DOI: 10.24850/j-tyca-2024-05-02

Articles

Trivariate flood frequency analysis (*Q, V, D*) through Copula Functions

Análisis de frecuencias de crecientes trivariado (*Q*, *V*, *D*) a través de funciones Cópula

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 *trivariate* flood maximum flow frequencies analysis (*Q*), the runoff volume (*V*) and the total duration (*D*), allow to estimate with greater accuracy the Design Flood's Hydrograph. To process available joint annual records of *Q* and *V*, it was proposed to estimate *D* as the duration of the Gamma hydrograph up to 0.1 % of *Q*. Then, each record of *Q*, *V* and *D* was searched for its ideal probability distribution to obtain the marginal functions. Next, the *Copula Function* (*CF*) that best represents the joint

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

variables *Q*–*V*, *Q*–*D* and *V*–*D* was adopted. For these searches and the subsequent trivariates, we worked with the Clayton, Frank, Gumbel-Hougaard and Joe *CF*s. In both cases, the selection of the best *CF* was based on the fit errors between the empirical and theoretical probabilities. For the triads of *Q*, *V* and *D data*, the symmetrical and asymmetrical best–fit *CF*s of the four families mentioned were sought. Next, the joint return periods of type OR, AND and Kendall were calculated. The latter allow the estimation of the *Q*, *V* and *D design events*. The trivariate frequency analysis is described for the 55 annual floods of the La Cuña hydrometric station from the Hydrological Region No. 12-3 (Santiago River), Mexico. Finally, the conclusions were formulated, which highlight the simplicity of the trivariate frequency analysis, when performed with *CF*.

Keywords: Copula functions, Kendall's tau ratio, observed dependence, symmetric multivariate Copula Functions, asymmetric trivariate Copula Functions, joint return periods, design events.

Resumen

El análisis de frecuencias de crecientes *trivariado*, del gasto máximo (*Q*), el volumen escurrido (*V*) y la duración total (*D*) permite estimar con mayor exactitud el hidrograma de la creciente de diseño. Para procesar registros anuales conjuntos de *Q* y *V* disponibles se propuso estimar *D* como la duración del hidrograma Gamma hasta el 0.1 % del *Q*. Después, a cada registro de *Q*, *V* y *D* se le busca su distribución de probabilidades idónea para obtener las funciones marginales. En seguida, se adopta la *función Cópula* (*FC*) que mejor representa a las variables conjuntas *Q*-*V*, *Q*-*D* y *V*-*D*. Para estas búsquedas y las trivariadas subsecuentes, se trabajó con las *FC* de Clayton, Frank, Gumbel-Hougaard y Joe. En ambos

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

casos, la selección de la mejor *FC* se basa en los errores de ajuste entre las probabilidades empíricas y teóricas. A las ternas de datos *Q*, *V* y *D* se les buscó las *FC* de mejor ajuste simétricas y asimétricas de las cuatro familias citadas. A continuación se calculan los periodos de retorno conjuntos de tipo OR, AND y de Kendall. Estos últimos permiten la estimación de los *eventos de diseño* de *Q*, *V* y *D*. Se describe el análisis de frecuencias trivariado para las 55 crecientes anuales de la estación hidrométrica La Cuña de la Región Hidrológica No. 12-3 (Río Santiago), México. Por último, se formulan las conclusiones, que destacan la sencillez de los análisis de frecuencias trivariados cuando se realizan con *FC*.

Palabras clave: funciones Cópula, cociente tau de Kendall, dependencia observada, funciones Cópula multivariadas simétricas, funciones Cópula trivariadas asimétricas, periodos de retorno conjuntos, eventos de diseño.

Received: 19/09/2022

Accepted: 04/04/2023

Published Online: 17/04/2023

Introduction

Generalities

The first records of *multivariate* flood frequency analyses date from the end of the last century. These studies processed three variables: maximum or peak flow (*Q*), runoff volume (*V*) and total duration (*D*), but accepted several restrictive considerations being the most common: (1)

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

consider that the variables *Q*, *V* and *D* are independent; (2) accept that the marginal distributions are equal and (3) adopt that such distributions are normal, or apply a transformation towards normality (Goel, Seth, & Chandra, 1998; Zhang & Singh, 2007).

Yue, Ouarda, Bobée, Legendre and Bruneau (1999) apply an extreme bivariate distribution, the Gumbel, to process two pairs of variables *Q*–*V* and *V*–*D*. This study considers the correlation between the variables, but their marginals are of the same type.

In fact, the variables *Q*, *V* and *D* of the annual floods are correlated and do not come from Normal distributions (Zhang & Singh, 2007). On the other hand, the *Design Hydrograph* that is derived from them must have a magnitude or probability of exceedance, which ensures safety of the reservoir to be designed or reviewed, in accordance with the established standards (Gräler *et al.*, 2013; Xu, Yin, Guo, Liu, & Hong, 2016).

In order to overcome the aforementioned disadvantages, inherent to the *probability distribution functions* (PDF), which are applied in the analysis of multivariate increasing frequencies, in recent years the *Copula Functions* (*CF*) have been used; mathematical models that are based on the correlation shown by the variables *Q*, *V* and *D* and which allow the construction of the multivariate PDF, exclusively from the previously adopted marginal functions (Salvadori & De Michele, 2004, 2007; Favre, El Adlouni, Perreault, Thiémonge, & Bobée, 2004; Meylan, Favre, & Musy, 2012; Genest & Chebana, 2017; Campos-Aranda, 2024).

Regarding the *trivariate* frequency analysis, its beginnings can be found in Grimaldi and Serinaldi (2006a), who used seven symmetric multivariate FCs to obtain a design hyetogram. Zhang and Singh (2007)

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

processed annual Q, V and D data based on the Gumbel-Hougaard *CF*, their results are contrasted against those of the trivariate Normal distribution. Grimaldi and Serinaldi (2006b) describe and apply the asymmetric multivariate *CF*, which is explained later. Ma, Song, Ren, Jiang and Song (2013) processed annual droughts with the trivariate approach, using *CF* Gaussian and Student's t. Finally, Zhang and Singh (2019) comprehensively address the trivariate flood frequencies analysis using *CF*.

Objectives

In the present study, the following seven objectives were formulated: (1) as Q and V joint data were available, their total duration (D) was introduced as the third random variable of the floods; (2) we worked with four Archimedean *CF* families for the Q–V, Q–D and V–D bivariate analyses: Clayton, Frank, Gumbel-Hougaard and Joe; (3) the aforementioned *CF* families were applied in the trivariate analyses, with their multivariate versions called symmetric; (4) the aforementioned nested or asymmetric *CF* families with two association parameters were also used in the trivariate analyses; (5) trivariate return periods of the OR, AND and Kendall types were estimated; (6) design events were estimated, based on Kendall's joint return period and (7) the exposed theory was applied to the joint record of Q and V of the 55 annual floods recorded at the La Cuña hydrometric station in the Hydrological Region. No. 12–3 (Santiago River), Mexico.

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

In Figure 1, the sequence of theoretical topics and their complementary calculations are schematized, with the idea of formulating a flowchart of the entire study.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

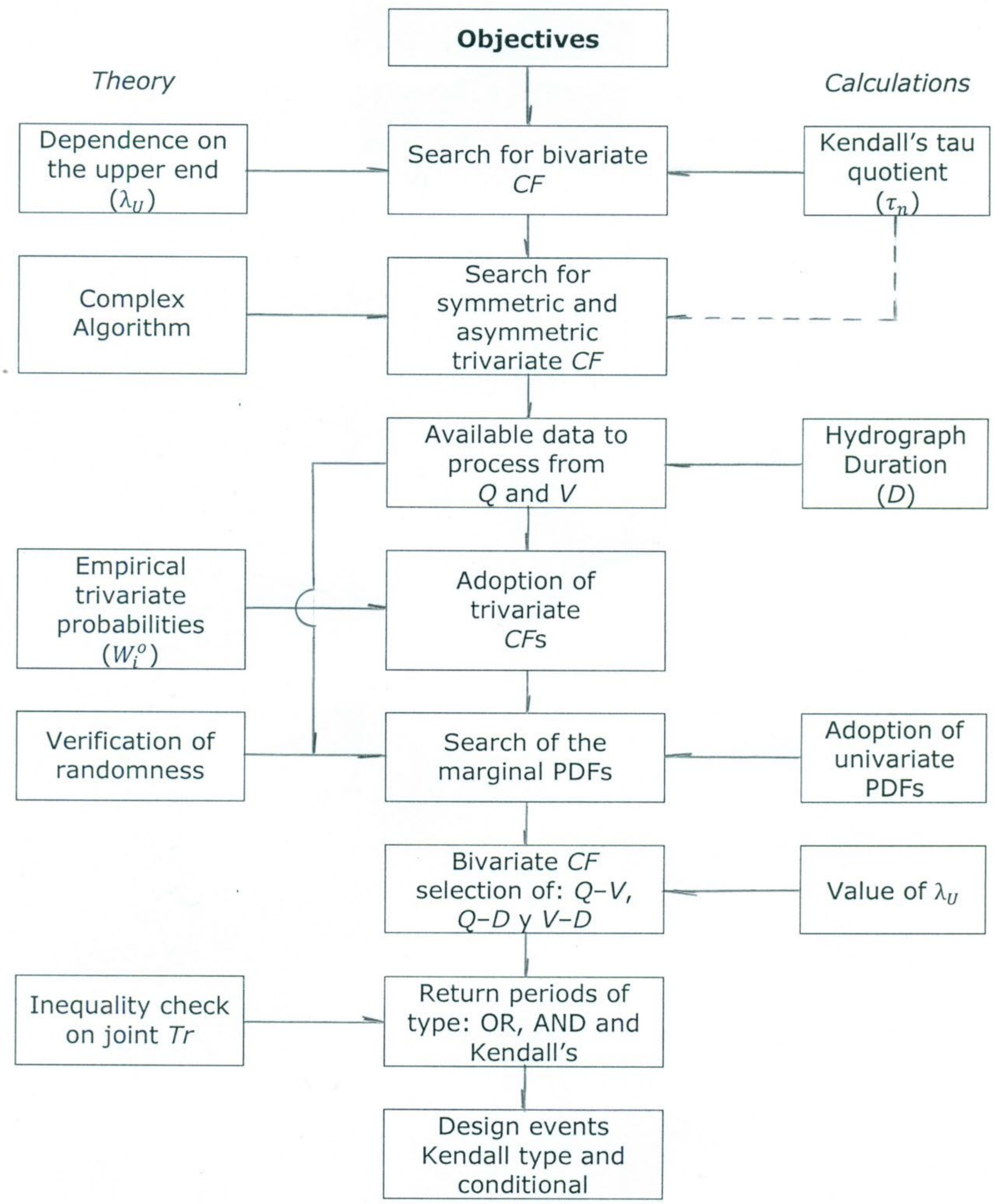

Figure 1. Flowchart of theoretical concepts and their complementary calculations, carried out in the study.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

Topics of the Copula Functions

Advantages and definition

The fundamental advantage of *Copula Functions* (*CF*) consists in allowing to form and express the *joint distribution* of random variables that are correlated, 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 CFs when forming multivariate distributions is the fact that they separate the effect of dependence between random variables from the effects of marginal distributions in joint modelling.

Therefore, the construction of the multivariate distribution is reduced to the study of the relationship between the correlated variables, when the univariate marginal distributions are known. The use of *CF* offers complete freedom to adopt or select the univariate marginal distributions that best represent the data (Shiau, Wang, & Tsai, 2006; Salvadori, De Michele, Kottegoda, & Rosso, 2007; Meylan *et al*., 2012; Zhang & Singh, 2019).

Since at the present study, the *CF*s will be applied in the *trivariate frequency analysis* of annual floods, the following definition refers to three correlated random variables X, Y and Z, where the joint cumulative probability distribution function is $F_{XYZ}(x,y,z)$ with univariate marginal distributions $F_X(x)$, $F_Y(y)$ and $F_Z(z)$; then the *CF* exists, is $C[\cdot]$ and is such that:

$$F_{XYZ}(x,y,z) = C[F_X(x), F_Y(y), F_Z(z)] \tag{1}$$

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

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

Copula Families

The *Copula functions* (*FC*) that have been developed have been classified into four classes: Archimedean, extreme value, elliptic, and miscellaneous. They are also classified into one-parameter or multi-parameter copulas, depending on the amplitude with which the structure of the dependency between the variables *Q*, *V* and *D* is defined (Meylan *et al.*, 2012; Chowdhary & Singh, 2019). Salvadori *et al.* (2007) present a comprehensive and useful summary of *CF* that have been applied in the field of hydrology.

Bivariate Archimedean Copulas

Archimedean copulas have found wide application due to their simple construction, single parameter, wide range, and acceptance of both types of dependency (positive and negative). By designating $FX(x)=u$, $F_Y(y)=v$ and θ the parameter that measures the dependence or association between u and v, the following four families of Archimedean copulas are obtained, which accept positive and/or negative dependence (Shiau *et al.*, 2006; Genest & Favre, 2007; Salvadori *et al.*, 2007; Zhang & Singh, 2019; Chen & Guo, 2019; Chowdhary & Singh, 2019).

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

1. **Clayton**. This Copula function is called Cook–Johnson by Zhang and Singh (2006). Its equation and variation space of θ are:

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

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

$$\tau_n = \frac{\theta}{\theta+2} \tag{3}$$

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

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

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

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

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

where D1(θ) is the Debye function of order 1, whose expression is:

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

in which, *s* is the unitary random variable $0 < s \leq 1$. The previous equation was estimated with numerical integration, based on Equation (13), ratifying its results with the values tabulated by Stegun (1972).

3. **Gumbel-Hougaard**, which accepts only positive dependence. 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 θ=1 there is independence between *u* and *v*. The relationship of θ to Kendall's tau ratio is as follows:

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

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

$$C(u, v) = 1 - \left[(1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta}(1-v)^{\theta}\right]^{1/\theta} \tag{9}$$

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

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

$$\tau_n = 1 + 4\int_0^1 \frac{\varphi(s)}{\varphi\prime(s)}\, ds \tag{10}$$

in which, 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):

$$\varphi(s) = -\ln\left[1 - (1 - s)^{\theta}\right] \tag{11}$$

$$\varphi'(s) = \frac{\theta(1-s)^{\theta-1}}{(1-s)^{\theta}-1} \tag{12}$$

Taking the value of Kendall's tau quotient as data, Equation (10) is integrated numerically based on Equation (13), to obtain by trial and error the value of θ that satisfies it.

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Numerical integration

To quantify the integrals, for example, equations (6) and (10), a numerical integration was performed based on the Gauss-Legendre quadrature method, whose operational equation is (Nieves & Domínguez, 1998; Campos-Aranda, 2003):

$$\int_a^b f(x)\,dx \cong \frac{b-a}{2}\sum_{i=1}^{np} w_i \cdot f\left[\frac{(b-a)h_i+b+a}{2}\right] \tag{13}$$

in which, w_i are the coefficients of the method where the abscissas are h_i and *np* the number of pairs in which the function *f*(*x*) is evaluated, with the argument indicated in *f* [·]. In Davis and Polonsky (1972) the 12 used pairs of w_i and h_i with 15 digits were obtained, which are accepted in the *Basic* language, as double precision variables.

Numeric indicator of association

Concordance

Since the *CF* characterizes the *dependency* between the random variables *u* and *v*, it is necessary to study the association measures, in order to have a method that allows estimating its parameter θ. In general terms, a 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).

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Some variables with direct linear correlation will be concordant, since when one increases the other also does so. Variables with inverse linear correlation will be *discordant*, as large values of one will correspond to small values of the other and vice versa. 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).

Kendall's tau quotient

It is a non–parametric numerical indicator that measures the probability of having concordant couples, which is why it is the *quotient* of $c-d$ between $c+d$. The expression to estimate it with bivariate data is (Zhang & Singh, 2006; Zhang & Singh, 2019):

$$\tau_n = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{signo}\left[(x_i - x_j)(y_i - y_j)\right] \quad (14)$$

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, which accepts the null hypothesis H_0 of independent X and Y and then its statistic has an approximately Normal distribution with mean zero and variance $2(2n+5)/[9n(n-1)]$. Therefore, H_0 will be rejected with a confidence level α = 5 % if:

$$\sqrt{\frac{9n(n-1)}{2(2n+5)}}\,|\tau_n| > Z_{\alpha/2} = 1.96 \quad (15)$$

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Dependence on the extremes of the bivariate CF's

Generalities

The most important criterion applied to select a bivariate *CF* is based on the magnitude of the upper tail dependence in the joint distribution, which has an impact on the veracity of the extreme predictions. 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 unity. The lower left tail dependency (λ_L) compares *Y* to be less than *X*, as *s* approaches zero (Chowdhary & Singh, 2019).

In relation to the bivariate *CF*s exposed, Frank's has non-significant dependencies in its extreme zones: therefore, λ_L=0 and λ_U=0. Clayton's has a significant dependence on its lower tail equal to: $\lambda_L = 2^{-1/\theta}$ In contrast, the Gumbel-Hougaard and Joe copulas have significant upper–tail dependency, equal to:

$$\lambda_U = 2 - 2^{1/\theta} \tag{16}$$

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 Clayton associate (*Survival Clayton*). Poulin, Huard, Favre and Pugin (2007) arrive to similar conclusions, when comparing the same six

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

families of copulas and the so–called A12, which has significant dependence on its right tail.

Estimate of the observed dependency

In order to approach the estimation of the upper tail dependence (λ_U) shown by the available data, the *Empirical Copula* must be defined first. Like the *CF* that characterizes the dependence between the random variables X and Y; then the pair of ranges R_i and S_i coming from such variables is 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:

$$C_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} 1\left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v\right) \tag{17}$$

In the above equation, $1(\cdot)$ indicates 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:

$$U_i = \frac{\text{Range } (X_i)}{n+1} = F_n(X_i)\ V_i = \frac{\text{Range } (Y_i)}{n+1} = F_n(Y_i) \tag{18}$$

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

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:

$$\lambda_U^{CFG} = 2 - 2\exp\left\{\frac{1}{n}\sum_{i=1}^{n} \ln\left[\sqrt{\ln\frac{1}{U_i}\cdot\ln\frac{1}{V_i}}/\ln\left(\frac{1}{\max(U_i,V_i)^2}\right)\right]\right\} \tag{19}$$

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

Trivariate Archimedean Copulas

Symmetrical Archimedean Copulas

Chen and Guo (2019) indicate for multivariate random variables, greater than two ($d \geq 3$) and correlated, that the family of Archimedean copulas are divided into symmetric and asymmetric. The former are easy to build and have a single association parameter (θ), which forces all pairs of variables to show the same structure and degree of dependency (Zhang & Singh, 2019).

The most common *symmetric multivariate* Archimedean copulas ($d \geq 3$) are four, for which their range of θ is indicated, their generating functions $\varphi(s)$ and their first and second derivatives $\varphi'(s)$, $\varphi''(s)$, where s

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

is a random variable in the interval from 0 to 1 (Grimaldi & Serinaldi, 2006a; Xu *et al.*, 2016; Chen & Guo, 2019; Zhang & Singh, 2019).

1. **Clayton's multivariate copula**, the range of θ is (0, +∞) and the limit of θ=0 corresponds to the independence condition in u_k:

$$C(u_1, u_2, \cdots, u_d) = \left(\sum_{k=1}^{d} u_k^{-\theta} - d + 1\right)^{-1/\theta} \tag{20}$$

$$\varphi(s) = \frac{1}{\theta}(s^{-\theta} - 1) \tag{21}$$

$$\varphi'(s) = -s^{-1-\theta} \tag{22}$$

$$\varphi''(s) = (\theta + 1)/s^{\theta+2} \tag{23}$$

2. **Frank's multivariate copula**, the range of θ is (0, +∞) and the value of θ=1 corresponds to the independence condition in u_k:

$$C(u_1, u_2, \cdots, u_d) = -\frac{1}{\theta}\ln\left[1 + \frac{\prod_{k=1}^{d}(e^{-\theta u_k} - 1)}{(e^{-\theta} - 1)^{d-1}}\right] \tag{24}$$

$$\varphi(s) = -\ln\left(\frac{e^{-\theta s} - 1}{e^{-\theta} - 1}\right) \tag{25}$$

$$\varphi'(s) = \frac{\theta}{1 - e^{\theta s}} \tag{26}$$

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

$$\varphi''(s) = \frac{\theta^2}{e^{\theta s} - 2 + e^{-\theta s}} \tag{27}$$

3. **Gumbel-Hougaard multivariate copula**, the range of θ is (1, $+\infty$) and the limit of θ=1 corresponds to the independence condition on u_k:

$$C(u_1, u_2, \cdots, u_d) = \exp\left\{-\left[\sum_{k=1}^{d}(-\ln u_k)^{\theta}\right]^{1/\theta}\right\} \tag{28}$$

$$\varphi(s) = [-\ln(s)]^{\theta} \tag{29}$$

$$\varphi'(s) = -\frac{\theta}{s}[-\ln(s)]^{\theta-1} \tag{30}$$

$$\varphi''(s) = \frac{\theta}{s^2}\left\{(\theta-1)[-\ln(s)]^{\theta-2} + [-\ln(s)]^{\theta-1}\right\} \tag{31}$$

4. **Joe's multivariate copula**, the range of θ is [1, $+\infty$) and the limit of θ=1 corresponds to the independence condition on u_k:

$$C(u_1, u_2, \cdots, u_d) = 1 - \left[1 - \prod_{k=1}^{d}\left(1 - (1-u_k)^{\theta}\right)\right]^{1/\theta} \tag{32}$$

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Tecnología y Ciencias del Agua
OPEN ACCESS

The equations of its generating function and its first derivative have already been exposed as equations (11) and (12). The equation of the second derivative of $\varphi(s)$ is the following:

$$\varphi''(s) = \frac{\theta(\theta-1)(1-s)^{\theta-2}+\theta(1-s)^{2\theta-2}}{\left[(1-s)^{\theta}-1\right]^{2}} \quad (33)$$

The application of the symmetric trivariate Archimedean *CF* is justified for comparison purposes, since the asymmetric ones should lead to better adjustments.

Asymmetric Archimedean Copulas

To model different dependency structures, in multivariate random variables, Chen and Guo (2019) resort to the approach by Grimaldi and Serinaldi (2006b), of applying *nested* Archimedean copulas. With such approach, the most common asymmetric trivariate Archimedean copulas of two association parameters ((θ_1 and θ_2) have the general formula: $C(u,v,w) = C_{\theta_1}\left(w, C_{\theta_2}(u,v)\right)$ and are the following four (Joe, 1993; Grimaldi & Serinaldi, 2006b; Xu *et al.*, 2016; Zhang & Singh, 2019; Chen & Guo, 2019):

1. **Trivariate asymmetric Clayton copula**, with $\theta_2 \geq \theta_1 \in [0,\infty$ and $\tau_{12}, \tau_{13}, \tau_{23} \in [0,1]$ for three random variables with positive dependence:

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

$$C(u,v,w) = \left[\left(u^{-\theta_2} + v^{-\theta_2} - 1\right)^{-\theta_1/\theta_2} + w^{-\theta_1} - 1\right]^{-1/\theta_1} \tag{34}$$

2. **Trivariate Frank's asymmetric copula**, with a range of association parameters identical to the previous ones:

$$C(u,v,w) = -\frac{1}{\theta_1}\ln\left\{1 - \mathrm{F}_1^{-1}(1 - \left[1 - \mathrm{F}_2^{-1}\left(1 - e^{-\theta_2 u}\right)\left(1 - e^{-\theta_2 v}\right)\right]^{\theta_1/\theta_2})\left(1 - e^{-\theta_1 w}\right)\right\} \tag{35}$$

being, $\mathrm{F}_1 = 1 - e^{-\theta_1}$ y $\mathrm{F}_2 = 1 - e^{-\theta_2}$.

3. **Trivariate asymmetric Gumbel-Hougaard copula**, with $\theta_2 \geq \theta_1 \in [1,\infty$ and $\tau_{12}, \tau_{13}, \tau_{23} \in [0,1]$ for three random variables with positive dependence:

$$C(u,v,w) = \exp\left\{-\left[\left((-\ln u)^{\theta_2} + (-\ln v)^{\theta_2}\right)^{\theta_1/\theta_2} + (-\ln w)^{\theta_1}\right]^{1/\theta_1}\right\} \tag{36}$$

4. **Joe trivariate asymmetric copula**, with range of association parameters identical to the previous ones:

$$C(u,v,w) = 1 - \left\{\left[(1-u)^{\theta_2}\left(1 - (1-v)^{\theta_2}\right) + (1-v)^{\theta_2}\right]^{\theta_1/\theta_2}\left(1 - (1-w)^{\theta_1}\right) + (1-w)^{\theta_1}\right\}^{1/\theta_1} \tag{37}$$

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Trivariate empirical probabilities

The empirical univariate and trivariate non–exceedance probabilities were estimated based on the Gringorten formula, which has been suggested by various authors in bivariate frequency analyses and by Zhang and Singh (2007) for trivariate ones; its expression is:

$$p = \frac{i-0.44}{n+0.12} \tag{38}$$

where *i* is the data number when they are ordered from lowest to highest and *n* is the total number or years of records of peak flow, volume and total annual duration.

In the case of trivariate probabilities, we worked in a three–dimensional space, with flow and volume in the *x*,*y* plane, and durations in the perpendicular (*z*) axis. The numerical process begins by saving the historical records of annual maximum flow (*Q*), volume (*V*) and duration (*D*) in files *Qh*, *Vh* and *Dh*; also, they were ordered in a progressive manner of magnitude in files *Qo*, *Vo* and *Do*. Next, each annual data is processed, to compare the historical value against the ordered one and the times that the second was less than or equal to it are counted and *NQ*, *NV* and *ND* are designated. The foregoing is equivalent to changing the original data, of each list of historical annual values, by its order number or range.

Afterwards, each historical range triplet is compared against all the others and the times in which the three ranges (AND condition) are lower are counted and such amount is called *NQVD*; that is, the number of

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

occurrences of minor q, v, and d combinations in three–dimensional space. Finally, the Gringorten graphic position formula is applied, for the trivariate case it is the following:

$$F_e(x, y, z) = P(Q \leq q, V \leq v, D \leq d) = \frac{NQVD_i - 0.44}{n + 0.12} \tag{39}$$

Selection of the Copula Function

A simple approach to selecting the Copula Function is based on the statistics of the fitting error, by comparing the observed empirical probabilities (W_i^o) with the calculated theoretical ones (W_i^c) with the Copula Function 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; Chen & Guo, 2019):

$$EME = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (W_i^o - W_i^c)^2} \tag{40}$$

$$EMA = \frac{1}{n} \sum_{i=1}^{n} |(W_i^o - W_i^c)| \tag{41}$$

$$EAM = \max_{i=1:n} |(W_i^o - W_i^c)| \tag{42}$$

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Estimation of the dependency parameter θ

The simplest method to estimate the parameter θ of the symmetric trivariate *CF*s (equations (20), (24), (28) and (32)), is by trial and error seeking that the fitting statistics (equations (40) to (42)) are minimal.

Estimation of the dependency parameters θ_1 and θ_2

The search for the minimum value of Equation (40) or mean standard error, for the fit of the asymmetric trivariate *CF*s of Clayton, Frank, Gumbel-Hougaard and Joe, defined by equations (34) to (37), was carried out based on the Complex algorith mof multiple restricted or bounded variables, to find the optimal values of θ_1 and θ_2, fulfilling the condition $\theta_2>\theta_1$.

The *Complex algorithm* is a local exploratory technique (Box, 1965), which is guided exclusively by what its found in its path; its background, a brief description of its operational process and its OPTIM code in *Basic* language can be consulted in Campos-Aranda (2003). See Bunday (1985) for another description and code of this search method.

The main designations in the OPTIM code are NX and NY, which define the number of decision and dependent variables, a function of the former; for the analyzed case two (θ_1,θ_2) and one ($\theta_2>\theta_1$).

An important advantage of the OPTIM code lies in allowing easy access to the limits (L = lower, U = upper), names and initial values of the variables, in the aforementioned subroutine, by means of the following designations: XL(I), XU(I), XN$(I), X(I), YL(J), YU(J), YN$(J), and Y(J). For the case studied, I varies from 1 to 2 and J=1. In all the

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

decision variables, 0.001 was used as the lower limit and 5 and 10 as the upper limit, for θ_1 and θ_2, in the Clayton and Gumbel-Hougaard *CF* and 10 and 20 in the Frank and Joe *CF*. The only dependent variable was defined by the ratio of θ_2 between θ_1, with a lower limit of one and an upper limit of 5; value that was arbitrarily adopted.

The objective function is called FO in the OPTIM code and is defined at the end of the program, it logically corresponds to Equation (40), named FO$="*EME*", mean standard error. For the convergence criteria of the absolute and relative deviations of the FO, the following values were used: 0.0002 and 0.00001.

Ratification of the selected Copula Function

This is the most important stage in the process of practical application of the *CF*, since it is verified that such a model faithfully reproduces the observed joint probabilities (Equation (39)). Yue (2000) indicates the simple and practical way to represent the empirical and theoretical joint probabilities. This consists of taking the first axis to the abscissa 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.

Yue and Rasmussen (2002) apply the Kolmogorov–Smirnov test with a significance level (α) of 5 %, to accept or reject the *absolute maximum difference (dma*, by its acronym in Spanish) between the empirical and theoretical joint probabilities. To evaluate the statistic (D_n) of the test, the expression that Meylan *et al.* (2012), for α = 5 % this is:

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

$$D_n = \frac{1.358}{\sqrt{n}} \quad (43)$$

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

Trivariate return periods

OR and AND types

The first *trivariate return period* of the event (X, Y, Z) is defined under the OR condition, which indicates that the limits x, y or z, or all three *can be* exceeded and then, the classical equation of the return period or inverse of the exceedance probability will be (Zhang & Singh, 2019):

$$T_{XYZ} = \frac{1}{P(X>x \text{ or } Y>y \text{ or } Z>z)} = \frac{1}{1-F_{XYZ}(x,y,z)} = \frac{1}{1-C[F_X(x),F_Y(y),F_Z(z)]} \quad (44)$$

in which, $C[F_X(x), F_Y(y), F_Z(z)] = C(u,v,w)$ is the selected or tested *CF*.

The second *trivariate return period* of the event (X, Y, Z) is associated to the case in which the three limits are exceeded ($X>x$, $Y>y$, $Z>z$) or AND condition, its equation is the (45) following (Zhang & Singh, 2019):

$$T'_{XYZ} = \frac{1}{P(X>x \text{ and } Y>y \text{ and } Z>z)} = \frac{1}{F'_{XYZ}(x,y,z)} = \frac{1}{1-u-v-w+C(u,v)+C(u,w)+C(v,w)-C(u,v,w)} \quad (45)$$

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Secondary or Kendall type

Salvadori and De Michele (2004) introduce in detail the concept of the *Secondary bivariate Return Period* (ζ), thus designated to emphasize that the joint return period T_{XY} is the primary one, from which it proceeds when using the *isolines* defined by the applied *CF*, whici is expressed as:

$$L_s = [(u, v) \in \mathbb{I}^2 : C(u, v) = s] \tag{46}$$

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

$$B_C(s) = \{(u, v) \in \mathbb{I}^2 : C(u, v) \leq s\} \tag{47}$$

In the *CF*s of Archimedean class, the univariate Kendall 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):

$$K_C(s) = s - \frac{\varphi(s)}{\varphi'(s)} \tag{48}$$

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

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

$$\zeta = \frac{1}{1-K_C(s)} \tag{49}$$

whose denominator is the probability of exceedance (*survival function*), which corresponds to probably destructive or dangerous events.

The Kendall parametric distribution (Equation (49)), for the symmetric *trivariate* Archimedean copulas is the following (Barbe, Genest, Ghoudi, & Rémillard, 1996; Grimaldi & Serinaldi, 2006a; Zhang & Singh, 2019):

$$K_C(s) = P[C(u,v,w) \leq s] = s - \frac{\varphi(s)}{\varphi'(s)} - \frac{\varphi^2(s)\cdot\varphi''(s)}{2[\varphi\prime(s)]^3} \tag{50}$$

By substituting equations (29) to (31) in (50), we obtain an expression for the Kendall distribution of the symmetric trivariate Gumbel-Hougaard *CF*, which is used later.

Graler *et al*. (2013) extend the inequality $T_{XYZ} \leq T'_{XYZ}$, which indicates that the OR type return period is less than the AND type and indicates that T_{KEN} is intermediate between the two mentioned. T_{KEN} is obtained with Equation (49). Thus obtaining:

$$T_{XYZ} \leq T_{KEN} \leq T'_{XYZ} \tag{51}$$

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Salvadori, De Michele and Durante (2011) highlight that the estimation of return periods and their design events in multivariate frequency analyses are a complex problem. To solve it, they establish a theoretical framework based on *CF* and the Kendall distribution, which they apply through numerical simulation.

Conditional type

Regarding the *conditional trivariate return periods*, Zhang and Singh (2007) in their appendix, present a summary of those they consider to be of practical application. Zhang and Singh (2019) present an exhaustive list of the different types; all of them studied by Serinaldi (2015). Ma *et al.* (2013) propose the following conditional probability:

$$P(Q \leq q, V \leq v | D \geq d) = \frac{C(u,v) - C(u,v,w)}{1-w} \qquad (52)$$

Grimaldi and Serinaldi (2006b) contrast the results of Equation (52) with the symmetric and asymmetric Frank copulas for the maximum values of volume and duration (V,D), given that magnitudes of maximum flow (Q) have occurred with various periods of univariate return. The asymmetrical Frank *CF* provides a better fit and larger values of V and D.

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Data to be processed

Joint record of *Q* and *V*

This record was integrated with the 55 data on maximum flow (*Q*) and runoff volume (*V*) during the annual floods recorded at the La Cuña hydrometric station, of the Río Verde in Hydrological Region No. 12–3 (Rio Santiago), Mexico; with a basin area of 19097 km^2. Such a record was exposed by Gomez, Aparicio and Patiño (2010), and by Campos-Aranda (2024) and is reproduced in Table 1.

Table 1. Joint records of peak flow (*Q*) and volume (*V*) of the annual floods registered at the hydrometric station La Cuña, Mexico.

Year	*Q* (m³/s)	*V* (Mm³)
1947	784.0	146.80
1948	736.8	155.12
1949	510.0	111.40
1950	461.0	94.06
1951	411.0	111.55
1952	326.0	70.82
1953	349.8	144.75
1954	130.4	23.22
1955	690.0	203.31
1956	266.0	106.76
1957	199.0	45.92

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Year	Q (m^3/s)	V (Mm^3)
1958	690.0	188.71
1959	340.6	47.91
1960	249.6	91.58
1961	350.0	130.68
1962	317.0	51.27
1963	732.6	127.90
1964	265.1	82.75
1965	743.6	295.34
1966	463.9	202.90
1967	1474.9	598.38
1968	323.0	118.25
1969	160.4	32.22
1970	763.8	187.75
1971	578.0	166.61
1972	191.8	26.39
1973	2440.0	920.30
1974	238.4	66.66
1975	622.1	249.07
1976	1374.0	527.96
1977	439.7	111.77
1978	280.2	66.23
1979	267.2	45.80
1980	287.3	99.60

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Year	*Q* (m³/s)	*V* (Mm³)
1981	280.7	28.70
1982	156.5	35.37
1986	698.2	193.51
1987	184.7	55.39
1988	595.2	242.21
1989	110.2	42.49
1990	523.9	248.07
1991	1636.3	443.30
1992	1168.0	172.49
1993	295.0	96.50
1994	212.8	53.55
1995	367.4	114.61
1996	144.6	57.43
1997	78.4	16.55
1998	261.9	66.17
1999	196.3	41.15
2000	46.8	18.62
2001	313.8	75.78
2002	319.6	153.79
2003	621.1	326.28
2004	824.5	384.45
media	499.87	154.84
MED	340.60	111.40

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Year	Q (m^3/s)	V (Mm^3)
D.E.	432.23	162.58
Cs	2.403	2.699
Ck	10.364	12.105

Estimation of the total duration (*D*)

For the estimation or assignment of the third random variable, related to the total duration (D) of each annual flood in hours, the representation of the hydrographs with the Gamma probability distribution of two adjustment parameters was used, which was proposed by Ponce (1989) and applied and described by Aldama-Rodriguez and Ramírez-Orozco (1998).

This method begins by estimating an annual peak time (t_p), with the following expression, which comes from the dimensionless unit hydrographs defined by Snider (1972), also cited by Aldama, Ramírez, Aparicio, Mejía-Zermeño and Ortega-Gil (2006):

$$t_p = ct\frac{3V}{4Q} \tag{53}$$

When the volume (V) of the annual flood is expressed in millions of m^3 (Mm^3) and its peak flow (Q) in m^3/s, the transformation coefficient (ct) of the above equation is 208.3333 with the peak time (t_p) in hours.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Ponce (1989) exposes the analytical equation of the Gamma hydrographs, function of the peak times (t_p) and to the centroid of the aforementioned hydrograph (t_g), this is:

$$Q = Q_p \left(\frac{t}{t_p}\right)^m \exp\left[\frac{t_p - t}{t_g - t_p}\right] \text{ para } t > 0 \tag{54}$$

in which, Q is the flow, Q_p is the peak flow, t_p is the peak time, t_g is the time to the centroid of the hydrograph ($t_g > t_p$) and the exponent $m = t_p/(t_g - t_p)$. The flows are expressed in units of m^3/s, the times in seconds and the volumes in m^3.

Integrating the previous equation, an expression for the volume (V) is obtained, which can be solved by trial and error to estimate the t_g, since the volume is given, this is (Aldama-Rodriguez & Ramirez-Orozco, 1998):

$$V = \int_0^\infty Q(t) \cdot dt = Q_p \cdot \exp(m) \left(\frac{1}{m}\right)^m (t_g - t_p) \cdot \Gamma(1 + m) \tag{55}$$

For the estimation of the Gamma function $\Gamma(\omega)$ the Stirling formula (Davis, 1972) was used:

$$\Gamma(\omega) \cong e^{-\omega} \cdot \omega^{\omega - \frac{1}{2}} \cdot (2\pi)^{1/2} \cdot F1 \tag{56}$$

being:

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

$$F1 = \left(1 + \frac{1}{12\cdot\omega} + \frac{1}{288\cdot\omega^2} - \frac{139}{51840\cdot\omega^3} - \frac{571}{2488320\cdot\omega^4} + \cdots\right) \quad (57)$$

Taking Q and V from Table 1 as data, the peak time (t_p) is estimated with Equation (53) and then by trial and error the t_g with Equation (55). Finally, estimates of the final time (t=t_f) are made, also by trial and error, up to 0.1 % of the peak flow (Q_p=Q) with Equation (54). Table 2 shows the results obtained for the data in Table 1.

Table 2. Peak times (t_p), to the centroid (t_g) and to the end (t_f) in the Gamma hydrographs that represent the annual floods of the hydrometric station La Cuña, Mexico.

Year	t_p hours	t_g hours	t_f=D hours	Q_f m³/s
1947	39.009	49.56	169.0	0.79
1948	43.861	55.73	190.2	0.74
1949	45.507	57.82	197.4	0.51
1950	42.507	54.01	184.5	0.46
1951	56.544	71.84	245.0	0.41
1952	45.258	57.50	196.0	0.33
1953	86.210	109.53	374.0	0.35
1954	37.097	47.13	161.3	0.13
1955	61.386	77.99	266.2	0.69
1956	83.615	106.23	362.5	0.27
1957	48.074	61.08	207.7	0.20
1958	56.978	72.39	247.1	0.69
1959	29.305	37.23	127.2	0.34

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Year	t_p hours	t_g hours	t_f=*D* hours	Q_f m^3/s
1960	76.439	97.12	331.8	0.25
1961	77.786	98.83	337.6	0.35
1962	33.695	42.81	146.2	0.32
1963	36.372	46.21	157.8	0.73
1964	65.030	83.03	286.4	0.26
1965	82.745	105.13	359.0	0.74
1966	91.121	115.77	395.6	0.46
1967	84.523	107.39	368.1	1.40
1968	76.271	96.90	331.3	0.32
1969	41.849	53.17	181.9	0.16
1970	51.211	65.06	222.1	0.76
1971	60.053	76.30	260.4	0.58
1972	28.665	36.42	124.6	0.19
1973	78.578	99.83	341.1	2.40
1974	58.253	74.01	252.8	0.24
1975	83.410	105.97	361.9	0.62
1976	80.052	101.71	347.2	1.37
1977	52.957	67.28	229.7	0.44
1978	49.243	62.56	213.7	0.28
1979	35.710	45.37	154.9	0.27
1980	72.224	91.76	313.3	0.29
1981	21.301	27.06	92.4	0.28
1982	47.085	59.82	204.2	0.16
1986	57.741	73.36	250.4	0.70

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Year	t_p hours	t_g hours	t_f=D hours	Q_f m³/s
1987	62.477	79.38	272.0	0.18
1988	84.779	107.71	368.0	0.59
1989	80.327	102.06	281.1	1.10
1990	98.647	125.33	428.2	0.52
1991	56.441	71.71	245.2	1.60
1992	30.767	39.09	133.4	1.17
1993	68.150	86.58	296.2	0.29
1994	52.426	66.61	227.9	0.21
1995	64.989	82.57	281.9	0.37
1996	82.743	105.12	360.6	0.14
1997	43.979	55.88	189.5	0.08
1998	52.636	66.87	228.6	0.26
1999	43.673	55.49	190.2	0.19
2000	82.888	105.30	360.4	0.05
2001	50.324	63.94	218.7	0.31
2002	100.249	127.37	435.1	0.32
2003	109.443	139.05	474.8	0.62
2004	97.142	123.42	421.5	0.82
MAX	109.443	139.05	474.8	2.40

Verification for randomness of records

The Wald-Wolfowitz Test is a non–parametric test that 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

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

annual flows (X_i). Therefore, it was proposed to apply this test to the joint records of Q, V and D, which must be *random* samples.

Fitting errors

The first criterion applied for seelecting 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 makes it possible to adopt an adequate PDF among the various models tested.

Changing in equations (40) and (41), the probabilities observed by the ordered data of the analyzed series (x_i, y_i or z_i) and the probabilities calculated by the estimated values with the inverse solution of the PDF that is tested or contrasted, the standard error of fit (*EEA*) and mean absolute error (*EAM*). The estimated values ($\hat{x}_i$, $\hat{y}_i$ y $\hat{z}_i$) are sought for the same probability of non–exceedance, assigned to the data with the empirical Gringorten formula (Equation (38)).

Zhang and Singh (2007) also apply two other quality-of-fit statistics, the first, when PDFs are fitted using the maximum likelihood method, they use the Akaike Information Criterion (AIC); the second was the Kolmogorov–Smirnov test, shown in Equation (43).

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Results and their discussion

Adopted marginal distributions

Verification for randomness of records

The joint records of maximum flow and annual volume from Table 1 and the final time ($t_f = D$, in hours) from columns 4 and 9 of Table 2, resulted in random series, whose statistic (U) from the Wald Test– Wolfowitz resulted in 0.284, 0.213 and 1.139, respectively.

PDF of maximum flows and anual volumes

Campos-Aranda (2024) processed the maximum flow and volume records in Table 1, with the three most convenient PDFs according to the L–moment Ratio Diagram of Hosking and Wallis (1997) and the Kappa and Wakeby distributions. All the applied PDFs were fitted with the L–moments method, except the Log–Pearson type III (LP3) which was fitted by moments in the logarithmic and real domains.

Campos-Aranda (2024) found for the maximum annual flows, that the two best distributions, the Log–Normal and the General Extreme Values (GVE), report the lowest and similar fit errors; but its predictions are high in the four extreme return periods (Tr >500 years). On the other hand, the fit errors of the Kappa distribution are almost identical to those of the Log–Normal, but with intermediate predictions and for this reason, it was adopted.

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

The location (u_1), scale (a_1) and shape (k_1 and h_1) parameters of the Kappa distribution selected for the maximum flows are: 258.7462, 228.381, −0.2685394 and 0.2888472, whose expression is:

$$F(x) = \left\{1 - h_1\left[1 - \frac{k_1(x-u_1)}{a_1}\right]^{1/k_1}\right\}^{1/h_1} \tag{58}$$

For the annual volumes in Table 1, Campos-Aranda (2024) found that the best distribution, LP3, leads to the lowest adjustment errors; but its predictions are much lower than those of the GVE. Instead, the Kappa distribution leads to low fit errors and its predictions are intermediate, therefore it was adopted.

The location (u_2), scale (a_2) and shape (k_2 and h_2) parameters of the Kappa distribution adopted for the annual volumes are: 60.39858, 78.31082, −0.3155518 and 0.4287021, which is expressed as:

$$F(y) = \left\{1 - h_2\left[1 - \frac{k_2(y-u_2)}{a_2}\right]^{1/k_2}\right\}^{1/h_2} \tag{59}$$

PDF of total durations

Applying the Weighted Absolute Distance criterion, exposed by Campos-Aranda (2024), to the record of total Durations (D) of the annual floods in Table 2, it was obtained that the three best PDFs were the Generalized Pareto (PAG), the LP3 and the GVE. In addition, Table 3 shows the results of the Log–Normal (LN3), Kappa and Wakeby models.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Table 3. Fit errors and predictions (m^3/s) of the three best PDFs and three of general application in the record of durations of the annual floods (*D*) of the hydrometric station La Cuña, Mexico.

PDF	***EEA*** **(m^3/s)**	***EAM*** **(m^3/s)**	**Return periods, in years**					
			50	**100**	**500**	**1 000**	**5 000**	**10 000**
PAG	10.7	8.7	441	448	456	457	459	459
LP3	13.1	9.4	483	517	587	614	671	694
GVE	12.4	9.8	477	507	563	584	622	636
LN3	12.9	10.2	479	513	587	616	682	709
Kappa	9.4	7.8	453	466	482	486	491	492
Wakeby	10.6	8.3	448	458	468	470	473	473

The PDF of Table 3 define for non–exceedance probability (p) of 1 %, the following values: 123.1, 93.2, 76.4, 74.4, 110.5 and 126.9 hours, with which, the lowest value of the record of total durations D=92.4 hours from Table 2, would lead to a negative probability value, except with the GVE and LN3 models. The GVE distribution is adopted because it has smaller fitting errors than the LN3 function.

The location (u_3), scale (a_3) and shape (k_3) parameters of the adopted GVE distribution are: 227.791, 86.95609, 0.1676378, with the following equation:

$$F(z) = \exp\left\{-\left[1 - k_3 \frac{(z-u_3)}{a_3}\right]^{1/k_3}\right\} \quad (60)$$

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

The positive value of k_3 indicates a Weibull distribution with an upper limit. Table 4 shows the predictions obtained with the PDFs adopted, for the *Q*, *V* and *D* records of Table 1 and Table 2.

Table 4. Predictions obtained with the PDF adopted in the records of *Q*, *V* and *D* of the annual floods recorded at the hydrometric station La Cuña, Mexico.

Record	PDF	Low return periods, in years					High return periods, in years				
		2	5	10	25	50	100	500	1 000	5 000	10 000
Peak flow (m^3/s)	Kappa	372	692	971	1419	1835	2335	3920	4844	7784	9489
Volume (Mm^3)	Kappa	104	217	321	495	663	873	1576	2007	3459	4350
Duration (hours)	GVE	259	343	391	443	477	507	563	584	622	636

Estimation of empirical probabilities

These probabilities are estimated based on the Gringorten formula (Equation (38)) applied in three-dimensional space, according to the process described. The results for Equation (39) are set forth in Table 5.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Table 5. Estimated duration (D) with the Gamma type hydrograph for the annual floods of the hydrometric station La Cuña, Mexico, its number of minor Q, V and D combinations in three–dimensional space and the observed empirical probability (Equation (39)).

No.	D (hours)	$NQVD_i$ (W_i^o)
1	169.0	8(0.1372)
2	190.2	12(0.2097)
3	197.4	12(0.2097)
4	184.5	8(0.1372)
5	245.0	17(0.3004)
6	196.0	9(0.1553)
7	374.0	28(0.5000)
8	161.3	1(0.0102)
9	266.2	24(0.4274)
10	362.5	17(0.3004)
11	207.7	7(0.1190)
12	247.1	21(0.3730)
13	127.2	3(0.0464)
14	331.8	12(0.2097)
15	337.6	25(0.4456)
16	146.2	3(0.0464)
17	157.8	6(0.1009)
18	286.4	13(0.2279)
19	359.0	37(0.6633)
20	395.6	34(0.6089)

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

No.	***D*** **(hours)**	***NQVDᵢ*** **(W_i^o)**
21	368.1	47(0.8447)
22	331.3	21(0.3730)
23	181.9	2(0.0283)
24	222.1	19(0.3367)
25	260.4	22(0.3911)
26	124.6	1(0.0102)
27	341.1	41(0.7358)
28	252.8	9(0.1553)
29	361.9	34(0.6089)
30	347.2	40(0.7177)
31	229.7	17(0.3004)
32	213.7	9(0.1553)
33	154.9	2(0.0283)
34	313.3	17(0.3004)
35	92.4	1(0.0102)
36	204.2	3(0.0464)
37	250.4	22(0.3911)
38	272.0	5(0.0827)
39	368.0	35(0.6270)
40	281.1	2(0.0283)
41	428.2	36(0.6451)
42	245.2	26(0.4637)
43	133.4	4(0.0646)
44	296.2	17(0.3004)

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

No.	***D*** **(hours)**	***NQVD$_i$*** **(W_i^o)**
45	227.9	8(0.1372)
46	281.9	20(0.3549)
47	360.6	5(0.0827)
48	189.5	1(0.0102)
49	228.6	9(0.1553)
50	190.2	5(0.0827)
51	360.4	1(0.0102)
52	218.7	11(0.1916)
53	435.1	25(0.4456)
54	474.8	40(0.7177)
55	421.5	47(0.8447)
$\overline{X}$	265.6	–
MED	250.4	–
D.E.	91.6	–
Cs	0.272	–
Ck	2.359	–

Selection and ratification of bivariate *CFs*

To apply Equation (45) to the trivariate AND type return period, the bivariate *CF*s are required: $C(u,v)$, $C(u,w)$ and $C(v,w)$. The upper part of Table 6 lists the association and concordance measures of the three bivariate variables and then the fit statistics of the Clayton, Frank, Gumbel-Hougaard and Joe *CF*s, defined in equations (2) to (12).

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Table 6. Fit errors of the four *CF* applied to the three indicated *bivariate* variables of the annual floods of the station hydrometric La Cuña, Mexico.

Association measuresand fitting errors	**Bivariate Variables**		
	Q–V	***Q–D***	***V–D***
Correlation coefficient (r_{xy})	0.9302	0.1840	0.4648
Kendall Tau quotient (τ_n)	0.7199	0.1367	0.4168
Spearman rho coefficient (ρ_n)	0.9088	0.2116	0.5742
Observed dependency (λ_U^{CFG})	0.7819	0.1835	0.4530
Clayton copula			
Association parameter(θ)	5.1394	0.3167	1.4296
Standard Mean Error (*EME*)	0.0245	0.0276	0.0273
Absolute Mean Error (*EAM*)	0.0190	0.0213	0.0215
Maximum Absolute Error (*EMA*)	0.0714	0.0821	0.0627
Upper dependency (λ_U)	0.0000	0.0000	0.0000
Frank copula			
Association parameter(θ)	12.3850	1.2490	4.3970
Standard Mean Error (*EME*)	*0.0227*	*0.0272*	*0.0258*
Absolute Mean Error (*EAM*)	*0.0169*	*0.0210*	*0.0203*
Maximum Absolute Error (*EMA*)	*0.0582*	0.0772	*0.0593*
Upper dependency (λ_U)	0.0000	0.0000	0.0000
Gumbel-Hougaard copula			
Association parameter(θ)	3.5697	1.1583	1.7148
Standard Mean Error (*EME*)	0.0255	0.0289	0.0297
Absolute Mean Error (*EAM*)	0.0192	0.0224	0.0223
Maximum Absolute Error (*EMA*)	0.0676	0.0741	0.0792
Upper dependency (λ_U)	0.7857	0.1808	0.5019
Joe Copula			
Association parameter(θ)	5.8866	1.2786	2.3104
Standard Mean Error (*EME*)	0.0289	0.0302	0.0340
Absolute Mean Error (*EAM*)	0.0225	0.0233	0.0264
Maximum Absolute Error (*EMA*)	0.0788	*0.0733*	0.0920
Upper dependency (λ_U)	0.8750	0.2804	0.6501

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Table 6 shows that Frank's *CF* leads to the best fit of the bivariate data and Joe's *CF* leads to the poorest fit. When taking into account the observed dependence (λ_U^{CFG}) of the bivariate variables, there is no doubt in selecting the Gumbel-Hougaard *CF*, which practically reproduces two of them and is slightly exceeded in the *V–D* variables.

Selection and ratification of the trivariate *CF*

Symmetric Copula Functions

The *joint* annual data of maximum flow (*Q*), runoff volume (*V*) and total duration (*D*) of the hydrograph of each flood, taken from Table 1 and Table 5, were adjusted by Clayton's trivariate *CF* (*d*=3), Frank, Gumbel-Hougaard and Joe, defined by equations (20), (24), (28) and (32). This fit was made by testing the value of its association parameter (θ), looking for the smallest fitting errors according to equations (40) to (42). The results obtained are shown in Table 7.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Table 7. Statistical indicators of the indicated *symmetric* trivariate fit of the copula functions, in the annual floods of the hydrometric station La Cuña, Mexico.

Copula	*θ*	*EME*	*EAM*	No. DP	No. DN	MDP	MDN
Clayton	1.959	0.0346	0.0275	18	37	0.1029	−0.0631
Frank	5.775	*0.0300*	*0.0222*	29	26	0.0832	−0.0659
G–H	2.100	0.0351	0.0246	31	24	0.0830	−0.0917
Joe	3.250	0.0419	0.0296	33	22	0.0692	−0.1196

DP, DN: positive and negative differences.

MDP, MDN: máximum positive and negative difference.

On the other hand, Equation (43) defines D_n=0.1831 and since the absolute maximum difference of the symmetric Frank *CF* in Table 7 is 0.0832, the Kolmogorov–Smirnov test confirms its adoption. The correlation coefficient (r_{xy}) between the empirical probabilities (Table 5) and the theoretical ones, estimated with the symmetric Frank *CF*, was 0.9927; therefore, it has a very good fit.

Asymmetric Copula Functions

The application of the trivariate asymmetric *CF*s, with two association parameters (θ_1,θ_2), to the data in Table 1 and Table 5, was carried out based on the Complex algorithm of multiple bounded variables. The initial values in Clayton's *CF* were 1.5 and 2.0; in Frank's 3 and 7; in Gumbel-Hougaard's 1.5 and 3.5 and in Joe's 1.5 and 4.5. The optimal values found for θ_1 and θ_2 and their fit indicators have been concentrated in Table 8.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Table 8. Statistical indicators of the indicated *asymmetric* trivariate fit of the copula functions, in the annual floods of the hydrometric station La Cuña, Mexico.

Copula	No. EFO	θ_1	θ_2	*EME*	*EAM*	No. DP	No. DN	MDP	MDN
Clayton	37	0.9698	4.8490	0.0281	0.0208	24	31	0.0781	−0.0585
Frank	80	3.2736	11.1166	*0.0255*	*0.0194*	32	23	0.0657	−0.0567
G–H	39	1.3805	6.9013	0.0304	0.0226	32	23	0.0582	−0.0799
Joe	33	1.7483	8.7409	0.0351	0.0260	35	20	0.0635	−0.0953

EFO: Objective function evaluations.

θ_1, θ_2: Association parameters of the asymmetric *CF*.

Again, Frank's *CF* defines the best fit. As already indicated, Equation (43) defines D_n=0.1831 and since the maximum absolute difference of the asymmetric Frank *CF* in Table 8 is 0.0657, the Kolmogorov–Smirnov test confirms its adoption. The correlation coefficient (r_{xy}) between the empirical probabilities (Table 5) and the theoretical ones, estimated with the asymmetric Frank *CF*, was 0.9949; therefore, it has an excellent fit.

Adoption of a trivariate *CF*

The result of Table 6, when Gumbel-Hougaard *CF* is adopted, due to the reproduction it makes of the value of λ_U^{CFG} for the three bivariates *Q*–*V*, *Q*–*D* y *V*–*D*; guides its selection for the trivariate case.

Such a selection is not considered inappropriate, since, as seen in Table 7 and Table 8, such a Gumbel-Hougaard *CF* shows quite similar fits

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

to those of the symmetric and asymmetric Frank *CF*. This was verified based on the correlation coefficients between the empirical probabilities (Table 5) and the trivariate theoretical probabilities of the symmetric and asymmetric Gumbel-Hougaard *CF*, whose values were: 0.9916 and 0.9939.

Table 9 and Table 10 show a part of the observed empirical trivariate non–exceedance probabilities (w_i^o), taken from Table 5 and the calculated theoretical ones (w_i^c) with the symmetric and asymmetric Gumbel-Hougaard *CF*. The maximum positive and negative differences are also indicated shaded.

Table 9. Partial data of the trivariate non–exceedance probabilities and their differences, calculated with the *symmetric* Gumbel-Hougaard *CF*, for the annual floods of the La Cuña hydrometric station, Mexico.

No.	w_i^o	w_i^c	**Differences**
1	0.1372	0.1435	−0.0063
7	0.5000	0.4170	0.0830
10	0.3004	0.2701	0.0303
15	0.4456	0.3955	0.0501
19	0.6633	0.7550	−0.0917
25	0.3911	0.4446	−0.0535
30	0.7177	0.8057	−0.0880
35	0.0102	0.0087	0.0015
40	0.0283	0.0281	0.0002
45	0.1372	0.1021	0.0351
50	0.0827	0.0570	0.0257
55	0.8447	0.8380	0.0067

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Table 10. Partial data of the trivariate non–exceedance probabilities and their differences, calculated with the *asymmetric* Gumbel-Hougaard *CF*, for the annual floods of the La Cuña hydrometric station, Mexico.

No.	w_i^o	w_i^c	Differences
1	0.1372	0.1264	0.0108
7	0.5000	0.4418	0.0582
10	0.3004	0.3018	−0.0014
15	0.4456	0.4154	0.0302
20	0.6089	0.5961	0.0128
25	0.3911	0.4142	−0.0231
29	0.6089	0.6888	−0.0799
35	0.0102	0.0044	0.0058
40	0.0283	0.0372	−0.0089
45	0.1372	0.1089	0.0283
50	0.0827	0.0549	0.0278
55	0.8447	0.8323	0.0124

As already indicated, Equation (43) defines D_n=0.1831 and since the maximum absolute differences in Table 9 and Table 10 are 0.0917 and 0.0799, the Kolmogorov–Smirnov test confirms the symmetric and asymmetric Gumbel-Hougaard *CF* adopted. The graphic contrast between both probabilities, to ratify its adoption, is shown in Figure 2 and Figure 3 for the complete data in Table 9 and Table 10.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

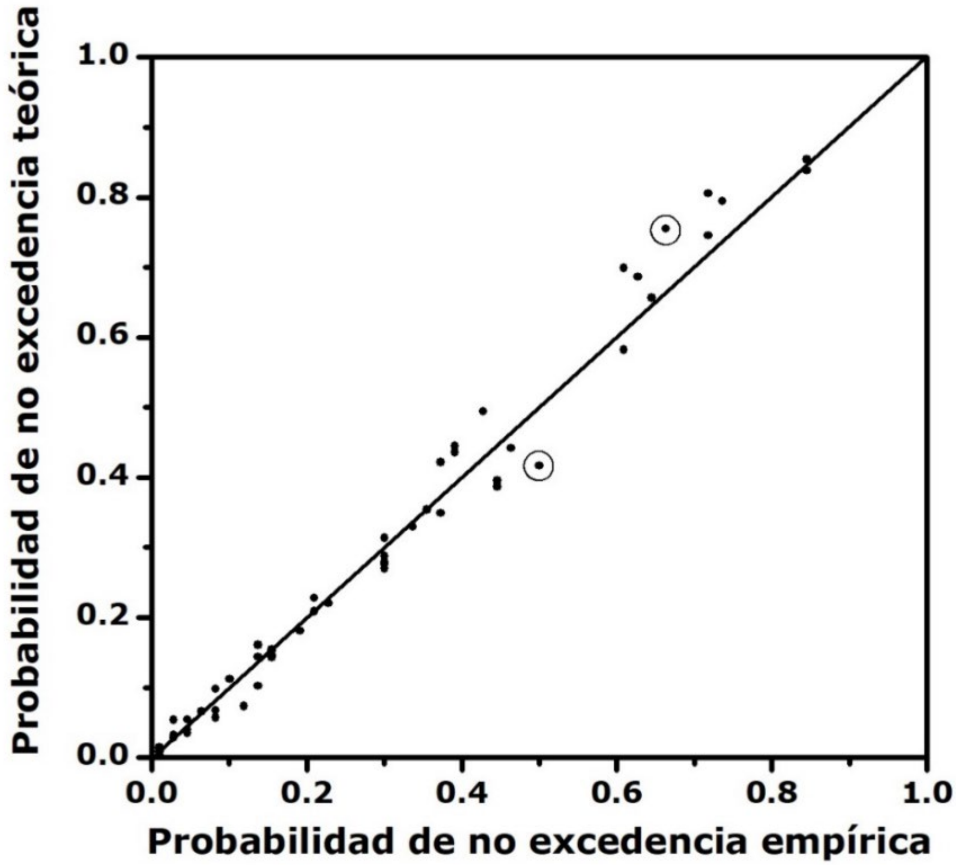

Figure 2. Graphical contrast of empirical and theoretical probabilities estimated with the *symmetric* trivariate Gumbel-Hougaard FC for the annual floods registered at the La Cuña station, Mexico.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

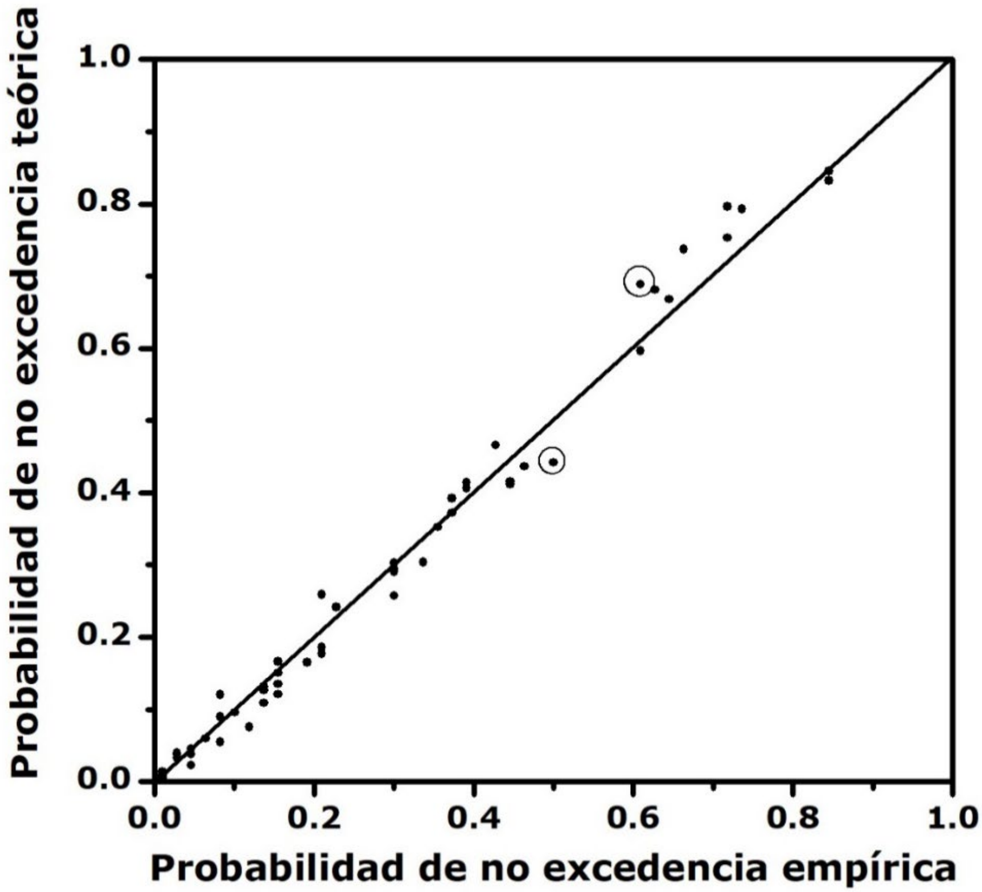

Figure 3. Graphic contrast of empirical and theoretical probabilities estimated with the *asymmetric* trivariate Gumbel-Hougaard *CF* for the annual floods registered at the La Cuña station, Mexico.

Estimation of trivariate return periods

Based on the list of predictions shown in Table 4, equations (44) and (45) were applied to estimate the OR joint return periods and AND type, shown in Table 11. In such expressions, the Gumbel-Hougaard bivariate *CF*s are applied, defined by Table 6 and the symmetric trivariate whose

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

association parameter (θ) is indicated in Table 7 and for *CF* asymmetric trivariate its parameters θ_1 and θ_2, were cited in Table 8. To estimate the return period secondary (T_{KEN}) Equation (50) was applied for Gumbel-Hougaard *CF*.

Table 11. Joint return periods of the OR, AND and secondary type estimated with the symmetric and asymmetric Gumbel-Hougaard trivariate *CF*s, for the annual floods of the hydrometric station La Cuña, Mexico.

***Tr* (years)**	**Type of Tr with symmetrical *CF***			**Asymmetric *CF***	
	OR	**Secondary**	**AND**	**OR**	**AND**
2	1.5	3.8	3.9	1.4	3.7
5	3.2	11.3	17.1	3.1	15.0
10	6.2	23.9	44.8	6.0	37.2
25	15.0	61.5	137.0	14.6	107.2
50	29.9	124.1	294.5	29.0	226.9
100	59.7	249.4	613.7	58.0	469.3
500	295.1	1251.8	3216.5	285.5	2358.0
1000	596.1	2504.1	6335.8	579.2	4836.3
5000	2957.4	12529.7	32202.0	2865.9	23899.2
10000	5938.8	25040.6	64035.2	5771.3	48771.0

The similarity between the *trivariate* or *joint* return periods OR-type in Table 11 is notable, with those estimated with the asymmetric *CF* being slightly lower; therefore, the most accurate. For those of the AND type,

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

there are greater differences and logically, the most reliable are the estimates with the asymmetric *CF*.

The similarity shown by the OR type trivariate return periods allows us to consider that the *symmetric* trivariate Gumbel-Hougaard *CF* offers an acceptable and reliable description of the trivariate joint data and therefore, the *secondary* return period can be used to obtain the joint design events, whose variables *Q*, *V* and *D* will be obtained with the marginal functions or distributions.

Estimation of design events

Assuming that in the downstream vicinity of the La Cuña hydrometric station and on the Verde River, dikes will be built to protect floodplains for agricultural and industrial purposes and a bridge to cross it, then it is necessary to estimate *design events* with periods joint or trivariate returns of 50, 100, 500 and 1000 years. In addition, a review of the hydrological safety of the projected reservoir will be carried out, at the site of the gauging station, with joint return periods of 5,000 and 10,000 years. Due to the above, it is necessary to estimate shortlist of *Q*, *V* and *D* with the six mentioned T_{KEN} joint return periods.

Based on Equation (50) of the trivariate Kendall distribution, established for the Gumbel-Hougaard *CF*, the univariate return period (*Tr*) and its respective probability of non–exceedance (*s*) were sought by trial and error, which define a return period *secondary* equal to the joint or design trivariate. Once such a value of the unitary variable (*s*) is found, the respective variables *Q*, *V* and *D* are obtained with the inverse solutions

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

of the marginal distributions (equations (58) to (60)). Table 12 indicates the results.

Table 12. Design events obtained for the indicated secondary return period, in the annual floods of the hydrometric station La Cuña, Mexico.

***Design Tr* secondary**	**Univariate *Tr***		**Design events**		
	(years)	**Prob. *s***	***Q* (m³/s)**	***V* (Mm³)**	***D* (hours)**
50	20	0.950	1300	448	431
100	40	0.9750	1693	605	466
500	200	0.9950	2935	1132	533
1000	400	0.9975	3657	1456	556
5000	1996	0.9994989	5952	2542	601
10000	3989	0.9997493	7288	3208	617

Estimation of conditional return periods

According to Grimaldi and Serinaldi (2006b) and Ma *et al.* (2013) an interesting *conditional probability* (*PC*) is defined by Equation (52) and for the case studied, its approach will be the following:

$$P(V \leq v, D \leq d | Q \geq q) = PC_{1,2} = \frac{C(v,w) - C(u,v,w)}{1-u} \tag{61}$$

This expression allows finding the volume (*V*) values and the duration (*D*) by trial and error, given that a maximum flow (*Q*) of a certain

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

return period has occurred, which must be equaled by the following *conditional return period*, $PC_{1,2}$ function:

$$Tr_{1,2} = 1/(1 - PC_{1,2}) \tag{62}$$

In Equation (61), *u*, *v*, and *w* are calculated for the *Q*, *V*, and *D* values with the marginal PDFs, according to equations (58) to (60). In Equation (61), $C(v,w)$ the bivariate Gumbel-Hougaard *CF* defined in Table 6 with θ=1.7148 and $C(u,v,w)$ are the symmetric (θ=2.1) and asymmetric (θ_1=1.3805 and θ_2=6.9013) trivariate Gumbel-Hougaard *CF*; which define the $PC_{1,2}$. The results for the four indicated return periods are shown in Table 13.

Table 13. Values of pairs of *V* and *D* conditioned to the occurrence of *Q*, whith the indicated return period, for the annual floods of the hydrometric station La Cuña, Mexico

Random variable:	**Return periods of the *Q*, in years.**			
	10	**25**	**50**	**100**
Q (m^3/s) assigned	971	1419	1835	2335
V (Mm^3) sought	1500	2500	3500	4500
D_1 (horas) sought	507	600	651	654
D_2 (horas) sought	494	594	649	651

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

Tecnología y Ciencias del Agua

In Table 13, the high values of duration (D_1) were estimated with the symmetric trivariate Gumbel-Hougaard *CF* and the low values (D_2) with the asymmetric *CF*.

Conclusions

The frequency analysis of *trivariate* floods, maximum flow variables(Q), runoff volume (V) and total duration (D), will allow a more accurate estimation of the Design Flood Hydrograph, associated with a joint return period.

In this study, to process the Q and V *joint annual records* available, D was assigned as the third estimated flood variable, based on the Gamma type hydrograph, for a final flow of 0.1 % of Q.

The use of the Copula Functions (*CF*) in the multivariate frequency analysis allows the construction of the multivariate distribution based on the marginal functions. Due to the above, the ideal probability distributions of Q, V and D are defined with the maximum possible precision and can be different and of any type.

The estimation of the trivariate return period of the AND type requires bivariate distributions; in the case studied of the joint variables Q–V, Q–D y V–D. Therefore, it is first sought that *CF*s reproduce the observed dependency (λ_U^{CFG}) and show a good fit with the aforementioned joint variables.

In the numerical application described, with 55 annual floods registered at the La Cuña gauging station, in Hydrological Region No. 12–

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

3 (Santiago River), Mexico; The following four CF families were used: Clayton, Frank, Gumbel-Hougaard, and Joe.

For the annual data triads of *Q*, *V* and *D*, symmetric multivariate Archimedean *CF*s were applied, with one association parameter (θ) and asymmetric trivariates, with two association parameters (θ_1, θ_2), of the aforementioned families. Finally, joint return periods of the OR, AND, and Kendall types were estimated. The latter allow obtaining the design events of *Q*, *V* and *D*, shown in Table 12.

The trivariate flood frequency analysis described are very simple and do not present computational complications, when they are performed based on the *CF*s.

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., 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.

Aldama-Rodríguez, A. A., & Ramírez–Orozco, A. I. (1998). Parametrización de hidrogramas mediante interpolantes hermitianos. *Ingeniería Hidráulica en México*, 13(3), 19-28.

Barbe, P., Genest, C., Ghoudi, K., & Rémillard, B. (1996). On Kendall's Process. *Journal of Multivariate Analysis*, 58(2), 197-229.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

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

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

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

Campos-Aranda, D. F. (2024). Análisis de frecuencias bivariado de Crecientes Anuales mediante enfoque práctico de las funciones Cópula. *Tecnología y ciencias del agua*, 15(2), 1-56. DOI: 10.24850/j-tyca-15-02-01

Campos-Aranda, D. F. (2003). Capítulo 7. Integración numérica y Capítulo 9. Optimización numérica. En: *Introducción a los métodos numéricos: software en Basic y aplicaciones en hidrología superficial* (pp. 137-153, 172-211). San Luis Potosí, México: Editorial Universitaria Potosina.

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, and Chapter 3. Copula–based Flood Frequency Analysis. *Copulas and its application in Hydrology and Water Resources* (pp. 13-38, 39-71). Gateway East, Singapore: Springer.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

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. (1972). Chapter 6. Gamma Function and related functions. In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of mathematical functions* (pp. 253-296). New York, USA: iDover Publications.

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). New York, USA: Dover Publications.

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)

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

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

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., & 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. DOI: 10.1061/(ASCE)0733–9429(1998)124:2(146)

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

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

Grimaldi, S., & Serinaldi, F. (2006a). Design hyetograph analysis with 3–copula function. *Hydrological Sciences Journal*, 51(2), 223-238. DOI: 10.1623/hysj.51.2.223

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Grimaldi, S., & Serinaldi, F. (2006b). Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources*, 29(8), 1155-1167. DOI: 10.1016/j.advwatres.2005.09.005

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. *Frequency and risk analyses in hydrology* (pp. 156-168). Fort Collins, USA: Water Resources Publications.

Ma, M., Song, S., Ren, L., Jiang, S., & Song, J. (2013). Multivariate drought characteristics using trivariate Gaussian and Student t copulas. *Hydrological Processes*, 27(8), 1175-1190. DOI: 10.1002/hyp.8432

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

Nelsen, R. B. (2006). Chapter 2. Definitions and Basic Properties. In: *An introduction to Copulas* (pp. 7-49), 2nd ed. New York, USA: Springer Series in Statistics.

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

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

Ponce, V. M. (1989). Section 2.4. Runoff. In: *Engineering hydrology. principles and practices* (pp. 62-84). Englewood Cliffs, USA: Prentice Hall.

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)

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

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

Salvadori, G., De Michele, C., & Durante, F. (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences*, 15(11), 3293-3305. DOI: 10.5194/hess–15–3293–2011

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)

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

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

Serinaldi, F. (2015). Dismissing return periods! *Stochastic Environmental Research and Risk Assessment*, 29(4), 1179-1189. DOI: 10.1007/s00477–014–0916–1

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

Snider, D. (1972). Chapter 16. Hydrographs. In: *National Engineering Handbook, Section 4: Hydrology* (pp. 16.1-16.26). Washington, DC, USA: Soil Conservation Service, U. S. Department of Agriculture.

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

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

Xu, C., Yin, J., Guo, S., Liu, Z., & Hong, X. (2016). Deriving design flood hydrograph based on conditional distribution: A case study of Danjiangkou reservoir in Hanjiang basin. *Mathematical Problems in Engineering*, 2016 (4319646), 1-16. DOI: 10.1155/2016/4319646

2024, Instituto Mexicano de Tecnología del Agua.
Open Access bajo la licencia CC BY-NC-SA 4.0
(https://creativecommons.org/licenses/by-nc-sa/4.0/)

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. (2000). 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., & 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

Zhang, L., & Singh, V. P. (2007). Trivariate flood frequency analysis using the Gumbel-Hougaard Copula. *Journal of Hydrologic Engineering*, 12(4), 431-439. DOI: 10.1061/(ASCE)1084–0699(2007)12:4(431)

Zhang, L., & Singh, V. P. (2019). Chapter 3. Copulas and their properties and Chapter 4. Symmetric Archimedean copulas. In: *Copulas and their applications in water resources engineering* (pp. 62-122, 123-171). Cambridge, United Kingdom. Cambridge University Press.

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)

2024, Instituto Mexicano de Tecnología del Agua. Open Access bajo la licencia CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)