DOI: 10.24850/j-tyca-2024-04-03

Articles

Bivariate Flood Frequency Analysis of occurrence dates and maximum flow through Copula functions

Análisis de Frecuencias de Crecientes bivariado de fechas de ocurrencia y gasto máximo 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

In the center and south of the Mexican Republic, each year the hurricanes of the Caribbean Sea and the Pacific Ocean cause *floods* that lead to a *wet season* and that generally increase in magnitude and danger as the cyclone season progresses. Both conditions allow *bivariate frequency analysis* of their dates of occurrence and their maximum flows (*Qm*). In this study, the bivariate distribution was formed based on the Gumbel-Hougaard *Copula function*, which satisfies the observed dependency condition (λ_U^{CFG}) and which combines the von Mises distributions as marginal distributions for the dates of occurrence in the year and for the *Qm* a suitable probabilistic function. The exposed theory is applied to the annual floods recorded at the *Guamúchil* gauging station of Hydrological Region No. 10 (Sinaloa), Mexico, in the period from 1940 to 1971. The von Mises distribution is fitted via numerical optimization with the de Rosenbrock algorithm and the ideal distribution of the *Qm* turned out to be the Kappa. The graph of *joint return periods* of the AND type of 50, 100 and 500 years was formed. In addition, *conditional* joint return periods of occurrence dates were estimated given that the *Qm* has the cited return periods. This allows estimates of the probability of exceedance of *Qm* in defined periods. The conclusions highlight the simplicity of these bivariate frequency analyses, by means of the Copula functions, and the practical importance of their *predictions*, according to the dates of occurrence.

Keywords: Dates of occurrence, von Mises distribution, Copula functions, Kendall's tau ratio, joint empirical probabilities, dependency on the extreme right, joint and conditional return periods.

Resumen

En el centro y sur de la república mexicana cada año los huracanes del mar Caribe y del océano Pacífico originan *crecientes* que definen una *estación húmeda*, y que en general aumentan en magnitud y peligrosidad conforme transcurre la temporada de ciclones. Ambas condiciones permiten el *análisis de frecuencias bivariado* de sus fechas de ocurrencia y sus gastos máximos (*Qm*). En este estudio, la distribución conjunta se formó con base en la *función Cópula* de Gumbel-Hougaard, que satisface la condición de dependencia (λ_U^{CFG}) observada y que combina como distribuciones marginales la de von Mises para las fechas de ocurrencia en el año y para los *Qm* una función probabilística idónea. La teoría expuesta se aplica a las crecientes anuales registradas en la estación de aforos *Guamúchil* de la Región Hidrológica No. 10 (Sinaloa), México, en el periodo de 1940 a 1971. La distribución de von Mises se ajusta vía optimización numérica con el algoritmo de Rosenbrock y la distribución idónea de los *Qm* fue la Kappa. Se formó la gráfica de *periodos de retorno conjuntos* de tipo AND de 50, 100 y 500 años. Además, se estimaron periodos de retorno conjuntos *condicionales* de fechas de ocurrencia, dado que el *Qm* tiene los periodos de retorno citados. Lo anterior permite estimaciones de la probabilidad de excedencia del *Qm* en lapsos definidos. Las conclusiones destacan la simplicidad de estos análisis de frecuencias bivariados por medio de las funciones Cópula y la importancia práctica de sus *predicciones*, según las fechas de ocurrencia.

Palabras clave: fechas de ocurrencia, distribución de von Mises, funciones Cópula, cociente tau de Kendall, probabilidades empíricas conjuntas, dependencia en el extremo derecho, periodos de retorno conjuntos y condicionales.

Received: 03/08/2022

Accepted: 21/12/2022

Published Online: 05/01/2023

Tecnología y Ciencias del Agua

Introduction

Generalities

In general terms, the risk of flooding and its damages are a direct function of the water volume that flows through the channel which exceeds its limit or capacity, overflowing and covering the floodplains. However, the *occurrence date* of the event is as important as its magnitude and in some cases greater, when they occur out of season or in the *wet season*, because it takes the population off guard, causing greater damage (Khedun, Singh, & Byrd, 2019).

Due to the above, the estimation of the *probability of flood occurrence* throughout the year, is vital for the elaboration of plans that have no hydraulic works of damage mitigation, which include the preparation for the event, with the purpose of reducing the exposure and vulnerability of the population, as well as optimizing the economic resources available for the emergency and accelerating post-event recovery (Durrans, Eiffe, Thomas, & Goranflo, 2003; Khedun *et al.*, 2019).

In general, understanding the seasonal behavior of floods is vital in the planning and management of the river's hydraulic resources, both for agricultural and hydroelectric uses, as well as for navigation, recreational uses, and other activities associated with waterbodies. Therefore, knowing the relationship between the maximum flow and its date of occurrence becomes extremely important, to the point of requiring its *joint bivariate study* (Chen, Guo, Yan, Liu, & Fang, 2010).

Bivariate flood frequency analysis began formally at the beginning of this century (Yue & Rasmussen, 2002) and was generally based on maximum flow and of annual flood volume, using the bivariate Normal distribution and the Logistic model which accepts as probability distribution functions (FDP, by its acronym in Spanish) *equal marginals* to the extreme value distributions, the most common being the Gumbel and the GVE (Escalante-Sandoval & Reyes-Chávez, 2002; Aldama, Ramírez, Aparicio, Mejía-Zermeño, & Ortega-Gil, 2006).

Bivariate flood analysis is the *simplest* multivariate approach, and yet it involves five mathematical complications: (1) a bivariate FDP must be used; (2) its validation requires the estimation of bivariate empirical probabilities; (3) now there are joint and conditional probabilities; (4) a joint return period must be defined, for which there are *infinite* pairs of values of X and Y that can satisfy it and (5) the critical or design events must be selected among the aforementioned pairs (Ramírez-Orozco & Aldama, 2000; Escalante-Sandoval & Reyes-Chávez, 2002; Volpi & Fiori, 2012; Requena, Mediero, & Garrote, 2013).

Currently, with the use of the mathematical tool known as "Copulas", the bivariate FDPs can be constructed with marginals of different types, given the fact that *copula functions* allow multivariate distributions to be represented from univariate or *marginal* FDPs, regardless of their shape or form type (Salvadori, De Michele, Kottegoda, & Rosso, 2007; Meylan, Favre, & Musy, 2012; Genest & Chebana, 2017; Zhang & Singh, 2019; Chowdhary & Singh, 2019).

By means of the Copula functions, the *bivariate* frequency analysis of the annual flood occurrence dates and their maximum flows is addressed. The former are represented by the von Mises distribution and the latter by a suitable PDF.

Objectives

The *objectives* of this study can be encompassed by the following six: (1) the directional statistics that represent the flood occurrence dates during the year are exposed; (2) the *von Mises distribution* (dvM, by its acronym in Spanish) and its adjustment via numerical optimization are described, which allows the probabilistic characterization of the dates of occurrence; (3) the basic characteristics of the Frank and Gumbel-Hougaard *copula functions* (*FC*, by its acronym in Spanish) are presented, including: Kendall's tau ratio, observed and *FC* dependence, estimation of joint empirical probabilities, and selection and ratification of the *FC*; (4) the selection and adoption of the ideal marginal FDP of the maximum annual flows is exposed; (5) the joint return periods of the OR, AND and conditional type are described and (6) as a numerical example, the application of the theory and procedures exposed to the 32 annual floods registered in the *Guamuchil* hydrometric station of the Hydrological Region No. 10 (Sinaloa), Mexico, is detailed.

Operative theory

Circular data and directional indicators

The dates of occurrence of the annual floods in Mexico, generally define a *wet season* that covers from June to October, which is the *season* of hurricane incidence that are generated in the Caribbean Sea and in the Pacific Ocean. Therefore, the dates of of the annual flood occurrence can be represented within a circle, which covers the 365 days of the year.

There are several conventions or ways of working in the circle, to graph the data (Ramírez-Orozco, Gutiérrez-López, & Ruiz-Silva, 2009). From now on, the Burn convention (Burn, 1997) will be used, due to its similarity with Cartesian quadrants. In such a scheme, the advance is counterclockwise, starting on the abscissa axis; for this reason, January 1 and December 31 coincide in such beginning.

Having several circular data drawn, it is possible to obtain its directional indicators, the most important are two, its *mean direction* ($\bar{\alpha}$) and its *seasonality index* ($\bar{r}$). The first defines the central tendency of the data and, therefore, is the average occurrence date of the annual floods and the second quantifies the dispersion of such values (Campos-Aranda, 2017).

Average address and seasonality index

To estimate these indicators, one begins by transforming each date of occurrence of the annual floods to a Julian day (D_i), that is, from 0 to 365; This implies not taking leap years into account. If a flood occurs on February 29, it is assigned the 28th. January dates remain the same, but February 31 is added, March 59, April 90, and so on until February. November which is added 304 and those of December are added 334, to obtain the Julian day. Next, the angle α_i in radians corresponding to the date *i* of each flood (D_i) is obtained, with the following expression (Burn, 1997; Cunderlik, Ouarda, & Bobée, 2004; Chen, Singh, Guo, Fang, & Liu, 2013; Campos-Aranda, 2023b):

$$\alpha_i = 2\pi \frac{D_i}{365} = X_i \text{ with } 0 \leq \alpha_i \leq 2\pi \quad (1)$$

in which

π = number pi with 3.141592654 as an approximate value

X_i = *random variable of the dates of occurrence*, in radians

Next, the *x* and *y* coordinates of the flood occurrence dates described by the angles α_i are estimated based on the cosines and sines and their average values are obtained through the following equations:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} \cos(\alpha_i) \quad (2)$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} \text{sen}\,(\alpha_i) \tag{3}$$

where

n = number of dates of occurrence of the annual floods analyzed

Now, the *mean direction* ($\overline{\alpha}$) of the average date of the floods will be:

$$\overline{\alpha} = \arctan\left(\frac{\overline{y}}{\overline{x}}\right) \tag{4}$$

The application of the previous equation is done by first obtaining the arc tangent of $\overline{y}$ between $\overline{x}$, both with a positive sign designated α, in radians; then if $\overline{x}$ and $\overline{y}$ are positive $\overline{\alpha} = \alpha$, if $\overline{x}$ < 0 and $\overline{y}$ > 0 $\overline{\alpha} = \mu - \alpha$, if both are negative $\overline{\alpha} = \pi + \alpha$ and finally, if $\overline{x}$ > 0 and $\overline{y}$ < 0 $\overline{\alpha} = 2\pi - \alpha$. The angles α_i and $\overline{\alpha}$ are converted to degrees (0º to 360º) by multiplying them by 57.295755.

The value of $\overline{\alpha}$ in Julian days is called *mean flood day* (*DMC*, by its acronym in Spanish), it is obtained by dividing by 2π and multiplying by 365. The *DMC* index indicates the average occurrence date of the maximum annual flows in a given basin. Basins with similar *DMC* values can be expected to show similarities in other important hydrological characteristics. Logically, the *DMC* will be related to the size of the basin and its geographical location within the studied hydrological region (Burn, 1997; Cunderlik *et al.*, 2004).

A measure of the variability of the n dates of occurrence of the floods, in relation to the DMC, can be estimated by calculating the resulting average, which is expressed as:

$$\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2} \quad (5)$$

The *seasonality index* $\bar{r}$ is a dimensionless measure of the data dispersion, it takes values between zero and one. A unit value indicates that all the floods occur on the same date, while a value close to zero implies great variability of occurrences throughout the year.

Ramirez-Orozco *et al*. (2009) establish the following five degrees of seasonality: (1) very strong, when $\bar{r} > 0.90$, (2) strong, when $\bar{r}$ fluctuates between 0.70 and 0.90, (3) medium, when $\bar{r}$ varies from 0.50 to 0.70, (4) low, when $\bar{r}$ changes from 0.10 to 0.50 and (5) very low or weak, when $\bar{r} < 0.10$. Chen *et al*. (2013) indicate that if $\bar{r}$ is close to unity, a single season or wet season can be expected to be dominant.

The von Mises distribution

This probabilistic model is commonly used to represent random variables that have two-dimensional direction and a single mode. For this reason, the *von Mises distribution* (dvM, by its acronym in Spanish) is considered the natural analogy of the Normal model for seasonal data. Its probability density function is the following equation (Metcalfe, 1997; Carta, Bueno, & Ramirez, 2008; Chen *et al*., 2010):

$$f(x) = \frac{\exp[\kappa \cdot \cos(x-\mu)]}{2\pi I_0(\kappa)} \text{ with } 0 \leq x \leq 2\pi, 0 < \mu \leq 2\pi, \kappa > 0 \quad (6)$$

The dvM is symmetric with its mode at $x = \mu$, which is also its *mean direction* ($\overline{\alpha}$) and the dispersion is given by the *concentration parameter* κ (kappa). The denominator of Equation (6) makes the area under the curve unitary and for this reason it is called the *normalization factor* (*FN*, by its acronym in Spanish); includes the modified Bessel function of the first type of order zero [$I_0(\kappa)$].

To estimate the non-exceedance probability of a value *x*, Equation (6) is numerically integrated, that is:

$$F(x) = \frac{1}{2\pi I_0(\kappa)} \int_0^x \exp[\kappa \cdot cos(x-\mu)] \quad (7)$$

The previous expression defines the PDF of the dvM. $I_0(\kappa)$ is estimated with the following ascending series, which comes from Olver (1972):

$$I_0(\kappa) = 1 + \frac{(\kappa|2/4)}{1} + \frac{(\kappa^2/4)^2}{4} + \frac{(\kappa^2/4)^3}{36} + \frac{(\kappa^2/4)^4}{576} + \frac{(\kappa^2/4)^5}{14400} + \cdots \quad (8)$$

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

Numerical integration of the dvM

To carry out such numerical integration of Equation (7), the Gauss-Legendre quadrature method was adopted, whose univariate operational equation is (Nieves & Domínguez, 1998; Campos-Aranda, 2003):

$$\int_a^b f(x)\,dx \cong \frac{b-a}{2}\sum_{i=1}^{np} w_i \cdot f\left[\frac{(b-a)h_i+b+a}{2}\right] \tag{9}$$

in which

w_i = coefficients of the method whose abscissas are h_i

np = number of pairs in which the function $f(x)$ is evaluated, with the argument indicated in $f(\cdot)$ of Equation (9).

In Davis and Polonsky (1972) the 12 used pairs of w_i and h_i with 15 digits were obtained, because the Basic language accepts 16 digits as double precision variables.

Fit of the dvM in the wet season

When the annual flood occurrence dates cover, for the most part, a quite defined period in months, for example, from June to October, then the application of Equation (7) is carried out via numerical optimization, to find the values of μ and κ that reduce the sum of the differences between theoretical probabilities [$F_T(x)$] and empirical probabilities [$F_E(x)$] to the square, that is:

$$\text{Minimize FO} = \sum_{i=1}^{n}[F_T(x) - F_E(x)]^2 \tag{10}$$

The minimization of the previous objective function will be carried out by means of the Rosenbrock algorithm. The empirical probabilities are defined with the Gringorten formula (Chen *et al.*, 2010), which is:

$$F_E(x) = \frac{m-0.44}{n+0.12} \tag{11}$$

where

m = order number of the data or occurrence date in radians ($x = \alpha_i$), when they are located in progressive magnitude

n = total number of data

Logically, the occurrence dates outside the main period or *wet season* of the floods are previously eliminated, to improve the fit and, therefore, the definition of the dvM. More details of the previous process can be consulted in Campos-Aranda (2023b).

Rosenbrock algorithm

It is a numerical direct search procedure that attempts to define the *minimum* of a nonlinear function of multiple unbounded random variables. *Rosenbrock's algorithm* assumes that the function is unimodal and begins by defining a straight line, or search direction, from a given starting point. Next, it evaluates the *objective function* (FO) at various points on the line and determines the optimum; when this has happened, a new search direction is selected and the process is repeated recursively in stages.

In this algorithm it is convenient to provide different starting points to search for the global minimum, based on the estimated local minima. A more detailed description of the process can be found in Rosenbrock (1960), Kuester and Mize (1973), and Campos-Aranda (2003).

Advantages of Copula Functions

As already indicated, the essential advantage of *Copula Functions* (*FC*, by its acronym in Spanish) consists in allowing to express a *joint distribution* of correlated random variables, as a function of their marginal distributions, previously adopted. So, a *FC* links or relates the univariate marginal distributions to form a multivariate distribution. Another basic advantage of *FC*s 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.

Due to the above, the construction of the multivariate distribution is reduced to the study of the relationship between the correlated variables, if the univariate marginal distributions are known. The use of *FC*s offers complete freedom to adopt or select the univariate marginal distributions that best represent the data (Meylan *et al.*, 2012; Zhang & Singh, 2019; Chowdhary & Singh, 2019).

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Copula families to apply

The *Copula functions* (*FC*, by its acronym in Spanish) 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 *X* and *Y* is defined (Meylan *et al.*, 2012; Chowdhary & Singh, 2019). Salvadori *et al.* (2007) present a comprehensive and useful summary of *FC* that have been applied in the field of hydrology.

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 and extreme values (Salvadori *et al.*, 2007; Zhang & Singh, 2019; Chen & Guo, 2019; Chowdhary & Singh, 2019).

1. Frank. Its equation and variation space of θ are:

$$C(u,v) = \frac{-1}{\theta}\ln\left[1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right] \quad (-\infty,\infty) \setminus \{0\} \qquad (12)$$

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

$$\tau_n = 1 + \frac{4}{\theta}[\mathrm{D}_1(\theta) - 1] \qquad (13)$$

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

$$D_1(\theta) = \frac{1}{\theta}\int_0^{\theta} \frac{s}{e^s - 1} ds \qquad (14)$$

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

2. Gumbel-Hougaard, which accepts only positive dependence. Its equation and variation space of θ are:

$$C(u, v) = \exp\left\{-\left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{1/\theta}\right\} \; [1, \infty) \qquad (15)$$

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

$$\tau_n = \frac{\theta - 1}{\theta} \qquad (16)$$

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

Association measures

Concordance

Since the *FC* 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).

Some variables with direct linear correlation will be concordant, since as one increases the other does too. Variables with inverse linear correlation will be *discordant*, since large values of one will correspond to small values of the other. 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 variable association. For comparison purposes, such measures range from zero to +1 or to -1, indicating perfect association positive at +1 or negative at -1. Kendall's tau ratio and Spearman's rho coefficient are two non-parametric measures that provide information about a special form of association or dependency, known as *concordance* (Salvadori et al., 2007; Chen & Guo, 2019).

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Kendall's tau ratio

It measures the probability of having matching pairs, 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):

$$\tau_n = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sign}\left[\left(x_i - x_j\right)\left(y_i - y_j\right)\right] \tag{17}$$

In the above equation:

n = number of observations

sign[·] = +1 if such pairs are concordant and -1 if they are discordant

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

$$\sqrt{\frac{9n(n-1)}{2(2n+5)}}\left|\tau_n\right| > Z_{\alpha/2} = 1.96 \tag{18}$$

Dependency parameter estimation

The simplest method to estimate the parameter θ of the *FC*s is similar to the method of moments and is based on the inversion of the equation that relates θ to the Kendall tau ratio or to Spearman's rho coefficient (Meylan *et al.*, 2012; Chowdhary & Singh, 2019; Zhang & Singh, 2019; Chen & Guo, 2019). To obtain θ, in Equation (13) we proceed by trial and error; on the other hand, in Equation (16) 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). Such a formula is:

$$p = \frac{i-0.44}{n+0.12} \tag{19}$$

in which:

i = number of each piece of data, when they are ordered progressively

n = total number of them, or the width of the processed record

The previous expression was applied in the two-dimensional plane, with the data ordered progressively; the dates of occurrence (X_i) in the rows and the maximum flows (Q_i) 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 (X_i and Q_i) is located in the two-dimensional plane and the cell 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* data pairs are drawn, year 1 is searched for and a rectangular or square area of minor *X* and *Q* values is defined, whose count of numbered cells within is NM_1 or minor *X* and *Y* combinations. Once the *n* values of NM_i have been calculated, Gringorten's graphical position formula is applied to estimate the joint or bivariate empirical probability:

$$F(x, y) = P(X \leq x, Q \leq q) = \frac{NM_i - 0.44}{n + 0.12} \quad (20)$$

Selection of the Copula Function

A simple approach to selecting the bivariate Copula function 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 *FC* being tested. The indicators applied are the *standard mean error* (*EME*), the *absolute mean error* (*EMA*) and the *maximum absolute error* (*EAM*); their expressions are (Chowdhary & Singh, 2019):

$$EME = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(w_i^o - w_i^c\right)^2} \quad (21)$$

$$EMA = \frac{1}{n}\sum_{i=1}^{n}|(w_i^o - w_i^c)| \quad (22)$$

$$EAM = \max_{i=1:n}|(w_i^o - w_i^c)| \quad (23)$$

Dependency at the top end

Generalities

The most important criterion applied to select a bivariate *FC* is based on the magnitude of the dependence at the upper end of 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 *Q* is greater than a certain percentile(*s*) of $F_Q(q)$, given that *X* is greater than that percentile in $F_X(x)$; as s approaches unity. The dependence on the lower-left-tail (λ_L), compares *Q* to be less than *X*, when *s* approaches zero (Chowdhary & Singh, 2019; Salvadori *et al.*, 2007).

In relation to the exposed *FC*s, Frank Copulas have no tail dependencies. In contrast, the Gumbel-Hougaard copula has significant upper-tail dependency, equal to:

$$\lambda_U = 2 - 2^{1/\theta} \quad (24)$$

Dupuis (2007) tested six Copula families and found that their ability to estimate extreme events ranges from poor to good, in the following

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

order: Clayton, Frank, Normal, *t*-Student, Gumbel-Hougaard, and Asociated Clayton (*Survival Clayton*). Poulin, Huard, Favre and Pugin (2007) reached similar conclusions, when comparing the same six families of copulas and the so-called A12, which has significant right tail dependency.

Estimation of the observed dependency

In order to approach the estimation of the dependence in the upper tail (λ_U) shown by the available data, the so-called *Empirical Copula* must first be defined. Since the *FC* characterizes the dependence between the random variables X and Q, then the pair of ranks 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:

$$C_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} 1\left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v\right) \quad (25)$$

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, related to the empirical probability integrals $F_n(X)$ and $F_n(Y)$, with the following equations:

$$U_i = \frac{\text{Rank}(X_i)}{n+1} = F_n(x_i) \quad V_i = \frac{\text{Rank}(Q_i)}{n+1} = F_n(q_i) \tag{26}$$

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:

$$\lambda_U^{CFG} = 2 - 2\exp\left\{\frac{1}{n}\sum_{i=1}^{n} \ln\left[\sqrt{\ln\frac{1}{U_i}\cdot\ln\frac{1}{V_i}}/\ln\left(\frac{1}{\max(U_i,V_i)^2}\right)\right]\right\} \tag{27}$$

This estimator accepts that the *FC* can be approximated by one of the extreme values class, it has the advantage of not requiring a threshold value for its estimation.

Ratification of the selected Copula function

This is the most important stage of the *FC* application process, since it is verified that such model faithfully reproduces the observed joint probabilities (Equation (20)). Yue (2000a) describes a simple and practical way of representing the empirical and theoretical joint probabilities. It consists of taking the first one to the abscissa axis and the second one to the ordinate axis; in such a graph, each data pair defines a point that coincides with or departs from the 45º line. 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 *absolute maximum difference* (*dma*, by its acronym in Spanish) between the joint probabilities. To evaluate the statistic (D_n) of the test, the expression that Meylan *et al.* (2012), for α = 5 % this is:

$$D_n = \frac{1.358}{\sqrt{n}} \tag{28}$$

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

Selection of marginal distributions

The approach for selecting the marginal distributions was very simple and consisted of applying the three FDPs 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 models were applied: the Generalized Pareto (PAG), the Kappa and the Wakeby. The first four mentioned FDPs have three fit parameters and the last two, four and five. With the exception of the LP3 that was applied with the method of moments in the logarithmic (WRC, 1977) and real (Bobée, 1975) domains, the rest were adjusted with the method of L moments (Hosking & Wallis, 1997).

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

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

Fitting errors

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

By changing in equations (20) and (21), the probabilities observed by the ordered data of the analyzed series (x_i or y_i) and the probabilities calculated by the estimated values with the FDP that is tested or contrasted, the standard error of fit (*EEA*) and the mean absolute error (*EAM*) are obtained. The values that are estimated ($\hat{x}_i$ or $\hat{y}_i$) are sought for the same probability of non-exceedance assigned to the data with the empirical Gringorten formula (Equation (19)).

Bivariate return periods

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

$$T_{XY}=\frac{1}{P(X>x\vee Y>y)}=\frac{1}{1-F_{X,Y}(x,y)}=\frac{1}{1-C[F_X(x),F_Y(y)]} \tag{29}$$

in which $C[F_X(x),F_Y(y)]$ is the selected *FC*.

The second *bivariate return period* of the event (X, Y) is associated to 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):

$$T'_{XY}=\frac{1}{P(X>x\wedge Y>y)}=\frac{1}{F'_{X,Y}(x,y)}=\frac{1}{1-F_X(x)-F_Y(y)+C[F_X(x),F_Y(y)]} \tag{30}$$

Aldama (2000) obtains the expression $F'_{XY}(x,y)$ of the bivariate probability of exceedance by means of a logical and simple probability reasoning applied in the Cartesian plane. Instead, Yue and Rasmussen (2002) resort to 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):

$$T_{XY} \leq \min[T_X, T_Y] \leq \max[T_X, T_Y] \leq T'_{XY} \tag{31}$$

being:

$$T_X = \frac{1}{F_X'(x)} = \frac{1}{1-F_X(x)} \tag{32}$$

$$T_Y = \frac{1}{F_Y'(y)} = \frac{1}{1-F_Y(y)} \tag{33}$$

The *conditional bivariate return periods* are based on the conditional distribution of *X* given that *Y*≤*y*, which is expressed as follows (Chen & Guo, 2019):

$$P(X \leq x|Y \leq y) = C(F_X(x)|F_Y(y) \leq y) = \frac{C[F_X(x), F_Y(y)]}{F_Y(y)} \tag{34}$$

A similar equation is obtained for Y given that *X*≤*x*. By similarity with the previous equation and Equation (30), the conditional distribution can be obtained for variables *X* and *Y* exceeding some limits (Chen & Guo, 2019), this is:

$$P(X > x|Y > y) = \frac{P(X>x,Y>y)}{P(Y>y)} = \frac{1-F_X(x)-F_Y(y)+C[F_X(x),F_Y(y)]}{1-F_Y(y)} \quad (35)$$

Data for processing

In this study, the record of the *Guamúchil* hydrometric station, from Hydrological Region No. 10 (Sinaloa), Mexico, was processed. The Guamúchil gauging station has a basin area of 1 645 km^2 and a 32-year record that began in 1940 and ended in 1971, when the Eustaquio Buelna dam was built (Aldama *et al.*, 2006). In this record, three years were eliminated: (1) 1949 whose flow of 375 m^3/s occurred on January 22; (2) 1960 with a rate of 1 373 m^3/s occurred on January 13, and (3) 1968 with a rate of 200 m^3/s that occurred on February 10. The record of the adopted *wet season* is shown in Table 1, in its columns 1 to 6.

Table 1. Flows of annual floods, dates of occurrence, theoretical and empirical non-exceedance probabilities and count for joint empirical probability at the *Guamúchil* hydrometric station in Hydrological Region No. 10 (Sinaloa), Mexico.

1	2	3	4	5	6	7	8	9	10
No.	Q_i (m^3/s)	Mes	Día	DJ	α_i (radians)		$F_T(x)$ ord	$F_E(x)$ ord	NM_i
					obs	ord			
1	255	AUG	4	216	3.718269	3.0641	0.041	0.019	3
2	65	SEP	22	265	4.561765	3.2879	0.086	0.054	1

1	2	3	4	5	6	7	8	9	10
No.	Q_i (m³/s)	**Mes**	**Día**	**DJ**	α_i **(radians)**		$F_T(x)$ **ord**	$F_E(x)$ **ord**	NM_i
					obs	**ord**			
3	445	OCT	2	275	4.733907	3.5117	0.165	0.088	12
4	1550	SEP	26	269	4.630622	3.5633	0.189	0.122	24
5	392	AUG	30	242	4.165838	3.5806	0.198	0.157	8
6	916	OCT	8	281	4.837192	3.5806	0.198	0.191	22
7	241	AUG	10	222	3.821554	3.6666	0.244	0.225	3
8	530	AUG	12	224	3.855983	3.6838	0.254	0.260	9
9	648	JUL	23	204	3.511698	3.7183	0.275	0.294	3
10	272	AUG	16	228	3.924839	3.7355	0.285	0.328	5
11	422	SEP	7	250	4.303552	3.8216	0.342	0.363	9
12	377	AUG	5	217	3.735483	3.8560	0.366	0.397	4
13	1173	SEP	17	260	4.475694	3.9076	0.402	0.431	21
14	219	JUL	10	191	3.287913	3.9248	0.415	0.466	2
15	3507	SEP	23	266	4.578979	3.9593	0.440	0.500	25
16	165	JUN	27	178	3.064129	4.0281	0.491	0.534	1
17	526	AUG	18	230	3.959268	4.0970	0.543	0.569	10
18	1014	SEP	20	263	4.527337	4.1658	0.593	0.603	20
19	1610	AUG	2	214	3.683840	4.3036	0.689	0.637	8
20	525	AUG	1	213	3.666626	4.3380	0.711	0.672	5

1	2	3	4	5	6	7	8	9	10
No.	Q_i **(m³/s)**	**Mes**	**Día**	**DJ**	α_i **(radians)**		$F_T(x)$ **ord**	$F_E(x)$ **ord**	NM_i
					obs	**ord**			
21	985	OCT	4	277	4.768336	4.4413	0.772	0.706	22
22	460	JUL	26	207	3.563341	4.4757	0.791	0.740	3
23	390	AGO	26	238	4.096982	4.5273	0.817	0.775	7
24	449	JUL	27	208	3.580555	4.5618	0.833	0.809	3
25	794	JUL	27	208	3.580555	4.5790	0.841	0.843	6
26	720	AUG	22	234	4.028124	4.6306	0.862	0.878	13
27	312	SEP	9	252	4.337981	4.7339	0.898	0.912	6
28	520	SEP	15	258	4.441266	4.7683	0.909	0.946	13
29	1045	AUG	15	227	3.907625	4.8372	0.927	0.981	12

DJ = Julian day.

Obs = observed values.

Ord = ordered values.

$F_T(x)$ = probability of theoretical non-exceedance (Equation (7)).

$F_E(x)$ = probability of empirical non-exceedance (Equation (19)).

NMi = no. of minor X_i and Q_i combinations (Equation (20)).

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Wald-Wolfowitz test

This non-parametric test has been exposed and applied 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 (Q_i). Therefore, it was proposed to apply this test to the records of α_i in radians and of maximum annual flows, which must be random samples.

Results and their discussion

Finding marginal distributions

Verification of randomness

First, the randomness of the records to process was verified, based on the Wald-Wolfowitz Test, whose statistic (U) led to values of -0.720 and 1.522, for the occurrence dates (α_i) in radians and for the flow maximums of Table 1. Since U is less than 1.96, both series or samples are random.

Distribution of annual occurrence dates

To apply the Rosenbrock algorithm, the data (columns 3 and 4) from Table 1 of the Guamúchil station, were adopted as initial values of μ and κ, 4.25 and 0.50, which define an initial FO of 1.040. After 15 stages and 83 evaluations of the FO, the following was obtained: FO = 0.0456, μ =

4.04104, κ = 3.79233, *FN* = 59.3770 and the results are concentrated in column 8 of Table 1 and Figure 1. This fit of the dvM covers the interval of occurrences from 3.0641 to 4.8372 radians, which correspond to the following dates: June 27 to October 8 (281 – 178 = 103 days).

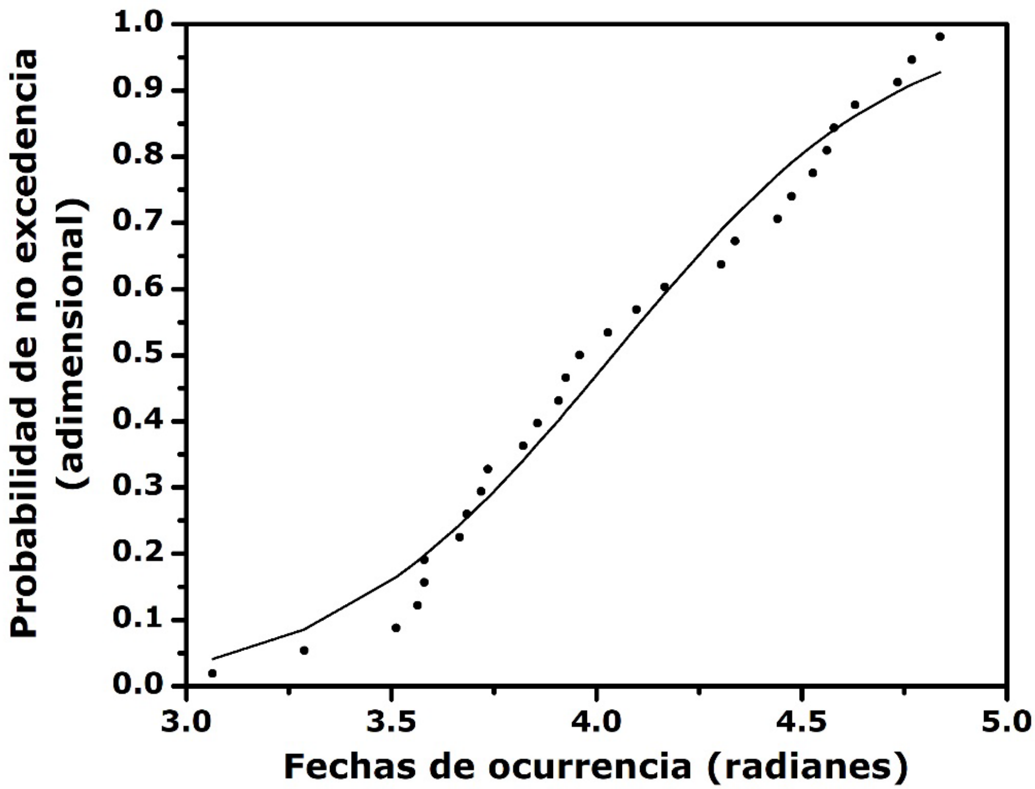

Figure 1. Fit of the von Mises distribution to the annual flood occurrence dates recorded at the *Guamúchil* hydrometric station, in Hydrological Region No. 10 (Sinaloa), Mexico. X axis: Occurrence dates (radians). Y axis: Non–exceedance probability (adimensional).

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

The estimation of dates corresponding to the return periods (*Tr*, years) that will contain the flood *predictions*, was explored by trial and error based on Equation (7), and the obtained values of μ = 4.04104 and κ = 3.79233 for recording Table 1. The month and day that define the most approximate values of non-exceedance probability [$F(x)$], related to the estimated *Tr*, are adopted. Table 2 shows the estimates obtained for the indicated *Tr*.

Table 2. Estimated occurrence dates with von Mises distribution for the indicated return periods, in the annual floods recorded at the *Guamúchil* hydrometric station, Mexico.

***Tr* (years) assigned**	**Obtained date**	**Day in radians**	***Estimated F(x)***	**Estimated 1-*F*(*x*)**	***Estimated Tr* (years)**
50	OCT 29	5.198691	0.97970	0.02030	49.3
100	NOV 8	5.370833	0.98969	0.01031	97.0
500	NOV 28	5.715117	0.99806	0.00194	515.5
1 000	DEC 4	5.818403	0.99904	0.00096	1041.7

Distribution of maximum annual flows

Table 3 shows the fit errors and predictions (m^3/s) obtained with the three reference distributions and the three in general use, applied to the maximum flow records in Table 1. The minimum value of the maximum flows of 65 m^3/s, was lower than the magnitudes with an exceedance probability of 1 %, which are estimated with the FDP applied, except with the Kappa distribution, which was adopted accordingly.

Table 3. Fit errors and predictions (m^3/s) of the six distributions applied in the record of maximum annual flows of the floods in the *Guamúchil* hydrometric station, of the Hydrological Region No. 10 (Sinaloa), Mexico.

FDP	***EEA*** **(m^3/s)**	***EAM*** **(m^3/s)**	**Return periods in years**					
			50	**100**	**500**	**1 000**	**5 000**	**10 000**
LP3	181.5	90.0	2649	3246	4855	5650	7747	8765
GVE	152.6	*71.1*	2749	3647	6852	8929	16361	21178
LOG	*140.4*	72.2	2702	3649	7283	9798	19499	26219
PAG	163.3	82.6	2738	3439	5493	6599	9837	11581
KAP	225.6	116.8	2430	3031	4946	6068	9664	11775
WAK	152.7	76.0	2789	3697	6839	8815	15651	19951

The location (u_2), scale (a_2) and shape (k_2 and h_2) parameters of the adopted Kappa distribution are: 578.6213, 265.6867, -0.275 and -1.0, expressed in the following equation:

$$F(y) = \left\{1 - h_2\left[1 - \frac{k_2(y-u_2)}{a_2}\right]^{1/k_2}\right\}^{1/h_2} \quad (36)$$

The Kappa distribution adopts a value of minus one in its second shape parameter (*h*), because the L ratios (τ_3 = 0.42744 and τ_4= 0.32260) define in the L Ratio Diagram, a point above the Generalized Logistic distribution (Campos-Aranda, 2023a).

Selection and ratification of the FC

The bivariate data processing in Table 1 resulted in the following two association indicators: r_{xy} = 0.3563 and τ_n = 0.2315. Equation (18), with n = 29 and the cited tau, gives a value of 0.9424; therefore, the Kendall ratio is not significant.

Although the value of tau is not statistically different from zero, being positive indicates a direct correspondence or concordance, although low, between the flood occurrence dates and their maximum flow value, for their annual magnitudes. Such relationship or dependency will be modeled by the *FC*.

Table 4 shows the statistical fit indicators that were obtained by applying the Frank and Gumbel-Hougaard (G-H) *FC*s. In equations (21) to (23), the empirical bivariate probabilities were estimated with Equation (20) and the theoretical ones with equations (12) and (15).

Table 4. Statistical indicators of the fit of the Copula Functions indicated in the annual floods of the *Guamúchil* hydrometric station, Mexico.

Copula	***θ***	***EME***	***EAM***	No. DP	No. DN	MDP	MDN	λ_U
Frank	2.1790	*0.0309*	*0.0237*	20	9	0.0682	-0.0368	0.0000
G-H	1.3013	0.0326	0.0246	18	11	0.0785	-0.0369	0.2965

DP, DN = positive and negative differences.

MDP, MDN = maximum positive and negative difference.

Since the application of Equation (27) led to a λ_U^{CFG} value of 0.3416, there is no difficulty in selecting *FC* Gumbel-Hougaard in Table 4.

On the other hand, since the value of λ_U^{CFG} was slightly higher than the λ_U of the *FC*, one should look for a *FC* with greater dependence on its right end, which was not necessary in this case, because predictions were not made for high return periods, but rather estimates of exceedance probabilities were formulated, according to occurrence dates, based on *conditional* return periods. Chen and Guo (2019) exclusively apply the Gumbel-Hougaard *FC*, in this type of frequency analysis.

Table 5 shows the observed empirical bivariate non-exceedance probabilities (w_i^o) calculated with Equation (20) and estimated theoretical (w_i^c) with the Gumbel-Hougaard *FC*. The maximum positive and negative differences are also indicated shaded.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Table 5. Joint non-exceedance probabilities and their differences, calculated with the Gumbel-Hougaard *FC*, for the annual floods of the *Guamúchil* station, Mexico.

No.	w_i^o	w_i^c	Differences
1	0.0879	0.0785	0.0094
2	0.0192	0.0561	-0.0369
3	0.3970	0.3529	0.0441
4	0.8091	0.8232	-0.0141
5	0.2596	0.2318	0.0279
6	0.7404	0.7201	0.0203
7	0.0879	0.0881	-0.0002
8	0.2940	0.2155	0.0785
9	0.0879	0.1215	-0.0335
10	0.1566	0.1165	0.0400
11	0.2940	0.2812	0.0128
12	0.1223	0.1229	-0.0007
13	0.7060	0.7126	-0.0066
14	0.0536	0.0250	0.0285
15	0.8434	0.8391	0.0043
16	0.0192	0.0103	0.0090
17	0.3283	0.2514	0.0769
18	0.6717	0.6919	-0.0202
19	0.2596	0.2485	0.0111

No.	w_i^o	w_i^c	Differences
20	0.1566	0.1486	0.0080
21	0.7404	0.7410	-0.0007
22	0.0879	0.1043	-0.0163
23	0.2253	0.2156	0.0097
24	0.0879	0.1061	-0.0182
25	0.1909	0.1628	0.0282
26	0.4313	0.3619	0.0694
27	0.1909	0.1998	-0.0089
28	0.4313	0.3865	0.0448
29	0.3970	0.3623	0.0346
-	-	-	-

On the other hand, Equation (28) defines D_n = 0.2522 and since the maximum absolute difference in Table 5 is 0.0785, the Kolmogorov-Smirnov test ratifies the adopted Gumbel-Hougaard (G-H) *FC*. The correlation coefficient (r_{xy}) between the empirical and theoretical probabilities, estimated with the *FC* of G-H, was 0.9931; therefore, the fit is excellent. The graphic contrast between both probabilities, to ratify the adoption of *FC* of G-H, is shown in Figure 2 for the data in Table 5.

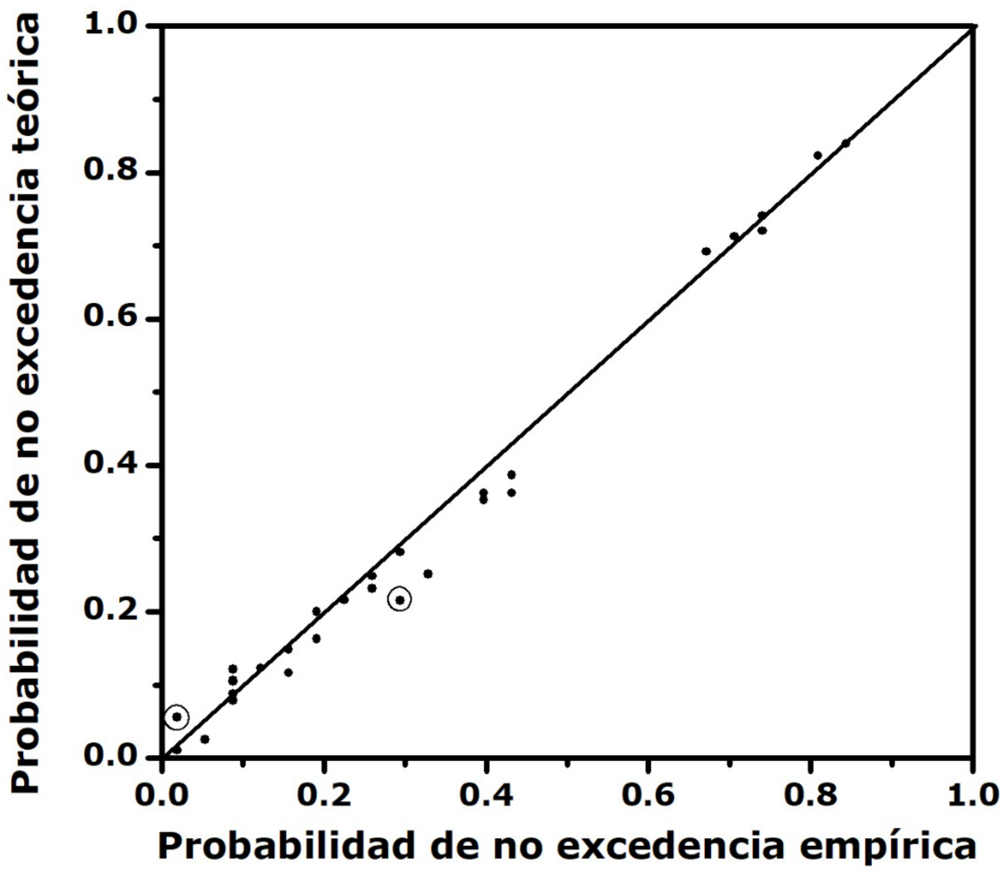

Figure 2. Graphic contrast of joint probabilities of the occurrence dates and the maximum flows of the floods recorded at the *Guamúchil* hydrometric station, Mexico; with the Gumbel-Hougaard *FC*. X axis: Empirical non–exceedance probability. Y axis: Theoretical non–exceedance probability.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Graphs of joint return periods

The bivariate return periods of the AND type are calculated with Equation (30). For their estimates with values of T'_{XY} of 50, 100 and 500 years, a date (month and day) is defined and its non-exceedance probability $F(x)$ with Equation (7) and the values of μ = 4.04104 and κ = 3.79233. Now by *trial and error*, a maximum flow ($y = q_{max}$) is assigned and its $F(y)$ is estimated with Equation (36). Both results are taken to Equation (15), with θ = 1.3013, to estimate the joint non-exceedance probability and thus obtaining with Equation (30), the T'_{XY} that must coincide with the searched value.

Dates of occurrence and maximum flows are selected arbitrarily, in order to define the curves of T'_{XY}. Table 6 shows the values adopted to define the three graphs in Figure 3.

Table 6. Pairs of dates of occurrence and maximum annual flow used to define the graphs of the AND type joint return period with the Gumbel-Hougaard *FC*, in the floods of the *Guamúchil* station, Mexico.

T'_{XY} = 50 years			T'_{XY} = 100 years			T'_{XY} = 500 years		
Date	**ND***	***Qm* M³/s**	**Date**	**ND***	***Qm* M³/s**	**Date**	**ND***	***Qm* M³/s**
Oct 29	121	117	Nov 8	131	255	Nov 28	151	0
Oct 27	119	569	Nov 6	129	717	Nov 27	150	607
Oct 25	117	814	Nov 5	128	880	Nov 26	149	1055

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

T'_{XY} = 50 years			T'_{XY} = 100 years			T'_{XY} = 500 years		
Date	**ND***	***Qm* M^3/s**	**Date**	**ND***	***Qm* M^3/s**	**Date**	**ND***	***Qm* M^3/s**
Oct 22	114	1097	Nov 3	126	1159	Nov 25	148	1452
Oct 20	112	1248	Oct 31	123	1496	Nov 24	147	1798
Oct 15	107	1530	Oct 28	120	1749	Nov 23	146	2095
Oct 10	102	1719	Oct 25	117	1941	Nov 22	145	2347
Oct 5	97	1854	Oct 20	112	2171	Nov 21	144	2560
Oct 1	93	1937	Oct 15	107	2331	Nov 20	143	2743
Sep 25	87	2034	Oct 10	102	2450	Nov 18	141	3036
Sep 20	82	2098	Oct 5	97	2543	Nov 15	138	3357
Sep 15	77	2152	Sep 30	92	2617	Nov 10	133	3707
Sep 15	72	2197	Sep 25	87	2679	Nov 5	128	3937
Sep 5	67	2235	Sep 20	82	2731	Oct 31	123	4103
Aug 31	62	2268	Sep 15	77	2776	Oct 25	117	4252
Aug 25	56	2302	Sep 10	72	2815	Oct 20	112	4349
Aug 20	51	2326	Sep 5	67	2849	Oct 15	107	4429
Aug 15	46	2346	Aug 31	62	2879	Oct 10	102	4496
Aug 10	41	2363	Aug 25	56	2910	Oct 5	97	4554
Aug 5	36	2377	Aug 20	51	2932	Sep 30	92	4604
Jul 31	31	2389	Aug 15	46	2951	Sep 25	87	4648
Jul 25	25	2400	Aug 10	41	2967	Sep 20	82	4687
Jul 13	13	2415	Aug 5	36	2980	Sep 15	77	4722

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

T'_{XY} = 50 years			T'_{XY} = 100 years			T'_{XY} = 500 years		
Date	**ND***	***Qm* M^3/s**	**Date**	**ND***	***Qm* M^3/s**	**Date**	**ND***	***Qm* M^3/s**
Jul 1	1	2430	Jul 31	31	2992	Sep 10	72	4754
			Jul 25	25	3002	Sep 5	67	4783
			Jul 20	20	3009	Aug 31	62	4808
			Jul 15	15	3015	Aug 20	51	4855
			Jul 10	10	3019	Aug 10	41	4886
			Jul 1	1	3031	Aug 5	36	4899
						Jul 31	31	4909
						Jul 25	25	4919
						Jul 20	20	4926
						Jul 1	1	4946

*ND = day number from the 1st. of July.

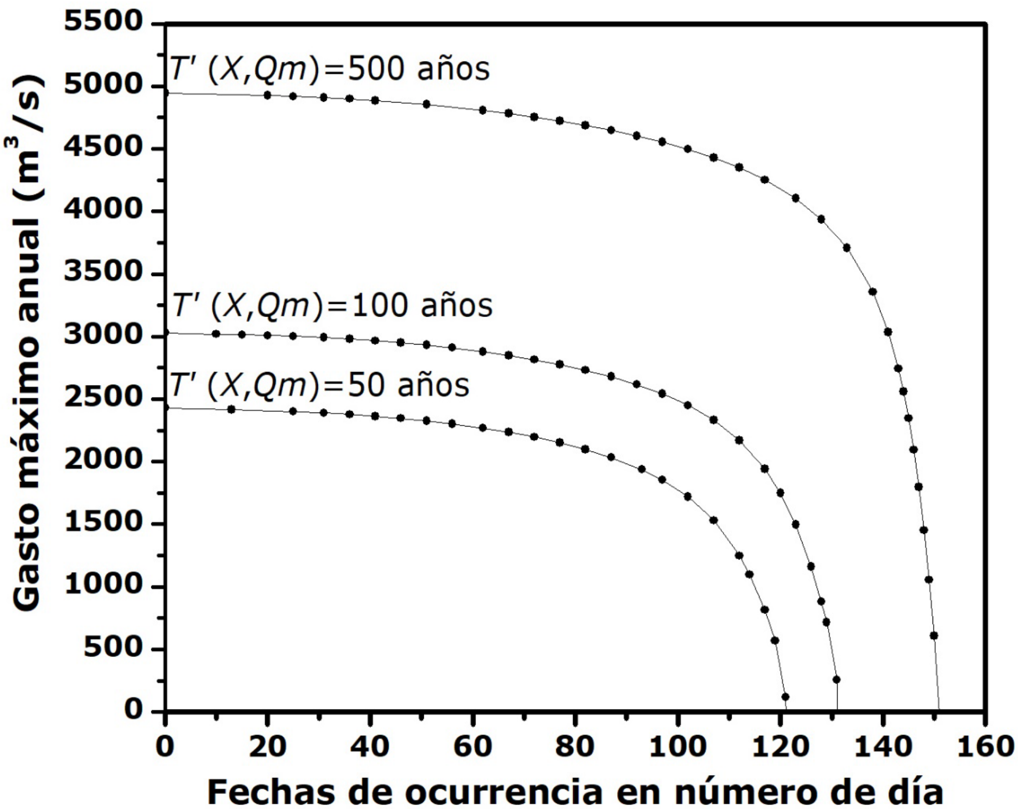

Figure 3. Graphs of the three design joint return periods T'_{XY} obtained with the Gumbel-Hougaard *FC*, in the floods of the *Guamúchil* hydrometric station, Mexico. X axis: dates of occurrence in day number. Y axis: maximum annual flow (m^3/s).

Probabilities of occurrence of design events

Table 7, Table 8 and Table 9 show the *conditional* exceedance probabilities $P(X>x|Qm>qo)$, calculated with Equation (35), when *qo* has a return period (*Tr*) of 50, 100 and 500 years, that is, when the marginal probability $v = F_Y(y)$ is equal to 0.98. 0.99 and 0.998 and the flow *Qm* exceeds 2 430, 3 031 and 4 946 m^3/s, according to Table 3.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Table 7. Calculations of the conditional probability of the occurrence dates provided that the maximum flow has a *Tr* = 50 years in the floods of the *Guamúchil* hydrometric station, Mexico.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Proposal PE (%)	99	95	90	80	70	60	50	40	30	20	10	5	1
Date obtained	JUN 3	JUN 30	JUL 13	JUL 27	AUG 6	AUG 15	AUG 23	AUG 30	SEP 8	SEP 18	OCT 2	OCT 15	NOV 8
α_i (radians) = X_i	2.6510	3.1158	3.3396	3.5806	3.7527	3.9076	4.0453	4.1658	4.3208	4.4929	4.7339	4.9577	5.3708
Real PE	0.9902	0.9509	0.8994	0.8025	0.7038	0.5977	0.4959	0.4068	0.3000	0.2002	0.1016	0.0488	0.0103
PNE of the *FC* GH	0.0097	0.0489	0.1002	0.1967	0.2949	0.4003	0.5014	0.5897	0.6954	0.7939	0.8902	0.9404	0.9739
Conditional PE	0.9985	0.9916	0.9813	0.9592	0.9333	0.9012	0.8652	0.8280	0.7726	0.7016	0.5853	0.4591	0.2106

PE = probability of exceedance, dimensionless.

PNE = probability of non-exceedance, dimensionless.

GH = Gumbel-Hougaard *FC*.

Table 8. Calculations of the conditional probability of the occurrence dates provided that the maximum flow has a *Tr* = 100 years in the floods of the *Guamúchil* hydrometric station, Mexico.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Proposal PE (%)	99	95	90	80	70	60	50	40	30	20	10	5	1
Date obtained	JUN 3	JUN 30	JUL 13	JUL 27	AUG 6	AUG 15	AUG 23	AUG 30	SEP 8	SEP 18	OCT 2	OCT 15	NOV 8
α_i (radians) = X_i	2.6510	3.1158	3.3396	3.5806	3.7527	3.9076	4.0453	4.1658	4.3208	4.4929	4.7339	4.9577	5.3708
Real PE	0.9902	0.9509	0.8994	0.8025	0.7038	0.5977	0.4959	0.4068	0.3000	0.2002	0.1016	0.0488	0.0103
PNE of the *FC* GH	0.0098	0.0490	0.1005	0.1972	0.2957	0.4015	0.5030	0.5918	0.6982	0.7974	0.8951	0.9467	0.9828
Conditional PE	0.9988	0.9932	0.9849	0.9671	0.9461	0.9202	0.8911	0.8609	0.8160	0.7582	0.6622	0.5543	0.3073

PE = probability of exceedance, dimensionless.

PNE = probability of non-exceedance, dimensionless.

GH = Gumbel-Hougaard *FC*.

Table 9. Calculations of the conditional probability of the occurrence dates provided that the maximum flow has a *Tr* = 500 years in the floods of the *Guamuchil* hydrometric station, Mexico.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Proposal PE (%)	99	95	90	80	70	60	50	40	30	20	10	5	1
Date obtained	JUN 3	JUN 30	JUL 13	JUL 27	AUG 6	AUG 15	AUG 23	AUG 30	SEP 8	SEP 18	OCT 2	OCT 15	NOV 8
α_i (radians) = X_i	2.6510	3.1158	3.3396	3.5806	3.7527	3.9076	4.0453	4.1658	4.3208	4.4929	4.7339	4.9577	5.3708
Real PE	0.9902	0.9509	0.8994	0.8025	0.7038	0.5977	0.4959	0.4068	0.3000	0.2002	0.1016	0.0488	0.0103
PNE of the *FC* GH	0.0098	0.0491	0.1006	0.1975	0.2961	0.4022	0.5039	0.5930	0.6997	0.7995	0.8980	0.9506	0.9888
Conditional PE	0.9993	0.9958	0.9907	0.9798	0.9670	0.9510	0.9332	0.9147	0.8870	0.8514	0.7919	0.7231	0.5423

PE = probability of exceedance, dimensionless.

PNE = probability of non-exceedance, dimensionless.

GH = Gumbel-Hougaard *FC*.

Then, for the case of *qo* = 2 430 m^3/s (Table 7), the following event probabilities can occur: (1) the *occurrence probability* of such flows after July 13 is 98.13 %; (2) the probability that such flows occur between July 13 and November 8 will be: 98.13 - 21.06 = 77.07 %, and (3) during the period from September 8 to 18 will be: 77.26 - 70.16 = 7.10 %.

Then, for the case of *qo* = 3 031 m^3/s (Table 8), the following event probabilities can occur: (1) the occurrence probability of such flows after July 13 is 98.49 %; (2) the probability that such flows occur between July 13 and November 8 will be: 98.49 - 30.73 = 67.76 %, and (3) during the period from September 8 to 18 will be: 81.60 - 75.82 = 5.78 %.

Finally, for the case of *qo* = 4946 m^3/s (Table 9), there are three following event probabilities (1) the probability of occurrence of such flows after July 13 is 99.07 %; (2) the probability that such flows occur between

July 13 and November 8 will be: 99.07 - 54.23 = 44.84 %, and (3) during the period from September 8 to 18 will be: 88.70 - 85.14 = 3.56 %.

The comparison between the results of the three limits established for the maximum flow (*qo*), indicates that the probability of occurrence after July 13 increases slightly as the *qo* grows, which is logical, given the positive correlation of r_{xy} = 0.3563 shown between the occurrence dates and the maximum flows. The above is also influenced by the denominator of Equation (35); which, as *qo* increases, is reduced from 0.02 to 0.002.

On the contrary, the two occurrence probabilities calculated for the two periods within the wet season decrease as *qo* increases, which is due to the lower probability of exceedance of each maximum flow, having 2, 1 and 0.2 %.

Contrast with results of a previous work

Chen and Guo (2019) present on their pages 42 to 44 and Table 3.2, as a case study of a bivariate flood frequency analysis of occurrence dates and maximum flow, its application in the multi-purpose *Geheyan* reservoir, with a basin of 17 000 km^2. which receives an average annual rainfall of 1 500 millimeters and whose flood season covers five months from May 1 to September 30 (153 days). Keeping the due proportions, the results of Table 3.2 coincide with those of Table 8, since both are for a maximum flow of a return period of 100 years.

Application in basins with two wet seasons

In general, in large basins, only the storms which result in cold fronts and hurricanes, are covered entirely with low intensity and long duration rains, which generate floods of great magnitude, due to their maximum flow, volume and duration. On the other hand, in such basins, convective storms are local, of great intensity and give rise to ordinary floods.

In Mexico, the large basins of the mountainous zone of Hydrological Region No. 10 (Sinaloa) present two dates of occurrence of their floods, in autumn and winter (November to March) and in summer (June to September). In these flood regimes there are two modes and their probabilistic characterization of their occurrence dates is carried out with a mixture of von Mises distributions (Carta *et al.*, 2008; Campos-Aranda, 2023b).

Chen *et al.* (2010), and Chen and Guo (2019) have presented floods *seasonal* analyzes to define times of occurrence, in basins with two or more wet seasons.

Application in large basins with a wet season

In large basins with flood regimes of a single wet season, the bivariate study of dates of occurrence and maximum flow, in their various tributaries or main rivers can help to understand the evolution or development of their floods and then make more reliable *forecasts* of arrival dates downstream.

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

Conclusions

The bivariate frequency analysis of dates of occurrence (X) and maximum flow (Qm) of annual floods is feasible due to the *Copula functions*; which allow building their *joint* probability distribution, based on some univariate marginals. The von Mises distribution is the one that characterizes the occurrence dates, and the Qm in particular is an ideal FDP.

In this study, a copula function (FC) of a single fitting parameter (θ) was used, which is estimated based on the Kendall tau ratio, which is calculated with the joint registration of X and Qm. Such an approach first estimates λ_U^{CFG}, or observed in the right tail dependency of the available record set. Afterwards, an FC is that reproduces such a value of λ_U^{CFG} is searched; in this study the Gumbel-Hougaard was chosen. An FC that does not have a significant right tail dependency, such as Frank's, is also applied to allow comparing and assessing the quality of the previously adopted FC fit.

The described numerical application, in the 32 annual data of occurrence dates and maximum flow (m^3/s) of the annual floods recorded in the *Guamúchil* hydrometric station of the Hydrological Region No. 10 (Sinaloa), Mexico; showed in Figure 2 a reliable reproduction of the empirical and theoretical bivariate probabilities, through the Gumbel-Hougaard FC, with a linear correlation coefficient of 0.9931.

On the other hand, in Figure 3, with regard to the joint return periods of AND type design, infinite pairs of critical X and Qm can be defined, since they are in the curved region of each graph.

Finally, Tables 7, 8 and 9 show the calculation of *conditional* probabilities, which allow estimating, for an arbitrarily adopted period, the exceedance probability of design events with return periods of 50, 100 and 500 years.

References

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

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

Bobée, B. (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.

Burn, D. H. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, 202(1-4), 212-230. DOI: 10.1016/S0022-1694(97)00068-1

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.

Campos-Aranda, D. F. (2017). Definición de tres épocas de crecientes utilizando estadísticos direccionales. *Tecnología y ciencias del agua*, 8(1), 155-165.

Campos-Aranda, D. F. (2023a). Análisis de Frecuencias comparativo con momentos L entre la distribución Kappa y seis de aplicación generalizada. *Tecnología y ciencias del agua*, 14(1), 200-250. DOI: 10.24850/j-tyca-14-01-05

Campos-Aranda, D. F. (2023b). Caracterización probabilística de las fechas de ocurrencia de las crecientes anuales mediante la distribución de von Mises. *Tecnología y ciencias del agua*, 14(2), 204-260. DOI: 10.24850/j-tyca-14-02-06

Carta, J. A., Bueno, C., & Ramírez, P. (2008). Statistical modelling of directional wind speeds using mixtures of von Mises distributions: Case study. *Energy Conversion and Management*, 49(5), 897-907. DOI: 10.1016/j.enconman.2007.10.017

Cunderlik, J. M., Ouarda, T. B. M. J., & Bobée, B. (2004). Determination of flood seasonality from hydrological records. *Hydrological Sciences Journal*, 49(3), 511-526. DOI: 10.1623/hysj.49.3.511.54351

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., Yan, B., Liu, P., & Fang, B. (2010). A new seasonal design flood method based on bivariate joint distribution of flood magnitude and date of occurrence. *Hydrological Sciences Journal*, 55(8), 1264-1280. DOI: 10.1080/02626667.2010.520564

Chen, L., Singh, V. P., Guo, S., Fang, B., & Liu, P. (2013). A new method for identification of flood seasons using directional statistics. *Hydrological Sciences Journal*, 58(1), 28-40. DOI: 10.1080/02626667.2012.743661

Chen, L., & Guo, S. (2019). Chapter 3. Copula-based Flood Frequency Analysis and Chapter 4. Copula-based Seasonal Design Flood Estimation. In: *Copulas and its application in Hydrology and Water Resources.* (pp. 39-71, 73-96). 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.

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)

Durrans, S. R., Eiffe, M. A., Thomas Jr., W. O., & Goranflo, H. M. (2003). Joint seasonal/annual flood frequency analysis. *Journal of Hydrologic Engineering*, 8(4), 181-189. DOI: 10.1061/(ASCE)1084-0699(2003)8:4(181)

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

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

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, England: Cambridge University Press.

Khedun, C. P., Singh, V. P., & Byrd, A. R. (2019). Joint probability of extreme streamflow and its day of occurrence. *Journal of Hydrologic Engineering*, 24(8), 06019005:1-8. DOI: 10.1061/(ASCE)HE.1943-5584.0001813

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

Kuester, J. L., & Mize, J. H. (1973). Chapter 9. Multivariable unconstrained methods. III. ROSENB algorithm. In: *Optimization techniques with Fortran* (pp. 320-330). New York, USA: McGraw-Hill Book Co.

Metcalfe, A. V. (1997). Theme 2.4.4. Vector variables and von Mises distribution. In: *Statistics in Civil Engineering* (pp. 31-32). London, England: Arnold Publishers.

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.

Nieves, A., & Domínguez, F. C. (1998). Sección 6.2. Cuadratura de Gauss. En: *Métodos numéricos. Aplicados a la ingeniería* (pp. 416-425). México, DF, México: Compañía Editorial Continental.

Olver, F. W. J. (1972). Chapter 9. Bessel Functions of Integer Order. In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of Mathematical Functions* (pp. 355-436) (9th print.). New York, USA: Dover Publications.

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., Gutiérrez-López, A., & Ruiz-Silva, H. L. (2009). Análisis de la ocurrencia en el tiempo de los gastos máximos en México. *Ingeniería Hidráulica en México*, 24(1), 115-124.

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. In: *Flood frequency analysis* (pp. 1-21). Boca Raton, USA: CRC Press.

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

Rosenbrock, H. H. (1960). An automatic method of finding the greatest or least value of a function. *Computer Journal*, 3(3), 175-184.

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.

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

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.

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

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

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. In: *Copulas and their applications in water resources engineering* (pp. 62-122). Cambridge, UK: Cambridge University Press.