

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

Artículos

**Selección y aplicación de funciones Cópula con
dependencia en su extremo derecho al análisis de
frecuencias conjunto (Q, V) de crecientes anuales**
**Selection and application of Copula functions with
dependence on its tail right to the joint frequency
analysis (Q, V) of annual floods**

Daniel Francisco Campos-Aranda¹, ORCID: <https://orcid.org/0000-0002-9876-3967>

¹Profesor jubilado de la Universidad Autónoma de San Luis Potosí, San Luis Potosí, México, campos_aranda@hotmail.com

Autor para correspondencia: Daniel Francisco Campos-Aranda, campos_aranda@hotmail.com

Resumen

El diseño hidrológico de los *embalses* por construir o la revisión de los ya existentes requiere la estimación del llamado *hidrograma de la creciente de diseño*. La manera más simple y aproximada para estimar tal hidrograma, para un determinado *periodo de retorno conjunto*, es a



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

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

Abstract

The hydrological design of the *reservoirs* to be built or the revision of the existing ones, requires the estimation of the so-called *Design Flood*



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

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

Recibido: 30/12/2021

Aceptado: 07/03/2022

Introducción

Generalidades

La república mexicana se localiza en la zona de influencia de los huracanes que se originan en el mar Caribe y en el océano Pacífico. Además, lo afectan otros fenómenos meteorológicos de gran alcance espacial, como son los frentes fríos y las tormentas convectivas extensas, algunas de origen orográfico. Los fenómenos atmosféricos citados producen lluvias de gran magnitud, las cuales generan *crecientes* o *avenidas máximas* que inundan extensas regiones del país. Tales inundaciones causan pérdida de vidas humanas, y enormes daños económicos y ambientales. Este panorama que se repite cada año en distintas zonas de México destaca la importancia del estudio de las crecientes, con la finalidad de entender su dinámica e intentar cuantificar sus magnitudes, con el propósito de diseñar y proteger las obras hidráulicas de protección y aprovechamiento que se construyen (Aldama, 2000; Aldama, Ramírez, Aparicio, Mejía-Zermeño, & Ortega-Gil, 2006; Gómez, Aparicio, & Patiño, 2010).

Las principales obras hidráulicas de protección son los diques y muros de contención que por lo general están relacionados con las rectificaciones y canalizaciones. Los puentes y otras obras del drenaje urbano también son estructuras de protección, cuyo diseño hidrológico se realiza con base en las *crecientes de diseño* (CD), que son gastos máximos del río asociados

con bajas probabilidades de ser excedidos. La estimación más confiable de las CD se realiza a través del *análisis de frecuencias de crecientes* (AFC), el cual intenta interpretar el registro de gastos máximos anuales disponible en términos de eventos futuros de una determinada probabilidad de ocurrencia (Bobée & Ashkar, 1991; Rao & Hamed, 2000; Meylan, Favre, & Musy, 2012).

Para que los resultados del AFC sean aproximados y relativamente confiables se deben cumplir los siguientes cuatro requisitos: (1) el registro de gastos máximos anuales debe ser aleatorio; (2) la *función de distribución de probabilidades* (FDP) que lo represente debe ser idónea; (3) el ajuste de las FDP probadas se tiene que hacer con métodos estadísticos eficientes, y (4) la selección de resultados requiere ser objetiva (Hosking & Wallis, 1997; Stedinger, 2017).

Por otra parte, el diseño hidrológico de los *embalses* o presas de almacenamiento necesita la estimación del llamado *hidrograma de la creciente de diseño*, ya que en tales obras hidráulicas el volumen de la creciente es sumamente importante para establecer su peligrosidad y definir su manejo y almacenamiento temporal sin riesgo. Aldama (2000) demostró que los embalses no son sensibles al tiempo que transcurre en llegar al gasto máximo, pero sí a su magnitud y al volumen de la creciente. Lo anterior permite simplificar de forma notable el *análisis multivariado* de las crecientes, pues el *enfoque bivariado* del gasto máximo (Q) y del volumen (V) anuales permite definir con suficiente aproximación el hidrograma de la creciente de diseño, pues V tiene correlación positiva con la duración total (D) del hidrograma.

Los *análisis bivariados de las crecientes* comenzaron a finales del siglo pasado, con los trabajos de Goel, Seth y Chandra (1998), de Yue (1999), y de Yue, Ouarda, Bobée, Legendre y Bruneau (1999), que procesaron los datos conjuntos de Q y V , y de V y D . Al inicio de este siglo, Yue y Rasmussen (2002) establecen los conceptos hidrológicos básicos del análisis probabilístico bivariado. Los primeros análisis bivariados de crecientes emplearon la distribución Normal bivariada y el llamado modelo Logístico, que acepta funciones marginales de valores extremos, como las distribuciones Gumbel y GVE. Estos modelos tienen dos limitaciones importantes: (1) sus distribuciones marginales son iguales, y (2) los parámetros de ajuste de tales marginales intervienen en la modelación conjunta del registro de Q y V (Favre, El Adlouni, Perreault, Thiémonge, & Bobée, 2004).

En la actualidad, las limitaciones citadas son evitadas al aplicar las llamadas *funciones Cópula* (FC), modelos matemáticos que aceptan cualquier FDP como marginal, para formar la distribución bivariada o multivariada, y que se basan en la correlación que existe entre Q y V (Shiau, Wang, & Tsai, 2006; Zhang & Singh, 2006). El uso de las FC se ha consolidado, como lo demuestran los textos especializados existentes (Salvadori, De Michele, Kottegodá, & Rosso, 2007; Zhang & Singh, 2019; Chen & Guo, 2019; Chowdhary & Singh, 2019).

Ramírez-Orozco y Aldama (2000), Escalante-Sandoval y Reyes-Chávez (2002), así como Volpi y Fiori (2012), y Requena, Mediero y Garrote (2013) indican que el AFC bivariado conduce a una infinidad de combinaciones de gasto pico y volumen para una probabilidad de excedencia *conjunta* adoptada. Lo anterior implica que para un mismo

periodo de retorno conjunto existen muchas crecientes o *hidrogramas* que producirán distintos efectos en el embalse que se diseña o revisa, seleccionando lógicamente el que genera las condiciones más severas en su almacenamiento y vertedor de excedencias.

Objetivos

Los trabajos de Dupuis (2007), y de Poulin, Huard, Favre y Pugin (2007) demostraron que un aspecto decisivo en la selección de la FC *idónea* por utilizar está relacionado con la *dependencia* en el extremo derecho de los datos y la que muestran ciertas familias de FC. Por lo anterior, los *objetivos básicos* de este estudio fueron los cuatro siguientes: 1) exponer cuatro familias de FC que tienen dependencia significativa en su cola derecha (λ_U); 2) describir cómo se estima la λ_U del registro de Q y V disponible (λ_U^{CFG}); 3) contrastar los valores de λ_U^{CFG} obtenidos en 16 registros reales de Q y V contra los que definen las cuatro FC expuestas, y 4) procesar como aplicación numérica el registro de 26 datos de Q y V anuales de las crecientes de entrada a la presa Adolfo López Mateos (*Humaya*) del estado de Sinaloa, México.

Funciones Cópula

Concepto y definición

Como ya se indicó, la ventaja esencial de las *funciones Cópula* (FC) consiste en permitir expresar una *distribución conjunta* de variables aleatorias correlacionadas, como una función de sus distribuciones marginales, previamente adoptadas. Entonces, una FC enlaza o relaciona las distribuciones marginales univariadas para formar una distribución multivariada. Otra ventaja básica de las FC al formar distribuciones multivariadas es el hecho de que ellas separan el efecto de la dependencia entre las variables aleatorias de los efectos de las distribuciones marginales en la modelación conjunta.

Debido a lo anterior, la construcción de la distribución multivariada se reduce al estudio de la relación que guardan las variables correlacionadas si las distribuciones marginales univariadas son conocidas. El uso de las FC ofrece entera libertad para adoptar o seleccionar las distribuciones marginales univariadas que mejor representen a los datos (Shiau *et al.*, 2006; Kottegoda & Rosso, 2008; Meylan *et al.*, 2012; Zhang & Singh, 2019).

Como en este estudio aplicarán las FC en el *análisis de frecuencias bivariado* de las crecientes anuales, la definición siguiente se refiere a dos variables aleatorias X y Y correlacionadas, cuya función de distribución de probabilidades acumuladas conjuntas es $F_{X,Y}(x,y)$, con distribuciones

marginales univariadas $F_X(x)$ y $F_Y(y)$; entonces la FC existe, es $C[\cdot]$ y es tal que:

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

La ecuación anterior define el concepto básico para el desarrollo de las FC y se conoce como Teorema de Sklar, expuesto en 1959 (Shiau *et al.*, 2006; Kottegodda & Rosso, 2008; Meylan *et al.*, 2012; Zhang & Singh, 2019).

Familias de Cópulas

Las *funciones Cópula* (FC) que han sido desarrolladas se han clasificado en cuatro clases: de Arquímedes, de valores extremos, elípticas y misceláneas. También se clasifican en Cópulas de un parámetro o de varios, dependiendo de la amplitud con la cual la estructura de la dependencia entre las variables Q y V es definida (Meylan *et al.*, 2012; Chowdhary & Singh, 2019).

Salvadori *et al.* (2007) exponen un resumen amplio y útil de FC que han sido aplicadas en el campo de la hidrología. Al respecto, en la relación siguiente no se han incluido las familias de Cópulas de Ali-Mikhail-Haq,

llamada Cópula AMH, debido a que limita la dependencia a un rango del cociente tau de Kendall de -0.1817 a 0.3333, y la familia de Farlie-Gumbel-Morgenstern, designada FGM, que lo limita aún más, al intervalo de ± 0.2222 .

Las Cópulas de Arquímedes han tenido aplicación amplia debido a su construcción simple, un único parámetro, rango amplio y aceptación de ambos tipos de dependencia (positiva y negativa). Designando a $F_X(x)=u$, $F_Y(y)=v$ y θ el parámetro que mide la dependencia o asociación entre u y v , se tienen las siguientes dos familias de Cópulas de Arquímedes, las cuales aceptan dependencia negativa y positiva (Shiau *et al.*, 2006; Genest & Favre, 2007; Salvadori *et al.*, 2007; Michiels & De Schepper, 2008; Zhang & Singh, 2019; Chen & Guo, 2019; Chowdhary & Singh, 2019).

1. **Clayton.** Esta función Cópula es llamada de Cook-Johnson por Zhang y Singh (2006). Sraj, Bezak y Brilly (2015) la citan con los dos nombres. Su ecuación y espacio de variación de θ son:

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

Para la dependencia positiva $\theta > 0$ y para la negativa $-1 \leq \theta < 0$, con $\theta = 0$ para la independencia entre u y v . La relación de θ con el cociente tau de Kendall es la siguiente:

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

2. **Frank.** Su ecuación y espacio de variación de θ son:

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

Para la dependencia negativa $0 \leq \theta < 1$ y para la positiva $\theta > 1$, con $\theta = 1$ para la independencia entre u y v . La relación de θ con el τ_n es la siguiente:

$$\tau_n = 1 + \frac{4}{\theta} [D_1(\theta) - 1] \quad (5)$$

siendo $D_1(\theta)$ la función Debye de orden 1, cuya expresión es:

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

La ecuación anterior se estimó con integración numérica, ratificando sus resultados con los valores tabulados por Stegun (1972).

Enseguida se cita una familia de Cópulas de valores extremos, la cual acepta sólo dependencia positiva.

3. **Gumbel-Hougaard**. Su ecuación y espacio de variación de θ son:

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

Con $\theta = 1$ se tiene independencia entre u y v . La relación de θ con el cociente tau de Kendall es la siguiente:

$$\tau_n = \frac{\theta-1}{\theta} \quad (8)$$

Por último, se citan dos familias de Cópulas de clase miscelánea.

4. **Plackett** acepta dependencia negativa ($\theta < 1$) y positiva ($\theta > 1$), con $\theta = 1$ para la independencia entre u y v :

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

La relación de θ con la rho de Spearman es la siguiente:

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

5. **Raftery** acepta sólo dependencia positiva, y θ varía entre 0 y 1, con $\theta = 0$ para la independencia entre u y v :

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

Las relaciones de θ con el cociente tau de Kendall y con el coeficiente rho de Spearman son las siguientes:

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

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

Las familias de las Cópulas elípticas son distribuciones y sus marginales que aceptan la propiedad de ser *elípticas*, como las distribuciones Normal, t de Student y Cauchy. Un vector aleatorio $X \in \mathbb{R}^d$ se dice que tiene una distribución elíptica si admite la siguiente representación estocástica: $X = \mu + RAU$; en la cual, $\mu \in \mathbb{R}^d$, R es una variable aleatoria independiente de U , que es un vector aleatorio uniforme en la esfera unitaria en \mathbb{R}^d y A es una matriz fija de $d \times d$, tal que $A \cdot A^T$ es no singular (Salvadori *et al.*, 2007; Chen & Guo, 2019).

Cuando las distribuciones marginales son de tipos diferentes y posiblemente no elípticas, las Cópulas elípticas se denominan *Cópulas Metaelípticas*, cuyas familias principales son la Gaussiana y la de la t de

Student (Chowdhary & Singh, 2019). La Cópula t de Student será descrita con detalle posteriormente.

Familias de Cópulas asociadas

Las cuatro clases de familias de Cópulas que existen y para el caso bivariado tienen tres *familias asociadas*, se definen por las siguientes transformaciones (Salvadori & De Michele, 2004; Michiels & De Schepper, 2008; Chowdhary & Singh, 2019):

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

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

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

Las dos primeras transformaciones cambian la naturaleza de la dependencia de negativa a positiva y viceversa. La tercera transformación se denomina Cópula de supervivencia (*Survival Copula*), por estar relacionada con la distribución conjunta de las probabilidades de excedencia, del par de variables aleatorias u y v cuya Cópula es C . Además, las tres familias de Cópulas asociadas tienen el mismo valor del cociente tau de Kendall; las dos primeras con su signo cambiado.

Dupuis (2007) destacó la importancia de la Cópula de **Clayton Asociada** por tener dependencia significativa en su cola derecha definida por la Ecuación (17), y cuya fórmula de acuerdo con las ecuaciones (2) y (16) es la Ecuación (18):

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

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

Integración numérica

Para cuantificar varias integrales, por ejemplo, la Ecuación (6), se realizó una integración numérica con base en el método de la cuadratura Gauss-Legendre, cuya ecuación operativa es (Nieves & Domínguez, 1998; Campos-Aranda, 2003):

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

en la cual w_i son los coeficientes del método cuyas abscisas son h_i y np el número de parejas en donde se evalúa la función $f(x)$, con el argumento indicado en $f[\cdot]$. En Davis y Polonsky (1972) se obtuvieron las 12 parejas

utilizadas de w_i y h_i con 15 dígitos, que son aceptables en el lenguaje *Basic* como variables de doble precisión.

La ecuación operativa de la integración bivariada es la integral doble siguiente (Nieves & Domínguez, 1998):

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

En Davis y Polonsky (1972) se obtuvieron las 20 parejas utilizadas en esta integral doble de w_i y h_i con 15 dígitos.

Medidas de asociación

Concordancia

Como la FC caracteriza la *dependencia* entre las variables aleatorias u y v , es necesario el estudio de las medidas de asociación para disponer de un método que permita estimar su parámetro θ . En términos generales, una variable aleatoria es *concordante* con otra cuando sus grandes valores están asociados con los grandes valores de la otra y los valores pequeños de una con los valores reducidos de la otra (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019).

Unas variables con correlación lineal directa serán concordantes, pues al aumentar una la otra también lo hace. Variables con correlación lineal

inversa serán *discordantes*, ya que a grandes valores de una le corresponderán pequeños valores de la otra. Entonces, se dice que (x_1, y_1) y (x_2, y_2) son concordantes si $x_1 < x_2$ y $y_1 < y_2$, o bien $x_1 > x_2$ y $y_1 > y_2$. Son discordantes si $x_1 < x_2$ y $y_1 > y_2$, o bien $x_1 > x_2$ y $y_1 < y_2$. Lo anterior implica que son *concordantes* (c) las parejas $(x_1-x_2)(y_1-y_2) > 0$ y *discordantes* (d) cuando $(x_1-x_2)(y_1-y_2) < 0$ (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019).

Una medida numérica de asociación es un estadístico que indica el grado de dependencia o asociación de las variables. Para fines de comparación, tales medidas varían de 0 a +1 o a -1, indicando la asociación perfecta positiva en +1 o negativa en -1. El cociente tau de Kendall y el coeficiente rho de Spearman son dos medidas no paramétricas que proporcionan información sobre una forma especial de asociación o dependencia, conocida como *concordancia* (Salvadori *et al.*, 2007; Chen & Guo, 2019).

Cociente tau de Kendall

Mide la probabilidad de tener parejas concordantes, por lo cual es el *cociente* de $c-d$ entre $c+d$. Su expresión para estimarlo con datos bivariados es (Zhang & Singh, 2006; Zhang & Singh, 2019):

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

En la ecuación anterior, n es el número de observaciones y el signo $[\cdot]$ es +1 si tales parejas son concordantes y -1 si son discordantes.

Genest y Favre (2007) exponen una prueba para el cociente tau, aceptando que la hipótesis nula H_0 indica que X y Y son independientes y entonces el estadístico tiene distribución aproximadamente Normal con media 0 y varianza $2(2n+5)/[9n(n-1)]$. Entonces, H_0 será rechazada con un nivel de confianza $\alpha = 5\%$ si:

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

Coeficiente rho de Spearman

Mide la correlación entre parejas de rangos (R_i , S_i) de las variables aleatorias X_i y Y_i . Por ello equivale al coeficiente de correlación de Pearson. Su expresión para estimarlo en un registro bivariado de tamaño n es la siguiente (Chowdhary & Singh, 2019; Zhang & Singh, 2019):

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

Salvadori *et al.* (2007) presentan otra fórmula para su cálculo a partir de un registro bivariado con n datos, la cual coincide con la utilizada por la WMO (1971):

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

Genest y Favre (2007) también presentan una prueba similar para el coeficiente rho, cuya distribución se aproxima a la Normal con media 0 y varianza $1/(n-1)$; por lo cual se rechaza H_0 con un nivel $\alpha = 5 \%$, si:

$$\sqrt{n-1} \cdot |\rho_n| > 1.96 \quad (25)$$

Estimación del parámetro de dependencia

El método más simple para estimar el parámetro θ de las FC se asemeja al método de momentos y se basa en la inversión de la ecuación que relaciona a θ con el cociente tau de Kendall o con el coeficiente rho de Spearman. Los otros dos procedimientos disponibles se denominan: método de máxima pseudo-verosimilitud y método exacto de máxima verosimilitud (Meylan *et al.*, 2012; Chowdhary & Singh, 2019; Zhang & Singh, 2019; Chen & Guo, 2019). Para obtener θ , en las ecuaciones (5), (10) y (13) se procede por tanteos; en cambio, en las ecuaciones (3), (8) y (12) se despeja su valor.

Estimación de probabilidades empíricas conjuntas

Las probabilidades empíricas bivariadas se estimaron con base en la fórmula de Gringorten, aplicada por Yue (2000b), Zhang y Singh (2019), y Chen y Guo (2019). Tal fórmula es:

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

En la cual, i es el número de cada dato cuando están ordenados de manera progresiva, y n es el número total de ellos o amplitud del registro procesado. La expresión anterior se aplicó en el plano bidimensional, con los datos ordenados en forma progresiva; los gastos máximos (Q) en los renglones y los volúmenes (V) en las columnas. El plano formado es un cuadrado de n por n casillas, con n casillas en su diagonal principal, cuando el número de orden del renglón es igual al de la columna. Después, cada pareja de datos anual (Q y V) se localiza en el plano bidimensional, y la casilla definida por la intersección del renglón y columna se identifica con el número i , que corresponde al año histórico dibujado.

Cuando las n parejas de datos están dibujadas, se busca el año 1 y se define un área rectangular o cuadrada de valores menores de Q y de V , cuyo *conteo* de casillas numeradas dentro es NM_1 o combinaciones de Q y V menores. Calculados los n valores de NM_i , se aplica la fórmula de

posición gráfica de Gringorten para estimar la probabilidad empírica conjunta o bivariada:

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

Selección de la Función Cópula

Un enfoque simple de selección de la función Cópula se basa en los estadísticos del error de ajuste, al comparar las probabilidades empíricas observadas (w_i^o) con las teóricas calculadas (w_i^c) con la FC que se prueba. Los indicadores aplicados son el *error medio estándar (EME)*, el *error medio absoluto (EMA)* y el *error absoluto máximo (EAM)*; sus expresiones son (Chowdhary & Singh, 2019):

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

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

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

Dependencia en el extremo superior

Generalidades

El criterio más importante que se aplica para seleccionar una FC es el basado en la magnitud de la dependencia en el extremo superior de la distribución conjunta, lo cual tiene impacto en la *veracidad* de las predicciones extremas. La dependencia en la cola superior derecha (λ_U) es la probabilidad condicional de que Y sea mayor que un cierto percentil (s) de $F_Y(y)$, dado que X es mayor que tal percentil en $F_X(x)$, conforme s se aproxima a la unidad. La dependencia en la cola inferior izquierda (λ_L) compara que Y sea menor que X , cuando s se aproxima a 0 (Chowdhary & Singh, 2019; Salvadori *et al.*, 2007).

En relación con las FC expuestas, las de Frank y Plackett tienen dependencias insignificantes en sus zonas extremas, por ello $\lambda_L = 0$ y $\lambda_U = 0$. Por el contrario, la Cópula Gumbel-Hougaard tiene dependencia significativa en la cola superior igual a:

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

Las Cópulas de Clayton y Raftery la tienen en su cola inferior e igual a:

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

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

Como ya se indicó, la Cópula de Clayton Asociada tiene $\lambda_U = 2^{-1/\theta}$ (ecuación (17)). Familias de Cópulas no expuestas, con $\lambda_U > 0$ son las de Galambos, Genest-Ghoudi y Hüsler-Reiss (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019).

Dupuis (2007) probó seis familias de Cópulas y encontró que su capacidad para estimar eventos extremos varía de mala a buena, con el siguiente orden: Clayton, Frank, Normal, *t*-Student, Gumbel-Hougaard y de Clayton Asociada (*Survival Clayton*). A conclusiones similares llegan Poulin *et al.* (2007), al comparar las mismas seis familias de Cópulas y la denominada A12, que tiene dependencia significativa en su cola derecha.

Estimación de la dependencia observada

Para abordar la estimación de la dependencia en la cola superior (λ_U) que muestran los datos disponibles, se debe definir primero la llamada *Cópula Empírica*. Como la FC caracteriza la dependencia entre las variables aleatorias X y Y , entonces el par de rangos R_i y S_i procedentes de las tales variables son el estadístico, que retiene la mayor cantidad de información y su escalamiento con el factor $1/(n+1)$ genera una serie de puntos en el cuadrado unitario $[0, 1]^2$, formando el dominio de la Cópula Empírica (Chowdhary & Singh, 2019), definida así:

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

En la ecuación anterior, $1(\cdot)$ indica una función de las variables aleatorias U y V , que son una transformación en aumento continuo de X y Y en relación con las integrales de probabilidad empírica $F_n(X)$ y $F_n(Y)$, cuyas ecuaciones son:

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

Poulin *et al.* (2007) utilizan el estimador propuesto por Frahm, Junker y Schmidt (2005), el cual se basa en una muestra aleatoria obtenida de la Cópula empírica; su expresión es:

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

Este estimador acepta que la FC se puede aproximar por una de la clase de valores extremos; tiene la ventaja de no requerir un valor umbral para su estimación, como los cuatro expuestos por AghaKouchak, Sellars y Sorooshian (2013).

Familias de Cópulas por contrastar

Justificación

Desde los inicios de las aplicaciones de las *funciones Cópula* (FC) en los análisis hidrológicos multivariados, Favre *et al.* (2004) formularon en sus conclusiones que la etapa crucial de tales análisis radica en la selección adecuada de la FC a los datos disponibles. Posteriormente, Requena *et al.* (2013) destacaron que una FC que muestra un buen ajuste al total de datos disponibles no garantiza una representación adecuada de los valores extremos de tales datos y, por ello, hay que estimar la dependencia observada λ_{ij}^{CFG} (ecuación (36)), para buscar una FC que la reproduzca.

Por lo anterior, cuando se requieren estimaciones de periodos de retorno conjuntos elevados, superiores a tres veces la amplitud del registro conjunto disponible de gastos y volúmenes máximos anuales, se debe buscar y seleccionar la FC que mejor reproduce la dependencia en la cola derecha de los datos (λ_{ij}^{CFG}). En este estudio se contrastan cuatro familias de Cópulas, cuya dependencia en su extremo derecho varía de *mayor a menor grado*, las cuales son las siguientes: Joe, Clayton Asociada, Gumbel-Hougaard y *t* de Student. Las tres primeras pertenecen a la clase de las Cópulas de Arquímedes y la última es de la clase elíptica.

Cópula de Joe

La ecuación básica de esta familia de FC, que sólo acepta dependencia positiva, es la siguiente (Joe, 1993; Salvadori *et al.*, 2007; Chowdhary & Singh, 2019):

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

cuyo parámetro de dependencia es $\theta \geq 1$, con $\theta = 1$, para el caso de independencia entre u y v . Como no tiene una expresión que relacione su parámetro θ con el cociente tau de Kendall, éste se estima en función de su ecuación del generador $\phi(s)$ y de su derivada $\phi'(s)$ de la FC, cuya expresión es (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019):

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

en la cual s es la variable aleatoria unitaria $0 < s \leq 1$. Las ecuaciones de su generador y de la derivada de éste son (Salvadori *et al.*, 2007; Zhang & Singh, 2019):

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

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

Teniendo como dato el valor de cociente tau de Kendall, la Ecuación (38) se integró numéricamente con base en la Ecuación (19), para obtener por tanteos el valor de θ que la cumple. Obtenido el valor de θ , la dependencia significativa en el extremo derecho (λ_U) de la Cópula de Joe, se obtiene por medio de la Ecuación (31) de la Cópula de Gumbel-Hougaard.

Cópula de Clayton Asociada

La Cópula definida con la Ecuación (18) será aplicada y contrastada debido a que Dupuis (2007) la expuso y mostró que tiene una fuerte dependencia significativa en su extremo derecho (λ_U).

Cópula t de Student

Demarta y McNeil (2005) indican que las Cópulas de Gumbel-Hougaard y Clayton Asociada son de aplicación más simple que las Cópulas extremas derivadas de la t -Student, pero esta última define un valor inferior de la dependencia, como lo señaló previamente Dupuis (2007) y lo ratificó Poulin *et al.* (2007). La ecuación de la Cópula t de Student es la siguiente (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019):

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

En la expresión anterior $v > 2$, θ varía entre -1 y 1, con $\theta = 0$ para la independencia entre u y v ; $t_v^{-1}(\cdot)$ es la inversa de la distribución t de Student univariada, con v grados de libertad, para la probabilidad indicada en el paréntesis.

Entonces, con la idea de acotar la dependencia en la cola derecha en un valor inferior se exponen las ecuaciones de la Cópula t -Student (Demarta & McNeil, 2005; Poulin *et al.*, 2007; Salvadori *et al.*, 2007; Genest, Favre, Béliveau, & Jacques, 2007):

$$\tau_n = \frac{2}{\pi} \arcsen \theta \quad (42)$$

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

El signo menos dentro del paréntesis indica que tal valor es la probabilidad de excedencia (*survival probability*) que debe definir la variable aleatoria con distribución t de Student y $v + 1$ grados de libertad. Para evitar el uso de algoritmos numéricos que estimen el valor de t_{v+1} en función de la probabilidad de excedencia, Demarta y McNeil (2005) exponen la siguiente fórmula alternativa:

$$\lambda = \frac{\int_{(\pi/2 - \arcsen \theta)/2}^{\pi/2} \cos^{\nu} s \cdot ds}{\int_0^{\pi/2} \cos^{\nu} s \cdot ds} \quad (44)$$

En la ecuación anterior, λ no lleva subíndice debido a que la Cópula t de Student tiene igual dependencia significativa en sus dos colas o extremos. La Ecuación (44) se integró numéricamente con base en la Ecuación (19) y con tales resultados se elaboró la Tabla 1, que es una ampliación de la expuesta por Demarta y McNeil (2005).

Tabla 1. Valor de la dependencia en los extremos de la Cópula t de Student para diversos valores del ν y θ .

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

ν = grados de libertad.

En la Tabla 1 se destaca que incluso para valores negativos y correlación 0, la Cópula t de Student tiene dependencia asintótica en sus extremos y que tal dependencia decrece con el aumento de los grados de libertad.

La aplicación o integración de la Cópula t de Student (Ecuación (41)) se realizó con base en la Ecuación (20), con 20 parejas de coeficientes w_i y abscisas h_i de 15 dígitos tomados de Davis y Polonsky (1972). Para la estimación de los límites superiores de las integrales se aplicó el algoritmo expuesto por Zelen y Severo (1972) para estimar la variable aleatoria t de Student con ν grados de libertad, en función de la variable Normal x_p , en la cual p es la probabilidad de excedencia:

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

siendo:

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

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

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

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

Cuando $p < 0.50$ hacer $t = -t$. El algoritmo para estimación de la variable Normal x_p también procede de Zelen y Severo (1972) y es el siguiente:



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

en la cual:

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

siendo:

$$c_0 = 2.515517$$

$$c_1 = 0.802853$$

$$c_2 = 0.010328$$

$$d_1 = 1.432788$$

$$d_2 = 0.189269$$

$$d_3 = 0.001308$$

Si $p > 0.50$ hacer $p = 1 - p$ y $x_p = -x_p$. Los algoritmos definidos por las ecuaciones (45) a (51) fueron probados con los valores numéricos que exponen Zhang y Singh (2019).

Para establecer los límites inferiores de las integrales de la Ecuación (41) se utilizó la tabulación de la distribución t de Student, expuesta por Ostle y Mensing (1975) como Apéndice 5. En la Tabla 1 se definen para una probabilidad de no excedencia de 0.9995 los siguientes valores de t con

ν grados de libertad de 2 a 10: 31.598, 12.941, 8.610, 6.859, 5.959, 5.405, 5.041, 4.781 y 4.587. Tales valores con signo menos se tomaron como límites inferiores según el grado de libertad asignado en la aplicación de la Ecuación (41), el cual, lógicamente, varió de 2 a 10; seleccionado la Cópula t de Student de mejor ajuste (ecuaciones (28) a (30)).

Ratificación de la función Cópula seleccionada

Esta es la etapa más importante del proceso de aplicación práctica de las FC, pues en ella se verifica que tal modelo reproduzca fielmente las probabilidades conjuntas observadas (Ecuación (27)). Yue (2000a) indica que la forma simple y práctica de representar las probabilidades conjuntas empíricas y teóricas consiste en llevar al eje de las abscisas la primera y al eje de las ordenadas la segunda; lógicamente, en tal gráfica, cada pareja de datos define un punto que coincide o se aleja de la recta a 45°. La inspección de la gráfica descrita y el valor del coeficiente de correlación, en estos casos, superior a 0.98, ratifican la validez del modelo probabilístico conjunto.

Yue (2000b), y Yue y Rasmussen (2002) aplican el test de Kolmogorov-Smirnov, con un nivel de significancia (α) del 5 %, para aceptar o rechazar la *diferencia máxima absoluta* (d_{ma}) entre las probabilidades conjuntas. Para evaluar la estadística (D_n) del test se utilizó la expresión que exponen Meylan *et al.* (2012), para $\alpha = 5$ %:

$$D_n = \frac{1.358}{\sqrt{n}} \quad (52)$$

n es el número de datos. Si la dma es menor que D_n se ratifica la FC adoptada.

Conceptos del análisis de frecuencias bivariado

Periodos de retorno bivariados

El primer *periodo de retorno bivariado* del evento (X, Y) se define bajo la condición OR, lo cual indica que los límites x o y , o ambos *pueden* ser excedidos, y entonces la ecuación clásica del periodo de retorno o inverso de la probabilidad de excedencia será (Shiau *et al.*, 2006; Genest & Chebana, 2017):

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

en la cual $C[F_X(x), F_Y(y)]$ es la FC seleccionada.

El segundo *periodo de retorno bivariado* del evento (X, Y) está asociado con el caso en que ambos límites *son* excedidos ($X > x, Y > y$) o condición AND; su ecuación es (Shiau *et al.*, 2006; Genest & Chebana, 2017):

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

Aldama (2000) obtiene la expresión $F'_{X,Y}(x,y)$ de la probabilidad bivariada de excedencia mediante un razonamiento de probabilidades lógico y simple aplicado en el plano cartesiano. En cambio, Yue y Rasmussen (2002) recurren al plano cartesiano para definir el evento bivariado (X, Y) conceptualmente, que puede ocurrir en alguno de los cuatro cuadrantes.

La relación entre los periodos de retorno bivariados y los univariados es la siguiente (Yue & Rasmussen, 2002; Shiau *et al.*, 2006; Vogel & Castellarin, 2017):

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

siendo:

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

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

Salvadori y De Michele (2004) introducen con detalle el concepto del *periodo de retorno secundario* (ζ), designado así para enfatizar que el

periodo de retorno conjunto T_{XY} es el primario, de donde procede al utilizar las *isolíneas* que define la FC aplicada, cuya expresión es:

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

siendo s la variable aleatoria unitaria $0 < s \leq 1$ y C la FC probada. Después se define una región $B_C(s)$ en el espacio unitario (\mathbb{I}^2) sobre la isolínea, debajo de ella y a la izquierda, que será:

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

En FC de la clase de Arquímedes, la distribución univariada de Kendall, designada $K_C(s)$, proporciona una medida de los eventos dentro de la $B_C(s)$; su ecuación es (Salvadori & De Michele, 2004; Salvadori & De Michele, 2007; Salvadori *et al.*, 2007; Gräler *et al.*, 2013):

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

en la cual $\phi(s)$ es el generador de la FC y $\phi'(s)$ su derivada. Por último, el periodo de retorno secundario (ζ) de los eventos fuera de $B_C(s)$ es:

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

cuyo denominador es la probabilidad de excedencia (*survival function*), la cual corresponde a eventos posiblemente destructivos o peligrosos. La relación que guarda ζ con los periodos de retorno *conjuntos* es la siguiente (Gräler *et al.*, 2013):

$$T_{XY} \leq \zeta \leq T'_{XY} \quad (62)$$

La aplicación de la Ecuación (61) requiere la estimación previa de la Ecuación (60) para las siguientes seis probabilidades de no excedencia (s): 0.98, 0.99, 0.998, 0.999, 0.9998 y 0.9999, que corresponden a los *periodos de retorno conjuntos* de 50, 100, 500, 1 000, 5 000 y 10 000 años. En las Cópulas de Gumbel-Hougaard y de Joe, las ecuaciones de su generador y su derivada son las siguientes (Zhang y Singh, 2019):

$$\phi(s) = (-\ln s)^\theta \quad (63)$$

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

$$\phi(s) = -\ln[1 - (1 - s)^\theta] \quad (65)$$

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

Eventos críticos o de diseño

Volpi y Fiori (2012) destacan que la gráfica del periodo de retorno conjunto de tipo AND, mostrada más adelante como Figura 2, presenta una severa inconsistencia al contener, en un contexto bivariado, umbrales críticos univariados. Debido a lo anterior, tal gráfica se considera integrada por dos porciones: las dos designadas *simples* (*naive part*) y la *correcta* (*proper part*). Las partes rectas son las colas o rectas asintotas a la parte curva. La probabilidad de ocurrencia de un evento o pareja de Q y V es variable en la parte curva y decrece a lo largo de la parte recta, aunque todos los valores definen el *mismo* periodo de retorno conjunto. En resumen, las parejas de valores de las rectas asintotas tienen probabilidades de ocurrencia bajas y por ello no deben ser incluidas en los análisis de búsqueda de las crecientes (Q y V) críticas o severas. Para fines prácticos, los puntos extremos de la parte curva se pueden definir según su distribución empírica o cercanos al inicio de las rectas asintotas.

Selección de las distribuciones marginales

El planteamiento para la selección de las distribuciones marginales fue muy simple y consistió en aplicar las tres FDP que se han establecido de referencia o de aplicación bajo precepto, que son la Log-Pearson tipo III (LP3), la General de Valores Extremos (GVE) y la Logística Generalizada (LOG). Además, se aplicaron tres distribuciones de uso generalizado: la

Pareto Generalizada (PAG), la Kappa y la Wakeby. Las primeras cuatro FDP citadas tienen tres parámetros de ajuste, y las dos últimas, cuatro y cinco. Con excepción de la LP3 que se aplicó con el método de momentos en los dominios logarítmico (WRC, 1977) y real (Bobée, 1975), el resto se ajustó con el método de momentos L (Hosking & Wallis, 1997).

La selección de la FDP más conveniente se basó en el valor que cada una genera con una probabilidad de no excedencia del 1 %; es decir, un valor muy bajo que no supere a los valores mínimos observados para evitar probabilidades marginales (u , v) negativas. Este criterio de selección es el requerido en registros de Q y V que presentan valores mínimos muy reducidos, en comparación con sus extremos máximos.

Además, se tomaron en cuenta a los errores estándar de ajuste (EEA) y absoluto medio (EAM), así como la magnitud de las *predicciones* en los periodos de retorno mayores a 500 años.

Errores de ajuste

El primer criterio aplicado para la selección de la mejor FDP a unos datos o serie disponible fueron los llamados *errores de ajuste* (Kite, 1977; Willmott & Matsuura, 2005; Chai & Draxler, 2014). Este criterio y el descrito para evitar probabilidades negativas permitirán adoptar una distribución adecuada entre los modelos: LP3, GVE, LOG, PAG, Kappa y Wakeby.

Cambiando en las ecuaciones (28) y (29) las probabilidades observadas por los datos ordenados de la serie analizada (x_i o y_i), y las probabilidades

calculadas por los valores estimados con la FDP que se prueba o contrasta, se obtienen el error estándar de ajuste (EEA) y el error absoluto medio (EAM). Los valores que se estiman (\hat{x}_i o \hat{y}_i) se buscan para la misma probabilidad de no excedencia, asignada a los datos por la fórmula empírica de Gringorten (Ecuación (26)).

Datos por procesar

Para realizar el contraste de las cuatro FC que tienen dependencia significativa en su extremo derecho (λ_U), se procesaron los 16 registros *conjuntos* de gasto máximo (Q) y volumen (V) de las crecientes anuales que expusieron Aldama *et al.* (2006), los cuales suman 620 parejas de datos en 15 embalses importantes de México y uno en proyecto (La Parota).

Además, se incluyó el registro conjunto de Q y V anuales de las crecientes registradas en la estación hidrométrica La Cuña, expuesto por Gómez *et al.* (2010) y Campos-Aranda (2022). También se procesó el registro de Q y V de las crecientes de entrada a la presa Malpaso, mostrado por Domínguez y Arganis (2012).

En la Tabla 2 se expone el registro de los 26 datos conjuntos de Q y V de las crecientes de entrada a la presa Adolfo López Mateos en el río Humaya del estado de Sinaloa, México. Tal registro procede de Aldama *et al.* (2006) y se procesó como aplicación numérica.

Tabla 2. Gastos máximos, volúmenes y sus números de orden bivariados en las 26 crecientes anuales de entrada a la presa Adolfo López Mateos (*Humaya*), México (*Aldama et al.*, 2006).

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

Test de Wald-Wolfowitz

Esta prueba no paramétrica ha sido utilizada por Bobée y Ashkar (1991), Rao y Hamed (2000), y Meylan *et al.* (2012) para verificar *independencia* y *estacionariedad* en registros de gastos máximos anuales (X_i). Por lo anterior, se propuso aplicar esta prueba a los registros conjuntos de Q y V , que deben ser muestras *aleatorias*.

Resultados y su discusión

Contraste de funciones Cópula con $\lambda_U > 0$

Verificación de la aleatoriedad

En la Tabla 3 se muestran en orden progresivo de tamaño de cuenca (A en km^2) los 16 registros *conjuntos* de Q y V procesados, tomados de Aldama *et al.* (2006) y, al final, los dos registros conjuntos también procesados. En la columna 4 se expone el número de datos (n) de cada registro, y en las columnas 5 y 6, el valor del estadístico U del test de Wald-Wolfowitz, cuyo valor absoluto debe ser menor de 1.96 para que tal registro sea *aleatorio*.

Tabla 3. Datos generales y resultados del test de Wald-Wolfowitz en los 18 registros conjuntos de Q y V procesados.

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

Se deduce de los valores citados que dos registros conjuntos no son aleatorios: los de las presas *Guamúchil* y Abelardo L. Rodríguez. A tales registros se les aplicaron pruebas estadísticas específicas y se encontró en la presa *Guamúchil* que las series de Q y V muestran persistencia y exceso de variabilidad por sus valores extremos. Además, la serie de Q muestra tendencia ascendente. En la presa Abelardo L. Rodríguez, ambas series de Q y V muestran persistencia, exceso de variabilidad y tendencia negativa. El registro de Q de la presa *El Sabino* puede ser aceptado aleatorio, ya que sólo presenta persistencia ligera. Por lo anterior, los registros conjuntos de las presas *Guamúchil* y Abelardo L. Rodríguez se eliminaron de los análisis posteriores.

Valores de las medidas de asociación

En la columna 2 de la Tabla 4 se muestran los valores del cociente tau de Kendall obtenidos con la Ecuación (21). Respecto al coeficiente rho de Spearman, mostrado en la columna 2 de la Tabla 5, los valores que no tienen paréntesis resultaron iguales al calcularse con las ecuaciones (23) y (24). En cambio, las magnitudes entre paréntesis provienen de la Ecuación (23), que conduce a valores ligeramente mayores.

Tabla 4. Valores del cociente tau de Kendall de la dependencia observada (λ_U^{CFG}), y de las Cópulas aplicadas de Joe y de Clayton Asociada al registro conjunto indicado.

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

Tabla 5. Valores del coeficiente rho de Spearman, de la dependencia observada (λ_U^{CFG}), y de las Cópulas de Gumbel-Hougaard y *t* de Student en el registro conjunto indicado.

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

Contraste de dependencias (λ_U)

En las columnas 3 de la Tabla 4 y Tabla 5 se exponen los valores de la dependencia observada (λ_U^{CFG}) estimada con la Ecuación (36). El criterio de selección de la FC fue que su dependencia significativa (λ_U) debería ser algo mayor o escasamente menor a la dependencia observada (λ_U^{CFG}). Con tal criterio, en la Tabla 4 se definieron las FC de Joe para el registro de la presa Adolfo López Mateos (*Humaya*) y la FC de Clayton Asociada para el registro de la presa Plutarco Elías Calles (*El Novillo*).

Por otra parte, en la Tabla 5, con el criterio de selección descrito, se definen 10 registros, cuya dependencia en su extremo derecho puede ser presentada por la FC de Gumbel-Hougaard (GH), y tres que están comprendidos entre las FC de Clayton Asociada y de GH, cuyos valores de la dependencia (λ_U) se han mostrado entre paréntesis. La FC de la *t* de Student reproduce la dependencia del registro de la presa Malpaso, con $\nu = 2$.

Búsqueda de las distribuciones marginales

Distribución de los gastos máximos anuales

En la Tabla 6 se exponen los errores de ajuste y predicciones (m^3/s) obtenidas con las seis distribuciones aplicadas al registro de gastos máximos de la Tabla 2, cuyo valor mínimo registrado es $127 m^3/s$. De estas seis FDP, las únicas que definen valores con probabilidad de no

excedencia del 1 % menores al gasto citado son la GVE y la LOG. Tomando en cuenta los errores de ajuste, se adopta la GVE.

Tabla 6. Errores de ajuste y predicciones (m³/s) de las seis distribuciones aplicadas en el registro de gastos máximos anuales de las crecientes de entrada a la presa Adolfo López Mateos (*Humaya*), México.

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

Los parámetros de ubicación (u_1), escala (a_1) y forma (k_1) de la distribución GVE adoptada son 447.7865, 372.8246 y -0.6236228, respectivamente, cuya ecuación es:

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

Distribución de los volúmenes anuales

En la Tabla 7, similar a la anterior para los volúmenes anuales de la Tabla 2, cuyo valor mínimo es 50 Mm³, nuevamente las únicas FDP que no generan probabilidades marginales negativas son la GVE y LOG. Se adopta la GVE por mostrar menores errores de ajuste.

Tabla 7. Errores de ajuste y predicciones (m³/s) de las seis distribuciones aplicadas al registro de volúmenes anuales de las crecientes de entrada a la presa Adolfo López Mateos (*Humaya*), México.

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

Los parámetros de ubicación (u_2), escala (a_2) y forma (k_2) de la distribución GVE adoptada son 193.4795, 167.2692 y -0.6288097, respectivamente, cuya expresión es:

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

Selección y ratificación de la FC

El procesamiento bivariado de los datos de la Tabla 2 condujo a los siguientes tres indicadores de asociación: $r_{xy} = 0.9670$, $\tau_n = 0.4523$ y $\rho_n = 0.6738$. La Ecuación (22), con $n = 26$ y el tau citado origina un valor de 3.24; por lo cual, el cociente de Kendall es significativo. La Ecuación (25) genera un valor de 2.26 y entonces el coeficiente de Spearman también es significativo.

Por otra parte, en la Tabla 8 se muestran los indicadores estadísticos de ajuste que se obtuvieron al aplicar las FC de Clayton, Frank, Plackett, *t*-Student, Gumbel-Hougaard (GH), Clayton Asociada y Joe. En las ecuaciones (28) a (30), las probabilidades bivariadas empíricas se estimaron con la Ecuación (27) y las teóricas con las ecuaciones (2), (4), (9), (41), (7), (18) y (37). Los valores reportados para la Cópula *t*-Student corresponden a un $\nu = 6$, donde ya estabilizaron los errores de ajuste con $\theta = \text{sen}\left(\frac{\pi \cdot \tau_n}{2}\right) = 0.6522$, según la Ecuación (42).

Tabla 8. Indicadores estadísticos del ajuste de las funciones Cópula indicadas en las crecientes anuales de entrada a la presa Adolfo López Mateos (*Humaya*), México.

Cópula	θ	<i>EME</i>	<i>EAM</i>	Núm. DP	Núm. DN	MDP	MDN	λ_U
Clayton (C.)	1.6517	0.0529	0.0365	14	12	0.1505	-0.0779	0.0000
Frank	4.9310	0.0461	0.0321	15	11	0.1334	-0.0701	0.0000
Plackett	11.079	<i>0.0441</i>	<i>0.0315</i>	11	15	0.1291	-0.0727	0.0000
<i>t</i> -Student	0.6522	0.0482	0.0347	13	13	0.1415	-0.0643	0.2641
GH	1.8258	0.0473	0.0354	14	12	0.1352	-0.0601	0.5383
C. Asociada	1.6517	0.0475	0.0362	14	12	0.1277	-0.0549	0.6573
Joe	2.5220	0.0477	0.0365	15	11	<i>0.1267</i>	<i>-0.0533</i>	0.6837

Significado de los nuevos acrónimos:

DP, DN = diferencias positivas y negativas.

MDP, MDN = máxima diferencia positiva y negativa.

Con base en los resultados de la Tabla 8, no existe dificultad para seleccionar como mejor FC la de Plackett, para los datos de la Tabla 2, debido a que reporta los indicadores estadísticos (*EME* y *EAM*) más reducidos (mostrados en cursivas).

Sin embargo, al tomar en cuenta que la dependencia observada (λ_U^{CFG}) alcanzó un valor de 0.6834, la FC que se debe seleccionar es la Joe, como ya se expuso en la Tabla 4. En la Tabla 8 se observa que el ajuste de la FC de Joe a los datos es mejor que el peor de todos, que lo definió FC de Clayton. Es destacable que la FC de Joe origina los valores más bajos de las diferencias máximas entre probabilidades empíricas y teóricas (mostradas en cursivas y sombreadas).

En la Tabla 9 se muestran las probabilidades de no excedencia bivariadas, empíricas observadas (w_i^o) y teóricas calculadas (w_i^c) con la FC de Joe. También se indican sombreadas las diferencias máximas positiva y negativa.

Tabla 9. Probabilidades de no excedencia conjuntas y sus diferencias calculadas con la FC de Joe para las crecientes de anuales de entrada a la presa Adolfo López Mateos (*Humaya*), México.

Núm.	w_i^o	w_i^c	Diferencias	Núm.	w_i^o	w_i^c	Diferencias
1	0.1363	0.1896	-0.0533	14	0.2129	0.1619	0.0509
2	0.1363	0.1300	0.0063	15	0.0980	0.1051	-0.0071
3	0.0980	0.0311	0.0669	16	0.2894	0.2392	0.0502
4	0.0214	0.0101	0.0114	17	0.8254	0.8643	-0.0389
5	0.1363	0.1378	-0.0015	18	0.3277	0.2369	0.0909
6	0.5574	0.5663	-0.0089	19	0.0980	0.1215	-0.0234
7	0.2511	0.2249	0.0263	20	0.2511	0.1955	0.0556
8	0.9786	0.9851	-0.0066	21	0.2129	0.2460	-0.0331
9	0.8637	0.8939	-0.0302	22	0.0214	0.0051	0.0163
10	0.0980	0.0973	0.0007	23	0.9020	0.9195	-0.0175
11	0.3277	0.3210	0.0067	24	0.3277	0.2666	0.0612
12	0.5957	0.4690	0.1267	25	0.8254	0.8767	-0.0513
13	0.7489	0.6690	0.0798	26	0.6340	0.6058	0.0282

Por otra parte, la Ecuación (52) define a $D_n = 0.2663$ y como la diferencia máxima absoluta de la Tabla 9 es 0.1267, el test de Kolmogorov-Smirnov ratifica a la FC de Joe adoptada. El coeficiente de correlación (r_{xy}) entre las probabilidades empíricas y las teóricas estimadas con la FC de Joe

resultó de 0.9897; por lo tanto, aceptable. El contraste gráfico entre ambas probabilidades para ratificar su adopción se expone en la Figura 1 para los datos de la Tabla 9.

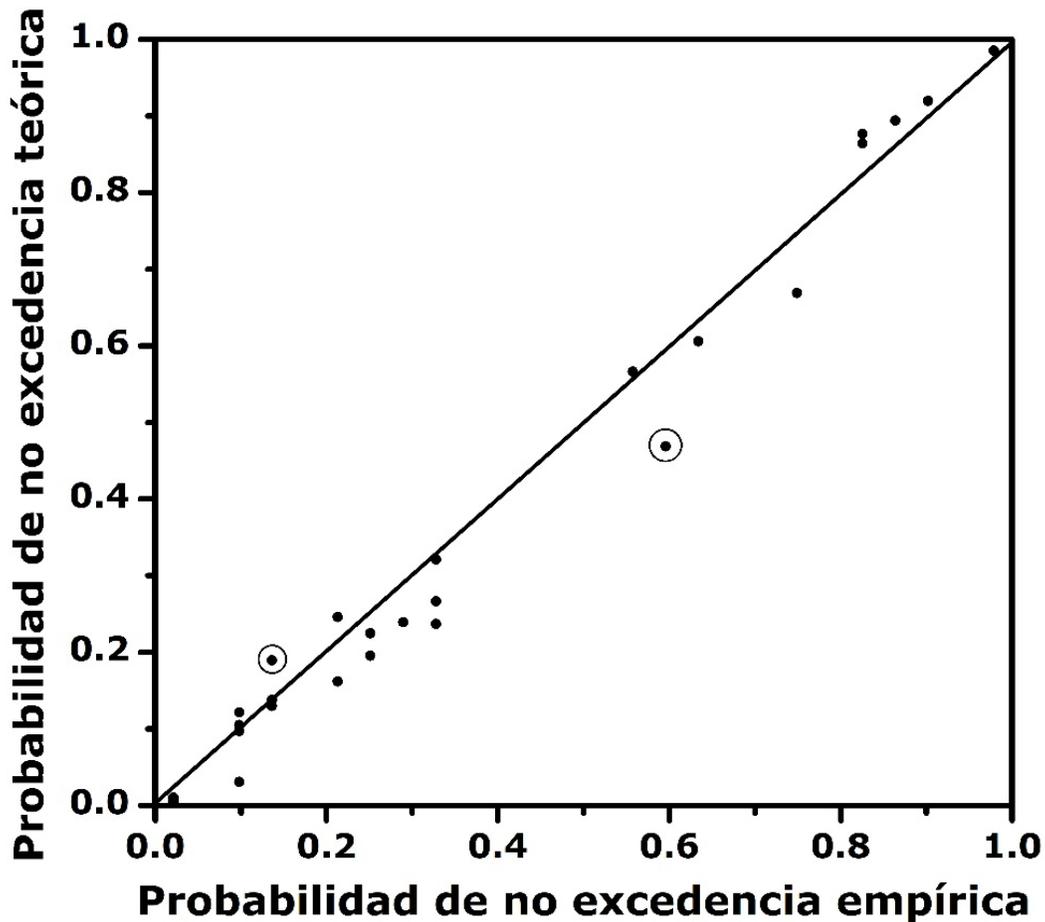


Figura 1. Contraste gráfico de probabilidades conjuntas (Q , V) calculadas con la FC de Joe para las crecientes anuales de entrada a la presa Adolfo López Mateos (*Humaya*), México.

Gráficas del periodo de retorno bivariado T'_{XY}

Los periodos de retorno bivariados de tipo AND se estiman con la Ecuación (54). Para sus estimaciones con valores del T'_{XY} de 500, 1 000, 5 000 y 10 000 años se aplicó la FC de Joe. De manera arbitraria se seleccionan volúmenes y gastos pico para obtener sus probabilidades de no excedencia marginales (ecuaciones (67) y (68)) y conjunta (Ecuación (54)). En la Tabla 10 se muestran resultados para definir las cuatro gráficas de la Figura 2.

Tabla 10. Parejas de gasto pico y volumen anual utilizadas para definir las gráficas del periodo de retorno conjunto tipo AND con la FC de Joe en las crecientes de entrada a la presa Adolfo López Mateos (*Humaya*), México.

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

T'_{XY} 500 años		T'_{XY} 1 000 años		T'_{XY} 5 000 años		T'_{XY} 10 000 años	
V Mm ³	Q m ³ /s	V Mm ³	Q m ³ /s	V Mm ³	Q m ³ /s	V Mm ³	Q m ³ /s
8 000	26 000	13 000	39 560	28 000	114 700	35 000	181 120
9 000	24 858	14 000	38 348	32 000	111 900	40 000	178 750
10 000	23 266	15 000	36 850	36 000	108 140	45 000	175 800
10 500	22 222	16 000	34 960	40 000	103 060	50 000	172 050
11 000	20 940	17 000	32 520	42 000	99 850	52 500	169 960
11 500	19 315	18 000	29 212	44 000	96 060	55 000	167 520
12 000	17 165	18 500	27 050	46 000	91 430	57 500	164 740
12 500	14 046	19 000	24 335	47 000	88 770	60 000	161 570
12 750	11 745	19 500	20 720	48 000	85 800	62 500	157 970
13 000	8 125	19 750	18 330	49 000	82 460	65 000	153 850
13 050	6 990	20 000	15 170	50 000	78 670	67 500	149 310
13 100	5 490	20 100	13 530	51 000	74 250	70 000	143 810
13 163	0	20 200	11 520	52 000	69 140	72 000	138 820
		20 300	8 730	53 000	62 840	74 000	133 200
		20 350	6 450	53 500	59 250	76 000	126 530
		20 401	0	54 000	54 985	78 000	118 720
				54 500	50 160	80 000	109 000
				55 000	44 100	81 000	103 580
				55 500	35 840	82 000	96 950
				56 000	23 600	83 000	89 650
				56 269	0	84 000	80 650

T'_{XY} 500 años		T'_{XY} 1 000 años		T'_{XY} 5 000 años		T'_{XY} 10 000 años	
V Mm ³	Q m ³ /s	V Mm ³	Q m ³ /s	V Mm ³	Q m ³ /s	V Mm ³	Q m ³ /s
						85 000	68 990
						86 000	53 500
						87 019	0

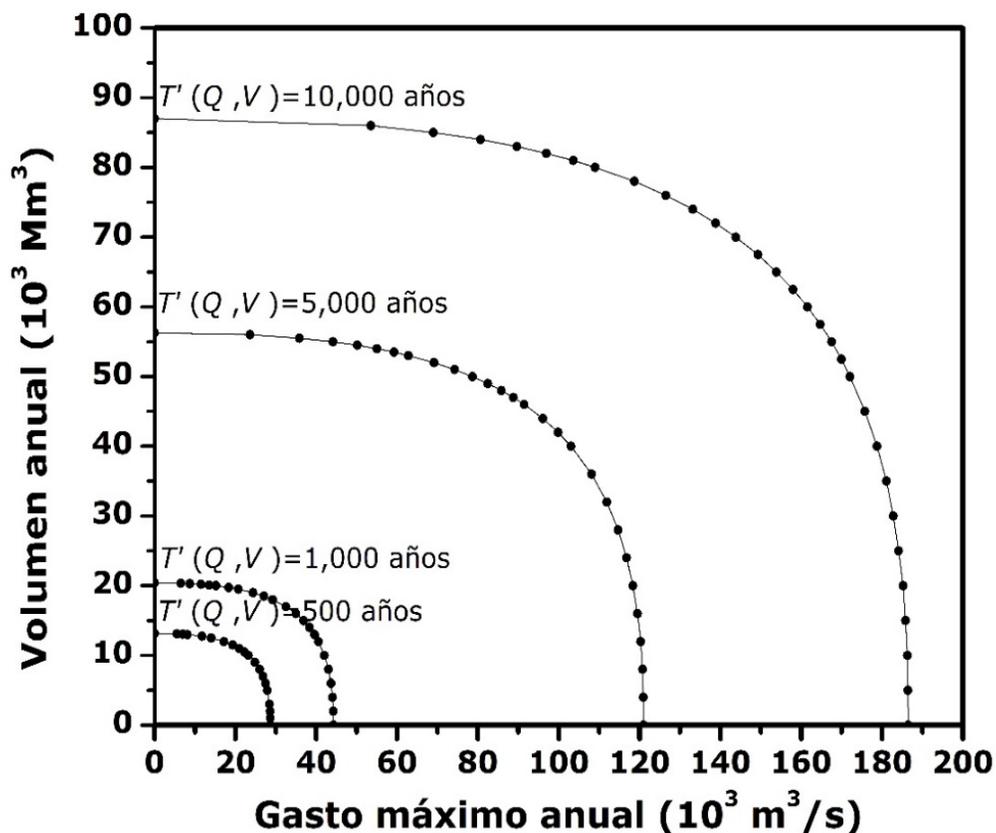


Figura 2. Gráficas de los cuatro periodos de retorno conjunto T'_{XY} de diseño obtenidas con la FC de Joe en las crecientes anuales de entrada a la Presa Adolfo López Mateos (*Humaya*), México.

Selección de los eventos de diseño

En la Figura 2 o en la Tabla 10 se pueden seleccionar infinitas parejas de Q y V , que satisfacen el periodo de retorno conjunto de diseño y que se definen como *subgrupo de parejas críticas* por estar dentro de la porción curva de cada gráfica del T'_{XY} fuera de las rectas asíntotas (Volpi & Fiori, 2012).

Las combinaciones de gasto pico y volumen que tienen el mismo *periodo de retorno bivariado* de diseño establecen crecientes o *hidrogramas* que producirán diferentes efectos en el embalse que se diseña o revisa, adoptando por seguridad, el que genera las condiciones más críticas. Para formar cada hidrograma de diseño existen métodos teóricos y empíricos (Aldama, 2000; Aldama *et al.*, 2006; Campos-Aranda, 2008; Gómez *et al.*, 2010; Gräler *et al.*, 2013).

Contraste de los periodos de retorno conjuntos

En la Tabla 11 y Tabla 12 se muestran los cálculos realizados para llevar a cabo la verificación de las ecuaciones (55) y (62) con base en las FC de Gumbel-Hougaard y de Joe. Se observa y comprueba en todos los casos que ambas FC cumplen con la Ecuación (55). Para el periodo de retorno secundario, la Ecuación (62) se cumple exclusivamente con la FC de Joe en los tres últimos periodos de retorno analizados.

Tabla 11. Periodos de retorno univariados y bivariados estimados con la FC de Gumbel-Hougaard en las crecientes de entrada a la presa Adolfo López Mateos (*Humaya*), México.

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

Tabla 12. Periodos de retorno univariados y bivariados estimados con la FC de Joe en las crecientes de entrada a la presa Adolfo López Mateos (*Humaya*), México.

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

Conclusiones

En México, los seis trabajos citados: Aldama (2000), Ramírez-Orozco y Aldama (2000), Escalante-Sandoval y Reyes-Chávez (2002), Aldama *et al.* (2006), Gómez *et al.* (2010), y Domínguez y Arganis (2012) ponen de manifiesto los esfuerzos realizados para desarrollar *métodos confiables de estimación de crecientes*, que brinden seguridad a las obras hidráulicas rurales y urbanas que son dimensionadas y/o revisadas con ellas, al buscar estimar crecientes bivariadas lo más aproximadas posible; este estudio se suma a los esfuerzos citados.

La ventaja fundamental de usar las funciones Cópula (FC) en el análisis de frecuencias de crecientes *bivariado* (AFCb) radica en construir fácilmente la distribución de probabilidades *conjunta* con base en las distribuciones univariadas marginales iguales o diferentes adoptadas, previa estimación de la dependencia entre las variables aleatorias: gasto máximo (Q) y volumen escurrido (V) de las crecientes anuales disponibles.

La aplicación de las FC en los AFCb ha mostrado que la selección de la *Cópula idónea* se basa en reproducir la dependencia del extremo derecho observada, lo cual garantiza una estimación más aproximada de las *predicciones* asociadas con altos periodos de retorno conjuntos. Por lo anterior, ahora se prueban FC que exhiben, de más a menos, varios grados de dependencia en su cola derecha, por las familias de las Cópulas de Joe, Clayton Asociada, Gumbel-Hougaard y t de Student.

El contraste realizado con 18 registros anuales de Q y V recopilados detectó dos no aleatorios, que fueron eliminados. Para el resto, la FC de Joe se sugiere en un registro procesado como aplicación numérica. La FC de Clayton Asociada se adopta en otro registro y la FC de Gumbel-Hougaard se propone en 10 registros. En tres registros su dependencia observada está comprendida entre la que exhiben estas dos FC citadas. Por último, se encontró que la FC de la t de Student reproduce la dependencia del registro de la Presa Malpaso, con $\nu = 2$ (ver Tabla 5).

Las FC que no muestran tal dependencia, como las Clayton, Frank y Plackett, también se aplican a los datos disponibles de Q y V , y permiten *validar* la calidad del ajuste logrado por la FC probada. Lo anterior está respaldado por el hecho de que todas las FC citadas y expuestas muestran un ajuste similar a los datos (ver Tabla 8).

El procesamiento de los 26 datos anuales de Q y V de las crecientes de entrada a la presa Adolfo López Mateos (*Humaya*) del estado de Sinaloa, México, permitió mostrar el desarrollo numérico de la teoría expuesta y describir con detalle las Figuras 1 y 2, que engloban la aceptación de la FC adoptada y la gráfica básica de los resultados o *predicciones* de Q y V del periodo de retorno *conjunto* tipo AND.

Agradecimientos

Se agradecen las sugerencias de los árbitros anónimos B, F e I, las cuales ayudaron a corregir errores de redacción, omisión y de la estructura general del artículo, para darle una secuencia acorde con la teoría y aplicación de las funciones Cópula en el análisis de frecuencias bivariado

de crecientes. Sus comentarios permitieron destacar los esfuerzos realizados en México para contar con métodos confiables de estimación de crecientes de diseño en grandes embalses.

Referencias

- AghaKouchak, A., Sellars, S., & Sorooshian, S. (2013). Chapter 6. Methods of tail dependence estimation. In: AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (eds.). *Extremes in a changing climate* (pp. 163-179). Dordrecht, The Netherlands: Springer.
- Aldama, A. A. (2000). Hidrología de avenidas. Conferencia Enzo Levi 1998. *Ingeniería Hidráulica en México*, 15(3), 5-46.
- Aldama, A. A., Ramírez, A. I., Aparicio, J., Mejía-Zermeño, R., & Ortega-Gil, G. E. (2006). *Seguridad hidrológica de las presas en México*. Jiutepec, México: Instituto Mexicano de Tecnología del Agua.
- Bobée, B. (1975). The Log-Pearson type 3 distribution and its application to Hydrology. *Water Resources Research*, 11(5), 681-689. DOI: 10.1029/WR011i005p00681
- Bobée, B., & Ashkar, F. (1991). Chapter 1. Data requirements for hydrologic frequency analysis. In: *The Gamma Family and derived distributions applied in Hydrology* (pp. 1-12). Littleton, USA: Water Resources Publications.

- Campos-Aranda, D. F. (2003). Capítulo 7. Integración numérica. En: *Introducción a los métodos numéricos: Software en Basic y aplicaciones en Hidrología Superficial* (pp. 137-153). San Luis Potosí, México: Editorial Universitaria Potosina.
- Campos-Aranda, D. F. (2008). Procedimiento para revisión (sin hidrometría) de la seguridad hidrológica de presas pequeñas para riego. *Agrociencia*, 42(5), 551-563.
- Campos-Aranda, D. F. (2022). Aplicación de la distribución GVE bivariada en el análisis de frecuencias conjunto de crecientes. *Tecnología y ciencias del agua*, 13(6), 534-602. DOI: <https://doi.org/10.24850/j-tyca-13-06-11>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247-1250. DOI: 10.5194/gmd-7-1247-2014
- Chen, L., & Guo, S. (2019). Chapter 2. Copula theory. Chapter 3: Copula-based flood frequency analysis. In: *Copulas and its application in Hydrology and Water Resources* (pp. 13-38, 39-71). Gateway East, Singapore: Springer.
- Chowdhary, H., & Singh, V. P. (2019). Chapter 11. multivariate frequency distributions in hydrology. In: Teegavarapu, R. S. V., Salas, J. D., & Stedinger, J. R. (eds.). *Statistical Analysis of Hydrologic Variables* (pp. 407-489). Reston, USA: American Society of Civil Engineers.

- Davis, P. J., & Polonsky, I. (1972). Chapter 25. Numerical interpolation, differentiation and integration. In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of mathematical functions* (pp. 875-926), 9th print. New York, USA: Dover Publications.
- Demarta, S., & McNeil, A. J. (2005). The *t* Copula and related copulas. *International Statistical Review*, 73(1), 111-129.
- Domínguez, R., & Arganis, M. L. (2012). Validation of method to estimate design discharge flow for dam spillways with large regulating capacity. *Hydrological Sciences Journal*, 57(3), 460-478. DOI: 10.1080/02626667.2012.665993
- Dupuis, D. J. (2007). Using Copulas in hydrology: Benefits, cautions, and issues. *Journal of Hydrologic Engineering*, 12(4), 381-393. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(381)
- Escalante-Sandoval, C., & Reyes-Chávez, L. (2002). Capítulo 9. Análisis conjunto de eventos hidrológicos. En: *Técnicas estadísticas en hidrología* (pp. 203-246). México, DF, México: Facultad de Ingeniería de la Universidad Nacional Autónoma de México.
- Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40(1), 1-12. DOI: 10.1029/2003WR002456
- Frahm, G., Junker, M., & Schmidt, R. (2005). Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 37(1), 80-100. DOI: 10.1016/j.insmatheco.2005.05.008

- Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about Copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4), 347-368. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(347)
- Genest, C., Favre, A. C., Béliveau, J., & Jacques, C. (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research*, 4(W09401), 1-12. DOI: 10.1029/2006WR005275
- Genest, C., & Chebana, F. (2017). Copula modeling in hydrologic frequency analysis. In: Singh, V. P. (ed.). *Handbook of Applied Hydrology* (pp. 30.1-30.10). 2nd ed. New York, USA: McGraw-Hill Education.
- Goel, N. K., Seth, S. M., & Chandra, S. (1998). Multivariate modeling of flood flows. *Journal of Hydraulic Engineering*, 124(2), 146-155.
- Gómez, J. F., Aparicio, M., & Patiño, C. (2010). Capítulo 6. Análisis de frecuencias bivariado para la estimación de avenidas de diseño. En: *Manual de análisis de frecuencias en hidrología* (pp. 106-127). Jiutepec, México: Instituto Mexicano de Tecnología del Agua.
- Gräler, B., van den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., & Verhoest, N. E. C. (2013). Multivariate return periods in hydrology: A critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, 17(4), 1281-1296. DOI: 10.5194/hess-17-1281-2013

- Hosking, J. R., & Wallis, J. R. (1997). Appendix: *L*-moments for some specific distributions, In: *Regional frequency analysis. An approach based on L-moments* (pp. 191-209). Cambridge, UK: Cambridge University Press.
- Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis*, 46(2), 262-282.
- Kite, G. W. (1977). Chapter 12. Comparison of frequency distributions. In: *Frequency and risk analyses in hydrology* (pp. 156-168). Fort Collins, Colorado, USA: Water Resources Publications.
- Kottegoda, N. T., & Rosso, R. (2008). Theme 3.5. Copulas. In: *Applied statistics for civil and environmental engineers* (pp. 154-157), 2nd. ed. Oxford, UK: Blackwell Publishing.
- Meylan, P., Favre, A. C., & Musy, A. (2012). Chapter 1. Introduction; Chapter 3. Selecting and checking data series; Theme 9.2. Multivariate frequency analysis using Copulas (pp.). In: *Predictive hydrology. A frequency analysis approach* (pp. 1-13, 29-70, 164-176). Boca Raton, USA: CRC Press.
- Michiels, F., & De Schepper, A. (2008). A Copula test space model. How to avoid the wrong copula choice. *Kybernetika*, 44(6), 864-878.
- Nieves, A., & Domínguez, F. C. (1998). Secciones 6.2 y 6.3. Cuadratura de Gauss e integrales múltiples. En: *Métodos numéricos. Aplicados a la Ingeniería* (pp. 416-434). México, DF, México: Compañía Editorial Continental.

- Ostle, B., & Mensing, R. W. (1975). Appendix 5. Cumulative t distribution. In: *Statistics in research* (pp. 544-545). 3rd ed. Ames, USA: Iowa State University Press.
- Poulin, A., Huard, D., Favre, A. C., & Pugin, S. (2007). Importance of tail dependence in bivariate frequency analysis. *Journal of Hydrologic Engineering*, 12(4), 394-403. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(394)
- Ramírez-Orozco, A. I., & Aldama, A. A. (2000). Capítulo 1. Teoría estadística y análisis de frecuencias conjunto. En: *Análisis de frecuencias conjunto para estimación de avenidas de diseño* (pp. 25-58). Avances en Hidráulica No. 7. México, DF, México: Asociación Mexicana de Hidráulica e Instituto Mexicano de Tecnología del Agua.
- Rao, A. R., & Hamed, K. H. (2000). Chapter 1. Introduction and Chapter 3. Probability weighted moments and L-moments. In: *Flood frequency analysis* (pp. 1-21, 53-72). Boca Raton, Florida, USA: CRC Press.
- Requena, A. I., Mediero, L., & Garrote, I. (2013). A bivariate return period based on copulas for hydrologic dam design: Accounting for reservoir routing in risk estimation. *Hydrology and Earth System Sciences*, 17(8), 3023-3038. DOI: 10.5194/hess-17-3023-2013
- Salvadori, G., & De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40(W12511), 1-17. DOI: 10.1029/2004WR003133

- Salvadori, G., & De Michele, C. (2007). On the use of Copulas in Hydrology: Theory and practice. *Journal of Hydrologic Engineering*, 12(4), 369-380. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(369)
- Salvadori, G., De Michele, C., Kottegoda, N. T., & Rosso, R. (2007). Chapter 3. Bivariate analysis via Copulas; Appendix B: Dependence and Appendix C: Families of Copulas. In: *Extremes in nature. An approach using Copulas* (pp. 131-175, 219-232, 233-269). Dordrecht, The Netherlands: Springer.
- Shiau, J. T., Wang, H. Y., & Tsai, C. T. (2006). Bivariate frequency analysis of floods using Copulas. *Journal of the American Water Resources Association*, 42(6), 1549-1564. DOI: 10.1111/j.1752-1688-2006.tb06020.x
- Sraj, M., Bezak, N., & Brilly, M. (2015). Bivariate flood frequency analysis using the copula function. A case study of the Litija station on the Sava River. *Hydrological Processes*, 29(2), 225-238. DOI: 10.1002/hyp.10145
- Stedinger, J. R. (2017). Flood frequency analysis. In: Singh, V. P. (ed.). *Handbook of applied hydrology* (pp. 76.1-76.8), 2nd ed. New York, USA: McGraw-Hill Education.
- Stegun, I. A. (1972). Chapter 27. Miscellaneous functions. In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of mathematical functions* (pp. 997-1010), 9th print. New York, USA: Dover Publications.

- Vogel, R. M., & Castellarin, A. (2017). Risk, reliability, and return periods and hydrologic design. In: Singh, V. P. (ed.). *Handbook of applied hydrology* (pp. 78.1-78.10), 2nd ed. New York, USA: McGraw-Hill Education.
- Volpi, E., & Fiori, A. (2012). Design event selection in bivariate hydrological frequency analysis. *Hydrological Sciences Journal*, 57(8), 1506-1515. DOI: 10.1080/02626667.2012.726357
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82. DOI: 10.3354/cr030079
- WMO, World Meteorological Organization. (1971). Annexed III: Standard tests of significance to be recommended in routine analysis of climatic fluctuations. In: *Climatic Change* (pp. 58-71) (Technical Note No. 79). Geneva, Switzerland: Secretariat of the World Meteorological Organization.
- WRC, Water Resources Council. (1977). *Guidelines for determining flood flow frequency* (revised edition) (Bulletin # 17A of the Hydrology Committee). Washington, DC, USA: Water Resources Council.
- Yue, S. (1999). Applying bivariate Normal distribution to flood frequency analysis. *Water International*, 24(3), 248-254.
- Yue, S., Ouarda, T. B. M. J., Bobée, B., Legendre, P., & Bruneau, P. (1999). The Gumbel mixed model for flood frequency analysis. *Journal of Hydrology*, 226(1-2), 88-100.

- Yue, S. (2000a). Joint probability distribution of annual maximum storm peaks and amounts as represented by daily rainfalls. *Hydrological Sciences Journal*, 45(2), 315-326. DOI: 10.1080/02626660009492327
- Yue, S. (2000b). The Gumbel mixed model applied to storm frequency analysis. *Water Resources Management*, 14(5), 377-389.
- Yue, S., & Rasmussen, P. (2002). Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14), 2881-2898. DOI:10.1002/hyp.1185
- Zelen, M., & Severo, N. C. (1972). Chapter 26. Probability Functions. In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of Mathematical Functions* (pp. 927-995), 9th print. New Work, USA: Dover Publications.
- Zhang, L., & Singh, V. P. (2006). Bivariate flood frequency analysis using the Copula method. *Journal of Hydrologic Engineering*, 11(2), 150-164. DOI: 10.1061/(ASCE)1084-0699(2006)11:2(150)
- Zhang, L., & Singh, V. P. (2019). Chapter 3. Copulas and their properties and Chapter 7. Non-Archimedean Copulas. Meta-Elliptical Copulas. In: *Copulas and their applications in water resources engineering* (pp. 62-122, 261-303). Cambridge, UK: Cambridge University Press.