

DOI: 10.24850/j-tyca-15-02-01

Artículos

Análisis de frecuencias bivariado de crecientes anuales mediante enfoque práctico de las funciones Cópula

Bivariate frequencies analysis of annual floods through practical approach of the Copula functions

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 estudio de la seguridad hidrológica en los *embalses* se realiza con el llamado *hidrograma de la creciente de diseño*. El proceso más sencillo y aproximado para estimar tal gráfica se basa en el *análisis de frecuencias bivariado* (AFB) para definir su gasto máximo (Q) y volumen (V), asociados con el periodo de retorno *conjunto* de diseño. Las *funciones Cópula* (FC) se fundamentan en la dependencia entre Q y V , y definen la

distribución bivariada por medio de las funciones univariadas marginales previamente adoptadas. El *enfoque práctico* adoptado utiliza *FC* de un solo parámetro de ajuste y selecciona la más adecuada a partir de la dependencia que muestra el registro conjunto de Q y V en el extremo derecho de su distribución empírica. Además, contrasta y ratifica la *FC* adoptada contra varias de uso común en los AFB. Lo anterior, por medio de los errores de ajuste entre las probabilidades bivariadas empíricas y teóricas. La búsqueda de las distribuciones marginales se realizó con base en el diagrama de cocientes L para adoptar las tres mejores y confrontarlas con las funciones Kappa y Wakeby de gran versatilidad. Se realizó el AFB de las 55 crecientes anuales registradas en la estación hidrométrica La Cuña de la Región Hidrológica No. 12-3 (río Santiago), México. Se definen cuatro periodos de retorno *conjuntos* de diseño y se detalla el cálculo de sus curvas de tipo AND. Por último, se citan varias conclusiones que destacan las ventajas del uso de las *FC* en los AFB de crecientes.

Palabras clave: familias de funciones Cópula, cociente tau de Kendall, coeficiente rho de Spearman, probabilidades empíricas conjuntas, dependencia en el extremo derecho, periodos de retorno conjuntos, eventos críticos.

Abstract

The study of hydrological safety in *reservoirs* is carried out with the so-called *Design Flood Hydrograph*. The simplest estimation of such a graph is based on the *bivariate frequency analysis* (BFA), by defining its maximum flow (Q) and volume (V), associated with a *joint* design return

period. The *Copula functions* (*CF*) are based on the dependence between Q and V , and define the bivariate distribution by means of the previously adopted marginal univariate functions. The *practical approach* adopted uses *CF* with a single fit parameter and selects the most appropriate, based on the dependence shown by the joint record of Q and V , at the far-right end of its empirical distribution. Moreover, the practical approach compares and ratifies the adopted *CF*, against several of common use in the BFA. The above, using the fitting errors between the empirical and theoretical bivariate probabilities. Due to their importance, the search for marginal distributions was carried out based on the L quotient diagram, adopting the best three and comparing them to the highly versatile Kappa and Wakeby functions. The BFA of the 55 annual floods registered in the La Cuña hydrometric station of the Hydrological Region No. 12-3 (Santiago River), Mexico was carried out. Four *joint* design return periods were defined and the calculation of their AND curves is detailed. Finally, several Conclusions are cited which highlight the advantages of the use of *CF* in flood BFAs.

Keywords: Families of Copula functions, Kendall's tau quotient, Spearman's rho coefficient, joint empirical probabilities, right-tail dependency, joint return periods, critical events.

Recibido: 16/11/2021

Aceptado: 05/07/2022

Publicado online: 18/07/2022

Introducción

Generalidades

En México, las inundaciones causadas por las *crecientes* son uno de los desastres naturales que más daños materiales, ambientales y sociales originan en los asentamientos humanos. Por ello, la reducción del riesgo mediante obras hidráulicas y la elaboración de los planes de mitigación de daños, sin tales estructuras, han llegado a ser vitales para la sociedad. La planeación y el diseño hidrológico de obras de protección, como diques y muros de contención, rectificaciones, canalizaciones, puentes y algunas obras del drenaje urbano, se dimensionan a partir de las llamadas *crecientes de diseño* (CD). Las CD son gastos máximos anuales del río donde se ubicará la obra hidráulica, asociadas con bajas probabilidades de ser excedidos (Aldama, Ramírez, Aparicio, Mejía-Zermeño, & Ortega-Gil, 2006; Rao & Hamed, 2000; Meylan, Favre, & Musy, 2012).

La estimación más confiable de las CD se realiza a través del llamado *análisis de frecuencias de crecientes (AFC) univariado*, el cual intenta representar al registro de gastos máximos anuales del río con una *función de distribución de probabilidades* (FDP) o modelo probabilístico, con el cual se realizan las *predicciones* buscadas o CD. Los AFC conducen a resultados confiables cuando se cumplen varias condiciones, entre ellas, que el registro disponible de gastos sea aleatorio, que la FDP sea la más representativa de tales datos, y que su ajuste se realice con un método eficiente y robusto (Hosking & Wallis, 1997; Meylan *et al.*, 2012; Stedinger, 2017).

Por otra parte, los *embalses* o almacenamientos se pueden clasificar de manera simple en grandes, medianos y pequeños. Los grandes embalses son de múltiples propósitos y operación multianual, englobando el aprovechamiento y la protección contra crecientes. En cambio, los embalses pequeños son de propósito único, por lo general el riego, y por ello guardan agua de la temporada de lluvias para utilizarla en el siguiente estiaje. Sean grandes, medianos o pequeños, el diseño de los embalses por seguridad hidrológica requiere la estimación del *hidrograma* de la CD (Aldama, 2000; Aldama *et al.*, 2006; Gómez, Aparicio, & Patiño, 2010).

El hidrograma es la gráfica de evolución de la creciente con el tiempo en las abscisas y el gasto en las ordenadas. Entonces, el área que define es el volumen escurrido (V) y sus amplitudes en el tiempo la duración total (D) y en las ordenadas el gasto máximo (Q), cuyo lapso transcurrido desde el inicio se llama tiempo pico (tp). El hidrograma establece claramente que las crecientes son un fenómeno hidrológico *multivariado*, cuyos elementos están correlacionados (Favre, El Adlouni, Perreault, Thiémonge, & Bobée, 2004; Aldama *et al.*, 2006; Gräler *et al.*, 2013; Chowdhary & Singh, 2019).

Aldama (2000) demostró que los embalses no son sensibles al valor del tp , y como el Q y el V están correlacionados y el V lo está con D , el estudio multivariado de las crecientes se pudo reducir a un *análisis bivariado* de Q y V para definir de manera aproximada el hidrograma de la CD. El AFC bivariado comenzó a finales del siglo pasado con los trabajos de Goel, Seth y Chandra (1998), y de Yue, Ouarda, Bobée, Legendre y Bruneau (1999), que procesaron datos de Q y V de las crecientes anuales, los primeros con base en la distribución Normal bivariada, al igual que

Yue (1999), y los segundos aplicando el llamado modelo logístico (Yue & Wang, 2004), con distribuciones marginales Gumbel.

Los primeros AFC bivariados estuvieron limitados por la escasa disponibilidad de FDP multivariadas, cuyos primeros modelos utilizados fueron extensiones de las distribuciones univariadas, por ejemplo, las bivariadas Normal, Log-normal, Exponencial, Pareto y Gamma. Favre *et al.* (2004) indican que este tipo de distribuciones tiene las tres desventajas siguientes: (1) sus distribuciones marginales pertenecen a la misma familia; (2) la extensión más allá del caso bivariado no es simple o no existe, y (3) los parámetros de ajuste de tales marginales también se usan para modelar la dependencia entre las variables aleatorias Q y V .

Las limitaciones y desventajas de las distribuciones bivariadas existentes son fácilmente eliminadas al aplicar las llamadas *funciones Cópula (FC)*, las cuales permiten modelar la dependencia entre las variables Q y V , además de *conjuntar* cualquier tipo de distribución marginal univariada adoptada previamente para construir la nueva distribución bivariada (Shiau, Wang, & Tsai, 2006; Sraj, Bezak, & Brilly, 2015; Genest & Chebana, 2017; Zhang & Singh, 2019).

Las primeras aplicaciones prácticas de las *FC* lógicamente fueron bivariadas y comienzan con los trabajos de Favre *et al.* (2004), y Salvadori y De Michele (2004). Después se tienen los estudios de Zhang y Singh (2006), y de Genest y Favre (2007). De manera simultánea llegan las aplicaciones trivariadas con los análisis de Grimaldi y Serinaldi (2006), y de Zhang y Singh (2007). En la actualidad, las aplicaciones de las *FC* en los análisis de frecuencias de los diversos datos hidrológicos se han consolidado, como lo demuestran los textos de Salvadori, De Michele,

Kottegoda y Rosso (2007), Zhang y Singh (2019), Chen y Guo (2019), y los capítulos de Salvadori y De Michele (2013), y de Chowdhary y Singh (2019).

En este estudio, con el *enfoque práctico de las FC*, se intenta una *similitud operativa* con los AFC univariados, al realizar los *bivariados*. Para lograr lo anterior, se utilizan *FC* de un solo parámetro de ajuste, que ya han sido probadas en los AFC bivariados y cuya selección se hace con base en indicadores de ajuste y no por medio de simulación aleatoria. Además, de inicio se define si la muestra disponible de datos conjuntos de Q y V muestra dependencia en la cola derecha de su distribución empírica y, de ser así, primero se buscan y después se prueban *FC* que muestren tal dependencia.

Objetivos

Los *objetivos* de este estudio son los siete siguientes: (1) enunciar de manera breve el concepto operativo de las *FC*; (2) exponer las *FC* de cada familia que han sido empleadas en los AFC; (3) describir el ajuste de la *FC* con base en las medidas de asociación; (4) explicar el concepto de dependencia en el extremo derecho tanto de forma empírica como teórica de cada *FC* expuesta; (5) seleccionar y probar las *FC* que son aplicables en el ejemplo numérico descrito; (6) estimar y contrastar los periodos de retorno univariados y los bivariados, y (7) realizar una aplicación numérica con los 55 gastos máximos y volúmenes de las crecientes anuales registradas en la estación hidrométrica La Cuña de la Región Hidrológica No. 12-3 (río Santiago), México.

Teoría operativa

Las funciones Cópula

Concepto y definición

La característica esencial de las *funciones Cópula* (*FC*) consiste en permitir expresar una distribución conjunta de variables aleatorias correlacionadas como una función de sus distribuciones marginales. Entonces, una *FC* enlaza o relaciona las distribuciones marginales univariadas para formar una distribución multivariada. Una ventaja básica de las *FC* para 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.

Debido a lo anterior, la construcción de la distribución multivariada se reduce al estudio de la relación que guardan las variables dependientes cuando las distribuciones marginales univariadas son conocidas. En resumen, el uso de las *FC* ofrece entera libertad para adoptar o seleccionar los mejores modelos marginales univariados que representen a los datos (Shiau *et al.*, 2006; Kottegodda & Rosso, 2008; Meylan *et al.*, 2012; Zhang & Singh, 2019; Chen & Guo, 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 *función Cópula C* existe y es tal que:

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

La función Cópula C es única si las variables aleatorias X y Y son continuas; en otro caso está solo determinada en el producto del rango de X por el rango de Y . Además, una Cópula es una función de distribución bivariada con marginales uniformes en el intervalo $(0, 1)$.

La Ecuación (1) define el concepto básico para el desarrollo de las funciones Cópula y se conoce como *Teorema de Sklar* expuesto en 1959 (Sklar, 1959; Shiau *et al.*, 2006; Salvadori *et al.*, 2007; Kottegoda & Rosso, 2008; Meylan *et al.*, 2012; Chen & Guo, 2019; Zhang & Singh, 2019).

Familias de Cópulas

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

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

Las Cópulas arquimedeanas han tenido aplicación amplia debido a su construcción simple, un único parámetro, rango amplio y aceptación de ambos tipos de dependencia. Haciendo $F_X(x) = u$, $F_Y(y) = v$ y θ , el parámetro que mide la dependencia o asociación entre u y v , se tienen las siguientes dos familias de Cópulas arquimedeanas, las cuales aceptan dependencia negativa y positiva (Shiau *et al.*, 2006; Genest & Favre, 2007; Salvadori *et al.*, 2007; 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 *et al.* (2015) la citan con los dos nombres. Su ecuación y espacio de variación de θ son (Nelsen, 2006):

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

para $\theta = 0$, $C(u, v) = \Pi(u, v) = uv$ es la cópula producto que define de manera única la independencia. 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 $\theta = 1$, $C(u, v) = \Pi(u, v) = uv$ se define la independencia. 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:

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

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

En seguida se cita una familia de Cópulas de valores extremos, la cual acepta solo dependencia positiva.

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

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

para $\theta = 1$, $C(u, v) = \Pi(u, v) = uv$ se define la independencia. 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 ($\theta > 1$):

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

Para $\theta = 1$, $C(u, v) = \Pi(u, v) = uv$ se define la independencia. 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:

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

para $\theta = 0$, $C(u, v) = \Pi(u, v) = uv$ se define la independencia. La relación de θ con el cociente tau de Kendall y con la rho de Spearman son las siguientes:

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

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

Familias de Cópulas asociadas

Las cuatro clases de familias de Cópulas que existen y para el caso bivariado tienen tres *familias asociadas*, definidas por las siguientes transformaciones estrictamente crecientes o decrecientes, según Teorema 2.4.4 en Nelsen (2006). Tales familias son (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 las medidas de concordancia; es decir, la tau de Kendall y la rho de Spearman de negativas a positivas 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 .

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

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

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

Medidas de asociación

Concordancia

Como la función Cópula (FC) caracteriza la *dependencia* entre las variables aleatorias 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. Unas variables con correlación lineal directa serán *concordantes*, pues al aumentar una, la otra también lo hace. Variables con una correlación lineal inversa serán *discordantes*, ya que a grandes valores de una le corresponden 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, indicado 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).

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):

$$\tau_n = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{signo}[(x_i - x_j)(y_i - y_j)] \quad (19)$$

en la ecuación anterior, n es el número de observaciones y el signo $[\]$ 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 cero y varianza $2(2n + 5) / [9n(n - 1)]$. Entonces, H_0 será rechazada con un nivel de confianza $\alpha = 5 \%$ si:

$$\sqrt{\frac{9n(n-1)}{2(2n+5)}} |\tau_n| > Z_{\alpha/2} = 1.96 \quad (20)$$

Coeficiente rho de Spearman

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

$$\rho_n = \frac{12}{n(n+1)(n-1)} \sum_{i=1}^n R_i \cdot S_i - 3 \frac{n+1}{n-1} \quad (21)$$

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:

$$\sqrt{n - 1} \cdot |\rho_n| > 1.96 \quad (22)$$

Estimación del parámetro de dependencia

El método más simple para estimar el parámetro θ de las funciones Cópula se asemeja al método de momentos y se basa en la inversión de la ecuación que relaciona θ con el cociente tau de Kendall o con el coeficiente rho de Spearman. Las otras dos técnicas disponibles se denominan: (1) método de máxima pseudo-verosimilitud y (2) método exacto de máxima verosimilitud (Meylan *et al.*, 2012; Chowdhary & Singh, 2019; Zhang & Singh, 2019; Chen & Guo, 2019). En las ecuaciones (5), (10), (12) y (13) se procede por tanteos para obtener θ ; en cambio, en las fórmulas (3) y (8) se despeja su valor.

Estimación de las probabilidades empíricas

Las probabilidades empíricas bivariadas se estimaron con base en la fórmula de Gringorten (Ecuación (23)), aplicada por Yue *et al.* (1999), Chen y Guo (2019), y Zhang y Singh (2019). Tal fórmula es:

$$p = \frac{i-0.44}{n+0.12} \quad (23)$$

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:

$$F(x, y) = P(Q \leq q, V \leq v) = \frac{NM_i - 0.44}{n + 0.12} \quad (24)$$

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 función Cópula 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; Chen & Guo, 2019):

$$EME = \sqrt{\frac{1}{n} \sum_{i=1}^n (w_i^o - w_i^c)^2} \quad (25)$$

$$EMA = \frac{1}{n} \sum_{i=1}^n |(w_i^o - w_i^c)| \quad (26)$$

$$EAM = \max_{i=1:n} |(w_i^o - w_i^c)| \quad (27)$$

Dependencia en los extremos de la FC

Generalidades

Otro criterio que se aplica para seleccionar una FC es el basado en la magnitud de la dependencia en la cola 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).

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:

$$\lambda_U = 2 - 2^{1/\theta} \quad (28)$$

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

$$\lambda_L = 2^{-1/\theta} \quad (29)$$

$$\lambda_L = \frac{2\theta}{\theta+1} \quad (30)$$

Como ya se indicó (Ecuación (17)), la Cópula de Clayton asociada tiene $\lambda_U = 2^{-1/\theta}$. Familias de Cópulas no expuestas con $\lambda_U > 0$ son las de Galambos, Genest-Ghoudi, Hüsler-Reiss, Joe y t -Student; esta última de la familia elíptica (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019; Chen & Guo, 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, Huard, Favre y Pugin (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:

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n 1\left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v\right) \quad (31)$$

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:

$$U_i = \frac{\text{Rango}(X_i)}{n+1} = F_n(x_i) \quad V_i = \frac{\text{Rango}(Y_i)}{n+1} = F_n(Y_i) \quad (32)$$

La Ecuación (31) es una función de $[0,1]^2$ a $(0, 1)$ que tiene saltos de magnitud $1/n$ en cada pareja de la forma de la Ecuación (32), para $i \in \{1,2, \dots, n\}$. En realidad, la Ecuación (31) no es una Cópula sino una *subcópula* que sirve para aproximar Cópulas (Nelsen, 2006).

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

$$\lambda_U^{CFG} = 2 - 2 \exp \left\{ \frac{1}{n} \sum_{i=1}^n \ln \left[\sqrt{\ln \frac{1}{U_i} \cdot \ln \frac{1}{V_i}} / \ln \left(\frac{1}{\max(U_i, V_i)^2} \right) \right] \right\} \quad (33)$$

Este estimador acepta que la *FC* se puede aproximar por una de la clase de valores extremos y tiene la ventaja de no requerir un valor umbral para su estimación.

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 (24)). Yue (2000) 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°.

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 (dma)* entre las probabilidades conjuntas empíricas y teóricas. Para evaluar la estadística (D_n) del test se utilizó la expresión que exponen Meylan *et al.* (2012), para $\alpha = 5 \%$, la cual es:

$$D_n = \frac{1.358}{\sqrt{n}} \quad (34)$$

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

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; Vogel & Castellarin, 2017):

$$T_{XY} = \frac{1}{P(X > x \text{ or } Y > y)} = \frac{1}{1 - F_{X,Y}(x,y)} = \frac{1}{1 - C[F_X(x), F_Y(y)]} \quad (35)$$

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 *serán* excedidos ($X > x, Y > y$) o condición AND. Su ecuación es (Shiau *et al.*, 2006; Vogel & Castellarin, 2017):

$$T'_{XY} = \frac{1}{P(X>x \text{ and } Y>y)} = \frac{1}{F'_{X,Y}(x,y)} = \frac{1}{1-F_X(x)-F_Y(y)+C[F_X(x),F_Y(y)]} \quad (36)$$

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):

$$T_{XY} \leq \min[T_X, T_Y] \leq \max[T_X, T_Y] \leq T'_{XY} \quad (37)$$

siendo:

$$T_X = \frac{1}{F'_X(x)} = \frac{1}{1-F_X(x)} \quad (38)$$

$$T_Y = \frac{1}{F'_Y(y)} = \frac{1}{1-F_Y(y)} \quad (39)$$

Eventos críticos o de diseño

Volpi y Fiori (2012) destacan que la gráfica del periodo de retorno conjunto de tipo AND —mostrada como Figura 1— 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 asíntotas 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 asíntotas 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 cercana al inicio de las rectas asíntotas.

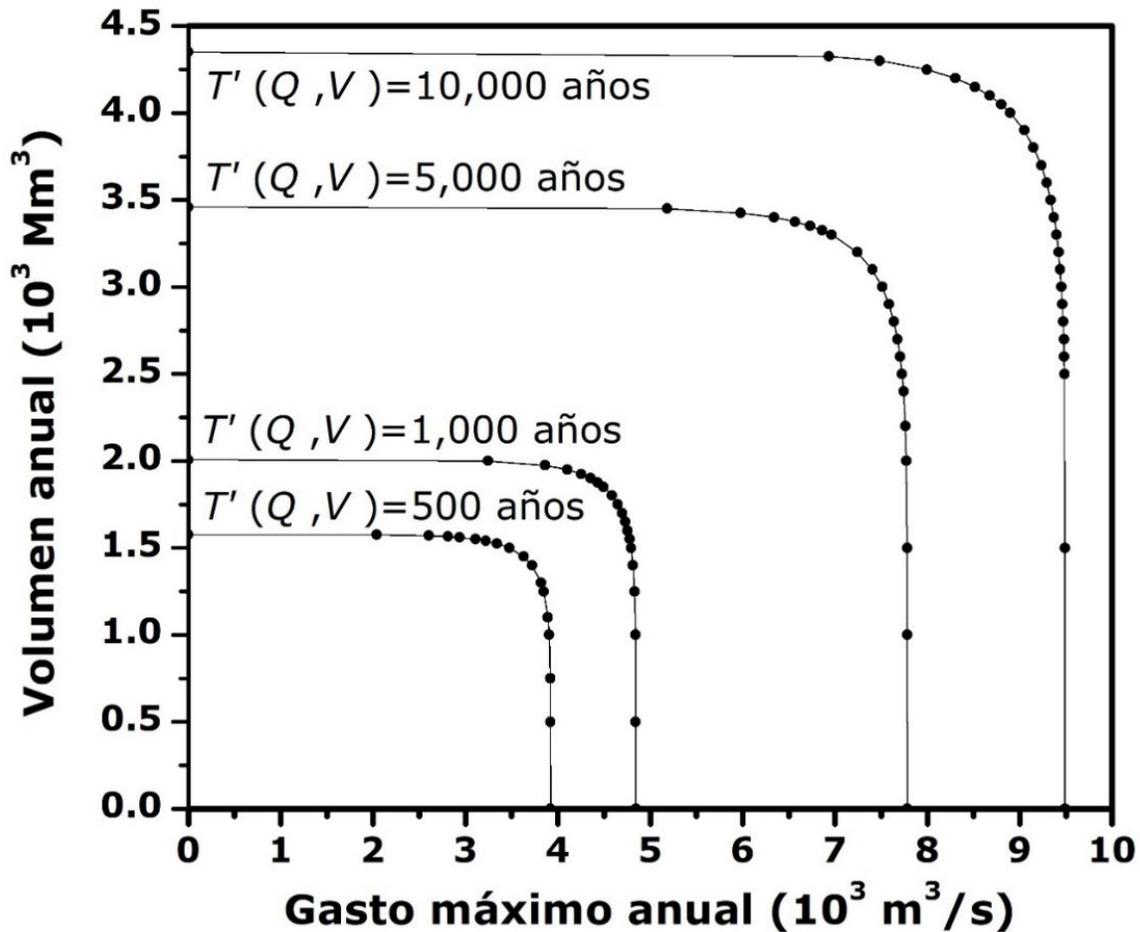


Figura 1. Gráficas de los cuatro periodos de retorno conjunto T'_{XY} de diseño obtenidas con la FC de Gumbel-Hougaard en las crecientes de la estación La Cuña, México.

Test de Wald-Wolfowitz

Esta prueba no paramétrica ha sido expuesta y aplicada 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 la prueba a los registros de gastos máximos y volúmenes anuales, que deben ser muestras aleatorias.

Selección de las distribuciones marginales

Ocho distribuciones de probabilidad aplicadas

El planteamiento de búsqueda de las distribuciones marginales idóneas tomó en cuenta las características estadísticas de los datos hidrológicos por procesar, gastos máximos y volúmenes de las crecientes anuales. Lo anterior, a través de los cocientes L de asimetría (t_3) y curtosis (t_4), que permiten definir en el diagrama de cocientes L de Hosking y Wallis (1997), las mejores distribuciones por su menor cercanía a las *cinco curvas* que se muestran en tal gráfico. Dichas distribuciones tienen tres parámetros de ajuste: logística generalizada (LOG), general de valores extremos (GVE), log-normal (LGN), Pearson tipo III (PE3) y Pareto generalizada (PAG). La curva de la PE3 permite probar la distribución log-Pearson tipo III (LP3).

Además, se aplicaron dos modelos probabilísticos que han mostrado gran versatilidad y universalidad para representar series de datos hidrológicos extremos debido a que tienen 4 y 5 parámetros de ajuste. Tales distribuciones son la Kappa (Hosking, 1994; Kjeldsen, Ahn, &

Prosdocimi, 2017; Campos-Aranda, 2023) y la Wakeby (Hosking & Wallis, 1997).

La distribución LP3 fue la única que se aplicó con el método de momentos en los dominios logarítmico (WRC, 1977) y real (Bobée, 1975), seleccionando la de mejor ajuste. Las siete restantes se ajustaron a los registros de gasto máximo y volumen anual de crecientes a través del método de momentos L , según procedimientos expuestos por Hosking y Wallis (1997).

Momentos L de la muestra

Los momentos L son combinaciones lineales de los momentos de probabilidad ponderada (b_r), definidos por Greenwood, Landwehr, Matalas y Wallis (1979); por lo anterior, son robustos ante los valores dispersos de muestra. Su cálculo comienza por ordenar las series disponibles de gastos máximos (x_i) y volúmenes anuales (y_i) de las crecientes de menor a mayor ($x_1 \leq x_2 \leq \dots \leq x_n$), y estimar los b_r con la ecuación siguiente (Hosking & Wallis, 1997; Rao & Hamed, 2000; Asquith, 2011; Stedinger, 2017):

$$b_r = \frac{1}{n} \sum_{i=r+1}^n \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} x_i \quad (40)$$

En la expresión anterior el número de orden r varía de 0 a 3 y n es el número de datos de las series anuales. Se deduce que b_0 es igual a la

media aritmética. Los momentos L de la muestra (l) y sus respectivos cocientes (t) de similitud con los coeficientes de variación, asimetría y curtosis son:

$$l_1 = b_0 \quad (41)$$

$$l_2 = 2 \cdot b_1 - b_0 \quad (42)$$

$$l_3 = 6 \cdot b_2 - 6 \cdot b_1 + b_0 \quad (43)$$

$$l_4 = 20 \cdot b_3 - 30 \cdot b_2 + 12 \cdot b_1 - b_0 \quad (44)$$

$$t_2 = l_2/l_1 \quad (45)$$

$$t_3 = l_3/l_2 \quad (46)$$

$$t_4 = l_4/l_2 \quad (47)$$

Diagrama de cocientes de momentos L

Tiene en el eje de las abscisas a t_3 y en el de las ordenadas a t_4 . Las FDP de tres parámetros de ajuste son líneas curvas con las ecuaciones de tipo polinomio siguientes (Hosking & Wallis, 1997):

Logística generalizada (LOG):

$$t_4^{\text{LOG}} = 0.16667 + 0.83333 \cdot t_3^2 \quad (48)$$

Pareto generalizada (PAG):

$$t_4^{\text{PAG}} = 0.20196 \cdot t_3 + 0.95924 \cdot t_3^2 - 0.20096 \cdot t_3^3 + 0.04061 \cdot t_3^4 \quad (49)$$

Log-normal (LGN):

$$t_4^{\text{LGN}} = 0.12282 + 0.77518 \cdot t_3^2 + 0.12279 \cdot t_3^4 - 0.13638 \cdot t_3^6 + 0.11368 \cdot t_3^8 \quad (50)$$

Pearson tipo III (PT3):

$$t_4^{\text{PT3}} = 0.12240 + 0.30115 \cdot t_3^2 + 0.95812 \cdot t_3^4 - 0.57488 \cdot t_3^6 + 0.19383 \cdot t_3^8 \quad (51)$$

General de valores extremos (GVE):

$$t_4^{GVE} = 0.10701 + 0.11090 \cdot t_3 + 0.84838 \cdot t_3^2 - 0.06669 \cdot t_3^3 + SF \quad (52)$$

siendo $SF = 0.00567 \cdot t_3^4 - 0.04208 \cdot t_3^5 + 0.03763 \cdot t_3^6$

Utilizando los logaritmos naturales de los datos en las ecuaciones (40) a (47) se obtienen los cocientes L logarítmicos y entonces se puede utilizar la Ecuación (51) para valorar a la distribución log-Pearson tipo III.

Distancia absoluta

Uno de los enfoques recientes para la selección de la mejor FDP de una serie de datos consiste en llevar al diagrama de cocientes L los valores de la muestra (t_3 y t_4) y definir su cercanía a alguna de las curvas para obtener el mejor modelo probabilístico. Para evitar la subjetividad en la selección de la FDP se ha propuesto evaluar la *distancia absoluta* (DA) con la expresión siguiente (Yue & Hashino, 2007):

$$DA = t_4(t_3^{obs}) - t_4^{obs} \quad (53)$$

donde t_3^{obs} y t_4^{obs} son los cocientes L de asimetría y curtosis de la serie analizada, y $t_4(t_3^{obs})$ es el valor teórico del cociente L de curtosis calculado con cada FDP (ecuaciones (48) a (52)), para el valor observado del cociente L de asimetría. Una FDP con el menor valor de la DA es la mejor para los datos procesados.

Errores de ajuste

El primer criterio aplicado para la selección de la mejor FDP a unos datos o serie disponible fue el llamado *error de ajuste* (Kite, 1977; Willmott & Matsuura, 2005; Chai & Draxler, 2014). Este criterio será aplicado después de seleccionar las tres mejores FDP en el diagrama de cocientes L de acuerdo con su distancia absoluta mínima, y de haber aplicado las distribuciones Kappa y Wakeby.

Cambiando en las ecuaciones (25) y (26) 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 (23)).

Datos por procesar

La hidrométrica La Cuña con clave 12054 y área de cuenca de 19 097 km², se ubica en el Río Verde de la Región Hidrológica No. 12-3 (río Santiago), México. Gómez *et al.* (2010) exponen el registro de gasto máximo en m³/s y volúmenes anuales en millones de m³ (Mm³) del periodo de 1947 a 2004, con 55 parejas de datos (n), ya que los años 1983, 1984 y 1985 no tienen información de volúmenes. Tales datos se exponen en la Tabla 1.

Tabla 1. Gastos máximos, volúmenes y sus números de orden bivariados de las crecientes anuales de la estación de aforos La Cuña, México (Gómez *et al.*, 2010).

Año	Q (m ³ /s)	V (Mm ³)	NM _i	Año	Q (m ³ /s)	V (Mm ³)	NM _i
1947	784.0	146.80	36	1975	622.1	249.07	40
1948	736.8	155.12	37	1976	1 374.0	527.96	52
1949	510.0	111.40	28	1977	439.7	111.77	28
1950	461.0	94.06	24	1978	280.2	66.23	15
1951	411.0	111.55	27	1979	267.2	45.80	9
1952	326.0	70.82	19	1980	287.3	99.60	20
1953	349.8	144.75	28	1981	280.7	28.70	5
1954	130.4	23.22	3	1982	156.5	35.37	4
1955	690.0	203.31	39	1986	698.2	193.51	38
1956	266.0	106.76	17	1987	184.7	55.39	7
1957	199.0	45.92	9	1988	595.2	242.21	38
1958	690.0	188.71	37	1989	110.2	42.49	3
1959	340.6	47.91	12	1990	523.9	248.07	37
1960	249.6	91.58	14	1991	1636.3	443.30	52
1961	350.0	130.68	28	1992	1168.0	172.49	40
1962	317.0	51.27	12	1993	295.0	96.50	20
1963	732.6	127.90	33	1994	212.8	53.55	10
1964	265.1	82.75	15	1995	367.4	114.61	27
1965	743.6	295.34	46	1996	144.6	57.43	5
1966	463.9	202.90	35	1997	78.4	16.55	1
1967	1 474.9	598.38	53	1998	261.9	66.17	13
1968	323.0	118.25	25	1999	196.3	41.15	7
1969	160.4	32.22	4	2000	46.8	18.62	1
1970	763.8	187.75	39	2001	313.8	75.78	18
1971	578.0	166.61	36	2002	319.6	153.79	25
1972	191.8	26.39	4	2003	621.1	326.28	40
1973	2 440.0	920.30	55	2004	824.5	384.45	50
1974	238.4	66.66	13	<i>Med</i>	340.6	111.40	-

Resultados y su discusión

Búsqueda de las distribuciones marginales

Verificación de la aleatoriedad

Primero se verificó la aleatoriedad de los registros por procesar con base en el Test de Wald-Wolfowitz, cuyo estadístico condujo a valores de 0.284 y 0.213 para los gastos y volúmenes de la Tabla 1.

Distancias absolutas calculadas

En la Tabla 2 se han concentrado los resultados de las ecuaciones (48) a (52), destacando las tres mejores FDP de cada registro de datos de la Tabla 1.

Tabla 2. Números de orden de la mejor FDP para los registros indicados de la estación hidrométrica La Cuña, México, según distancia absoluta en el diagrama de cocientes L (Hosking & Wallis, 1997).

Gastos máximos anuales (m^3/s)		Volúmenes anuales (Mm^3)	
$t_3 = 0.38713$	$\ln t_3 = 0.01249$	$t_3 = 0.43326$	$\ln t_3 = 0.01382$
$t_4 = 0.25523$	$\ln t_4 = 0.15953$	$t_4 = 0.28466$	$\ln t_4 = 0.12466$
Distribución LOG	0.0363 (3)	Distribución LOG	0.0384 (5)
Distribución GVE	0.0179 (2)	Distribución GVE	0.0240 (3)
Distribución LGN	0.0139 (1)	Distribución LGN	0.0128 (2)
Distribución PT3	0.0680 (6)	Distribución PT3	0.0755 (6)
Distribución PAG	0.0440 (5)	Distribución PAG	0.0320 (4)
Distribución LP3	0.0371 (4)	Distribución LP3	0.0022 (1)

Distribución de los gastos máximos anuales

En la Tabla 3 se exponen los errores de ajuste y predicciones (m^3/s) obtenidas con las tres mejores distribuciones, y la Kappa y Wakeby aplicadas al registro de gastos máximos de la Tabla 1. Se observa que las dos mejores distribuciones, la LGN y la GVE, reportan los errores de ajuste más bajos y similares, pero sus predicciones son bastante diferentes en los cuatro periodos de retorno elevados. Por otra parte, los errores de ajuste de la distribución Kappa son casi idénticos a los de la LGN, pero con predicciones intermedias, y por ello fue la adoptada.

Tabla 3. Errores de ajuste y predicciones (m^3/s) de las tres mejores FDP, la Wakeby y la Kappa en el registro de gastos máximos anuales de la estación hidrométrica La Cuña, México.

FDP	EEA (m^3/s)	EAM (m^3/s)	Periodos de retorno en años					
			50	100	500	1 000	5 000	10 000
LOG (3)	61.3	39.4	1 816	2 393	4 504	5 902	11 036	14 441
GVE (2)	55.7	35.8	1 838	2 372	4 160	5 254	8 926	11 172
LGN (1)	55.8	36.5	1 839	2 293	3 592	4 270	6 167	7 139
KAPPA	55.7	36.7	1 835	2 335	3 920	4 844	7 784	9 489
WAKEBY	35.2	55.7	1 844	2 304	3 646	4 367	6 469	7 599

Los parámetros de ubicación (u_1), escala (a_1) y forma (k_1 y h_1) de la distribución Kappa seleccionada son 258.7462, 228.381, -0.2685394 y 0.2888472 , respectivamente, cuya expresión es:

$$F(x) = \left\{ 1 - h_1 \left[1 - \frac{k_1(x-u_1)}{a_1} \right]^{1/k_1} \right\}^{1/h_1} \quad (54)$$

Distribución de los volúmenes anuales

En la Tabla 4 —similar a la anterior para los volúmenes anuales de la Tabla 1— se observa que la mejor distribución, la LP3, conduce a los errores de ajuste más bajos, pero sus predicciones son bastante menores que las de la GVE. En cambio, la distribución Kappa conduce a errores de ajuste bajos y sus predicciones son intermedias, por ello fue adoptada. En este caso, la distribución Wakeby ratifica la selección.

Tabla 4. Errores de ajuste y predicciones (Mm^3) de las tres mejores FDP, la Wakeby y la Kappa en el registro de volúmenes anuales de la estación hidrométrica La Cuña, México.

FDP	EEA (Mm^3)	EAM (Mm^3)	Periodos de retorno en años					
			50	100	500	1 000	5 000	10 000
GVE (3)	16.6	9.0	662	888	1 702	2 234	4 156	5 411
LGN (2)	14.5	7.6	670	863	1 439	1 752	2 661	3 142
LP3 (1)	10.5	6.3	694	907	1 574	1 952	3 101	3 736
KAPPA	15.5	8.1	663	873	1 576	2 007	3 459	4 350
WAKEBY	15.7	8.3	664	872	1 561	1 978	3 358	4 193

Los parámetros de ubicación (u_2), escala (a_2) y forma (k_2 y h_2) de la distribución Kappa adoptada son 60.39858, 78.31082, -0.3155518 y 0.4287021, respectivamente, cuya expresión es:

$$F(y) = \left\{ 1 - h_2 \left[1 - \frac{k_2(y-u_2)}{a_2} \right]^{1/k_2} \right\}^{1/h_2} \quad (55)$$

Selección y ratificación de la función Cópula

El procesamiento bivariado de los datos de la Tabla 1 condujo a los siguientes tres indicadores de asociación: $r_{xy} = 0.9302$, $\tau_n = 0.7199$ y $\rho_n = 0.9088$. La Ecuación (20), con $n = 55$ y el tau citado, origina un valor de 7.76; por lo cual, el cociente de Kendall es significativo. La Ecuación (22) genera un valor de 6.68 y entonces el coeficiente de Spearman también es significativo. La Ecuación (33) aporta un valor para la dependencia observada en la cola derecha de $\lambda_{ij}^{CFG} = 0.782$.

Por otra parte, en la Tabla 5 se muestran los indicadores estadísticos de ajuste que se obtuvieron al aplicar las seis FC expuestas de Clayton, Frank, Gumbel-Hougaard (GH), Plackett, Raftery y Clayton asociada. En las ecuaciones (25) a (27), las probabilidades bivariadas empíricas se estimaron con la Ecuación (24), y las teóricas con las ecuaciones (2), (4), (7), (9), (11) y (18).

Tabla 5. Indicadores estadísticos del ajuste de las funciones Cópula indicadas en las crecientes anuales de la estación hidrométrica La Cuña, México.

Cópula	θ	EME	EAM	No. DP	No. DN	MDP	MDN	λ_U
Clayton (C.)	5.1394	0.0245	0.0189	28	27	0.0714	-0.0424	0.000
Frank	12.386	0.0227	0.0169	29	26	0.0581	-0.0561	0.000
GH	3.5697	0.0255	0.0192	28	27	0.0541	-0.0676	0.786
Plackett	75.040	0.0247	0.0187	23	32	0.0495	-0.0647	0.000
Raftery	0.7940	0.0250	0.0193	27	28	0.0746	-0.0457	0.000
C. asociada	5.1394	0.0288	0.0225	29	26	0.0477	-0.0788	0.874

Significado de los nuevos acrónimos:

DP, DN: diferencias positivas y negativas.

MDP, MDN: máxima diferencia positiva y negativa.

Como los valores de Q y V de las crecientes anuales registradas en la estación hidrométrica La Cuña muestran una magnitud de λ_{ij}^{CFG} de 0.782 para la dependencia observada en su cola derecha, no existe dificultad para seleccionar como mejor FC la de Gumbel-Hougaard para los datos de la Tabla 1 debido a que reporta un valor para la dependencia significativa en su extremo derecho, prácticamente igual al observado o mostrado por los datos.

Cuando el valor de λ_{ij}^{CFG} resulte bastante inferior al de la FC de Gumbel-Hougaard, se puede recurrir a aplicar la FC de clase elíptica llamada de t -Student, la cual —de acuerdo con los resultados de Dupuis

(2007) y de Poulin *et al.* (2007)— tiene una dependencia igual en ambas colas, pero de menor valor.

En la Tabla 5 los indicadores estadísticos más bajos del *EME* y *EAM*, y las diferencias máximas entre las probabilidades bivariadas se han mostrado en cursivas. Se observa que la *FC* adoptada de Gumbel-Hougaard tiene indicadores de ajuste ubicados entre los más reducidos obtenidos con la *FC* de Frank y los más grandes generados por la *FC* de Clayton asociada.

En la Tabla 6, Tabla 7 y Tabla 8 se muestra una parte de las probabilidades de no excedencia bivariadas, empíricas observadas (w_i^o) y teóricas calculadas (w_i^c) con la *FC* de Frank, de Gumbel-Hougaard y de Clayton asociada. También se indican sombreadas las diferencias máximas positiva y negativa.

Tabla 6. Probabilidades de no excedencia conjuntas y sus diferencias para una parte de las crecientes de la estación La Cuña, México, con la *FC* de Frank.

No.	w_i^o	w_i^c	Diferencias	No.	w_i^o	w_i^c	Diferencias
1	0.6451	0.6477	-0.0026	30	0.9354	0.9344	0.0011
5	0.4819	0.4872	-0.0053	35	0.0827	0.0783	0.0044
10	0.3004	0.3139	-0.0134	40	0.0464	0.0375	0.0089
12	0.6633	0.7194	-0.0561	45	0.1734	0.1704	0.0226
20	0.6270	0.6096	0.0174	50	0.1190	0.1169	0.0021
25	0.6451	0.6609	-0.0157	55	0.8991	0.8410	0.0581

Tabla 7. Probabilidades de no excedencia conjuntas y sus diferencias para una parte de las crecientes de la estación La Cuña, México, con la FC de Gumbel-Hougaard.

No.	w_i^o	w_i^c	Diferencias	No.	w_i^o	w_i^c	Diferencias
1	0.6451	0.6512	-0.0061	30	0.9354	0.9520	-0.0166
5	0.4819	0.4765	0.0053	35	0.0827	0.0777	0.0050
12	0.6633	0.7309	-0.0676	40	0.0464	0.0401	0.0063
15	0.5000	0.4459	0.0541	45	0.1734	0.1624	0.0111
20	0.6270	0.6116	0.0154	50	0.1190	0.1143	0.0048
25	0.6451	0.6661	-0.0210	55	0.8991	0.8557	0.0435

Tabla 8. Probabilidades de no excedencia conjuntas y sus diferencias para una parte de las crecientes de la estación La Cuña, México, con la FC de Clayton asociada.

No.	w_i^o	w_i^c	Diferencias	No.	w_i^o	w_i^c	Diferencias
1	0.6451	0.6540	-0.0088	30	0.9354	0.9548	-0.0194
5	0.4819	0.4863	-0.0044	35	0.0827	0.0727	0.0100
12	0.6633	0.7421	-0.0788	40	0.0464	0.0286	0.0178
15	0.5000	0.4523	0.0477	45	0.1734	0.1454	0.0280
20	0.6270	0.6170	0.0100	50	0.1190	0.0944	0.0246
25	0.6451	0.6798	-0.0346	55	0.8991	0.8578	0.0414

Por otra parte, la Ecuación (34) define a $D_n = 0.1831$, y como las diferencias máximas absolutas de la Tabla 6, Tabla 7 y Tabla 8 son 0.0581, 0.0676 y 0.0788, el test de Kolmogorov–Smirnov ratifica las tres *FC* contrastadas. Los coeficientes de correlación (r_{xy}) entre las 55 probabilidades empíricas y las teóricas —estimadas con las *FC* de Frank, Gumbel-Hougaard y Clayton asociada— son los siguientes: 0.9968, 0.9964 y 0.9962; por lo tanto, se corresponden, en orden de magnitud, con las diferencias máximas observadas.

El contraste gráfico entre ambas probabilidades, para ratificar la adopción de la *FC* de Gumbel–Hougaard, se expone en la Figura 2 para los datos completos de la Tabla 7.

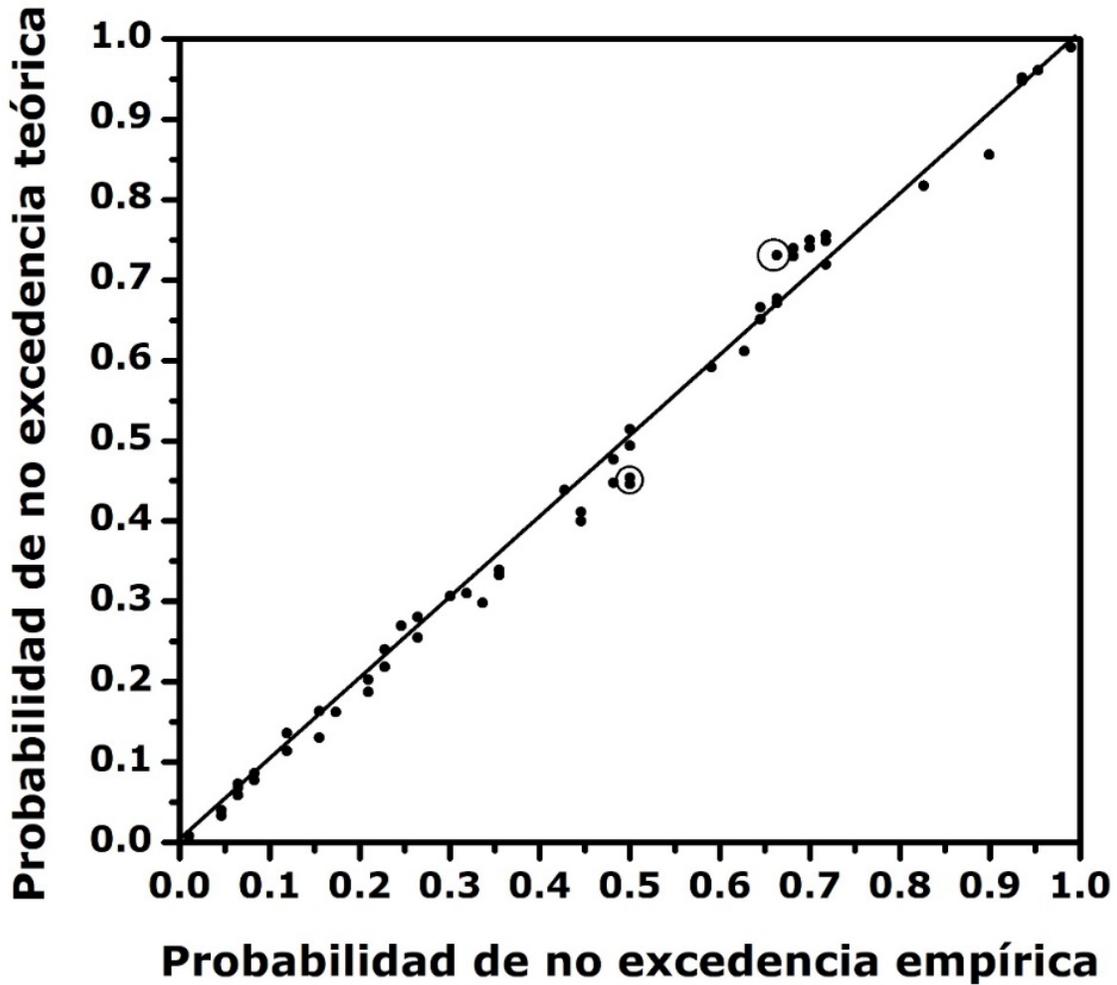


Figura 2. Contraste gráfico de probabilidades conjuntas del gasto pico y volúmenes de las crecientes de la estación La Cuña, México, con la *FC* de Gumbel-Hougaard.

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

Considerando que en el cauce y cerca del sitio de la estación de aforos La Cuña se va a construir un embalse que de acuerdo con la capacidad que se le asigne puede ser pequeño, mediano o grande, entonces se requieren los periodos de retorno bivariados de tipo AND de 500, 1 000, 5 000 y 10 000 años (Campos-Aranda, 2008). Tales magnitudes se estiman con la Ecuación (36). Para ello, de manera arbitraria se seleccionan volúmenes y gastos pico, a fin de obtener sus probabilidades de no excedencia marginales (ecuaciones (54) y (55)) y conjunta (Ecuación (7)) con la FC de Gumbel-Hougaard. En la Tabla 9 se muestran resultados para definir las cuatro gráficas de la Figura 1.

Tabla 9. Parejas de gasto máximo y volumen anual utilizadas para definir las gráficas del periodo de retorno conjunto tipo AND con la *FC* de Gumbel-Hougaard, en las crecientes de la estación hidrométrica La Cuña, México.

T'_{XY} 500 años		T'_{XY} 1 000 años		T'_{XY} 5 000 años		T'_{XY} 10 000 años	
<i>V</i> Mm ³	<i>Q</i> m ³ /s						
0	3 920	0	4 844	0	7 784	0	9 489
500	3 919	500	4 843	1 000	7 782	1 500	9 489
750	3 918	1 000	4 841	1 500	7 782	2 500	9 485
1 000	3 906	1 250	4 830	2 000	7 773	2 600	9 480
11 00	3 892	1 400	4 813	2 200	7 763	2 700	9 478
12 50	3 847	1 500	4 791	2 400	7 742	2 800	9 469
13 00	3 818	1 550	4 776	2 500	7 727	2 900	9 459
1 400	3 720	1 600	4 756	2 600	7 705	3 000	9 449
1 450	3 630	1 650	4 729	2 700	7 678	3 100	9 436
1 500	3 475	1 700	4 695	2 800	7 638	3 200	9 421
1 525	3 342	1 750	4 649	2 900	7 585	3 300	9 400
1 540	3 222	1 800	4 586	3 000	7 511	3 400	9 370
1 550	3 108	1 850	4 495	3 100	7 405	3 500	9 335
1 560	2 938	1 875	4 433	3 200	7 241	3 600	9 290
1 565	2 809	1 900	4 355	3 300	6 963	3 700	9 232
1 570	2 604	1 925	4 252	3 325	6 860	3 800	9 148
1 575	2 039	1 950	4 104	3 350	6 731	3 900	9 049
1 576	0	1 975	3 862	3 375	6 566	4 000	8 898
		2 000	3 244	3 400	6 340	4 050	8 800
		2 007	0	3 425	5 981	4 100	8 674
				3 450	5 183	4 150	8 515
				3 459	0	4 200	8 306
						4 250	7 996
						4 300	7 482
						4 325	6 935
						4 350	0

Contraste de los periodos de retorno bivariados

En la Tabla 10 se muestran los cálculos realizados para llevar a cabo la verificación de la Ecuación (37). Se observa y comprueba que en todos los casos el T_{XY} es menor de los periodos de retorno univariados (T) y por el contrario el T'_{XY} siempre es mayor.

Tabla 10. Periodos de retorno univariados y bivariados estimados con la FC de Gumbel-Hougaard en las crecientes de la estación hidrométrica La Cuña, México.

$T = T_x = T_y$	Q_T	V_T	$C[F_x(x), F_y(y)]$	T_{XY}	T'_{XY}
50	1 835	663	0.9757031	41.2	63.3
100	2 335	873	0.9878718	82.5	127.1
500	3 920	1 576	0.9975724	411.9	636.3
1 000	4 844	2 007	0.9987859	823.7	1 272.5
5 000	7 783	3 459	0.9997571	4 116.9	6 364.6
10 000	9 489	4 350	0.9998783	8 216.9	12 700.4

Selección de los eventos de diseño

En la Figura 1 o en la Tabla 9 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; se adopta por seguridad el que genera las condiciones más críticas o severas. Para formar cada hidrograma de diseño existen métodos teóricos y empíricos (Aldama, 2000; Aldama *et al.*, 2006; Campos-Aranda, 2008; Xiao, Guo, Liu, & Fang, 2008; Gómez *et al.*, 2010; Gräler *et al.*, 2013).

Contraste con resultados de la GVE bivariada

Campos-Aranda (2022) procesó los registros de la Tabla 1 con base en la distribución GVE bivariada, ajustada por medio de máxima verosimilitud; por lo cual, sus funciones marginales son de tipo GVE y sus parámetros de ajuste se involucran en el mejor ajuste logrado. En la Tabla 11, en los renglones 1 y 2, se muestran las predicciones univariadas conjuntas que se obtuvieron.

Tabla 11. Predicciones obtenidas con las distribuciones GVE bivariada y Kappa como marginales de las crecientes anuales de la estación hidrométrica La Cuña, México.

Renglón	Reg.	FDP	Periodos de retorno univariados en años					
			50	100	500	1 000	5 000	10 000
1	Q	GVEb	1 550	1 958	3 254	4 009	6 418	7 823
2	V	GVEb	568	739	1 317	1 673	2 877	3 619
3	Q	Kappa	1 835	2 335	3 920	4 844	7 784	9 489
4	V	Kappa	663	873	1 576	2 007	3 459	4 350

Por otra parte, en los renglones 3 y 4 de la Tabla 11 se tienen las predicciones adoptadas en la Tabla 3 y Tabla 4 anteriores. Para los periodos de retorno indicados se observa que todas las nuevas parejas de predicciones de Q y V son mayores, lo anterior repercute en la definición de las parejas críticas en la Figura 1, que ahora son más grandes, en comparación con las obtenidas al aplicar la distribución GVE bivariada.

Conclusiones

La ventaja fundamental de usar las *funciones Cópula (FC)* en el análisis de frecuencias de crecientes bivariado consiste en poder construir fácilmente la distribución de probabilidades *conjunta* con base en unas distribuciones univariadas marginales iguales o diferentes, incluso de poblaciones mezcladas, previa estimación de la dependencia entre las variables aleatorias gasto máximo (Q) y volumen (V) de las crecientes anuales.

El *enfoque práctico de aplicación* de las *FC* requiere de una minuciosa selección de las distribuciones marginales y por ello se empleó el diagrama de cocientes L , para buscar las tres mejores distribuciones de probabilidad y luego contrastarlas con base en sus errores de ajuste y magnitud de sus predicciones, y así adoptar las más representativas para los registros de Q y V disponibles. En tal contraste se incluyeron dos distribuciones de gran versatilidad, por tener 4 y 5 parámetros de ajuste: los modelos Kappa y Wakeby.

Este *enfoque práctico de aplicación* de las *FC* utiliza Cópulas de un solo parámetro de ajuste (θ), que se estima a partir del cociente tau de Kendall o del coeficiente rho de Spearman, que se calculan con el registro conjunto de Q y V . Tal enfoque primero estima a λ_U^{CFG} o dependencia observada en la cola derecha del registro disponible. Después se busca una *FC* que reproduzca tal valor de λ_U^{CFG} . También se aplican *FC* que no tienen dependencia significativa en su cola derecha, como las de Clayton, Frank, Plackett y Raftery, para poder comparar y juzgar la calidad del ajuste de la *FC* adoptada previamente.

La aplicación numérica descrita en los 55 datos anuales de Q y V de las crecientes registradas en la estación hidrométrica La Cuña de la Región Hidrológica No. 12-3 de México mostró una reproducción fidedigna de las probabilidades bivariadas empíricas y teóricas a través de la *FC* de Gumbel-Hougaard, con un coeficiente de correlación lineal de 0.9964. Por otra parte, en la Figura 1, relativa a los *periodos de retorno conjuntos de diseño* de tipo AND, se pueden definir infinitas parejas de Q y V críticas, por estar en la región curva de cada gráfica.

Agradecimientos

Se agradecen a los árbitros anónimos F, G y H sus observaciones y correcciones sugeridas, las cuales permitieron enmendar omisiones importantes teóricas y prácticas. Sus comentarios ayudaron a mejorar la redacción general y a destacar las aportaciones del enfoque propuesto para la aplicación de las funciones Cópula en los análisis de frecuencias bivariados de crecientes anuales.

Referencias

- 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.
- Asquith, W. H. (2011). Chapter 6. L-moments. In: *Distributional analysis with L-moments statistics using the R environment for statistical computing* (pp. 87-122). Lubbock, USA: Copyright by William H. Asquith.
- Bobée, B. (1975). The Log-Pearson type 3 distribution and its application to Hydrology. *Water Resources Research*, 11(5), 681-689.
- 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. (2008). Procedimiento para revisión (sin hidrometría) de la seguridad hidrológica de presas pequeñas para riego. *Agrociencia*, 42(5), 551-563.
- Campos-Aranda, D. F. (noviembre-diciembre, 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: 10.24850/j-tyca-13-6-11.
- Campos-Aranda, D. F. (enero-febrero, 2023). Análisis de frecuencias comparativo con momentos L entre la distribución Kappa y seis de aplicación generalizada. *Tecnología y ciencias del agua*, 14(1), 432-469. DOI: <https://doi.org/10.24850/j-tyca-14-01-05>
- 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.
- Dupuis, D. J. (2007). Using Copulas in hydrology: Benefits, cautions, and issues. *Journal of Hydrologic Engineering*, 12(4), 381-393. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(381)

- Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40(1), 1-12. DOI: 10.1029/2003WR002456
- Frahm, G., Junker, M., & Schmidt, R. (2005). Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 37(1), 80-100. DOI: 10.1016/j.insmatheco.2005.05.008
- Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about Copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4), 347-368. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(347)
- Genest, C., & Chebana, F. (2017). Copula modeling in hydrologic frequency analysis. In: Singh, V. P. (ed.). *Handbook of applied hydrology* (2nd ed.) (pp. 30.1-30.10). 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
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5), 1049-1054.
- Grimaldi, S., & Serinaldi, F. (2006). Design hyetograph analysis with 3-copula function. *Hydrological Sciences Journal*, 51(2), 223-238. DOI: 10.1623/hysj.51.2.223
- Hosking, J. R. M. (1994). The four-parameter Kappa distribution. *IBM Journal of Research and Development*, 38(3), 251-258.
- Hosking, J. R., & Wallis, J. R. (1997). Appendix: *L*-moments for some specific distributions. In: *Regional frequency analysis. An approach based on L-moments*. (pp. 191-209). Cambridge, UK: Cambridge University Press.
- Kite, G. W. (1977). Chapter 12. Comparison of frequency distributions. In: *Frequency and risk analyses in hydrology* (pp. 156-168). Fort Collins, USA: Water Resources Publications.
- Kjeldsen, T. R., Ahn, H., & Prosdocimi, L. (2017). On the use of a four-parameter kappa distribution in regional frequency analysis. *Hydrological Sciences Journal*, 62(9), 1354-1363. DOI: 10.1080/02626667.2017.1335400

- Kottegoda, N. T., & Rosso, R. (2008). Theme 3.5. Copulas. In: *Applied statistics for civil and environmental engineers* (2nd ed.) (pp. 154-157). 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. 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.
- Nelsen, R. B. (2006). *An introduction to Copulas. Lecture notes in statistics 139*. (2nd ed.). New York, USA: Springer.
- Poulin, A., Huard, D., Favre, A. C., & Pugin, S. (2007). Importance of tail dependence in bivariate frequency analysis. *Journal of Hydrologic Engineering*, 12(4), 394-403. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(394)
- Rao, A. R., & Hamed, K. H. (2000). Chapter 1. Introduction. Chapter 3. Probability weighted moments and L-moments. In: *Flood frequency analysis* (pp. 1-21, 53-72). Boca Raton, USA: CRC Press.
- 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. (2013). Multivariate extreme value methods. In: AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (eds.). *Extremes in a changing climate* (pp. 115-162). London, UK: Springer.
- Salvadori, G., De Michele, C., Kottegoda, N. T., & Rosso, R. (2007). Appendix B. Dependence. Appendix C: Families of Copulas (pp. 219-232, 233-269). In: *Extremes in nature. An approach using Copulas*. 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.
- Sklar, A. (1959). Functions de repartition à n dimensions et leur marges. *Publications de l'Institut de statistique de l'Université de Paris*, 8, 229-231.
- 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* (2nd ed.) (pp. 76.1-76.8). 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). 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* (2nd ed.). (pp. 78.1-78.10). New York, USA: McGraw-Hill Education.
- Volpi, E., & Fiori, A. (2012). Design event selection in bivariate hydrological frequency analysis. *Hydrological Sciences Journal*, 57(8), 1506-1515. DOI: 10.1080/02626667.2012.726357
- 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
- WRC, Water Resources Council. (1977). *Guidelines for determining flood flow frequency* (revised edition) (Bulletin #17A) Hydrology Committee. Washington, DC, USA: Water Resources Council.
- Xiao, Y., Guo, S., Liu, P., & Fang, B. (2008). A new design flood hydrograph method based on bivariate joint distribution. In: *Hydrological sciences for managing water resources in the Asian developing world* (pp. 75-82). UK: IAHS Publication.
- 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. (2000). Joint probability distribution of annual maximum storm peaks and amounts as represented by daily rainfalls. *Hydrological Sciences Journal*, 45(2), 315-326. DOI: 10.1080/02626660009492327
- Yue, S., & Rasmussen, P. (2002). Bivariate frequency analysis: Discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14), 2881-2898. DOI:10.1002/hyp.1185
- Yue, S., & Hashino, M. (2007). Probability distribution of annual, seasonal and monthly precipitation in Japan. *Hydrological Sciences Journal*, 52(5), 863-877. DOI: 10.1623/hysj.52.5.863
- Yue, S., & Wang, C. Y. (2004). A comparison of two bivariate extreme value distributions. *Stochastic Environmental Research and Risk Assessment*, 18(2), 61-66. DOI. 10.1007/s00477-003-0124-x
- 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. (2007). Trivariate flood frequency analysis using the Gumbel-Hougaard Copula. *Journal of Hydrologic Engineering*, 12(4), 431-439. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(431)
- Zhang, L., & Singh, V. P. (2019). Chapter 3. Copulas and their properties. Chapter 11. Flood frequency analysis. In: *Copulas and their applications in water resources engineering* (pp. 62-122, 396-440). Cambridge, UK: Cambridge University Press.