]}} \tag{39}$$

The value of U follows a Normal distribution (0,1) and can be used to test the independence of the series data with a level of significance α, commonly 5 %. In a two-tailed test, the standardized normal variable is $Z_{\alpha/2} \cong 1.96$; then, when the absolute value of U is less than 1.96, the series will be made up of independent values (random sample).

Results and their discussion

Univariate Predictions

Based on the statistics at the end of Table 1, equations (5) and (6) were applied to estimate the location and scale parameters of the marginal Gumbel distributions (expressions (3) and (4)), whose formulas obtained are:

$$x_p = 967.390 + 883.361\{-\ln[-\ln(p)]\} \quad (40)$$

$$y_p = 102.736 + 96.974\{-\ln[-\ln(p)]\} \quad (41)$$

From Equation (21) we get:

$$p = 1 - \frac{1}{T_X} \quad (42)$$

in which, T_X is the return period in years, same for the variables X and Y. Table 3 shows the predictions estimated with the three previous equations and the eight indicated return periods.

Table 3. Univariate predictions of the Gumbel marginal distribution in flood records of inlet at the Adolfo Ruiz Cortines dam (*Mocúzari*), Mexico.

Random variable	Return periods in years							
	10	**25**	**50**	**100**	**500**	**1 000**	**5 000**	**10 000**
Qp (m^3/s)	2 955	3 793	4 414	5 031	6 456	7 069	8 491	9 103
V (Mm^3)	321	413	481	549	705	773	929	996

Marginal tests

For the specific case of the records in Table 1, $|Z|$ = 2.695 for *Qp* and $|Z|$ = 1.436 for *V* were obtained from Equation (31). Therefore, the annual values of peak flow do not strictly accept the Gumbel distribution and the volume values can be represented by the Gumbel model. However, it should be noted that for *Qp* the absolute value of *Z* does not notably exceed the value of 1.645, as it does in records with extreme values, where the *k* value is close to -0.55; even in short records with *n*=30, the absolute value of *Z* exceeds 4.00.

Based on equations (40) and (41), the values of M_i to estimate r_x and r_y were calculated with equations (32) and (33). Values of 0.9588 was obtained for the peak flow and 0.9912 for the volumes. On the other hand, applying Equation (34) with n = 61, a PC = 30.9 was obtained, therefore, r_c = 0.9691. So, the record of peak flow does not come strictly from a Gumbel distribution ($r_x < r_c$) whereas the volume record does ($r_y > r_c$).

Randomness verification

Equations (35) to (39) were applied to the records of peak flow and annual volume in Table 1, to test their independence and stationarity. Both series were found to be *random*, with the following values for the U statistic: 0.147 and 1.189.

Graphic acceptance of marginals

First, the peak flows and volumes in Table 1 were ordered from lowest to highest. Then, their *theoretical* non-exceedance probabilities were calculated with equations (3) and (4), using the location and scale parameters of expressions (40) and (41). The *empirical* non-exceedance probabilities of both series (Qp and V) were estimated with the Equation (28).

Figure 2 and Figure 3 show the graphic contrast of probabilities for each ordered series. A much better fit is seen in the volume record. The maximum absolute differences between empirical and theoretical probabilities of peak flows and volumes were 0.123 and 0.062; the first occurred in ordered data number 49 and the second in 31. As both differences are less than D_n = 0.174 obtained with Equation (30), it is accepted that the records of *Qp* and *V* in Table 1 follow the Gumbel distribution.

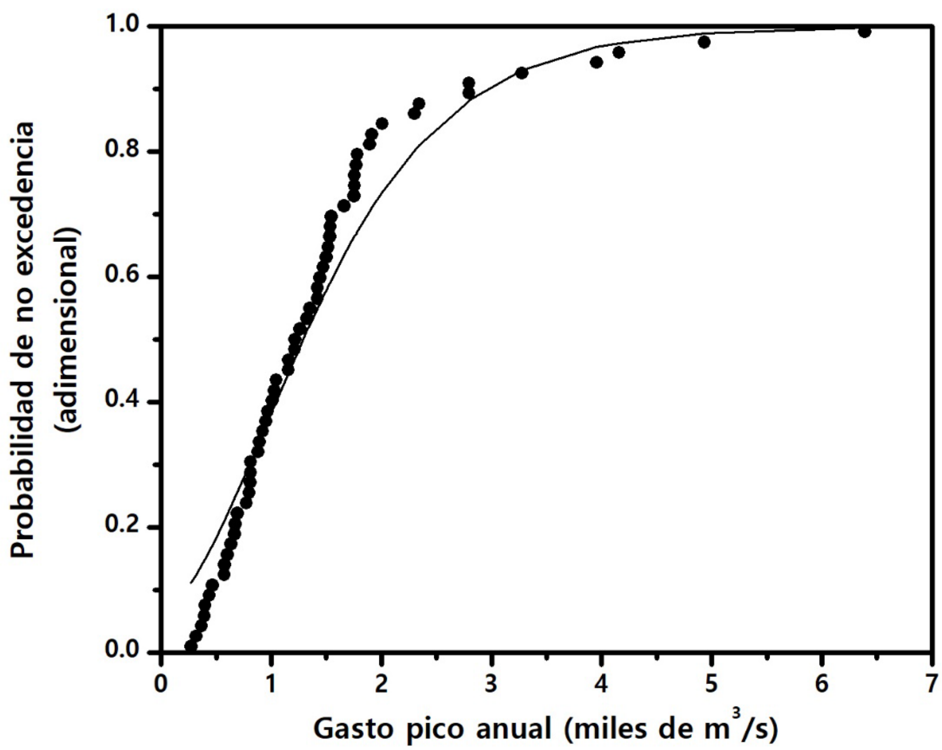

Figure 2. Gumbel marginal distribution of the annual peak flow of the flood inlet to the Adolfo Ruiz Cortines dam (*Mocúzari*), Mexico. Abscissa axis: Peak Flow (thousands of m^3/s); ordinate axis: Probability of non–exceedance (dimensionless).

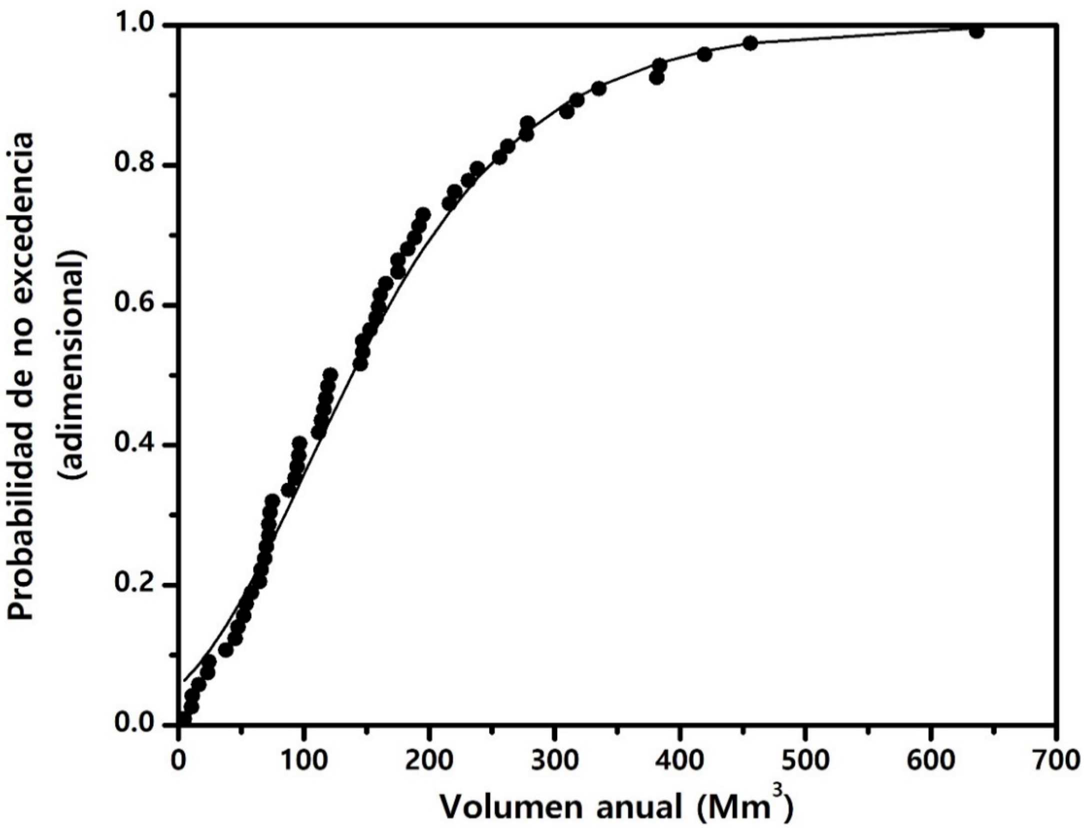

Figure 3. Gumbel marginal distribution of the inflow annual volume of the flood to the Adolfo Ruiz Cortines dam (*Mocúzari*), Mexico. Abscissa axis: Annual volumen (Mm^3); ordinate axis: Probability of non–exceedance (dimensionless).

Joint probability test

The estimation of the probability of *theoretical* joint non-exceedance was carried out by applying Equation (2), calculating beforehand the correlation (ρ) based on Equation (11), between the variables *Qp* and *V*, whose value of 0.8942 resulted in an association parameter *m* of 3.0744, according to Equation (10).

On the other hand, the *empirical* joint non-exceedance probability was estimated using Equation (29) and its described graphical procedure. Table 4 shows the results of both joint probabilities and their differences (empirical less theoretical); the maximum negative and positive differences are also shaded. From these values, it is deducted that *dif* = 0.1007.

Table 4. Joint non-exceedance probabilities and their differences for the annual floods at the entrance to the Adolfo Ruiz Cortines dam (*Mocúzari*), Mexico.

No.	*F*(*x*,*y*) theoretical	*F*(*x*,*y*) empirical	Differences	No.	*F*(*x*,*y*) theoretical	*F*(*x*,*y*) empirical	Differences
1	0.1531	0.1401	-0.0130	32	0.4945	0.5000	0.0055
2	0.1436	0.1237	-0.0199	33	0.4733	0.5000	0.0267
3	0.0981	0.0910	-0.0071	34	0.3735	0.4673	0.0938
4	0.0888	0.0746	-0.0142	35	0.1797	0.1891	0.0094
5	0.0627	0.0419	-0.0208	36	0.2697	0.3364	0.0667
6	0.2789	0.3200	0.0411	37	0.3188	0.4018	0.0830

No.	*F*(*x*,*y*) theoretical	*F*(*x*,*y*) empirical	Differences	No.	*F*(*x*,*y*) theoretical	*F*(*x*,*y*) empirical	Differences
7	0.0522	0.0092	-0.0430	38	0.7712	0.8109	0.0397
8	0.2396	0.3037	0.0641	39	0.6017	0.6145	0.0128
9	0.9958	0.9908	-0.0050	40	0.4321	0.4673	0.0352
10	0.0938	0.0746	-0.0192	41	0.5535	0.5818	0.0283
11	0.0493	0.0092	-0.0401	42	0.4931	0.5164	0.0233
12	0.1196	0.0910	-0.0286	43	0.8786	0.8927	0.0141
13	0.0544	0.0092	-0.0452	44	0.7883	0.8436	0.0553
14	0.0923	0.0092	-0.0831	45	0.5591	0.5818	0.0227
15	0.3824	0.4673	0.0849	46	0.4004	0.4346	0.0342
16	0.1225	0.0910	-0.0315	47	0.1311	0.1073	-0.0238
17	0.5062	0.5000	-0.0062	48	0.5121	0.5491	0.0370
18	0.3107	0.3855	0.0748	49	0.1416	0.1237	-0.0179
19	0.5713	0.5818	0.0105	50	0.9589	0.9581	-0.0008
20	0.9417	0.9254	-0.0163	51	0.5641	0.5818	0.0177
21	0.5758	0.6145	0.0387	52	0.4303	0.5000	0.0697
22	0.2596	0.3200	0.0604	53	0.2591	0.3200	0.0609
23	0.5723	0.5818	0.0095	54	0.5541	0.5818	0.0277
24	0.4450	0.4509	0.0059	55	0.8884	0.8927	0.0043
25	0.5333	0.5327	-0.0006	56	0.5334	0.5491	0.0157
26	0.5053	0.5000	-0.0053	57	0.5230	0.5164	-0.0066
27	0.6611	0.7618	0.1007	58	0.1990	0.2709	0.0719
28	0.2648	0.2709	0.0061	59	0.7245	0.8109	0.0864
29	0.1990	0.2709	0.0719	60	0.1939	0.2382	0.0443
30	0.2698	0.3364	0.0666	61	0.1749	0.2055	0.0306
31	0.9736	0.9745	0.0009	-	-	-	-

Validation of the probabilistic model

Figure 4 shows both joint probabilities of non-exceedance (theoretical and empirical from Table 4), observing a predominance of positive differences, that is, of points below the 45° line. The correlation coefficient was 0.990 and the value of the Kolmogorov-Smirnov test statistic is 0.1739 (Equation (30)), for which the logistic model is accepted as the joint distribution of the data in Table 1, since $dif = 0.1007 < D_n = 0.1739$.

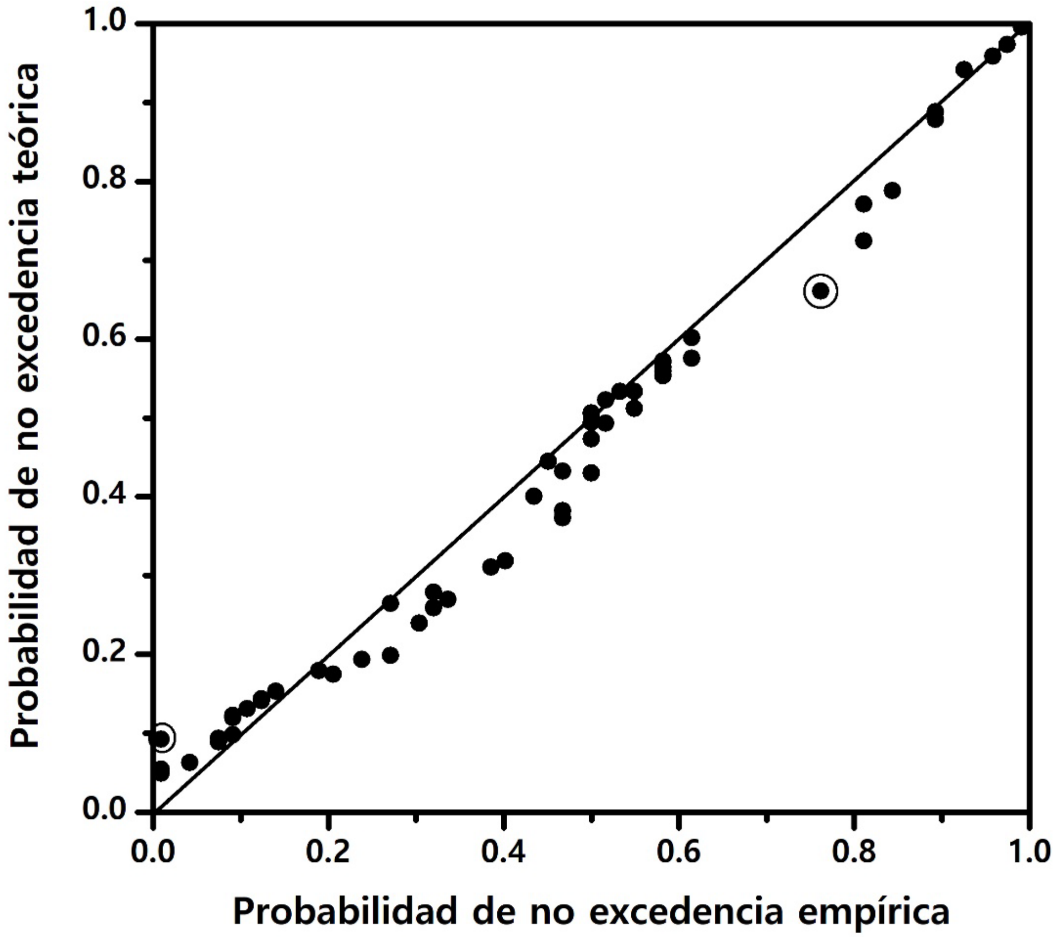

Figure 4. Graphical contrast of joint probabilities of the peak flow and volumes of the annual floods entering the Adolfo Ruiz Cortines Dam (*Mocúzari*), Mexico. Abscissa axis: Empirical non–exceedance probability; ordinate axis: Theoretical non–exceedance probability.

Probability restrictions verification

Before proceeding to estimate the joint design return periods $T'(Q,V)$, it is convenient to verify Equation (12), which establishes the probability restrictions. This is shown in Table 5 for a reduced number of historical data pairs.

Table 5. Verification of the restriction of probabilities of peak flow and annual volume of floods entering the Adolfo Ruiz Cortines Dam (*Mocúzari*), Mexico.

1	2	3	4	5	6	7
No.	***Qp***	***V***	***F*(*x*)**	***F*(*y*)**	***F*(*x*)·*F*(*y*)**	***F*(*x*,*y*)**
1	602.00	64.55	0.220	0.227	0.050	0.153
10	437.00	37.99	0.162	0.142	0.023	0.094
20	3 955.00	381.46	0.967	0.945	0.914	0.942
30	967.00	94.66	0.368	0.337	0.124	0.270
40	1 217.30	159.71	0.471	0.574	0.270	0.432
50	4 158.90	419.64	0.973	0.965	0.937	0.959
61	692.50	68.91	0.255	0.242	0.062	0.175

It was verified, in the complete Table 5, that the values of column 7 are greater than those of 6 and less than the smallest of columns 4 or 5.

Tecnología y Ciencias del Agua

Design events for hydraulic works

Assuming that in the vicinity of the entrance to the Adolfo Ruiz Cortines dam and on the Mayo River, dikes will be built to protect the flood plains and a bridge to cross it, then it is necessary to estimate design events with joint return periods of 500 and 1000 years. In addition, a review of the hydrological safety of such a reservoir will be carried out with joint return periods of 5,000 and 10,000 years. Due to the above, it is necessary to estimate peak flows and annual volumes with the 4 joint return periods mentioned.

Design events obtained with regression

The scatterplot of the 61 original data pairs (Table 1) indicated a linearly trending point cloud with a linear correlation coefficient (r_{xy}) of 0.8942. The linear regression equation that represents it is the following (Campos-Aranda, 2003):

$$V = 13.6962 + 0.09816 \cdot Qp \tag{43}$$

From Table 3 the following predictions are obtained for the peak flow (Qp) and the four joint design return periods: 6456, 7069, 8491 and 9103 m^3/s, respectively. Based on Equation (43), the following annual volumes are defined: 647, 708, 847 and 907 Mm^3, for the joint design floods.

Like Serinaldi and Grimaldi (2011), a great similarity is found between the volumes estimated with regression and their predictions in Table 3, which are: 705, 773, 929 and 996 Mm^3; that is, of the order of 9.5 % larger.

Conditional design events of type *T*(*V*|*Q*)

They are defined by Equation (23). The application of such a formula uses Equation (14) and auxiliary equations (15) to (20). For the four defined design joint return periods, the following four peak flows are obtained from Table 3: 6 456, 7 069, 8 491 and 9 103 m^3/s. Adopting such flows as *conditioning* values ($X = x$), we proceeded by trial and error of the volume (y) to estimate, with Equation (14), the respective probability of conditional non-exceedance and with it, in Equation (23), the conditional return period that must equal that of peak flow. The estimated volumes were: 911, 978, 1 185 and 1 274 Mm^3.

Plots of the joint return period *T′*(*Q*,*V*)

The joint return periods of the AND type are estimated based on Equation (25). Once the four joint design return periods have been defined, peak flows and volumes are arbitrarily selected to obtain their marginal and joint probabilities of non-exceedance. The former are estimated with equations (3) and (4) and the latter with Equation (2), from the logistic model.

Table 6 shows the couples of peak flow and annual volume used to define the four graphs of Figure 1, relative to the joint design return periods of the numerical application described.

Table 6. Couples of peak flow and annual volume used to define the graphs of the joint return period (Figure 1), in the floods at the entrance to the Adolfo Ruiz Cortines Dam (*Mocúzari*), 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**
400	6456	400	7069	400	8491	400	9103
500	6453	550	7065	500	8491	800	9099
525	6450	600	7062	650	8490	825	9095

***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**
550	6446	625	7056	700	8489	850	9091
575	6438	650	7047	750	8485	875	9081
600	6423	675	7030	800	8471	900	9062
625	6397	700	6997	825	8457	910	9051
650	6344	725	6929	850	8429	920	9038
675	6217	740	6847	875	8373	930	9018
680	6170	750	6751	880	8356	940	8992
685	6108	760	6570	890	8308	950	8956
690	6022	765	6392	900	8237	960	8904
695	5890	770	5976	905	8185	970	8823
700	5648	772	5340	910	8118	980	8679
705	4000	773	4000	915	8020	985	8552
				920	7864	990	8332
				923	7706	993	8049
				925	7537	995	7595
				929	4000	996	4000

Selection of design events

Table 7 shows the peak flows and annual volumes defined with the hybrid univariate return periods. Such pairs of *Qp* and *V* show the amplitude of variation that the annual volume can have, since they were defined by adopting the peak flow as the preponderant or dominant variable.

Table 7. Couples of design events obtained with the hybrid univariate return periods in the floods at the entrance to the Adolfo Ruiz Cortines Dam (*Mocúzari*), Mexico.

Design event	Joint design return period, in years							
	500		1000		5000		10000	
	Qp	***V***	***Qp***	***V***	***Qp***	***V***	***Qp***	***V***
with regression	6456	647	7069	708	8491	847	9103	907
conditional	6456	911	7069	978	8491	1185	9103	1274

On the other hand, in Figure 1 or in Table 6 many pairs of *Qp* and *V* can be selected, which satisfy the joint design return period. In relation to the above, Volpi and Fiori (2012) define as a *subgroup of critical couples* those that are inside the curve of each graph of $T'(Q,V)$; that is, those that are not defined in the asymptote lines.

From the **Introduction** it was indicated that the bivariate FFA leads, through Equation (25) of the $T'(Q,V)$, to an infinity of combinations of peak flow and volume that generate the same *joint return period* and therefore, there are many floods or *hydrographs* that will produce different effects in the reservoir that is designed or revised; adopting for security, the one that generates the most critical, severe or unfavorable conditions. The foregoing is incorporating in the hydrological design, the physical characteristics of the spillway and lake of the reservoir in project or under review.

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

Conclusions

The application of the *logistic model* (ML) or bivariate Gumbel distribution, in the joint frequency analysis of the 61 peak flows and annual volumes of the floodwaters that enter the Adolfo Ruiz Cortines dam (*Mocúzari*), in Sonora, Mexico, was presented. The ML fit allows the calculation of univariate, joint and conditional probabilities. This part of the process concludes with the contrast of the marginal functions and the validation of the ML, after graphical estimation of the univariate and bivariate empirical probabilities.

Then, the calculation of the hybrid and joint univariate return periods was addressed. The former uses a dominant design variable, for example, peak flow, and define the associated volume by regression; or they estimate it through conditional probability (equations (14) and (23)). These pairs of peak flow and volume that are defined are unique for each design return period. On the other hand, the search for the infinite pairs of peak flow and volume (*Qp* and *V*) that generate the same *joint return period* with Equation (25), allow the construction of its graph, shown in Figure 1.

In the bivariate AFC, dozens of pairs of *Qp* and *V* are defined, which will generate *critical hydrographs* that will produce different effects in the reservoir that is designed or revised; adopting for security, the one that generates the most severe or unfavorable conditions. In this way, the physical characteristics of the reservoir and its spillway under study or review are being incorporated into the hydrological design.

The ML is a bivariate Gumbel distribution suitable for simply and jointly processing peak flow records and flood volumes, which do not present scattered extreme values (*outliers*) or mixed populations. Its universality lies in the aforementioned limitation, to process floods of medium and small basins, of regions with unique meteorological mechanisms of formation of the floods.

Due to its explicit nature, ML is extremely simple to apply and leads to important results in joint frequency analyses. Due to its limitation of having Gumbel marginals, it would be convenient to identify in which regions of the country, the records of *Qp* and *V* of its annual floods, can be modeled by such FDP and thus have an idea of its potential application in Mexico.

Acknowledgments

The observations and corrections suggested by the anonymous referees B and D and from the editor are appreciated, which allowed us to avoid omissions of background and theory, in relation to the bivariate frequency analyses. Such corrections also helped define the scope of the study and its possible future lines of research.

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.

Anderson, R L. (1942). Distribution of the serial correlation coefficient. *The Annals of Mathematical Statistics*, 13(1), 1-13. DOI: 10.1214/aoms/1177731638

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.

Campos-Aranda, D. F. (2003). Capítulo 5. Ajuste de curvas. En: *Introducción a los métodos numéricos: software en Basic y aplicaciones en hidrología superficial* (pp. 93-127). 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.

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

Domínguez, R., & Arganis, M. L. (2012). Validation of method to estimate design discharge flow for dam spillways with large regulating capacity. *Hydrological Sciences Journal*, 57(3), 460-478. DOI: 10.1080/02626667.2012.665993

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. (2005). Análisis de eficiencia de la distribución Bi-Gumbel. *Ingeniería. Investigación y Tecnología*, 6(1), 13-17.

Goel, N. K., Seth, S. M., & Chandra, S. (1998). Multivariate modeling of flood flows. *Journal of Hydraulic Engineering*, 124(2), 146-155.

Gräler, B., Van-den-Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De-Baets, B., & Verhoest, N. E. C. (2013). Multivariate return periods in hydrology: A critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, 17(4), 1281-1296. DOI: 10.5194/hess-17-1281-2013

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.

Kite, G. W. (1977). Chapter 8. Type I extreme distribution and Chapter 12. Comparison of frequency distributions In: *Frequency and risk analyses in hydrology* (pp. 87-104, 156-168). Fort Collins, USA: Water Resources Publications.

Meylan, P., Favre, A. C., & Musy, A. (2012). Chapter 3. Selecting and checking data series. In: *Predictive hydrology. A frequency analysis approach* (pp. 29-70). Boca Raton, USA: CRC 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 (pp. 12-21) and Theme 7.2. The Extreme Value type I EV(1) distribution. In: *Flood frequency analysis* (pp. 229-241). 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

Salas, J. D., Obeysekera, J., & Vogel, R. M. (2018). Techniques for assessing water infrastructure for nonstationary extreme events: A review. *Hydrological Sciences Journal*, 63(3), 325-352. DOI: 10.1080/02626667.2018.1426858

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

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.

Vogel, R. M. (1986). The probability plot correlation coefficient test for Normal, lognormal and Gumbel distributional hypotheses. *Water Resources Research*, 22(4), 587-590.

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

Wald, A. & Wolfowitz, J. (1943). An exact test for randomnesss in the non–parametric case based on serial correlation coefficient. *The Annals of Mathematical Statistics*, 14(4), 378–388. DOI: 10.1214/aoms/1177731358

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.

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 lognormal 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