DOI: 10.24850/j-tyca-14-05-03

Articles

Selection and application of Copula functions with dependence on its tail right to the joint frequency analysis (*Q, V*) of annual floods

Selección y aplicación de funciones Cópula con dependencia en su extremo derecho al análisis de frecuencias conjunto (Q, V) de crecientes anuales

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

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

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

Abstract

The hydrological design of the *reservoirs* to be built or the revision of the existing ones, requires the estimation of the so-called *Design Flood Hydrograph*. The simplest and most approximate way to estimate such a hydrograph, for a determined *joint return period*, is through the *bivariate*

Frequency Analysis of the maximum flow (*Q*) and the annual runoff volume (*V*) of the registered floods. *Copula Functions* (FC) are probabilistic models based on the dependence between *Q* and *V*, which easily establish their bivariate distribution, based on previously adopted marginal functions or distributions of any type, equal or different. The application of the FC in the hydrological estimates showed that a decisive aspect in their ideal selection is related to the *dependence* on the extreme right of the data (λ_U^{CFG}) and that which have (λ_U) certain FC. Therefore, in this study the FC are exposed: Student's *t*, Gumbel-Hougaard, Clayton Associate and Joe, which show increasing values λ_U. The values of λ_U are contrasted against the λ_U^{CFG} obtained in 16 random real records of *Q* and *V*, to establish the applicability of each cited FC. In addition, the record of 26 annual *Q* and *V* data of the inflow floods to the Adolfo López Mateos Dam (*Humaya*), in the state of Sinaloa, Mexico, is processed as a numerical application. Finally, the Conclusions are presented, which highlight the advantages of FC in the bivariate frequency analysis of floods.

Keywords: Copula Functions, Kendall's tau quotient, Spearman's rho coefficient, dependence at the upper tail, observed dependence, joint return periods, secondary return period.

Resumen

El diseño hidrológico de los *embalses* por construir o la revisión de los ya existentes requiere la estimación del llamado *hidrograma de la creciente de diseño*. La manera más simple y aproximada para estimar tal hidrograma, para un determinado *periodo de retorno conjunto*, es a través del *análisis de frecuencias bivariado* del gasto máximo (*Q*) y el volumen escurrido (*V*) anuales de las crecientes registradas. Las *funciones Cópula* (FC) son modelos probabilísticos basados en la dependencia entre *Q* y *V*, que establecen fácilmente su distribución bivariada con base en las funciones marginales previamente adoptadas o distribuciones de cualquier tipo, iguales o diferentes. La aplicación de las FC en las estimaciones hidrológicas mostró que un aspecto decisivo en su selección idónea está relacionado con la *dependencia* en el extremo derecho de los datos (λ_U^{CFG}) y la que tienen (λ_U) ciertas FC. Por lo anterior, en este estudio se exponen las FC: *t* de Student, Gumbel-Hougaard, Clayton Asociada y Joe, que muestran valores de λ_U que van en aumento. Se contrastan los valores de λ_U contra los λ_U^{CFG} obtenidos en 16 registros reales aleatorios de *Q* y *V*, para establecer la aplicabilidad de cada FC citada. Además, se procesa como aplicación numérica el registro de 26 datos de *Q* y *V* anuales de las crecientes de entrada a la presa Adolfo López Mateos (*Humaya*) del estado de Sinaloa, México. Por último, se exponen las conclusiones, las cuales destacan las ventajas de las FC en los análisis de frecuencias bivariados de crecientes.

Palabras clave: funciones Cópula, cociente tau de Kendall, coeficiente rho de Spearman, dependencia en el extremo superior, dependencia observada, periodos de retorno conjuntos, periodo de retorno secundario.

Received: 30/12/2021

Accepted: 07/03/2022

Introduction

Generalities

The Mexican Republic is located within the area of hurricane influence which originate in the Caribbean Sea and the Pacific Ocean. In addition, it is affected by other meteorological phenomena of great spatial scope, such as cold fronts and extensive convective storms, some of orographic origin. The aforementioned atmospheric phenomena produce rains of great magnitude, which generate *floods* or *maximum flows* that inundate extensive regions of the country. Such floods can cause loss of human lives and tremendous economic and environmental damages. This scenario that repeats every year, in different areas of Mexico, emphasizes the importance of studying the floods, in order to understand their dynamics and try to quantify their magnitudes, with the purpose of designing building and protecting hydraulic works for protection and utilization (Aldama, 2000; Aldama, Ramírez, Aparicio, Mejía-Zermeño, & Ortega-Gil, 2006; Gómez, Aparicio, & Patiño, 2010).

The main hydraulic protection works are the dams and flood bariers, which are generally related to rectifications and channeling. Bridges and other urban drainage works are also protection structures, whose hydrological design is based on the *Design Floods* (DF), which are maximum river flows associated with low probabilities of being exceeded. The most reliable estimation of the DFs is made through the *Flood*

Frequencies Analysis (AFC, by its acronym in Spanish), which tries to interpret the record of maximum annual flows available, in terms of future events of a certain probability of occurrence (Bobée & Ashkar, 1991; Rao & Hamed, 2000; Meylan, Favre, & Musy, 2012).

For the results of the AFC to be approximate and relatively reliable, the following four requirements must be met: (1) the recording of maximum annual flows must be random; (2) the *probability distribution function* (PDF) that represents it must be suitable; (3) the fit of the tested PDFs must be carried out with efficient statistical methods and (4) the selection of results must be objective (Hosking & Wallis, 1997; Stedinger, 2017).

On the other hand, the hydrological design of *reservoirs* or storage dams requires the estimation of the so–called *Hydrograph of the Design Flood*, since in such hydraulic works the volume of the flood is extremely important to establish its dangerousness and define its management and temporary storage without risk. Aldama (2000) indicated that the reservoirs are not sensitive to the time that elapses in reaching the maximum flow, but to its magnitude and the volume of the flood. The later makes it possible to significantly simplify the *multivariate analysis* of the floods, since the *bivariate approach* of the annual maximum flow (Q) and volume (V) allows the hydrograph of the design flood to be defined with sufficient approximation, since V has a positive correlation with the total duration (D) of the hydrograph.

Bivariate flood analysis began at the end of the last century, with the work of Goel, Seth and Chandra (1998), Yue (1999), and Yue, Ouarda, Bobée, Legendre and Bruneau (1999), who processed the data sets of Q

and *V* and of *V* and *D*. At the beginning of this century, Yue and Rasmussen (2002) established the basic hydrological concepts of bivariate probabilistic analysis. The first flood bivariate analysis used the bivariate Normal distribution and the Logistic model that accepts marginal functions of extreme values such as the Gumbel and GVE distributions. These models have two important constraints: (1) their marginal distributions are equal and (2) the adjustment parameters of such marginals intervene in the joint modeling of the *Q* and *V* register (Favre, El Adlouni, Perreault, Thiémonge, & Bobée, 2004).

Currently, such constraints are avoided by applying the *Copula Functions* (*CF*), mathematical models that accept any PDF as marginal, to form the bivariate or multivariate distribution and which are based on the correlation that exists between *Q* and *V* (Shiau, Wang, & Tsai, 2006; Zhang & Singh, 2006). The use of *CF*s has been consolidated, as proven by specialized texts (Salvadori, De Michele, Kottegoda, & Rosso, 2007; Zhang & Singh, 2019; Chen & Guo, 2019; Chowdhary & Singh, 2019).

Ramírez-Orozco and Aldama (2000), Escalante-Sandoval and Reyes-Chávez (2002), as well as Volpi and Fiori (2012), and Requena, Mediero and Garrote (2013) indicate that the bivariate AFC leads to an infinity of peak flow and volume combinations for an adopted *joint* exceedance probability. This implies that for the same *joint return period* there are many floods or *hydrographs* that will produce different effects on the reservoir under design or revision; logically deciding for the one that generates the most severe conditions in its storage and spillway of excesses.

Objectives

The works of Dupuis (2007), and Poulin, Huard, Favre and Pugin (2007) showed that a decisive aspect in the selection of the *ideal CF* to use is related to the *dependence* on the right extreme of the data and the one shown by certain *CF* families. Therefore, the *basic objectives* of this study were the following four: [1] expose four families of *CF* that have significant dependence on their right tail (λ_U); [2] describe how the λ_U is estimated from the available *Q* and *V* record (λ_U^{CFG}); [3] to contrast the values of λ_U^{CFG} obtained in 16 real *Q* and *V records*, against those which define the four exposed *CF* and [4] to process as a numerical application of 26 data records of annual *Q* and *V* of floods at the entrance to the Adolfo López Mateos Dam (*Humaya*), in the state of Sinaloa, Mexico.

Copula Functions

Concept and definition

As already indicated, the key advantage of the *Copula Functions* (*CF*) consists in allowing to express a joint distribution of correlated random variables, as a function of their previously adopted marginal distributions. So, a *CF* links or relates the univariate marginal distributions to form a multivariate distribution. Another basic advantage of *CF*s when forming multivariate distributions is the fact that they separate the effects of dependency among the random variables from those of the marginal distributions in joint modeling.

Due to the above, the construction of the multivariate distribution is reduced to the study of the relationship between the correlated variables, provided the univariate marginal distributions are known. The use of *CF* offers complete freedom to adopt or select the univariate marginal distributions which best represent the data (Shiau *et al.*, 2006; Kottegoda & Rosso, 2008; Meylan *et al.*, 2012; Zhang & Singh, 2019).

Since in this study *CF* will be applied in the *bivariate frequency analysis* of annual floods, the following definition refers to two correlated random variables X and Y, whose joint cumulative probability distribution function is $F_{X,Y}(x,y)$ with univariate marginal distributions $F_X(x)$ and $F_Y(y)$; then the *CF* exists, it is $C[\cdot]$ and it is such that:

$$F_{X,Y}(x,y) = C[F_X(x), F_Y(y)] \tag{1}$$

The above equation defines the basic concept for the development of *CF*s and is known as Sklar's Theorem exposed in 1959 (Shiau *et al.*, 2006; Kottegoda & Rosso, 2008; Meylan *et al.*, 2012; Zhang & Singh, 2019).

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Copula Functions Families

The *Copula functions* (*CF*) that have been developed have been classified into four classes: Archimedean, extreme value, elliptic, and miscellaneous. They are also classified as single–parameter or multi-parameter copulas, depending on the extent to which the dependency structure between the variables *Q* and *V* is defined (Meylan *et al.*, 2012; Chowdhary & Singh, 2019).

Salvadori *et al.* (2007) provide a comprehensive and useful summary of *CF*s that have been applied in the field of hydrology. In this regard, in the following relationship the families of Ali–Mikhail–Haq Copulas, called AMH Copula, have not been included, since they restrict the dependence to a Kendall tau quotient which ranges from −0.1817 to 0.3333 and the Farlie–Gumbel–Morgenstern family, designated FGM, which restricts it further, to the interval of ±0.2222.

Archimedean copulas have found broad application due to their simple construction, single parameter, wide range, and acceptance of both types of dependency (positive and negative). Designating $F_X(x)=u$, $F_Y(y)=v$ and θ the parameter that measures the dependence or association between u and v, we have the following two families of Archimedean Copulas, which accept negative and positive dependence (Shiau *et al.*, 2006; Genest & Favre, 2007; Salvadori *et al.*, 2007; Michiels & De Schepper, 2008; Zhang & Singh, 2019; Chen & Guo, 2019; Chowdhary & Singh, 2019).

1. **Clayton**. This Copula function is called Cook–Johnson by Zhang and Singh (2006). Sraj, Bezak and Brilly (2015) cite it with both names. Its equation and variation space of θ are:

$$C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta} \quad [-1,\infty) \setminus \{0\} \qquad (2)$$

For the positive dependence $\theta > 0$ and for the negative $-1 \leq \theta < 0$, with $\theta = 0$ for the independence between *u* and *v.* The relationship of θ with the Kendall tau quotient is as follows:

$$\tau_n = \frac{\theta}{\theta+2} \qquad (3)$$

2. **Frank**. Its equation and variation space of θ are:

$$C(u,v) = -\frac{1}{\theta}\ln\left[1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right] \quad (-\infty,\infty) \setminus \{0\} \qquad (4)$$

For the negative dependence $0 \leq \theta < 1$ and for the positive $\theta > 1$, with $\theta = 1$ for the independence between *u* and *v*. The relationship of θ with τ_n is as follows:

$$\tau_n = 1 + \frac{4}{\theta}[\mathrm{D}_1(\theta) - 1] \qquad (5)$$

where $D_1(\theta)$ is the Debye function of order 1, whose expression is:

$$D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{s}{e^s - 1} ds \tag{6}$$

The previous equation was estimated by means of numerical integration, ratifying its results with the values tabulated by Stegun (1972).

Next, a family of Extreme Value Copulas is cited, which accepts only positive dependence.

3. **Gumbel–Hougaard**. Its equation and variation space of θ are:

$$C(u, v) = \exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right\} \ [1, \infty) \tag{7}$$

With $\theta = 1$ we have independence between u and v. The relationship of θ to Kendall's tau ratio is as follows:

$$\tau_n = \frac{\theta - 1}{\theta} \tag{8}$$

Lastly, two families of Miscellaneous class Copulas are cited.

4. **Plackett** accepts negative ($\theta < 1$) and positive ($\theta > 1$) dependence, with $\theta = 1$ for independence between *u* and *v*:

$$C(u,v) = \frac{1+(\theta-1)(u+v)}{2(\theta-1)} - \frac{\sqrt{[1+(\theta-1)(u+v)]^2 - 4uv\theta(\theta-1)}}{2(\theta-1)} \text{ con } 0 \leq \theta \leq \infty \quad (9)$$

The relationship of θ to Spearman's rho is as follows:

$$\rho_n = \frac{(\theta+1)}{(\theta-1)} - \frac{2\theta\, ln(\theta)}{(\theta-1)^2} \quad (10)$$

5. **Raftery** accepts only positive dependence and θ varies between 0 and 1, with $\theta = 0$ for independence between *u* and *v*:

$$C(u,v) = min(u,v) + \frac{1-\theta}{1+\theta}(uv)^{1/(1-\theta)}\left\{1 - [max(u,v)]^{-(1+\theta)/(1-\theta)}\right\} \quad (11)$$

The relationship of θ to Kendall's tau quotient and to Spearman's rho coefficient are as follows:

$$\tau_n = \frac{2\theta}{3-\theta} \quad (12)$$

$$\rho_n = \frac{\theta(4-3\theta)}{(2-\theta)^2} \quad (13)$$

The families of Elliptic Copulas are distributions and their marginals that accept the property of being elliptic, such as the Normal, Student's *t* and Cauchy distributions. A random vector $X \in R^d$ is said to have an elliptic distribution if it admits the following stochastic representation: $X = \mu + RAU$; in which, $\mu \in R^d$, *R* is an independent random variable of ***U***, which is a uniform random vector on the unit sphere in R^d and ***A*** is a fixed *dxd* matrix, such that $\boldsymbol{A}\cdot\boldsymbol{A}^T$ is nonsingular (Salvadori *et al.*, 2007; Chen & Guo, 2019).

When the marginal distributions are of different types and possibly not elliptic, the elliptic Copulas are called *Meta–elliptic Copulas*, whose main families are Gaussian and Student's *t* (Chowdhary & Singh, 2019). The Student's *t*–Copula will be described in detail later.

Families of associated Copulas

The four existing classes of Copula families and in the case of the bivariate ones, have three *associated families*, defined by the following transformations (Salvadori & De Michele, 2004; Michiels & De Schepper, 2008; Chowdhary & Singh, 2019):

$$C'(u,v) = u - C(u,1-v) \quad (14)$$

$$C''(u,v) = v - C(1-u,v) \quad (15)$$

$$\hat{C}(u,v) = u + v - 1 + C(1-u, 1-v) \quad (16)$$

The first two transformations change the nature of the dependency from negative to positive and vice versa. The third transformation is called the Copula of Survival (*Survival Copula*), because it is related to the joint distribution of the exceedance probabilities, of the pair of random variables *u* and *v* whose Copula is *C*. Furthermore, all three families of associated Copulas have the same value of Kendall's tau ratio; the first two with their sign changed.

Dupuis (2007) highlighted the importance of the **Associated Clayton** Copula, as it has a significant dependence on its right tail defined by the following Equation (17) and whose formula, according to equations (2) and (16), is Equation (18):

$$\lambda_U = 2^{-1/\theta} \quad (17)$$

$$\hat{C}(u,v) = u + v - 1 + \left[(1-u)^{-\theta} + (1-v)^{-\theta} - 1\right]^{-1/\theta} \quad (18)$$

Numerical integration

To quantify various integrals, for example, Equation (6), a numerical integration was performed based on the Gauss–Legendre quadrature method, whose operating equation is (Nieves & Domínguez, 1998; Campos-Aranda, 2003):

$$\int_a^b f(x)\,dx \cong \frac{b-a}{2}\sum_{i=1}^{np} w_i \cdot f\left[\frac{(b-a)h_i+b+a}{2}\right] \tag{19}$$

in which, w_i are the method coefficients whose abscissas are h_i and np the number of pairs in which the function $f(x)$ is evaluated, with the argument indicated in $f\,[\cdot]$. Davis and Polonsky (1972) obtained the 12 used pairs of w_i and h_i with 15 digits, which are acceptable in the Basic language, as double precision variables.

The operational equation of the bivariate integration is the following double integral (Nieves & Domínguez, 1998):

$$\int_c^d \int_a^b f(x,y)dxdy \cong \frac{(b-a)(d-c)}{4}\sum_{j=1}^{np}\sum_{i=1}^{np} w_j\, w_i f\left(\frac{(b-a)h_i+b+a}{2},\frac{(d-c)h_j+c+d}{2}\right) \tag{20}$$

Davis and Polonsky (1972) obtained the 20 pairs used in this double integral of w_i and h_i with 15 digits.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Association measures

Concordance

As the *CF* characterizes the *dependence* between the random variables u and v, it is necessary to study the measures of association, in order to have a method that allows estimating its parameter θ. In general terms, one random variable is *concordant* with another, when its large values are associated with the large values of the other and the small values of one with the small values of the other (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019).

Some variables with direct linear correlation will be concordant, since when one increases the other also does. Variables with inverse linear correlation will be *discordant*, since large values of one will correspond to small values of the other. Then, (x_1,y_1) and (x_2,y_2) are said to be concordant if $x_1<x_2$ and $y_1<y_2$, or $x_1>x_2$ and $y_1>y_2$. They are discordant if $x_1<x_2$ and $y_1>y_2$, or $x_1>x_2$ and $y_1<y_2$. This implies that the pairs $(x_1-x_2)(y_1-y_2)>0$ are *concordant* (c) and *discordant* (*d*) when $(x_1-x_2)(y_1-y_2)<0$ (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019).

A numerical measure of association is a statistic that indicates the degree of dependence or association of the variables. For comparison purposes, such measures range from zero to +1 or −1, indicating the perfect positive association at +1 or negative at −1. Kendall's tau quotient and Spearman's rho coefficient are two nonparametric measures that provide information on a special form of association or dependence, known as *concordance* (Salvadori *et al.*, 2007; Chen & Guo, 2019).

Kendall's tau quotient

It measures the probability of having concordance pairs, so it is the *quotient* of *c*–*d* to *c*+*d*. The expression to estimate it with bivariate data is (Zhang & Singh, 2006; Zhang & Singh, 2019):

$$\tau_n = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{signo}\left[\left(x_i - x_j\right)\left(y_i - y_j\right)\right] \quad (21)$$

in the above equation, *n* is the number of observations and the sign[·] is +1 if such pairs are concordant and −1 if they are discordant.

Genest and Favre (2007) present a test for the tau quotient, accepting that the null hypothesis H_0 indicates that *X* and *Y* are independent and then the statistic has an approximately Normal distribution with zero mean and variance 2(2*n*+5)/[9*n*(*n*–1)]. Then, H_0 will be rejected with a confidence level *α* = 5 % if:

$$\sqrt{\frac{9n(n-1)}{2(2n+5)}}\left|\tau_n\right| > Z_{\alpha/2} = 1.96 \quad (22)$$

Spearman's rho coefficient

It measures the correlation between pairs of ranks (R_i,S_i) of the random variables X_i and Y_i. Therefore, it is equivalent to the Pearson correlation coefficient. Its expression to estimate it in a bivariate record of size *n* is as follows (Chowdhary & Singh, 2019; Zhang & Singh, 2019):

$$\rho_n = \frac{12}{n(n+1)(n-1)} \sum_{i=1}^{n} R_i \cdot S_i - 3\frac{n+1}{n-1} \tag{23}$$

Salvadori *et al.* (2007) present another formula for its calculation from a bivariate record with *n* data; which coincides with that used by the WMO (1971):

$$\rho_n = 1 - \frac{6\sum_{i=1}^{n}(R_i - S_i)^2}{n^3 - n} \tag{24}$$

Genest and Favre (2007) also present a similar test for the rho coefficient, whose distribution is close to Normal with zero mean and variance $1/(n-1)$; therefore, H_0 is rejected with a level α = 5 %, if:

$$\sqrt{n-1} \cdot |\rho_n| > 1.96 \tag{25}$$

Estimation of the dependency parameter

The simplest method to estimate the *CF* parameter θ is similar to the method of moments and is based on inverting the equation that relates θ to Kendall's tau quotient or Spearman's rho coefficient. The other two available procedures are called: (1) maximum pseudo–likelihood method and exact maximum likelihood method (Meylan *et al.*, 2012; Chowdhary & Singh, 2019; Zhang & Singh, 2019; Chen & Guo, 2019). To obtain θ, in equations (5), (10) and (13) we proceed by trial and error; on the other hand, in equations (3), (8) and (12) its value is cleared.

Estimation of joint empirical probabilities

The bivariate empirical probabilities were estimated based on the Gringorten formula, applied by Yue (2000b), Zhang and Singh (2019), and Chen and Guo (2019). The formula is:

$$p = \frac{i - 0.44}{n + 0.12} \qquad (26)$$

Where *i* is the number of each data, when they are ordered progressively and *n* is the total number of them, or amplitude of the processed register. The previous expression was applied in the two–dimensional plane, with the data ordered progressively; the maximum flows (*Q*) in the rows and the 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. Then each pair of annual data (*Q* 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 historical year drawn.

When the *n* pairs of data are drawn, year 1 is located and a rectangular or square area of minor *Q* and *V values* is defined, in which the count of numbered cells within, is NM_1 or combinations of minor *Q* and *V*. Once the *n* values of NM_i have been calculated, the Gringorten graphical position formula is applied to estimate the joint or bivariate empirical probability:

$$F(x,y) = P(Q \leq q, V \leq v) = \frac{NM_i - 0.44}{n + 0.12} \quad (27)$$

Selection of the Copula Function

A simple Copula function selection approach is based on the fit error statistics, by comparing the observed empirical probabilities (w_i^o) with the calculated theoretical ones (w_i^c) with the *CF* being tested. The indicators applied are the *standard mean error* (*EME*, by its acronym in Spanish), the *absolute mean error* (*EMA*) and the *maximum absolute error* (*EAM*); their expressions are (Chowdhary & Singh, 2019):

$$EME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(w_i^o - w_i^c\right)^2} \tag{28}$$

$$EMA = \frac{1}{n}\sum_{i=1}^{n}\left|\left(w_i^o - w_i^c\right)\right| \tag{29}$$

$$EAM = \max_{i=1:n}\left|\left(w_i^o - w_i^c\right)\right| \tag{30}$$

Dependency at the top end

Generalities

The most important criterion applied to select a *CF* is based on the dependency's magnitude at the upper end of the joint distribution, which has an impact on the extreme predictions' *veracity*. The upper right tail dependency (λ_U) is the conditional probability that *Y* is greater than a certain percentile (*s*) of $F_Y(y)$, given that *X* is greater than that percentile in $F_X(x)$; as *s* approaches the unit. The lower left tail dependence (λ_L), compares that *Y* is less than *X*, when *s* approaches zero (Chowdhary & Singh, 2019; Salvadori *et al.*, 2007).

In relation to the exposed *CF*s, those of Frank and Plackett, have insignificant dependencies in their extreme zones: therefore, $\lambda_L = 0$ and $\lambda_U = 0$. In contrast, the Gumbel-Hougaard Copula has significant dependence on the upper tail, equal to:

$$\lambda_U = 2 - 2^{1/\theta} \qquad (31)$$

Clayton and Raftery Copulas have it in their lower tail and equal to:

$$\lambda_L = 2^{-1/\theta} \qquad (32)$$

$$\lambda_L = \frac{2\theta}{\theta+1} \qquad (33)$$

As already indicated, the Associated Clayton Copula has $\lambda_U = 2^{-1/\theta}$ (Equation (17)). Families of unexposed copulas with $\lambda_U>0$ are those of Galambos, Genest-Ghoudi and Hüsler-Reiss (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019).

Dupuis (2007) tested six Copula families and found that their ability to estimate extreme events ranges from poor to good, in the following order: Clayton, Frank, Normal, *t*–Student, Gumbel–Hougaard, and Associated Clayton (*Survival Clayton*). Similar conclusions are reached by Poulin *et al.* (2007), when comparing the same six families of Copulas and the one called A12, which has significant right tail dependence.

Estimation of observed dependency

To address the upper tail dependence (λ_U) estimation shown by the available data, the *Empirical Copula* must be defined first. Since the *CF* characterizes the dependence between the random variables X and Y, then the pair of ranges R_i and S_i from these variables are the statistic that retains the greatest amount of information and its scaling with the factor $1/(n+1)$ generates a series of points in the unit square $[0,1]^2$, forming the domain of the Empirical Copula (Chowdhary & Singh, 2019), defined as follows:

$$C_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} 1\left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v\right) \tag{34}$$

In the above equation, $1(\cdot)$ denotes a function of the random variables U and V, which are a continuously increasing transformation of X and Y, relative to the empirical probability integrals $F_n(X)$ and $F_n(Y)$, whose equations are:

$$U_i = \frac{\text{Rango } (X_i)}{n+1} = F_n(x_i) \; V_i = \frac{\text{Rango } (Y_i)}{n+1} = F_n(Y_i) \tag{35}$$

Poulin *et al.* (2007) use the estimator proposed by Frahm, Junker and Schmidt (2005), which is based on a random sample obtained from the Empirical Copula, its expression is:

$$\lambda_U^{CFG} = 2 - 2\exp\left\{\frac{1}{n}\sum_{i=1}^{n} \ln\left[\sqrt{\ln\frac{1}{U_i}\cdot\ln\frac{1}{V_i}}/\ln\left(\frac{1}{\max(U_i,V_i)^2}\right)\right]\right\} \tag{36}$$

This estimator accepts that the *CF* can be approximated by one of the extreme values class, it has the advantage of not requiring a threshold value for its estimation, like the four exposed by AghaKouchak, Sellars and Sorooshian (2013).

Families of Copulas to be contrasted

Justification

Since the initial applications of *Copula Functions* (*CF*) in multivariate hydrological analyses, Favre *et al.* (2004) formulated in their Conclusions, that the crucial stage of such analysis lies in the adequate selection of the *CF* to the available data. Subsequently, Requena *et al.* (2013), highlighted that a *CF* that shows a good fit to the total available data, does not guarantee an adequate representation of the extreme values of such data and therefore, it is necessary to estimate the observed dependence λ_U^{CFG} (Equation (36)), for look for a *CF* that reproduces it.

Therefore, when estimates of high joint return periods are required, greater than three times the amplitude of the available joint record of flows and maximum annual volumes, the *CF* that best reproduces the dependency on the right tail of the data (λ_U^{CFG}) must be sought and selected. In this study, four families of Copulas are contrasted, whose dependence on their right end varies from *greater to lesser degree*, these are: Joe, Associated Clayton, Gumbel–Hougaard and Student's *t*. The first three belong to the class of Archimedean Copulas and the last one is of the elliptic class.

Joe's Copula

The basic equation of this *CF* family, which only accepts positive dependence, is the following (Joe, 1993; Salvadori *et al.*, 2007; Chowdhary & Singh, 2019):

$$C(u,v) = 1 - \left[(1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta}(1-v)^{\theta}\right]^{1/\theta} \quad (37)$$

where the dependency parameter is $\theta \geq 1$, with $\theta = 1$, for the case of independence between *u* and *v*. It does not have an expression that relates its parameter θ to Kendall's tau quotient, therefore, it is estimated based on its generator equation $\phi(s)$ and their derivative $\phi'(s)$ of the *CF*, whose expression is (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019):

$$\tau_n = 1 + 4\int_0^1 \frac{\phi(s)}{\phi'(s)} ds \quad (38)$$

in which, *s* is the unitary random variable $0 < s \leq 1$. The equations of its generator and its derivative are (Salvadori *et al.*, 2007; Zhang & Singh, 2019):

$$\phi(s) = -\ln\left[1 - (1-s)^{\theta}\right] \quad (39)$$

$$\phi'(s) = \frac{\theta(1-s)^{\theta-1}}{-1+(1-s)^{\theta}} \quad (40)$$

Taking the value of Kendall's tau quotient as data, Equation (38) was numerically integrated based on Equation (19), to obtain by trial and error the value of θ that satisfies it. Obtained the value of θ, the significant dependence on the right end (λ_U) of the Joe's Copula, is obtained by means of Equation (31) of the Gumbel-Hougaard Copula.

Associated Clayton's Copula

The Copula defined with Equation (18) will be applied and contrasted, since Dupuis (2007) exposed it and showed that it has a strong significant right end dependence (λ_U).

Student'*t* Copula

Demarta and McNeil (2005) indicate that the Gumbel-Hougaard and Associated Clayton Copulas are simpler to apply than the extreme copulas derived from Student's *t*, but the latter defines a lower value of the dependency, as previously pointed out by Dupuis (2007) and confirmed by Poulin *et al.* (2007). The Student's *t*-Copula equation is as follows (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019):

$$C(u,v) = \int_{-\infty}^{t_v^{-1}(u)} \cdot \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \left(1 + \frac{x^2 - 2\theta \cdot x \cdot y + y^2}{v(1-\theta^2)}\right)^{-(v+2)/2} dx \cdot dy \tag{41}$$

In the above expression $v > 2$, θ varies between −1 and 1, with $\theta = 0$ for independence between *u* and *v*. $t_v^{-1}(\cdot)$ is the inverse of the univariate Student's *t*-distribution with v degrees of freedom, for the probability indicated in parentheses.

Then, with the idea of limiting the dependency in the right tail to a lower value, the Student *t*–Copula equations are exposed (Demarta & McNeil, 2005; Poulin *et al.*, 2007; Salvadori *et al.*, 2007; Genest, Favre, Beliveau, & Jacques, 2007):

$$\tau_n = \frac{2}{\pi} \text{arcsen}\theta \tag{42}$$

$$\lambda_U = 2t_{v+1}\left(\frac{-\sqrt{v+1}\sqrt{1-\theta}}{\sqrt{1+\theta}}\right) \tag{43}$$

The minus sign inside the parentheses indicates that such value is the exceedance probability (*survival probability*), which must define the random variable with Student's *t* distribution and $v+1$ degrees of freedom. To avoid the use of numerical algorithms that estimate the value of $t_{\nu+1}$ based on the probability of exceedance, Demarta and McNeil (2005) present the following alternative formula:

$$\lambda = \frac{\int_{(\pi/2-\text{arcsen}\theta)/2}^{\pi/2} \cos^{\nu} s \cdot ds}{\int_{0}^{\pi/2} \cos^{\nu} s \cdot ds} \tag{44}$$

In the above equation, λ does not have a subscript, because the Student's *t*–Copula has the same significant dependence on its two tails or extremes. Equation (44) was numerically integrated based on Equation (19) and with these results, Table 1 was prepared, which is an extension of the one presented by Demarta and McNeil (2005).

Table 1. Tail dependence values of the Student's *t*–Copula, for various values of ν and θ.

ν	**Association parameter value (θ)**															
	-0.40	-0.50	-0.75	0.00	0.40	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	0.98
2	0.0773	0.0577	0.0195	0.1817	0.3393	0.3910	0.4195	0.4502	0.4833	0.5195	0.5594	0.6042	0.6557	0.7177	0.7995	0.8729
3	0.0378	0.0257	0.0061	0.1161	0.2606	0.3125	0.3419	0.3739	0.4091	0.4481	0.4918	0.5415	0.5995	0.6702	0.7648	0.8505
4	0.0189	0.0117	0.0020	0.0756	0.2031	0.2532	0.2822	0.3144	0.3503	0.3907	0.4366	0.4896	0.5523	0.6298	0.7349	0.8311
6	0.0049	0.0025	0.0002	0.0331	0.1269	0.1705	0.1970	0.2275	0.2625	0.3031	0.3506	0.4071	0.4758	0.5630	0.6845	0.7980
8	0.0013	0.0006	0.0000	0.0150	0.0811	0.1173	0.1405	0.1679	0.2004	0.2393	0.2861	0.3434	0.4151	0.5086	0.6424	0.7699
10	0.0004	0.0001	0.0000	0.0069	0.0527	0.0819	0.1015	0.1255	0.1549	0.1911	0.2360	0.2925	0.3652	0.4627	0.6059	0.7452

ν = degrees of freedom.

Table 1 highlights that even for negative values and zero correlation, the Student's *t*–Copula has asymptotic dependence at its extremes and that such dependence decreases with increasing degrees of freedom.

The application or integration of the Student's *t*–Copula (Equation (41)) was carried out based on Equation (20), with 20 pairs of coefficients w_i and abscissa h_i of 15 digits taken from Davis and Polonsky (1972) for the estimation of the integrals' upper limits, the algorithm exposed by Zelen and Severo (1972) was applied to estimate the random variable Student's *t* with ν degrees of freedom, based on the Normal variable x_p; in which, *p* is the probability of exceedance:

$$t\ x_p + \frac{g_1(x_p)}{\nu} + \frac{g_2(x_p)}{\nu^2} + \frac{g_3(x_p)}{\nu^3} + \frac{g_4(x_p)}{\nu^4} + \cdots \quad (45)$$

being:

$$g_1(x) = \frac{1}{4}(x^3 + x) \quad (46)$$

$$g_2(x) = \frac{1}{96}(5x^5 + 16x^3 + 3x) \quad (47)$$

$$g_3(x) = \frac{1}{384}(3x^7 + 19x^5 + 17x^3 - 15x) \quad (48)$$

$$g_4(x) = \frac{1}{92160}(79x^9 + 776x^7 + 1482x^5 - 1920x^3 - 945x) \quad (49)$$

When $p < 0.50$ make $t = -t$. The algorithm for estimating the Normal variable x_p also comes from Zelen and Severo (1972) and is as follows:

$$x_p = \delta - \frac{c_0 + c_1 \cdot \delta + c_2 \cdot \delta^2}{1 + d_1 \cdot \delta + d_2 \cdot \delta^2 + d_3 \cdot \delta^3} \quad (50)$$

in which:

$$\delta = \sqrt{ln\left(\frac{1}{p^2}\right)} \quad (51)$$

being:

$c_0 = 2.515517$

$c_1 = 0.802853$

$c_2 = 0.010328$

$d_1 = 1.432788$

$d_2 = 0.189269$

$d_3 = 0.001308$

If $p > 0.50$, make $p = 1 - p$ and $x_p = -x_p$. The algorithms defined by equations (45) to (51) were tested with the numerical values exposed by Zhang and Singh (2019).

To establish the integrals' lower limits of Equation (41), the tabulation of the Student's t distribution was used, as indicated by Ostle and Mensing (1975) as Appendix 5. In said Table, the following values of t with ν degrees of freedom from 2 to 10: 31.598, 12.941, 8.610, 6.859, 5.959, 5.405, 5.041, 4.781 and 4.587. Such values with a minus sign were taken as lower limits according to the degree of freedom assigned in the application of Equation (41); which, logically, varied from 2 to 10; selecting the Student's t–Copula of best fit (equations (28) to (30)).

Ratification of the selected Copula function

This is the most important stage in the process of the *CF* practical application, since it is verified that the model faithfully reproduces the observed joint probabilities (Equation (27)). Yue (2000a) indicates that the simple and practical way of representing the empirical and theoretical joint probabilities consists of taking the first one to the abscissa axis and the second to the ordinate axis; logically, in such a graph, 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.

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 absolute difference (*dma*) between the joint probabilities. To evaluate the statistic (D_n) of the test, the expression proposed by Meylan *et al.* (2012), for α = 5 % this is:

$$D_n = \frac{1.358}{\sqrt{n}} \tag{52}$$

n is the number of data. If the *dma* is less than D_n, the adopted *CF* is ratified.

Bivariate frequency analysis concepts

Bivariate return period

The first *bivariate return period* of the event (*X*, *Y*) is defined under the OR condition, which indicates that the limits *x* ó *y*, or both *can be* exceeded and then, the classical equation of the return period or inverse probability of exceedance will be (Shiau *et al.*, 2006; Genest & Chebana, 2017):

$$T_{XY} = \frac{1}{P(X>x \vee Y>y)} = \frac{1}{1-F_{X,Y}(x,y)} = \frac{1}{1-C[F_X(x),F_Y(y)]} \tag{53}$$

where $C[F_X(x), F_Y(y)]$ is the selected *CF*.

The second *bivariate 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 (Shiau *et al.*, 2006; Genest & Chebana, 2017):

$$T'_{XY} = \frac{1}{P(X>x \wedge Y>y)} = \frac{1}{F'_{X,Y}(x,y)} = \frac{1}{1-F_X(x)-F_Y(y)+C[F_X(x),F_Y(y)]} \quad (54)$$

Aldama (2000) obtains the expression $F'_{X,Y}(x,y)$ of the bivariate probability of exceedance through a simple and logical reasoning of probabilities applied in the Cartesian plane. Instead, Yue and Rasmussen (2002) use the Cartesian plane to define the bivariate event (*X*, *Y*) conceptually, which can occur in any of the four quadrants.

The relationship between bivariate and univariate return periods is as follows (Yue & Rasmussen, 2002; Shiau *et al.*, 2006; Vogel & Castellarin, 2017):

$$T_{XY} \leq \min[T_X, T_Y] \leq \max[T_X, T_Y] \leq T'_{XY} \quad (55)$$

being:

$$T_X = \frac{1}{F'_X(x)} = \frac{1}{1-F_X(x)} \quad (56)$$

$$T_Y = \frac{1}{F'_Y(y)} = \frac{1}{1-F_Y(y)} \quad (57)$$

Salvadori and De Michele (2004) introduce in detail the concept of the *Secondary Return Period* (ζ), designated in this way to emphasize that the joint return period T_{XY} is the primary one, from where it comes when using the *isolines* defined by the applied *CF*, whose expression is:

$$L_s = [(u, v) \in I^2 : C(u, v) = s] \tag{58}$$

where *s* is the unitary random variable $0 < s \leq 1$ and *C* is the *CF* tested. Then, a region $B_C(s)$ is defined in the unit space (I^2) above the isoline, below it and to the left, which will be:

$$B_C(s) = \{(u, v) \in I^2 : C(u, v) \leq s\} \tag{59}$$

In a *CF* of the Archimedean class, the Kendall univariate distribution, designated $K_C(s)$, provides a measure of the events within the $B_C(s)$; its equation is (Salvadori & De Michele, 2004; Salvadori & De Michele, 2007; Salvadori *et al.*, 2007; Gräler *et al.*, 2013):

$$K_C(s) = s - \frac{\phi(s)}{\phi'(s)} \tag{60}$$

in which, $\phi(s)$ is the generator of the *CF* and $\phi'(s)$ its derivative. Finally, the secondary return period (ζ) of the events outside of $B_C(s)$, is:

$$\zeta = \frac{1}{1 - K_C(s)} \tag{61}$$

where the denominator is the probability of exceedance (*survival function*), which corresponds to possibly of destructive or dangerous events. The relationship between ζ and the joint return periods is as follows (Gräler *et al.*, 2013):

$$T_{XY} \leq \zeta \leq T'_{XY} \tag{62}$$

The application of Equation (61) requires the prior estimation of Equation (60), for the following six non–exceedance probabilities (*s*): 0.98, 0.99, 0.998, 0.999, 0.9998 and 0.9999, which correspond to the *joint return periods* of 50, 100, 500, 1 000, 5 000 and 10 000 years. In the Gumbel–Hougaard and Joe Copulas, the equations of its generator and its derivative are the following (Zhang and Singh, 2019):

$$\phi(s) = (-lns)^{\theta} \tag{63}$$

$$\phi'(s) = \frac{-\theta}{s}(-lns)^{\theta-1} \tag{64}$$

$$\phi(s) = -\ln\left[1-(1-s)^{\theta}\right] \tag{65}$$

$$\phi'(s) = \frac{-\theta(1-s)^{\theta-1}}{-1+(1-s)^{\theta}} \tag{66}$$

Critical or design events

Volpi and Fiori (2012) highlight that the plot of the AND–type joint return period, shown later as Figure 2, presents a severe inconsistency as it contains, in a bivariate context, univariate critical thresholds. Due to the above, such a graph is considered to be made up of two portions, the two designated *simple* (*naive part*) and the *correct* one (*proper part*). The straight parts are the tails or straight lines asymptotes to the curved part. The probability of an event occurrence or pair of Q and V, is variable in the curved part and decreases along the straight part, although all the values define the *same* joint return period. In summary, the pairs of values of the asymptote lines have low probabilities of occurrence and therefore should not be included in the analysis of the search for critical or severe floods (Q and V). For practical purposes, the extreme points of the curved part can be defined according to their empirical distribution or close to the beginning of the asymptotic lines.

Selection of marginal distributions

The approach for the selection of the marginal distributions was very simple and consisted of applying the three PDFs that have been established as reference or applicable under precept, which are the Log–Pearson type III (LP3), the General Extreme Values (GVE) and Generalized Logistics (LOG). In addition, three widely used distributions were applied: the Generalized Pareto (PAG), the Kappa and the Wakeby. The first four mentioned PDFs have three fit parameters and the last two, four and five. With the exception of LP3, which was applied with the moments method, in the logarithmic (WRC, 1977) and real (Bobée, 1975) domains, the rest were fitted with the L–moments method (Hosking & Wallis, 1997).

The selection of the most convenient FDP was based on the value that each one generates with a non–exceedance probability of 1 %; that is, a very low value that does not exceed the minimum values observed, to avoid negative marginal probabilities (u,v). This selection criterion is the one required in records of Q and V that present very low minimum values, compared to their maximum extremes.

In addition, the standard errors of fit (*EEA*) and mean absolute errors (*EAM*) were taken into account, as well as the *predictions'* magnitude in return periods greater than 500 years.

Fitting errors

The first criteria applied for the selection of the best PDF to some available data or series were the *fitting errors* (Kite, 1977; Willmott & Matsuura, 2005; Chai & Draxler, 2014). This criterion and the one described to avoid negative probabilities allow an adequate distribution to be adopted between the models: LP3, GVE, LOG, PAG, Kappa and Wakeby.

Changing the probabilities observed by the ordered data of the analyzed series (x_i or y_i) in equations (28) and (29), and the probabilities calculated by the estimated values with the PDF that is tested or contrasted, the standard error of fit (*EEA*) and the mean absolute error (*EAM*) are obtained. The estimated values ($\hat{x}_i$ or $\hat{y}_i$) are searched for the same probability of non–exceedance, assigned to the data by Gringorten's empirical formula (Equation (26)).

Data to be processed

In order to compare the four *CF*s that have a significant right tail dependence (λ_U), the 16 *joint* records of maximum flow (*Q*) and volume (*V*) of the annual floods reported by Aldama *et al.* (2006), which add up to 620 pairs of data, in 15 important reservoirs in Mexico and one in project (La Parota).

In addition, the joint record of annual *Q* and *V* of the floods recorded at the La Cuña hydrometric station, exposed by Gomez *et al.* (2010) and Campos-Aranda (2022). The record of *Q* and *V* of the floods at the entrance to the Malpaso Dam, shown by Domínguez and Arganis (2012), was also processed.

Table 2 shows the record of the 26 joint data of *Q* and *V* of the floods at the entrance to the Adolfo Lopez Mateos Dam on the Humaya River in the state of Sinaloa, Mexico. Such record comes from Aldama *et al.* (2006) and processed as a numerical application.

Table 2. Maximum flows, volumes and their bivariate order numbers of the 26 annual floods at the entrance to the Adolfo López Mateos Dam (*Humaya*), Mexico (Aldama *et al.*, 2006).

Año	***Q*** **(m^3/s)**	***V*** **(Mm^3)**	***NM_i***	**Año**	***Q*** **(m^3/s)**	***V*** **(Mm^3)**	***NM_i***
1974	600	140	4	1987	401	160	6
1975	290	200	4	1988	305	150	3
1976	316	70	3	1989	581	170	8
1977	242	50	1	1990	2 035	1 150	22
1978	248	420	4	1991	488	190	9
1979	891	400	15	1992	242	310	3
1980	461	190	7	1993	487	160	7
1981	9 245	4 580	26	1994	933	150	6
1982	2 388	1 410	23	1995	127	70	1
1983	277	160	3	1996	4 490	1 240	24
1984	481	290	9	1997	904	160	9
1985	614	450	16	1998	2 529	1 020	22
1986	1 064	550	20	1999	1 193	380	17

Wald-Wolfowitz Test

This nonparametric test has been used by Bobée and Ashkar (1991), Rao and Hamed (2000), and Meylan *et al.* (2012) to verify *independence* and *stationarity* in records of maximum annual flows (X_i). Therefore, it was proposed to apply this test to the joint records of Q and V, which must be *random* samples.

Results and their discussion

Contrast of Copula functions with λ_U>0

Randomness verification

Table 3 shows, in progressive order of basin size (A in km^2), the 16 *joint* records of Q and V processed, taken from Aldama *et al.* (2006) and at the end, the two joint records also processed. Column 4 shows the number of data (n) of each record and columns 5 and 6, the value of the U statistic of the Wald-Wolfowitz Test; whose absolute value must be less than 1.96 for such registration to be *random*.

Table 3. General data and results of the Wald-Wolfowitz Test in the 18 joint records of *Q* and *V* processed.

No.	**Name of the dam or record**	***A* (km²)**	***n***	**WW Test**	
				Q	***V***
1	Madin	171	30	−0.757	−0.335
2	Eustaquio Buelna (*Guamúchil*)	1 630	37	(4.156)	(2.597)
3	Josefa Ortiz de Domínguez (*El Sabino*)	2 268	34	(1.986)	1.529
4	Abelardo L. Rodríguez	2 430	44	(2.220)	(2.346)
5	Sanalona	3 250	38	0.337	−1.555
6	Ignacio Allende (*La Begoña*)	4 984	27	0.287	1.241
7	La Parota	7 067	31	−0.760	−0.532
8	Benito Juárez (*El Marqués*)	9 697	26	−0.601	0.427
9	Adolfo Ruiz Cortines (*Mocúzar*i)	10 719	61	0.147	1.189
10	Adolfo López Mateos (*Humaya*)	10 972	26	0.677	1.252
11	Belisario Domínguez (*La Angostura*)	18 537	33	0.955	0.247
12	La Boquilla	21 003	64	−0.084	−0.251
13	Luis Donaldo Colosio (*Huites*)	26 020	52	−0.091	−0.602
14	Venustiano Carranza (*Don Martín*)	31 034	52	0.726	1.054
15	Plutarco Elías Calles (*El Novillo*)	58 280	22	0.637	0.496
16	Adolfo López Mateos (*Infiernillo*)	108 000	43	−0.707	−0.022
17	Estación hidrométrica *La Cuña*	19 097	55	0.284	0.213
18	Presa Malpaso (*Netzahualcóyotl*)	34 800	47	−0.666	0.558

It can be deduced from the cited values that two joint records are not random, those of the *Guamúchil* and Abelardo L. Rodríguez dams. Specific statistical tests were applied to such records and it was found in the *Guamúchil* dam that the *Q* and *V* series show persistence and excess variability due to their extreme values. In addition, the *Q* series shows an upward trend. In the Abelardo L. Rodriguez dam, both series of *Q* and *V* show persistence, excess variability and a negative trend. The *Q* record of the *El Sabino* dam can be accepted randomly, since it only shows slight persistence. Due to the above, the joint records of the *Guamúchil* and Abelardo L. Rodríguez dams were eliminated from the subsequent analyses.

Values of association measures

In column 2 of Table 4, the Kendall's tau quotient values obtained with Equation (21) are displayed. Regarding Spearman's rho coefficient, shown in column 2 of Table 5, the values that do not have parentheses were equal to the calculated with equations (23) and (24). Instead, the quantities in parentheses come from Equation (23), which leads to slightly larger values.

Table 4. Kendall's tau quotient values of the observed dependency (λ_U^{CFG}) and of the applied Joe and Clayton Copulas Associated with the indicated joint record.

Dam or record	τ_n	λ_U^{CFG}	Joe Copula		Clayton Asoc.	
			θ	λ_U	θ	λ_U
Madin	0.6598	0.7356	4.687	0.8406	3.8784	0.8363
El Sabino	0.5936	0.6776	3.748	0.7969	2.9211	0.7888
Sanalona	0.5733	0.6053	3.525	0.7827	2.6867	0.7726
La Begoña	0.6866	0.7831	5.177	0.8567	4.3818	(0.8537)
La Parota	0.5226	0.6724	3.039	0.7438	2.1892	(0.7286)
El Marqués	0.5385	0.6462	3.179	0.7564	2.3333	0.7430
*Mocúzar*i	0.7093	0.8024	5.630	0.8690	4.8797	(0.8676)
Humaya	0.4523	0.6834	2.522	0.6837	1.6517	0.6573
La Angostura	0.4318	0.4730	2.397	0.6647	1.5200	0.6338
La Boquilla	0.6091	0.6440	3.938	0.8075	3.1168	0.8006
Huites	0.4495	0.5655	2.504	0.6811	1.6329	0.6541
Don Martín	0.6456	0.7446	4.455	0.8317	3.6426	0.8267
El Novillo	0.4286	0.6166	2.378	0.6616	1.5000	0.6300
Infiernillo	0.4064	0.4714	2.253	0.6398	1.3694	0.6028
La Cuña	0.7199	0.7819	5.887	0.8750	5.1394	0.8738
Presa Malpaso	0.3747	0.4159	2.092	0.6072	1.1982	0.5608

Table 5. Spearman's rho coefficient values of the observed dependency (λ_U^{CFG}) and of the Gumbel–Hougaard Copulas and Student's t in the indicated set of records.

Dam or record	ρ_n	λ_U^{CFG}	**Gumbel-Hougaard**		***t* de Student**	
			θ	λ_U	$\lambda_U(\nu = 2)$	$\lambda_U(\nu = 10)$
Madin	(0.8603)	0.7356	2.9392	0.7340	0.6678	0.3834
El Sabino	0.6883	0.6776	2.4605	0.6746	0.6072	0.2965
Sanalona	0.7595	0.6053	2.3433	0.6558	0.5889	0.2725
La Begoña	0.8107	0.7831	3.1909	(0.7574)	0.6929	0.4225
La Parota	0.6968	0.6724	2.0946	(0.6077)	0.5443	0.2184
El Marqués	0.7648	0.6462	2.1667	0.6230	0.5581	0.2345
*Mocúzar*i	(0.8827)	0.8024	3.4399	(0.7768)	0.7142	0.4570
Humaya	(0.7258)	0.6834	1.8258	0.5383	0.4849	0.1563
La Angostura	0.5836	0.4730	1.7600	0.5173	0.4681	0.1409
La Boquilla	0.7922	0.6440	2.5584	0.6888	0.6212	0.3157
Huites	0.5980	0.5655	1.8164	0.5354	0.4825	0.1541
Don Martín	(0.8303)	0.7446	2.8213	0.7215	0.6546	0.3636
El Novillo	0.5788	0.6166	1.7500	0.5140	0.4654	0.1385
Infiernillo	(0.5707)	0.4714	1.6847	0.4910	0.4476	0.1233
La Cuña	(0.9088)	0.7819	3.5697	0.7857	0.7243	0.4737
Presa Malpaso	(0.5510)	0.4159	1.5991	0.4574	0.4226	0.1038

Dependency contrast (λ_U)

Columns 3 of Table 4 and Table 5 show the values of the observed dependence (λ_U^{CFG}), estimated with Equation (36). The *CF* selection criterion was that its significant dependence (λ_U), should be somewhat larger or slightly smaller than the observed dependency (λ_U^{CFG}). With this criterion in Table 4, the Joe *CF*s were defined for the record of the Adolfo Lopez Mateos Dam (*Humaya*) and the Associated Clayton *CF* for the Plutarco Elias Calles Dam (*El Novillo*) record.

On the other hand, in Table 5, with the selection criteria described, 10 records are defined whose dependence on their right tail can be presented by the Gumbel–Hougaard *CF* (GH) and three that are included between the Associated Clayton *CF* and GH, whose dependency values (λ_U) are shown in parentheses. The *CF* of the Student's *t* reproduces the dependence of records from the Malpaso Dam, with $\nu = 2$.

Search of the marginal distributions

Distribution of maximum annual flow

Table 6 shows the fit errors and predictions (m^3/s) obtained with the six distributions applied to the record of maximum flows in Table 2, whose minimum value recorded is 127 m^3/s. Of these six PDFs, the only ones that define values with a probability of non–exceedance of 1 % that are lower than the aforementioned flow are the GVE and the LOG. Taking into account the fit errors, the GVE is adopted.

Table 6. Fit errors and predictions (m^3/s) of the six distributions applied in the record of maximum annual flows of the floods entering the Adolfo Lopez Mateos Dam (*Humaya*), Mexico.

PDF	*EEA* (m^3/s)	*EAM* (m^3/s)	Return periods in years					
			50	100	500	1 000	5 000	10 000
LP3	*230.8*	*132.6*	9 062	15 192	48 746	79 778	247 228	400 939
GVE	675.5	279.8	6 663	10 381	28 653	44 244	121 000	186 452
LOG	745.8	301.9	6 364	9 961	28 041	43 766	123 060	192 052
PAG	631.9	255.6	6 824	10 412	26 828	40 002	100 278	148 626
KAP	585.1	243.0	7 072	10 697	26 739	39 253	94 582	137 656
WAK	676.7	286.5	6 770	10 379	27 111	40 694	103 672	154 793

The location (u_1), scale (a_1) and shape (k_1) parameters of the adopted GVE distribution are: 447.7865, 372.8246, −0.6236228, where the equation is:

$$F(x) = \exp\left\{-\left[1 - k_1 \frac{(x-u_1)}{a_1}\right]^{1/k_1}\right\} \quad (67)$$

Distribution of annual volumes

In Table 7, similar to the previous one for the annual volumes of Table 2, whose minimum value is 50 Mm^3. Again, the only PDFs that do not generate negative marginal probabilities are the GVE and LOG. The GVE is adopted because it shows less fit errors.

Table 7. Fit errors and predictions (m^3/s) of the six distributions applied to the record of annual volumes of the floods at the entrance to the Adolfo López Dam Mateos (*Humaya*), Mexico.

PDF	***EEA*** **(m^3/s)**	***EAM*** **(m^3/s)**	**Return periods in years**					
			50	**100**	**500**	**1 000**	**5 000**	**10 000**
LP3	*189.8*	*96.6*	4 061	6 726	20 745	33 252	97 487	153 958
GVE	370.6	138.8	3 021	4 726	13 163	20 401	56 269	87 019
LOG	400.4	144.6	2 881	4 528	12 856	20 134	57 053	89 331
PAG	355.1	133.1	3 091	4 739	12 345	18 494	46 869	69 796
KAP	338.6	135.1	3 207	4 875	12 327	18 186	44 336	64 877
WAK	375.3	146.4	3 075	4 729	12 428	18 697	47 873	71 630

The location (u_2), scale (a_2) and shape (k_2) parameters of the adopted GVE distribution are: 193.4795, 167.2692 and −0.6288097, where the expression is:

$$F(y) = \exp\left\{-\left[1 - k_2 \frac{(y-u_2)}{a_2}\right]^{1/k_2}\right\} \qquad (68)$$

Selection and ratification of the *CF*

The bivariate data processing in Table 2 led to the following three indicators of association: r_{xy} = 0.9670, τ_n = 0.4523 and ρ_n = 0.6738. Equation (22), with n = 26 and the cited tau, gives a value of 3.24; therefore, the Kendall quotient is significant. Equation (25) generates a value of 2.26, so Spearman's coefficient is also significant.

On the other hand, Table 8 shows the statistical fit indicators that were obtained by applying the Clayton, Frank, Plackett, *t*–Student, Gumbel–Hougaard (GH), associated Clayton and Joe *CF*s. In equations (28) to (30), the empirical bivariate probabilities were estimated with Equation (27) and the theoretical ones with equations (2), (4), (9), (41), (7), (18) and (37). The values reported for the Copula *t*–Student correspond to a ν = 6, where the fit errors have already stabilized with $\theta = sen\left(\frac{\pi \cdot \tau_n}{2}\right) = 0.6522$, according to Equation (42).

Table 8. Statistical indicators of the fit of the Copula functions indicated in the annual input floods to the Adolfo López Mateos Dam (*Humaya*), Mexico.

Copula	***θ***	***EME***	***EAM***	**No. DP**	**No. DN**	**MDP**	**MDN**	λ_U
Clayton(C.)	1.6517	0.0529	0.0365	14	12	0.1505	−0.0779	0.0000
Frank	4.9310	0.0461	0.0321	15	11	0.1334	−0.0701	0.0000
Plackett	11.079	*0.0441*	*0.0315*	11	15	0.1291	−0.0727	0.0000
t−Student	0.6522	0.0482	0.0347	13	13	0.1415	−0.0643	0.2641
GH	1.8258	0.0473	0.0354	14	12	0.1352	−0.0601	0.5383
C. Asociada	1.6517	0.0475	0.0362	14	12	0.1277	−0.0549	0.6573
Joe	2.5220	0.0477	0.0365	15	11	*0.1267*	*−0.0533*	0.6837

New acronyms:

DP, DN = positive and negative differences.

MDP, MDN = maximum positive and negative difference.

Based on the results in Table 8, there is no difficulty in selecting the Plackett *CF* as the best *CF* for the data in Table 2, since it reports the smallest statistical indicators (*EME* and *EAM*) (shown in italics).

However, when taking into account that the observed dependency (λ_U^{CFG}) reached a value of 0.6834, the *CF* that should be selected is Joe, as already stated in Table 4. Table 8 shows that the fit of Joe's *CF* to the data is better than the poorest of all, which was defined by Clayton's *CF*. It is noteworthy that Joe's *CF* gives rise to the lowest values of the maximum differences between empirical and theoretical probabilities (shown in italics and shaded).

Table 9 shows the bivariate probabilities of non–exceedance, observed empirical (w_i^o) and calculated theoretical (w_i^c) with Joe's *CF*. The maximum positive and negative differences are also shaded.

Table 9. Joint non–exceedance probabilities and their differences, calculated with Joe's *CF*, for the floods annual entrance to the Dam Adolfo Lopez Mateos (*Humaya*), Mexico.

No.	w_i^o	w_i^c	Differences	No.	w_i^o	w_i^c	Differences
1	0.1363	0.1896	−0.0533	14	0.2129	0.1619	0.0509
2	0.1363	0.1300	0.0063	15	0.0980	0.1051	−0.0071
3	0.0980	0.0311	0.0669	16	0.2894	0.2392	0.0502
4	0.0214	0.0101	0.0114	17	0.8254	0.8643	−0.0389
5	0.1363	0.1378	−0.0015	18	0.3277	0.2369	0.0909
6	0.5574	0.5663	−0.0089	19	0.0980	0.1215	−0.0234
7	0.2511	0.2249	0.0263	20	0.2511	0.1955	0.0556
8	0.9786	0.9851	−0.0066	21	0.2129	0.2460	−0.0331
9	0.8637	0.8939	−0.0302	22	0.0214	0.0051	0.0163
10	0.0980	0.0973	0.0007	23	0.9020	0.9195	−0.0175
11	0.3277	0.3210	0.0067	24	0.3277	0.2666	0.0612
12	0.5957	0.4690	0.1267	25	0.8254	0.8767	−0.0513
13	0.7489	0.6690	0.0798	26	0.6340	0.6058	0.0282

On the other hand, Equation (52) defines D_n = 0.2663 and since the maximum absolute difference in Table 9 is 0.1267, the Kolmogorov–Smirnov test ratifies the adopted Joe's *CF*. The correlation coefficient (r_{xy}) between the empirical and theoretical probabilities, estimated with Joe's *CF*, was 0.9897; therefore acceptable. The graphic contrast between both probabilities, to ratify its adoption, is shown in Figure 1 for the data in Table 9.

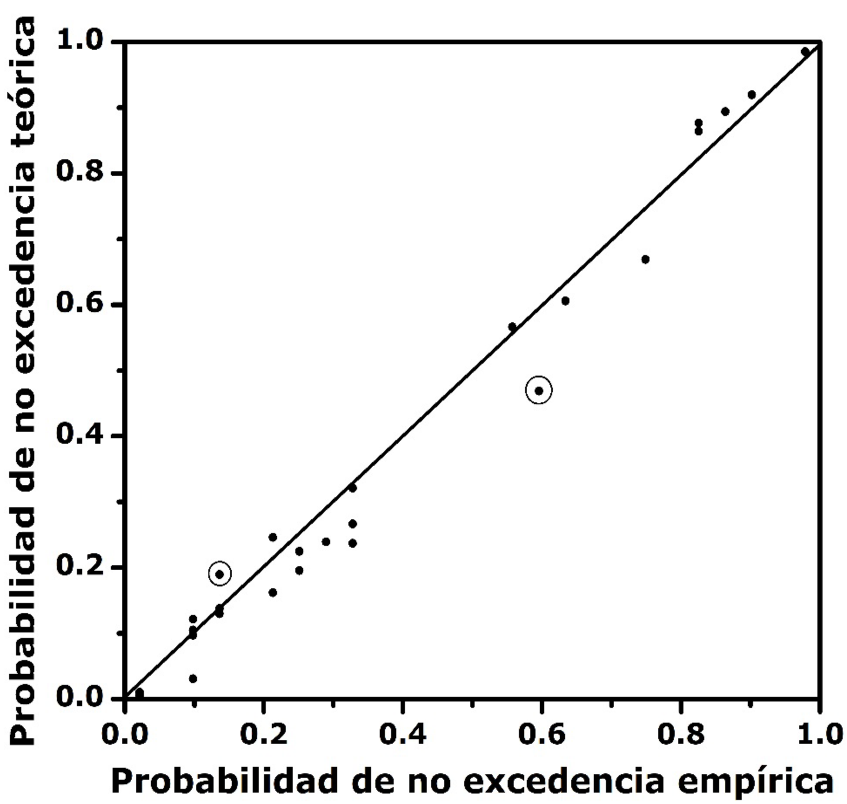

Figure 1. Graphical contrast of joint probabilities (*Q*, *V*) calculated with Joe *FC* de Joe, for annual floods at the entrance of the Adolfo Lopez Mateos (*Humaya*) dam, Mexico.

Bivariate return period graphs T'_{XY}

The bivariate return periods of the AND type are estimated with Equation (54). For their estimations with values of T'_{XY} of 500, 1 000, 5 000 and 10 000 years, Joe's *CF* was applied. Volumes and peak flows are arbitrarily selected to obtain their marginal (equations (67) and (68)) and joint (Equation (54)) probabilities of non–exceedance. Table 10 shows results to define the four graphs of Figure 2.

Table 10. Couples of peak flow and annual volume used to define the graphs of the joint return period type AND with Joe's *CF*, in the floods input to the Adolfo López Mateos Dam (*Humaya*), Mexico.

T'_{XY} **500 years**		T'_{XY} **1 000 years**		T'_{XY} **5 000 years**		T'_{XY} **10 000 years**	
V **Mm³**	***Q*** **m³/s**	***V*** **Mm³**	***Q*** **m³/s**	***V*** **Mm³**	***Q*** **m³/s**	***V*** **Mm³**	***Q*** **m³/s**
0	28 653	0	44 244	0	121 000	0	186 452
1 000	28 636	2 000	44 200	4 000	120 940	5 000	186 420
2 000	28 571	4 000	44 020	8 000	120 735	10 000	186 300
3 000	28 439	6 000	43 642	12 000	120 290	15 000	185 840
5 000	27 903	8 000	43 002	16 000	119 500	20 000	185 250
6 000	27 453	10 000	42 014	20 000	118 390	25 000	184 150
7 000	26 836	12 000	40 542	24 000	116 830	30 000	182 800

T'_{XY} 500 years		T'_{XY} 1 000 years		T'_{XY} 5 000 years		T'_{XY} 10 000 years	
V Mm³	**Q** m³/s	**V** Mm³	**Q** m³/s	**V** Mm³	**Q** m³/s	**V** Mm³	**Q** m³/s
8 000	26 000	13 000	39 560	28 000	114 700	35 000	181 120
9 000	24 858	14 000	38 348	32 000	111 900	40 000	178 750
10 000	23 266	15 000	36 850	36 000	108 140	45 000	175 800
10 500	22 222	16 000	34 960	40 000	103 060	50 000	172 050
11 000	20 940	17 000	32 520	42 000	99 850	52 500	169 960
11 500	19 315	18 000	29 212	44 000	96 060	55 000	167 520
12 000	17 165	18 500	27 050	46 000	91 430	57 500	164 740
12 500	14 046	19 000	24 335	47 000	88 770	60 000	161 570
12 750	11 745	19 500	20 720	48 000	85 800	62 500	157 970
13 000	8 125	19 750	18 330	49 000	82 460	65 000	153 850
13 050	6 990	20 000	15 170	50 000	78 670	67 500	149 310
13 100	5 490	20 100	13 530	51 000	74 250	70 000	143 810
13 163	0	20 200	11 520	52 000	69 140	72 000	138 820
		20 300	8 730	53 000	62 840	74 000	133 200
		20 350	6 450	53 500	59 250	76 000	126 530
		20 401	0	54 000	54 985	78 000	118 720
				54 500	50 160	80 000	109 000
				55 000	44 100	81 000	103 580
				55 500	35 840	82 000	96 950
				56 000	23 600	83 000	89 650

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

T'_{XY} **500 years**		T'_{XY} **1 000 years**		T'_{XY} **5 000 years**		T'_{XY} **10 000 years**	
V **Mm³**	***Q*** **m³/s**	***V*** **Mm³**	***Q*** **m³/s**	***V*** **Mm³**	***Q*** **m³/s**	***V*** **Mm³**	***Q*** **m³/s**
				56 269	0	84 000	80 650
						85 000	68 990
						86 000	53 500
						87 019	0

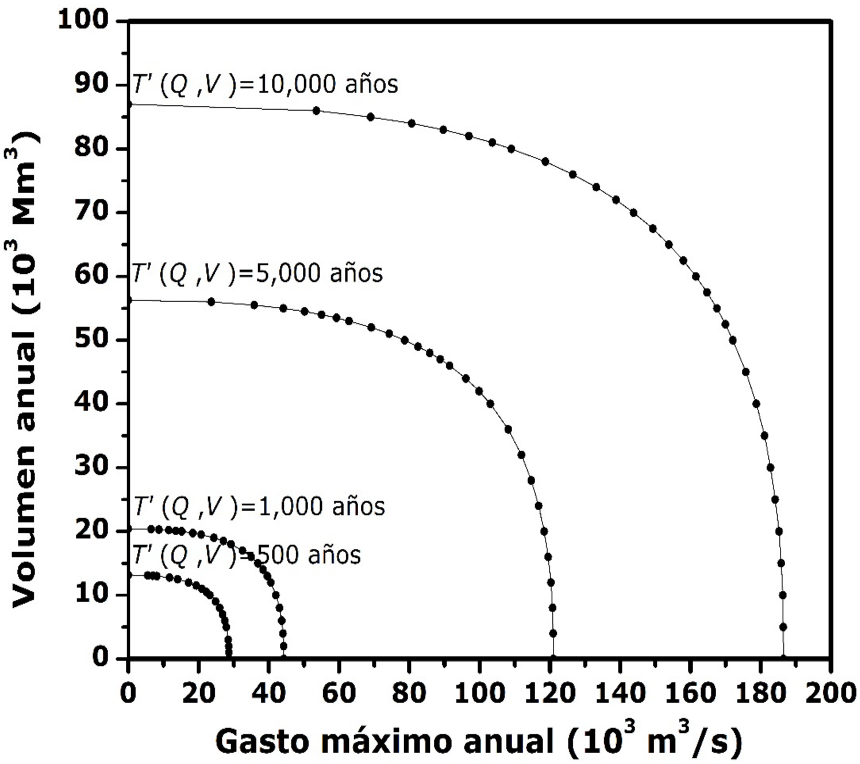

Figure 2. Graphs of the four design joint return periods T'_{XY} obtained with Joe's *CF*, in the annual floods at the entrance to the Adolfo Lopez Dam Mateos (*Humaya*), Mexico.

Selection of design events

In Figure 2 or in Table 10, infinite pairs of Q and V can be selected, which satisfy the joint design return period and which are defined as a *subgroup of critical pairs*, as they are within the curved portion of each T'_{XY} graph, outside the asymptote lines (Volpi & Fiori, 2012).

The combinations of peak flow and volume that have the same design *bivariate return period*, establish 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 conditions. To form each design hydrograph, there are theoretical and empirical methods (Aldama, 2000; Aldama *et al.*, 2006; Campos-Aranda, 2008; Gómez *et al.*, 2010; Gräler *et al.*, 2013).

Contrast of the Joint Return Periods

Table 11 and Table 12 show the calculations made to verify equations (55) and (62), based on the Gumbel–Hougaard and Joe *CF*s. It is observed and verified in all cases that both *CF*s comply with Equation (55). For the secondary return period, Equation (62) is exclusively fulfilled with Joe's *CF* in the last three return periods analyzed.

Table 11. Estimated univariate and bivariate return periods with the Gumbel–Hougaard *CF* in the floods at the entrance to the Adolfo Lopez Mateos Dam (*Humaya*), Mexico.

T_X y T_Y	**Gumbel–Hougaard CF**			T_{XY} **secondary**	
	$C[F_X(x),F_Y(y)]$	T_{XY}	T'_{XY}	$K_C(s)$	ζ
50	0.9709003	34.4	91.7	0.990844	109.2
100	0.9854162	68.6	184.6	0.995450	219.8
500	0.9970778	342.2	927.9	0.999094	1 104.2
1 000	0.9985386	684.3	1 856.7	0.999547	2 209.6
5 000	0.9997076	3 420.4	9 294.9	0.999910	11 059.5
10 000	0.9998538	6 836.7	18 600.0	0.999955	22 104.4

Table 12. Estimated univariate and bivariate return periods with the Joe *CF* in the floods at the entrance to the Adolfo Lopez Mateos Dam (*Humaya*), Mexico.

T_X y T_Y	**Joe *CF***			T_{XY} **secondary**	
	$C[F_X(x),F_Y(y)]$	T_{XY}	T'_{XY}	$K_C(s)$	ζ
50	0.9736739	38.0	73.1	0.987932	82.9
100	0.9868368	76.0	146.3	0.993994	166.5
500	0.9973674	379.9	731.4	0.999022	1 022.7
1 000	0.9986837	759.7	1 462.7	0.999000	1 000.0
5 000	0.9997368	3 799.2	7 313.5	0.999800	5 000.7
10 000	0.9998683	7 594.9	14 627.0	0.999900	9 998.3

Conclusions

In Mexico, the six works cited: Aldama (2000), Ramírez-Orozco and Aldama (2000), Escalante–Sandoval and Reyes-Chávez (2002), Aldama *et al.* (2006), Gómez *et al.* (2010), and Domínguez and Arganis (2012), highlight the efforts made to develop *reliable flood estimation methods* that provide security to rural and urban hydraulic works that are dimensioned and/or revised with them. When seeking to estimate bivariate floods, as approximate as possible; this study adds to the aforementioned efforts.

The fundamental advantage of using the *Copula Functions* (*CF*) in *bivariate* floods frequency analysis (AFCb, by its acronym in Spanish), lies in easily constructing the *joint* probability distribution, based on the same or different marginal univariate distributions adopted; prior estimation of the dependence between the random variables: maximum flow (Q) and runoff volume (V) from the available annual floods.

The application of the *CF* in the AFCb has shown that the selection of the *ideal Copula* is based on reproducing the dependence of the extreme right observed; which guarantees a more approximate estimate of the *predictions* associated with high joint return periods. Therefore, *CF*s that exhibit, from plus to minus, various degrees of right tail dependency, are now tested by the families of Joe's Copulas, Associated Clayton, Gumbel–Hougaard, and Student's t.

The contrast made with 18 annual records of Q and V collected, detected two non–random ones, which were eliminated. For the rest, Joe's

CF is suggested in a record, processed as a numerical application. The Associated Clayton *CF* is adopted in another registry and the Gumbel–Hougaard *CF* is proposed in 10 registries. In three registers, its observed dependency is between the one exhibited by these two mentioned *CF*s. Finally, it was found that the *CF* of the Student's *t* reproduces the dependence of the Malpaso Dam record, with $\nu = 2$ (see Table 5).

The *CF*s that do not show such dependence, such as the Clayton, Frank and Plackett, are also applied to the available *Q* and *V* data and allow *validating* the quality of the fit achieved by the tested *CF*. The foregoing is supported by the fact that all the *CF*s cited and exposed show a similar fit to the data (see Table 8).

The processing of the 26 annual data of *Q* and *V* of the floods at the entrance to the Adolfo Lopez Mateos Dam (*Humaya*), in the state of Sinaloa, Mexico; It allowed to show the numerical development of the exposed theory and to describe in detail Figure 1 and Figure 2, which include the acceptance of the *CF* adopted and the basic graph of the results or *predictions* of Q and V of the *joint* return period type AND.

Acknowledgments

The suggestions of the anonymous referees B, F and I are appreciated, as they helped to correct drafting errors, omissions and the general structure of the article, to provide a sequence according to the theory and application of the Copula functions in the flood bivariate frequency analysis. His comments made it possible to highlight the efforts made in

Mexico to count with reliable methods for estimating design floods in large reservoirs.

References

AghaKouchak, A., Sellars, S., & Sorooshian, S. (2013). Chapter 6. Methods of tail dependence estimation. In: AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (eds.). *Extremes in a changing climate* (pp. 163-179). Dordrecht, The Netherlands: Springer.

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. (1975). The Log-Pearson type 3 distribution and its application to Hydrology. *Water Resources Research*, 11(5), 681-689. DOI: 10.1029/WR011i005p00681

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 7. Integración numérica. En: *Introducción a los métodos numéricos: Software en Basic y aplicaciones en Hidrología Superficial* (pp. 137-153). 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.

Campos-Aranda, D. F. (2022). Aplicación de la distribución GVE bivariada en el análisis de frecuencias conjunto de crecientes. *Tecnología y ciencias del agua*, 13(6), 534-602. DOI: https://doi.org/10.24850/j-tyca-13-06-11

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247-1250. DOI: 10.5194/gmd-7-1247-2014

Chen, L., & Guo, S. (2019). Chapter 2. Copula theory. Chapter 3: Copula-based flood frequency analysis. In: *Copulas and its application in Hydrology and Water Resources* (pp. 13-38, 39-71). Gateway East, Singapore: Springer.

Chowdhary, H., & Singh, V. P. (2019). Chapter 11. multivariate frequency distributions in hydrology. In: Teegavarapu, R. S. V., Salas, J. D., & Stedinger, J. R. (eds.). *Statistical Analysis of Hydrologic Variables* (pp. 407-489). Reston, USA: American Society of Civil Engineers.

Davis, P. J., & Polonsky, I. (1972). Chapter 25. Numerical interpolation, differentiation and integration. In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of mathematical functions* (pp. 875-926), 9th print. New York, USA: Dover Publications.

Demarta, S., & McNeil, A. J. (2005). The *t* Copula and related copulas. *International Statistical Review*, 73(1), 111-129.

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

Dupuis, D. J. (2007). Using Copulas in hydrology: Benefits, cautions, and issues. *Journal of Hydrologic Engineering*, 12(4), 381-393. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(381)

Escalante-Sandoval, C., & 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 de la Universidad Nacional Autónoma de México.

Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40(1), 1-12. DOI: 10.1029/2003WR002456

Frahm, G., Junker, M., & Schmidt, R. (2005). Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 37(1), 80-100. DOI: 10.1016/j-insmatheco.2005.05.008

Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about Copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4), 347-368. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(347)

Genest, C., Favre, A. C., Béliveau, J., & Jacques, C. (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research*, 4(W09401), 1-12. DOI: 10.1029/2006WR005275

Genest, C., & Chebana, F. (2017). Copula modeling in hydrologic frequency analysis. In: Singh, V. P. (ed.). *Handbook of Applied Hydrology* (pp. 30.1-30.10). 2nd ed. New York, USA: McGraw-Hill Education.

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

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.

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.

Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis*, 46(2), 262-282.

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

Kottegoda, N. T., & Rosso, R. (2008). Theme 3.5. Copulas. In: *Applied statistics for civil and environmental engineers* (pp. 154-157), 2nd. ed. Oxford, UK: Blackwell Publishing.

Meylan, P., Favre, A. C., & Musy, A. (2012). Chapter 1. Introduction; Chapter 3. Selecting and checking data series; Theme 9.2. Multivariate frequency analysis using Copulas (pp.). In: *Predictive hydrology. A frequency analysis approach* (pp. 1-13, 29-70, 164-176). Boca Raton, USA: CRC Press.

Michiels, F., & De Schepper, A. (2008). A Copula test space model. How to avoid the wrong copula choice. *Kybernetika*, 44(6), 864-878.

Nieves, A., & Domínguez, F. C. (1998). Secciones 6.2 y 6.3. Cuadratura de Gauss e integrales múltiples. En: *Métodos numéricos. Aplicados a la Ingeniería* (pp. 416-434). México, DF, México: Compañía Editorial Continental.

Ostle, B., & Mensing, R. W. (1975). Appendix 5. Cumulative *t* distribution. In: *Statistics in research* (pp. 544-545). 3rd ed. Ames, USA: Iowa State University Press.

Poulin, A., Huard, D., Favre, A. C., & Pugin, S. (2007). Importance of tail dependence in bivariate frequency analysis. *Journal of Hydrologic Engineering*, 12(4), 394-403. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(394)

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 e Instituto Mexicano de Tecnología del Agua.

Rao, A. R., & Hamed, K. H. (2000). Chapter 1. Introduction and Chapter 3. Probability weighted moments and L-moments. In: *Flood frequency analysis* (pp. 1-21, 53-72). Boca Raton, Florida, 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

Salvadori, G., & De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40(W12511), 1-17. DOI: 10.1029/2004WR003133

Salvadori, G., & De Michele, C. (2007). On the use of Copulas in Hydrology: Theory and practice. *Journal of Hydrologic Engineering*, 12(4), 369-380. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(369)

Salvadori, G., De Michele, C., Kottegoda, N. T., & Rosso, R. (2007). Chapter 3. Bivariate analysis via Copulas; Appendix B: Dependence and Appendix C: Families of Copulas. In: *Extremes in nature. An approach using Copulas* (pp. 131-175, 219-232, 233-269). Dordrecht, The Netherlands: Springer.

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

Sraj, M., Bezak, N., & Brilly, M. (2015). Bivariate flood frequency analysis using the copula function. A case study of the Litija station on the Sava River. *Hydrological Processes*, 29(2), 225-238. DOI: 10.1002/hyp.10145

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

Stegun, I. A. (1972). Chapter 27. Miscellaneous functions. In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of mathematical functions* (pp. 997-1010), 9th print. New York, USA: Dover Publications.

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

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82. DOI: 10.3354/cr030079

WMO, World Meteorological Organization. (1971). Annexed III: Standard tests of significance to be recommended in routine analysis of climatic fluctuations. In: *Climatic Change* (pp. 58-71) (Technical Note No. 79). Geneva, Switzerland: Secretariat of the World Meteorological Organization.

WRC, Water Resources Council. (1977). *Guidelines for determining flood flow frequency* (revised edition) (Bulletin # 17A of the Hydrology Committee). Washington, DC, USA: Water Resources Council.

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., & 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

Zelen, M., & Severo, N. C. (1972). Chapter 26. Probability Functions. In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of Mathematical Functions* (pp. 927-995), 9th print. New Work, USA: Dover Publications.

Zhang, L., & Singh, V. P. (2006). Bivariate flood frequency analysis using the Copula method. *Journal of Hydrologic Engineering*, 11(2), 150-164. DOI: 10.1061/(ASCE)1084-0699(2006)11:2(150)

Zhang, L., & Singh, V. P. (2019). Chapter 3. Copulas and their properties and Chapter 7. Non-Archimedean Copulas. Meta-Elliptical Copulas. In: *Copulas and their applications in water resources engineering* (pp. 62-122, 261-303). Cambridge, UK: Cambridge University Press.