



Aplicación práctica de las funciones Cópula en el Análisis de Frecuencias Bivariado (Q, V) de Crecientes Anuales

Practical application of Copula functions in Bivariate Frequencies Analysis (Q, V) of Annual Floods

Campos-Aranda Daniel Francisco

Profesor Jubilado de la UASLP

Correo: campos_aranda@hotmail.com

<https://orcid.org/0000-0002-9876-3967>

Resumen

El estudio por seguridad hidrológica de los *embalses* requiere la estimación del hidrograma de su creciente de diseño. La manera más simple y aproximada para definir tal gráfica, es a través del análisis de frecuencias bivariado (AFB) para obtener el gasto máximo (Q) y el volumen (V), asociados a un cierto periodo de retorno *conjunto* de diseño. Las *funciones Cópula* modelan la dependencia entre Q y V , a la vez que permiten construir la distribución bivariada con base en las distribuciones univariadas marginales previamente adoptadas. El uso de las funciones Cópula implica dos ventajas, separan el efecto de la dependencia entre Q y V , de las influencias de las distribuciones marginales y son de aplicación simple. Se realizó el AFB de los 52 datos de Q y V anuales de las crecientes de entrada a la Presa Venustiano Carranza (*Don Martín*), del estado de Coahuila, México; aplicando las funciones Cópula de Clayton, Frank, Gumbel-Hougaard y Plackett. El enfoque práctico adoptado utiliza funciones Cópula de un solo parámetro de ajuste y selecciona la más adecuada con base en los errores de ajuste entre probabilidades bivariadas. Debido a su importancia, la búsqueda de las funciones marginales se realizó con la distancia absoluta mínima, a las seis distribuciones que expone el diagrama de cocientes L . Se detalla el cálculo de la gráfica del periodo de retorno conjunto de tipo AND. Por último, se citan varias conclusiones que destacan las ventajas del uso de las funciones Cópula en los AFB de crecientes.

Descriptores: Funciones Cópula, cociente tau de Kendall, coeficiente rho de Spearman, error medio estándar, error absoluto medio, periodos de retorno conjuntos, eventos críticos.

Abstract

The hydrological safety study of *reservoirs* requires the estimation of the hydrograph of its design flood. The simplest and most approximate way to define such graph is through the bivariate frequency analysis (AFB acronym in Spanish) to obtain the maximum flow (Q) and the volume (V), associated with a certain *joint* design return period. *Copula functions* model the dependence between Q and V , at the same time, they allow the construction of the bivariate distribution based on the previously adopted marginal univariate distributions. The use of Copula functions implies two advantages, they separate the effect of the dependence between Q and V from the influence of the marginal distributions, and they are simple to apply. The AFB of the 52 annual Q and V data of the inflow floods to the Venustiano Carranza Dam (*Don Martín*), from the state of Coahuila, Mexico, was carried out; applying the Copula functions of Clayton, Frank, Gumbel-Hougaard and Plackett. The practical approach adopted uses Copula functions of a single fit parameter and selects the most appropriate one based on the fit errors between bivariate probabilities. Due to their importance, the search for the marginal functions was carried out with the minimum absolute distance, for the six distributions that the diagram of quotients L exposes. The calculation of the graph of the joint return period of type AND is detailed. Finally, conclusions highlight the advantages of using Copula functions in flood AFBs.

Keywords: Copula functions, Kendall's tau quotient, Spearman's rho coefficient, mean standard error, mean absolute error, joint return periods, critical events.

INTRODUCCIÓN

GENERALIDADES

Los *embalses* son quizás la obra hidráulica más importante, ya sean de aprovechamiento o de control de crecientes. Su diseño hidrológico por seguridad, requiere estimar el *hidrograma de su Creciente de Diseño*, el cual muestra cómo evolucionó el gasto máximo durante tal evento y por ello, está definido por cuatro variables: el tiempo transcurrido hasta el gasto máximo (*tp*), el valor de este (*Q*), el volumen escurrido (*V*) y su duración total (*D*). Aldama (2000) demostró que los embalses no son sensibles al valor del *tp* y como el *V* y la *D* están correlacionados, 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 diseño.

Los análisis bivariados de las crecientes comenzaron a inicios de este siglo, con los trabajos de Yue (2000) y de Yue & Rasmussen (2002) y estuvieron limitados por la escasa disponibilidad de *funciones de distribución de probabilidades* (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) han destacado que este tipo de distribuciones tienen las siguientes tres desventajas: 1) sus distribuciones marginales pertenecen a la misma familia; 2) la extensión más allá del caso bivariado, no son simples o no existen y 3) los parámetros de ajuste de las distribuciones marginales también se usan para modelar la dependencia entre las variables aleatorias *Q* y *V*.

Las limitaciones que tienen las distribuciones bivariadas existentes, son fácilmente superadas usando las llamadas *funciones Cópula* (FC), las cuales permiten modelar la dependencia entre las variables *Q* y *V*, además de conjuntar las distribuciones marginales univariadas adoptadas previamente, para construir la nueva distribución bivariada (Shiau *et al.*, 2006; Sraj *et al.*, 2015; Genest & Chebana, 2017; Zhang & Singh, 2019).

Con el *enfoque de aplicación práctica* de las FC, se intenta una similitud operativa con los análisis de frecuencias de crecientes (AFC) bivariados, pero suprimiendo las desventajas citadas, al poder utilizar distribuciones marginales diferentes y de cualquier tipo; además, de tener enorme facilidad para construir la distribución bivariada con las FC. Lo anterior se logra utilizando FC que ya han sido probadas en los AFC, y no se aplica la simulación aleatoria para contraste de la FC que se prueba, sino que se selecciona con base en los indicadores de ajuste y su dependencia en el extremo derecho del ajuste.

OBJETIVOS

Los objetivos de este estudio son los cuatro siguientes:

- 1) Exponer de manera breve los conceptos operativos de las FC
- 2) Enlistar las FC de cada familia que han sido utilizadas en los AFC
- 3) Usar y probar las FC que son aplicables en el ejemplo numérico descrito y
- 4) Estimar y contrastar los periodos de retorno univariados y los bivariados

La aplicación numérica se realiza para los 52 gastos máximos y volúmenes de las crecientes de entrada a la Presa Venustiano Carranza (*Don Martín*), del estado de Coahuila, 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. Como en este estudio se 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:

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

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; Meylan *et al.*, 2012; Genest & Chebana, 2017; Zhang & Singh, 2019; Chowdhary & Singh, 2019).

FAMILIAS DE CÓPULAS

Son conjuntos de estas, en los cuales, las Cópulas que los integran se asemejan en cuanto a su forma. Las FC que han sido desarrolladas se han clasificado en cuatro clases (Meylan *et al.*, 2012; Chowdhary & Singh, 2019): de Arquímedes, de valores extremos, elípticas y misce-

láneas. También se clasifican en Cópulas de un parámetro o de varios, dependiendo de la exhaustividad con la cual la estructura de la dependencia entre las variables está definida.

Las Cópulas de Arquímedes han tenido aplicación extensa debido a su construcción simple, un solo parámetro, rango amplio y aceptación de varios 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 tres Cópulas de Arquímedes, las cuales aceptan dependencia negativa y positiva (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019).

1. Ali-Mikhail-Haq: Esta Cópula algunas veces se abrevia AMH y tiene un rango limitado para la dependencia, que se restringe a τ_n variando de -0.1817 a $1/3$:

$$C(u, v) = \frac{u \cdot v}{1 - \theta(1-u)(1-v)} \quad \text{con } -1 \leq \theta \leq 1 \quad (2)$$

La relación de θ con el cociente tau de Kendall es la siguiente:

$$\tau_n = 1 - \frac{2[\theta + (1-\theta)^2 \ln(1-\theta)]}{3\theta^2} \quad (3)$$

2. Frank:

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

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 evaluó con integración numérica, ratificando sus resultados con los valores tabulados por Stegun (1972).

3. Clayton: Esta función Cópula es llamada de Cook-Johnson por Zhang y Singh (2006). Sraj *et al.* (2015) la citan con los dos nombres.

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \quad \text{con } \theta \geq 0 \quad (7)$$

La relación de θ con el cociente tau de Kendall es la siguiente:

$$\tau_n = \frac{\theta}{\theta + 2} \quad (8)$$

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

4. Gumbel-Hougaard:

$$C(u, v) = \exp \left\{ -[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta} \right\} \quad \text{con } \theta \geq 1 \quad (9)$$

La relación de θ con el cociente tau de Kendall es la siguiente:

$$\tau_n = \frac{\theta - 1}{\theta} \quad (10)$$

Por último, se citan dos Cópulas de clase miscelánea, las cuales aceptan dependencia positiva y negativa, pero la primera tiene un rango bastante limitado.

5. Farlie-Gumbel-Morgenstern: Esta Cópula algunas veces se abrevia FGM, tiene un rango limitado para la dependencia, que se restringe a τ_n variando de $-2/9$ a $2/9$:

$$C(u, v) = u \cdot v + \theta \cdot u \cdot v(1-u)(1-v) \quad \text{con } -1 \leq \theta \leq 1 \quad (11)$$

La relación de θ con el cociente tau de Kendall y con la rho de Spearman son las siguientes:

$$\tau_n = \frac{2\theta}{9} \quad (12)$$

$$\rho_n = \frac{\theta}{3} \quad (13)$$

6. Plackett: Cuando la dependencia es positiva $\theta > 1$ cuando es negativa $\theta < 1$:

$$C(u, v) = \frac{1 + (\theta - 1)(u + v)}{2(\theta - 1)} - \frac{\sqrt{[1 + (\theta - 1)(u + v)]^2 - 4uv\theta(\theta - 1)}}{2(\theta - 1)} \quad \text{con } 0 \leq \theta \leq \infty \quad (14)$$

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

$$\rho_n = \frac{(\theta+1)}{(\theta-1)} - \frac{2\theta}{(\theta-1)^2} \ln(\theta) \quad (15)$$

MEDIDAS DE ASOCIACIÓN

CONCORDANCIA

Como la 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* (*c*) con otra, cuando sus grandes valores están asociados a 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; en cambio, unas variables con correlación lineal inversa serán *discordantes* (*d*), pues a grandes valores de una, le corresponden pequeños valores de la otra (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019).

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, la *concordancia* (Salvadori *et al.*, 2007).

COCIENTE TAU DE KENDALL

Mide la probabilidad de tener parejas concordantes; por lo cual, es el *cociente* del número de casos favorables entre el número de casos posibles. El numerador es *c-d* y el denominador *c+d*; siendo, *c* las parejas concordantes y *d* las discordantes, evaluadas en la ecuación 16. Su expresión para estimarlo con observaciones bivariadas es (Zhang & Singh, 2006, 2019):

$$\tau_n = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{signo}[(x_i - x_j)(y_i - y_j)] \quad (16)$$

en la ecuación anterior, *n* es el número de observaciones y el signo[·] es +1 si tales parejas son concordantes, es decir > 0 y -1 si son discordantes, o sea < 0.

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 coefi-

ciente de correlación (r_{xy}). Su expresión para estimarlo en un registro bivariado de tamaño *n* es la siguiente (Chowdhary & Singh, 2019; Zhang & Singh, 2019):

$$\rho_n = \frac{12}{n(n+1)(n-1)} \sum_{i=1}^n R_i \cdot S_i - 3 \frac{n+1}{n-1} \quad (17)$$

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 o con el coeficiente rho. 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). En las ecuaciones (3), (5) y (15) se procede por tanteos, para obtener θ ; en cambio, en las expresiones (8), (10), (12) y (13) se despeja su valor.

ESTIMACIÓN DE LAS PROBABILIDADES EMPÍRICAS

Las probabilidades empíricas bivariadas se estimaron con base en la fórmula de Gringorten, aplicada por Yue y Rasmussen (2002) y Zhang & Singh (2019). Tal fórmula es:

$$p = \frac{i-0.44}{n+0.12} \quad (18)$$

en la cual, *i* es el número de cada dato, cuando están ordenados de manera progresiva y *n* su número total. 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. 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 procesado.

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_i , o combinaciones de *Q* y *V* menores. Calculados los *n* valores de NM_i , se aplica la ecuación anterior para estimar la probabilidad empírica bivariada observada:

$$F_E(x, y) = w_i^0 = \frac{NM_i - 0.44}{n + 0.12} \quad (19)$$

SELECCIÓN DE LA FUNCIÓN CÓPULA

ERRORES ESTADÍSTICOS DE AJUSTE

Permiten aplicar un enfoque simple de selección de la FC, al comparar las probabilidades empíricas observadas (w_i^o) con las teóricas calculadas (w_i^c) con tal 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):

$$EME = \sqrt{\frac{1}{n} \sum_{i=1}^n (w_i^o - w_i^c)^2} \tag{20}$$

$$EMA = \frac{1}{n} \sum_{i=1}^n |w_i^o - w_i^c| \tag{21}$$

$$EAM = \max_{i=1:n} |w_i^o - w_i^c| \tag{22}$$

DEPENDENCIA EN EL EXTREMO SUPERIOR

Otro criterio que se aplica para seleccionar una FC, es el basado en la magnitud de la dependencia en la cola superior, 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 cero (Chowdhary & Singh, 2019).

En relación con las FC expuestas, las de AMH, Frank, FGM y Plackett, tienen dependencias insignificantes en sus colas: por ello, $\lambda_L = 0$ y $\lambda_U = 0$. En cambio, la Cópula Gumbel-Hougaard tiene dependencia significativa en la cola superior, igual a: $\lambda_U = 2 - 2^{1/\theta}$ y la Cópula de Clayton la tiene en su cola inferior e igual a: $\lambda_L = 2^{-1/\theta}$. Cópulas no expuestas, con $\lambda_U > 0$ son las de Galambos, Genest-Ghoudi, Hüsler-Reiss y Joe (Salvadori *et al.*, 2007).

Para intentar la estimación del valor de λ_U que muestran los datos disponibles, se define primero la *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 tales variables y su escalamiento con el factor $1 / (n+1)$ genera una serie de puntos U_i y V_i en el cuadrado unitario $[0,1]^2$, formando el dominio de la Cópula Empírica (Chowdhary & Singh, 2019). Poulin *et al.* (2007), proponen el siguiente estimador basado en la Cópula empírica, su expresión es:

$$\lambda_U^{CFG} = 2 - 2 \exp \left\{ \frac{1}{n} \sum_{i=1}^n \ln \left[\sqrt{\ln \frac{1}{U_i} \cdot \ln \frac{1}{V_i}} / \ln \left(\frac{1}{\max(U_i, V_i)^2} \right) \right] \right\} \tag{23}$$

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 ó 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):

$$T_{XY} = \frac{1}{P(X>x \text{ or } Y>y)} = \frac{1}{1 - F_{X,Y}(x,y)} = \frac{1}{1 - C[F_X(x), F_Y(y)]} \tag{24}$$

en la cual, $F_{X,Y}(x,y)$ es la FDP bivariada y $C[F_X(x), F_Y(y)]$ es la FC seleccionada.

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

$$T'_{XY} = \frac{1}{P(X>x \text{ and } Y>y)} = \frac{1}{F'_{X,Y}(x,y)} = \frac{1}{1 - F_X(x) - F_Y(y) + C[F_X(x), F_Y(y)]} \tag{25}$$

Aldama (2000) obtiene la expresión $F'_{XY}(x, y)$ de la probabilidad de excedencia bivariada tipo AND, 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 numéricamente un evento bivariado (X, Y), 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):

$$T_{XY} \leq \min[T_X, T_Y] \leq \max[T_X, T_Y] \leq T'_{XY} \tag{26}$$

Siendo:

$$T_X = \frac{1}{F'_X(x)} = \frac{1}{1 - F_X(x)} \tag{27}$$

$$T_Y = \frac{1}{F'_Y(y)} = \frac{1}{1 - F_Y(y)} \tag{28}$$

En las ecuaciones 27 y 28, $F'_X(x)$ y $F'_Y(y)$ son las probabilidades de excedencia univariadas.

EVENTOS CRÍTICOS DEL T'_{xy}

Volpi y Fiori (2012) destacan que la gráfica del periodo de retorno bivariado de tipo AND, mostrada posteriormente 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* y la *correcta*. 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.

SELECCIÓN DE LAS DISTRIBUCIONES MARGINALES

MOMENTOS L Y SUS COCIENTES

Por limitaciones de espacio no se exponen las ecuaciones de los momentos L, así como las expresiones de los cocientes L de asimetría (t_3) y de curtosis (t_4); pero ambas se pueden consultar en Campos-Aranda (2014) y en Hosking & Wallis (1997) en su capítulo 2.

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):

$$t_4^{LOG} = 0.16667 + 0.83333 \cdot t_3^2 \tag{29}$$

Pareto Generalizada (PAG):

$$t_4^{PAG} = 0.20196 \cdot t_3 + 0.95924 \cdot t_3^2 - 0.20096 \cdot t_3^3 + 0.04061 \cdot t_3^4 \tag{30}$$

Log-Normal (LGN):

$$t_4^{LGN} = 0.12282 + 0.77518 \cdot t_3^2 + 0.12279 \cdot t_3^4 - 0.13638 \cdot t_3^6 + 0.11368 \cdot t_3^8 \tag{31}$$

Pearson tipo III (PT3):

$$t_4^{PT3} = 0.12240 + 0.30115 \cdot t_3^2 + 0.95812 \cdot t_3^4 - 0.57488 \cdot t_3^6 + 0.19383 \cdot t_3^8 \tag{32}$$

General de Valores Extremos (GVE):

$$t_4^{GVE} = 0.10701 + 0.11090 \cdot t_3 + 0.84838 \cdot t_3^2 - 0.06669 \cdot t_3^3 + SF \tag{33}$$

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

Utilizando los logaritmos de los datos se obtienen los cocientes L logarítmicos y entonces se puede utilizar la expresión 32 para valorar la cercanía 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):

$$DA = t_4(t_3^{obs}) - t_4^{obs} \tag{34}$$

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 29 a 33), 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.

TRES DISTRIBUCIONES DE PROBABILIDAD SELECCIONADAS

Con el criterio anterior se definieron las tres mejores distribuciones para cada registro de gasto máximo y volumen anuales. Con excepción de la distribución LP3, que se ajustó con el método de momentos en el dominio logarítmico (WRC, 1977), las otras cinco se ajustaron con el método de los momentos L (Hosking & Wallis, 1997).

ERRORES DE AJUSTE

Este criterio (Chai & Draxler, 2014) complementa al anterior. Cambiando en las ecuaciones 20 y 21, las probabilidades observadas por los datos ordenados de la serie analizada (x_i, y_i) y las probabilidades calculadas por los valores estimados (\hat{x}_i, \hat{y}_i) 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 se obtienen para la probabilidad de no excedencia, asignada a los datos con la ecuación 18.

Tabla 1. Gastos máximos, volúmenes y sus números de orden bivariados de las crecientes que ingresaron a la Presa Venustiano Carranza (*Don Martín*), 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
1930	241.0	38.76	19	1956	68.6	13.75	3
1931	89.2	53.59	8	1957	451.2	83.73	36
1932	1071.2	403.78	49	1958	1342.1	529.11	51
1933	203.2	62.10	21	1959	521.6	94.18	38
1934	90.7	11.48	3	1960	99.5	33.31	10
1935	431.2	75.40	34	1961	569.4	161.03	42
1936	62.0	64.75	4	1962	92.8	12.03	4
1937	138.3	19.64	11	1963	340.8	33.13	15
1938	166.3	35.01	17	1964	586.4	156.79	42
1939	134.2	16.99	9	1965	214.6	38.78	19
1940	182.1	21.18	12	1966	76.8	19.12	4
1941	252.1	43.43	21	1967	425.6	215.79	38
1942	339.7	63.62	28	1968	119.6	33.24	11
1943	284.8	35.02	19	1969	50.9	25.81	3
1944	655.9	269.58	46	1970	511.3	246.39	42
1945	146.7	28.04	13	1971	4320.7	983.02	52
1946	243.4	93.00	27	1972	214.2	129.87	25
1947	339.4	63.61	27	1973	449.2	65.18	33
1948	238.4	99.71	27	1974	756.5	244.05	46
1949	97.7	55.93	11	1975	751.6	169.43	44
1950	280.2	71.81	28	1976	614.3	466.05	46
1954	115.5	34.09	12	1995	80.6	13.48	3
1952	114.7	14.87	7	1996	47.6	4.69	2
1953	238.4	99.71	27	1997	87.4	15.96	5
1954	367.0	39.17	23	1998	29.5	3.75	1
1955	213.5	56.00	21	1999	85.0	64.22	7

CRECIENTES DE ENTRADA A LA PRESA VENUSTIANO CARRANZA, MÉXICO

Aldama *et al.* (2006) muestran los 52 gastos máximos y sus volúmenes anuales de las crecientes que entran a la Presa Venustiano Carranza (*Don Martín*), en el estado de Coahuila, México. Su área de cuenca es de 31034 km². Tal información se tiene en la Tabla 1.

RESULTADOS Y SU VERIFICACIÓN

SELECCIÓN 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 (Rao &

Hamed, 2000; Meylan *et al.*, 2012), cuyo estadístico *U* condujo a valores de 0.726 y 1.054, para los gastos y volúmenes de la Tabla 1.

DISTANCIAS ABSOLUTAS CALCULADAS

En la Tabla 2 se han concentrado los resultados de las ecuaciones 29 a 33, para las seis FDP aplicadas a cada registro de datos de la Tabla 1.

DISTRIBUCIÓN DE LOS GASTOS MÁXIMOS ANUALES

En la Tabla 3 se exponen errores de ajuste y predicciones obtenidas con las tres mejores FDP aplicadas al registro de gastos máximos de la Tabla 1. Se observa que las distribuciones GVE y LOG, conducen a los errores de ajuste más bajos del tipo *EAM* y relativamente reducidos del *EEA*. Por otra parte, la tercera mejor distribu-

Tabla 2. Distancia absoluta en el diagrama de cocientes L (Hosking & Wallis, 1997) y números de orden de la mejor distribución para las series indicadas

Gastos máximos anuales (m ³ /s)		Volúmenes anuales (Mm ³)	
$t_3 = 0.55039$	$\ln t_3 = 0.05374$	$t_3 = 0.57344$	$\ln t_3 = 0.04605$
$t_4 = 0.41275$	$\ln t_4 = 0.10787$	$t_4 = 0.38702$	$\ln t_4 = 0.15555$
Distribución LOG	0.0064 (2)	Distribución LOG	0.0537 (5)
Distribución GVE	0.0006 (1)	Distribución GVE	0.0493 (4)
Distribución LGN	0.0467 (5)	Distribución LGN	0.0005 (1)
Distribución PT3	0.1256 (6)	Distribución PT3	0.0802 (6)
Distribución PAG	0.0408 (4)	Distribución PAG	0.0107 (2)
Distribución LP3	0.0154 (3)	Distribución LP3	0.0325 (3)

Tabla 3. Errores de ajuste y predicciones (m³/s) de las tres mejores distribuciones en el registro de gastos máximos anuales de las crecientes de entrada a la Presa Venustiano Carranza (Don Martin), México

FDP ajustada:	EAM (m ³ /s)	EEA (m ³ /s)	Periodos de retorno, en años					
			50	100	500	1000	5000	10000
LOG (2)	59	247	1843	2715	6616	9695	23525	34449
GVE (1)	60	240	1900	2773	6534	9403	21781	31224
LP3 (3)	64	223	2049	2941	6378	8723	17447	23247

ción, la LP3 condujo al EEA más bajo, pero sus predicciones se consideran reducidas. Por lo anterior, se adoptó a la distribución GVE.

Los parámetros de ajuste de ubicación (u_1), escala (a_1) y forma (k_1) de la distribución GVE adoptada son: 160.2069, 137.6956, -0.5178409. La expresión de su FDP es la siguiente (Hosking & Wallis, 1997):

$$F(x) = \exp \left\{ - \left[1 - k_1 \left(\frac{x-u_1}{a_1} \right) \right]^{1/k_1} \right\} \quad (35)$$

DISTRIBUCIÓN DE LOS VOLÚMENES ANUALES

En la Tabla 4, similar a la Tabla 3, se observa que la mejor distribución la LGN, conduce a errores de ajuste bajos, con predicciones reducidas. La tercera mejor opción, la LP3 reporta los errores de ajuste más bajos,

pero sus predicciones son también reducidas. Por lo anterior, se adoptó a la distribución PAG.

Los parámetros de ajuste de ubicación (u_2), escala (a_2) y forma (k_2) de la distribución PAG adoptada son: 6.141945, 56.37617 y -0.4577991. La expresión de su FDP es la siguiente:

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

La aplicación de la ecuación anterior, conduce a valores negativos en el paréntesis rectangular en las parejas de datos números 49 y 51 de la Tabla 1. Lo anterior, significa que el parámetro de ubicación es elevado y por ello, existe el método modificado de momentos L, que corrige tal anomalía (Rao & Hamed, 2000; Campos-Aranda, 2014); el cual conduce a estos nuevos parámetros de

Tabla 4. Errores de ajuste y predicciones (m³/s) de las distribuciones indicadas con el registro de volúmenes anuales de las crecientes de entrada a la Presa Venustiano Carranza (Don Martin), México

FDP ajustada:	EAM (m ³ /s)	EEA (m ³ /s)	Periodos de retorno, en años					
			50	100	500	1000	5000	10000
LGN (1)	11	24	654	926	1877	2463	4385	5516
PAG (2a)	13	28	621	897	2001	2792	5962	8321
LP3 (3)	9	17	685	988	2113	2850	5442	7073
PAG (2b)	14	30	615	864	1798	2429	4797	6397

ajuste: 2.5067, 64.14948 y -0.4038805 ; con las predicciones que se muestran en el renglón final de la Tabla 4.

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: $r_{xy} = 0.9188$, $\tau_n = 0.6471$ y $\rho_n = 0.8274$. Por otra parte, en la Tabla 5 se muestran los indicadores estadísticos de ajuste que se obtuvieron al aplicar las FC de Clayton, Frank, Gumbel-Hougaard (GH) y Plackett. En las ecuaciones 20 a 22, las probabilidades bivariadas empíricas se estimaron con la ecuación 19 y las teóricas con las expresiones 7, 4, 9 y 14.

Con base en los resultados de la Tabla 5, no existe dificultad para seleccionar como mejor FC la Gumbel-Hougaard, para los datos de la Tabla 1, debido a que reporta los indicadores estadísticos más reducidos (mostrados en cursivas), es la única que tiene dependencia en su cola derecha y esta se aproxima a la observada (ecuación 23) de $\lambda_{ii}^{CFG} = 0.744$.

En la Tabla 6 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 Gumbel-Hougaard. También se indican sombreadas las diferencias máximas negativa y positiva. Yue (2000) sugirió un contraste gráfico entre ambas probabilidades,

para ratificar su adopción, este se expone en la Figura 1 para los datos completos de la Tabla 6.

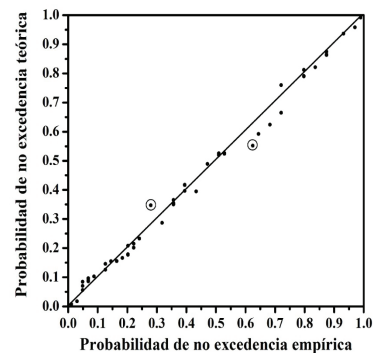


Figura 1. Contraste gráfico de probabilidades conjuntas del gasto pico y volúmenes de las crecientes anuales de entrada a la Presa Venustiano Carranza, México

GRÁFICAS DEL PERIODO DE RETORNO BIVARIADO T'_{XY}

Los periodos de retorno bivariado de tipo AND se estiman con base en la ecuación 25. Se consideró conveniente realizar estimaciones para valores del T'_{XY} de 500, 1000, 5000 y 10000 años. De manera arbitraria se seleccionan volúmenes y gastos pico, para obtener sus probabilidades de no excedencia marginales (ecuaciones 35 y 36) y conjunta (ecuación 9). En la Tabla 7 se muestran resultados para definir las cuatro gráficas de la Figura 2.

Tabla 5. Indicadores estadísticos del ajuste de las funciones Cópula aplicadas a las crecientes de entrada a la Presa Venustiano Carranza (Don Martín), México

Cópula	θ	EME	EAM	EMA	NDP	NDN	MDP	MDN	λ_{ii}
Clayton	3.6667	0.0380	0.0290	0.0873	22	30	0.0873	-0.0699	0.000
Frank	9.3400	0.0281	0.0220	0.0744	28	24	0.0744	-0.0703	0.000
GH	2.8333	0.0254	0.0188	0.0729	27	25	0.0729	-0.0677	0.723
Plackett	30.275	0.0280	0.0210	0.0786	27	25	0.0786	-0.0647	0.000

Significado de los nuevos acrónimos: NDP, NDN número de diferencias positivas y negativas; MDP, MDN máxima diferencia positiva y negativa

Tabla 6. Probabilidades de no excedencia conjuntas y sus diferencias para una parte de las crecientes anuales de entrada a la Presa Venustiano Carranza, México

Núm.	w_i^o	w_i^c	Diferencias	Núm.	w_i^o	w_i^c	Diferencias
1	0.3561	0.3651	-0.0090	30	0.7206	0.6648	0.0559
5	0.0491	0.0843	-0.0351	34	0.2794	0.3471	-0.0677
10	0.1642	0.1556	0.0086	40	0.0491	0.0563	-0.0071
15	0.8741	0.8642	0.0099	44	0.6247	0.5518	0.0729
20	0.2026	0.1790	0.0236	49	0.0299	0.0175	0.0125
25	0.4328	0.3950	0.0378	52	0.1259	0.1460	-0.0202

Tabla 7. Parejas de gasto pico 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 entrada a la Presa Venustiano Carranza, México

T'_{xy} 500 años		T'_{xy} 1000 años		T'_{xy} 5000 años		T'_{xy} 10000 años	
Vol. Mm ³	Qp m ³ /s	Vol. Mm ³	Qp m ³ /s	Vol. Mm ³	Qp m ³ /s	Vol. Mm ³	Qp m ³ /s
0	6534	0	9403	0	21781	0	31224
400	6530	800	9385	1250	21768	2000	31202
500	6526	900	9374	1500	21754	3000	31030
600	6518	1000	9358	1750	21727	3500	30799
800	6486	1200	9308	2000	21688	3750	30633
1000	6416	1400	9219	2250	21625	4000	30432
1100	6356	1600	9066	2500	21534	4200	30244
1200	6272	1700	8954	2750	21404	4400	29982
1250	6217	1800	8808	3000	21216	4600	29663
1300	6152	1900	8612	3250	20948	4800	29270
1400	5979	2000	8351	3400	20742	5000	28783
1500	5721	2100	7986	3500	20576	5100	28481
1600	5312	2150	7747	3600	20390	5200	28149
1650	5002	2200	7450	3750	20042	5300	27789
1700	4558	2250	7073	3900	19611	5400	27358
1750	3816	2300	6568	4000	19252	5500	26897
1775	3133	2350	5828	4100	18836	5600	26342
1790	2328	2400	4486	4200	18328	5700	25659
1795	1754	2410	4000	4300	17714	5800	24895
1798	0	2415	3705	4400	16920	5900	23997
		2420	3260	4500	15885	6000	22840
		2429	0	4600	14430	6050	22147
				4650	13438	6100	21394
				4700	12070	6150	20469
				4725	11164	6200	19347
				4750	9985	6250	17975
				4775	8000	6300	16058
				4797	0	6350	13255
						6397	0

En la Figura 2 o en la Tabla 7 se pueden seleccionar infinitas parejas de Qp 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 de 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).

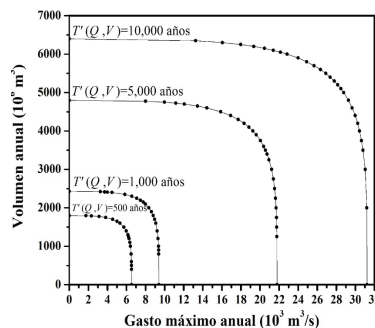


Figura 2. Gráficas de los cuatro periodos de retorno conjunto T'_{xy} de diseño, de las crecientes de entrada a la presa Venustiano Carranza, México

CONTRASTE DE LOS PERIODOS DE RETORNO BIVARIADOS

En la Tabla 8 se muestran los cálculos realizados para llevar a cabo la verificación de la ecuación 26. Se observa y comprueba que en todos los casos el T_{XY} es menor de los T y por el contrario, el T'_{XY} siempre es mayor.

Tabla 8. Diferencias entre los Periodos de retorno univariados y bivariados estimados con la FC de Gumbel-Hougaard, en las crecientes de entrada a la Presa Venustiano Carranza, México

$T = T_x = T_y$	Q_T	V_T	$C[F_x(x), F_y(y)]$	T_{XY}	T'_{XY}
100	2773	864	0.9872451	78.4	138.0
500	6534	1798	0.9974464	391.6	691.4
1000	9403	2429	0.9987227	782.9	1382.9
5000	21781	4797	0.9997446	3915.4	6915.6
10000	31224	6397	0.9998721	7818.6	13831.2

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 *diferentes*, 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 una acuciosa selección de las distribuciones marginales y para ello, se empleó el diagrama de cocientes L, para contrastar las tres mejores distribuciones de probabilidad, según sus errores de ajuste y magnitud de sus predicciones y así adoptar las más representativas para Q y V . Este *enfoque práctico de aplicación* de las FC, utiliza Cópulas de un solo parámetro de ajuste (θ), que se estima con base en el cociente tau de Kendall o con el coeficiente rho de Spearman, observados entre Q y V . Tal enfoque aplica las FC de Clayton, Frank, Gumbel-Hougaard y Plackett, seleccionando la de mejor ajuste a los datos y que además reproduzca la dependencia (λ_U) observada para los registros disponibles. Bajo tales condiciones se adoptó la FC de Gumbel-Hougaard.

La aplicación numérica descrita, con la muestra de 52 datos de las crecientes anuales de entrada a la Presa Venustiano Carranza (*Don Martín*), del estado de Coahuila, México; mostró en la Figura 1, una reproducción fidedigna de las probabilidades conjuntas empíricas y teóricas. La Figura 2, relativa a los *periodos de retorno conjuntos de diseño* de tipo AND, permitirá definir infinitas parejas de Q y V críticas, por estar en la región curva de cada gráfica.

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, Morelos*. México: Instituto Mexicano de Tecnología del Agua. 188 pp.

Campos-Aranda, D. F. (2014). Predicción de crecientes usando la distribución Pareto Generalizada ajustada con tres métodos simples. *Revista Digital Tláloc*, (65), octubre-diciembre.

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. <https://doi.org/10.5194/gmd-7-1247-2014>

Chowdhary, H., & Singh, V. P. (2019). Chapter 11: Multivariate frequency distributions in hydrology, 407-489. En: *Statistical Analysis of Hydrologic Variables*. Teegavarapu, R. S. V., Salas, J. D., & Stedinger, J. R. (eds.). Reston, Virginia, U.S.A. American Society of Civil Engineers. 548 p.

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. <https://doi.org/10.1029/2003WR002456>

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

Hosking, J. R., & Wallis, J. R. (1997). *Regional frequency analysis. An approach based on L-moments*. Chapter 2: L-moments, 14-43, and Appendix: L-moments for some specific distributions, 191-209. Cambridge, England: Cambridge University Press. 224 p.

Meylan, P., Favre, A. C., & Musy, A. (2012). *Predictive hydrology. A frequency analysis approach*. Boca Raton, Florida: CRC Press. 212 p.

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. <https://doi.org/10.1061/ASCE1084-0699200712:4394>
- Rao, A. R., & Hamed, K. H. (2000). *Flood frequency analysis*. Introduction, 1-21, and The generalized Pareto distribution, 271-290. Boca Raton, Florida: CRC Press. 350 p.
- Salvadori, G., De Michele, C., Kottegoda, N. T., & Rosso, R. (2007). *Extremes in nature. An approach using Copulas*. Appendix B: Dependence, 219-232, and Appendix C: Families of Copulas, 233-269. Dordrecht, The Netherlands: Springer. 292 p.
- 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. <https://doi.org/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. <https://doi.org/10.1002/hyp.10145>
- Stegun, I. A. (1972). Miscellaneous functions (Chapter 27, 997-1010). En: *Handbook of Mathematical Functions*. Abramowitz, M., & Stegun, I. A. (eds.). New York: Dover Publications. Ninth printing. 1046 p.
- Volpi, E., & Fiori, A. (2012). Design event selection in bivariate hydrological frequency analysis. *Hydrological Sciences Journal*, 57(8), 1506-1515. <https://doi.org/10.1080/02626667.2012.726357>
- WRC. (1977). *Guidelines for determining flood flow frequency* (Revised edition). Bulletin #17A of the Hydrology Committee. Washington, D. C., U.S.A: Water Resources Council.
- 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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1623/hysj.52.5.863>
- Zhang, L. & Singh, V. P. (2006). Bivariate flood frequency analysis using the Copula method. *Journal of Hydrologic Engineering*, 11(2), 150-164. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:2\(150\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:2(150))
- Zhang, L. & Singh, V. P. (2019). *Copulas and their applications in water resources engineering*. Cambridge, United Kingdom: Cambridge University Press. 603 p.

Cómo citar:

Campos-Aranda, D. F. (2022). Aplicación práctica de las funciones Cópula en el Análisis de Frecuencias Bivariado (Q,V) de Crecientes Anuales. *Ingeniería Investigación y Tecnología*, 23 (03), 1-12. <https://doi.org/10.22201/fi.25940732e.2022.23.3.023>