

Análisis de frecuencias de crecientes trivariados con dependencia regional y funciones Cópula

Trivariate flood frequencies analysis with regional dependence and 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

Las *crecientes de diseño* (CD) permiten dar dimensión por seguridad hidrológica a las obras hidráulicas de protección. Su estimación más confiable se obtiene con el *análisis de frecuencias* (AF) univariado, el cual representa los gastos máximos anuales disponibles, con una función de distribución de probabilidades (FDP) idónea, para estimar las *predicciones* buscadas. En este estudio, el AF se realiza con el enfoque *trivariado*, procesando un registro base de gastos QX y otros dos auxiliares, QY y

QZ , que están correlacionados con el primero y tienen igual amplitud. Se describe con detalle cómo se verifica que los gastos QX , QY y QZ sean *simultáneos*, es decir, que pertenezcan al mismo evento analizado. La FDP *conjunta* trivariada de gastos se obtuvo mediante la *función Cópula* de Gumbel-Hougaard, que mostró excelente ajuste y reprodujo la dependencia observada en los gastos. La aplicación numérica expuesta procesó 43 gastos anuales y se realizó en las estaciones hidrométricas Tempoal como base, y El Cardón y Terrerillos como auxiliares del sistema del río Tempoal de la Región Hidrológica No. 26 (Pánuco), México. Para la búsqueda de las FDP *marginales* idóneas se utilizó el diagrama de cocientes de momentos L , y además se aplicaron para contraste de predicciones las FDP Kappa y Wakeby. Por último, se formulan las conclusiones, las cuales destacan la importancia del enfoque trivariado, basado en la dependencia regional, para validar el comportamiento en magnitudes de las CD estimadas con el enfoque *univariado*.

Palabras clave: funciones Cópula (FC), FC de Frank y Gumbel-Hougaard, FC multivariadas simétricas, FC trivariadas asimétricas, cociente tau de Kendall, dependencia en el extremo superior y observada, periodo de retorno secundario, eventos de diseño.

Abstract

Design floods (DF) give dimension for hydrological security to the hydraulic protection works. The most reliable estimate is obtained through the univariate *frequency analysis* (FA), which represents the maximum annual flows available, with an appropriate probability distribution function (PDF), to estimate the *predictions* sought. In this study, the FA

is carried out with the *trivariate* approach, processing a base record of flows QX and two other auxiliaries, QY and QZ , which are correlated to the first and have the same amplitude. The verification of the *simultaneous* character of the QX , QY and QZ flows (that they belong to the same event analyzed) is described in detail. The *joint* trivariate PDF of flows was obtained using the Gumbel-Hougaard *Copula function*, which showed an excellent fit and reproduced the observed dependency on flows. A numerical application exposed here processed 43 annual flows and was carried out at the hydrometric stations, Tempoal as base, and El Cardón and Terrerillos as auxiliaries of the Tempoal river system of Hydrological Region No. 26 (Pánuco), Mexico. In order to obtain the ideal *marginal* PDFs, the Moment Ratios Diagram L was used and, in addition, the Kappa and Wakeby PDFs were applied to contrast predictions. Finally, conclusions are formulated, which highlight the importance of the trivariate approach, based on regional dependence, to validate the behavior in magnitudes of the DF estimated with the *univariate* approach.

Keywords: Frank and Gumbel-Hougaard *CF*, symmetric multivariate *CF*, asymmetric trivariate *CF*, Kendall's tau ratio, upper tail and observed dependences, secondary return period, design events.

Recibido: 15/02/2023

Aceptado: 16/09/2023

Publicado Online: 11/10/2023

Introducción

Generalidades

La república mexicana se localiza en la zona de influencia de los huracanes o ciclones que se originan en el mar Caribe y en el océano Pacífico, los cuales generan tormentas convectivas locales y extensas tormentas orográficas. Además, la afectan los frentes fríos que son fenómenos meteorológicos de gran alcance espacial. Estos eventos atmosféricos originan *crecientes* o *avenidas máximas* que inundan diversas regiones del país, las cuales producen pérdida de vidas humanas, y enormes daños económicos y ambientales (Aldama, Ramírez, Aparicio, Mejía-Zermeño, & Ortega-Gil, 2006).

El estudio hidrológico básico de estimación de las crecientes se denomina *análisis de frecuencias* (AF), el cual define las *crecientes de diseño* (CD), que son gastos máximos del río asociados con bajas probabilidades de ser excedidos. Las CD permiten dimensionar por seguridad hidrológica diversas obras hidráulicas de protección como muros y diques de contención, puentes, rectificaciones y canalizaciones de ríos, y el drenaje urbano.

Para que la estimación de las CD sea confiable, el registro de gastos máximos anuales procesado debe ser aleatorio, la función de distribución de probabilidades (FDP) utilizada para obtener las *predicciones* buscadas debe ser idónea y el método empleado para su ajuste eficiente. Además, la selección de resultados debe ser objetiva.

Los AF comenzaron a mediados del siglo pasado e inicialmente fueron de carácter *univariado*, procesando por lo general el gasto máximo anual.

Hacia finales de tal siglo comenzaron los AF *bivariados*, con dos enfoques. El primero utilizó otras variables de las crecientes anuales, como su volumen escurrido y la duración total (Goel, Seth, & Chandra, 1998; Yue, Ouarda, Bobée, Legendre, & Bruneau, 1999). El segundo enfoque empleó registros auxiliares, con cercanía regional que mostraban dependencia o correlación. Estos AF se llevaron hasta el nivel *trivariado* (Escalante-Sandoval & Raynal-Villaseñor, 1994).

Los primeros AF bivariados se basaron en una FDP conjunta, que tenía distribuciones marginales iguales (normales o Gumbel) y periodo común de registro en sus variables. Los primeros AF trivariados también aplicaron una FDP conjunta con marginales iguales, pero procesaron sus variables con periodos de registro diferentes. El ajuste de tal distribución conjunta se realizó por máxima verosimilitud, con un proceso algebraico complejo que se resolvió con optimización numérica (Escalante-Sandoval & Raynal-Villaseñor, 2008).

En este estudio, el AF se aborda bajo el enfoque *trivariado*, con un registro base (QX) y dos auxiliares (QY y QZ), que muestran dependencia, es decir, están correlacionados y tienen igual número de años de registro. La FDP que va a representar a la terna de registros se construye con base en sus distribuciones univariadas idóneas previamente adoptadas, por medio de una *función Cópula* (FC). Las predicciones estimadas con la FC adoptada se contrastan contra las obtenidas con una FDP univariada, ajustada al registro completo QX .

Objetivos

Para este estudio se formularon los siguientes cinco *objetivos*: (1) para los análisis trivariados se trabajó con dos familias de *FC* de la clase de Arquímedes: Frank y Gumbel-Hougaard; (2) las familias de *FC* citadas se aplicaron con sus versiones multivariadas denominadas simétricas; (3) también se utilizaron las familias de *FC* seleccionadas, de tipo anidado o asimétricas con dos parámetros de asociación; (4) se estiman eventos de diseño, a partir del periodo de retorno conjunto de Kendall, y (5) la teoría expuesta se aplicó en el sistema del río Tempoal, de la Región Hidrológica No. 26 (Pánuco), México. Se procesó como estación base Tempoal y como auxiliares El Cardón, Los Hules y Terrerillos, con un registro conjunto de 43 crecientes anuales.

Organización del estudio

Debido a lo extenso y diverso de los conceptos teóricos y cálculos involucrados, es conveniente describir su organización para una mejor comprensión del estudio o investigación. Para ello, se tienen tres capítulos: (1) el de funciones Cópula y datos procesados; (2) el de resultados y su discusión, y (3) las conclusiones.

El capítulo siguiente expone un resumen de los aspectos teóricos aplicados en el estudio, por ello comienza con un inciso que cita las ventajas y aspecto operativo de las funciones Cópula (*FC*). Después se presentan las *FC* bivariadas que serán utilizadas, cuyo ajuste emplea el cociente tau de Kendall y su selección se basa en la dependencia

probabilística de su extremo derecho. Esta primera parte teórica concluye con la descripción de las *FC* trivariadas que serán aplicadas.

Enseguida se abordan tres conceptos del análisis de frecuencias, cuya aplicación permite el ajuste y selección de las *FC*: (a) la estimación de las probabilidades empíricas univariadas y trivariadas; (b) la búsqueda de las FDP marginales óptimas, con base en los errores de ajuste, y (c) conceptos y ecuaciones de los periodos de retorno trivariados de tipo OR, AND y secundario, o de Kendall. Por último, en este capítulo se exponen los gastos de las crecientes máximas anuales que serán procesados y se describe cómo se verificó la simultaneidad de tales eventos regionales.

Respecto al capítulo de resultados y su discusión (Figura 1), se esquematiza la secuencia de temas teóricos y sus cálculos con la idea de formular un diagrama de flujo explicativo.

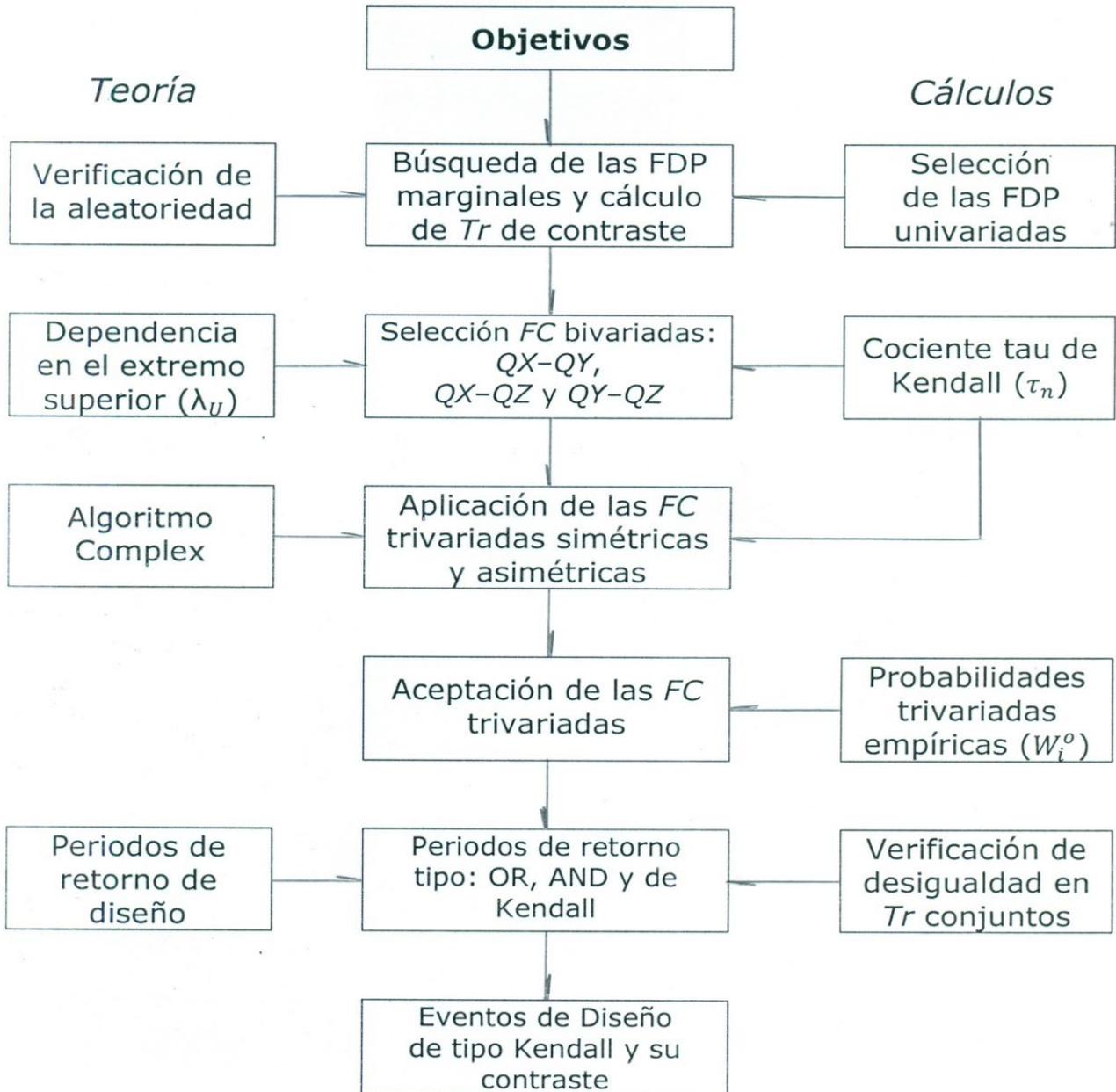


Figura 1. Diagrama de flujo de conceptos teóricos y cálculos, realizados en el capítulo de resultados y su discusión.

Funciones Cópula y datos procesados

Ventajas

La ventaja básica de las *funciones Cópula (FC)* consiste en permitir formar y expresar la *distribución conjunta o multivariada* de variables aleatorias que están correlacionadas, como una función de sus distribuciones marginales, previamente adoptadas. Por lo anterior, una *FC* enlaza o relaciona las distribuciones marginales univariadas para formar la *distribución multivariada*.

Otra ventaja de las *FC* al formar distribuciones multivariadas es el hecho de que ellas separan el efecto de la dependencia o correlación entre las variables aleatorias de los efectos de las distribuciones marginales en la modelación conjunta.

La aplicación de las *FC* ofrece entera libertad para adoptar o seleccionar las distribuciones marginales univariadas que mejor representen a los datos (Salvadori, De Michele, Kottegod, & Rosso, 2007; Meylan, Favre, & Musy, 2012; Genest & Chebana, 2017; Zhang & Singh, 2019).

Familias de Cópulas

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

2012; Genest & Chebana, 2017; Chowdhary & Singh, 2019). Salvadori *et al.* (2007) exponen un resumen amplio y útil de *FC*, que ha sido aplicada en el campo de la hidrología.

Cóputas de Arquímedes bivariadas

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

La primera *FC* se seleccionó para servir de contraste, de un buen ajuste a los datos (Chowdhary & Singh, 2019), pero carece de capacidad para reproducir la dependencia observada en la cola derecha de tales datos. Con tal planteamiento, se podrían haber utilizado las *FC* de Clayton, Planckett o Raftery, que son más fáciles de ajustar. La segunda *FC* seleccionada sí tiene tal capacidad.

(1) Familia de *FC* de Frank

Su ecuación y espacio de variación de θ son:

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

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

$$\tau_n = 1 + \frac{4}{\theta} [D_1(\theta) - 1] \quad (2)$$

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

$$D_1(\theta) = \frac{1}{\theta} \int_0^\theta \frac{s}{e^s - 1} ds \quad (3)$$

La ecuación anterior se evaluó con integración numérica, ratificando sus resultados contra los valores tabulados por Stegun (1972). Se aplicó el método de la cuadratura Gauss-Legendre, cuya ecuación operativa es (Nieves & Domínguez, 1998; Campos-Aranda, 2003):

$$\int_a^b f(x) dx \cong \frac{b-a}{2} \sum_{i=1}^{np} w_i \cdot f \left[\frac{(b-a)h_i + b+a}{2} \right] \quad (4)$$

en la cual:

w_i = coeficientes del método

h_i = abscisas

np = número de parejas en donde se evalúa la función $f(x)$, con el argumento indicado en $f[\cdot]$

En Davis y Polonsky (1972) se obtuvieron las 12 parejas utilizadas de w_i y h_i con 15 dígitos, que son aceptables en el lenguaje *Basic* como variables de doble precisión.

(2) La FC de Gumbel-Hougaard

También pertenece a la familia de valores extremos y solo acepta dependencia positiva. Su ecuación y espacio de variación de θ son:

$$C(u, v) = \exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right\} [1, \infty) \quad (5)$$

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

$$\tau_n = \frac{\theta-1}{\theta} \quad (6)$$

Indicador numérico de asociación

Concordancia

Como la *FC* caracteriza la *dependencia* entre las variables aleatorias u y v , es necesario el estudio de las medidas de asociación para disponer de un método que permita estimar su parámetro θ . En términos generales, una variable aleatoria es *concordante* con otra cuando sus grandes valores están asociados con las grandes magnitudes de la otra, y los

valores pequeños de una con los valores reducidos de la otra (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019).

Unas variables con correlación lineal directa serán concordantes, ya que al aumentar una la otra también lo hace. Variables con correlación lineal inversa serán *discordantes*, pues a grandes valores de una le corresponderán valores pequeños de la otra. Lo anterior implica que son *concordantes* (c) las parejas $(x_i - x_j)(y_i - y_j) > 0$ y *discordantes* (d) cuando $(x_i - x_j)(y_i - y_j) < 0$ (Salvadori *et al.*, 2007; Chowdhary & Singh, 2019).

Cociente tau de Kendall

Es un indicador numérico no paramétrico que mide la probabilidad de tener parejas concordantes; su expresión para estimarlo con datos bivariados es (Zhang & Singh, 2006; Zhang & Singh, 2019):

$$\tau_n = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{signo}[(x_i - x_j)(y_i - y_j)] \quad (7)$$

En la ecuación anterior:

n = número de observaciones

$\text{signo}[\cdot]$ = +1 si tales parejas son concordantes y -1 si son discordantes

Genest y Favre (2007) exponen una prueba para el cociente tau, que acepta la hipótesis nula H_0 de X y e *independientes*, y entonces su estadístico tiene distribución aproximadamente normal con media cero y varianza $2(2n + 5) / [9n(n - 1)]$. Por lo tanto, H_0 será rechazada con un nivel de confianza $\alpha = 5 \%$ si:

$$\sqrt{\frac{9n(n-1)}{2(2n+5)}} |\tau_n| > Z_{\alpha/2} = 1.96 \quad (8)$$

Dependencia en los extremos de las *FC* bivariadas

Generalidades

El criterio más importante que se aplica para seleccionar una *FC* bivariada 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 cero (Chowdhary & Singh, 2019).

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

$$\lambda_U = 2 - 2^{1/\theta} \quad (9)$$

Dupuis (2007) probó seis familias de *FC* bivariadas 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 (Nelsen, 2006), 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 muestra los datos disponibles, se debe definir primero la llamada *Cópula empírica*. Como la *FC* que 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:

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n 1\left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v\right) \quad (10)$$

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:

$$U_i = \frac{\text{Rango}(X_i)}{n+1} = F_n(X_i) \quad V_i = \frac{\text{Rango}(Y_i)}{n+1} = F_n(Y_i) \quad (11)$$

Poulin *et al.* (2007), y Requena, Mediero y Garrote (2013) utilizan el estimador propuesto por Frahm, Junker y Schmidt (2005), el cual está basado en una muestra aleatoria obtenida de la Cópula empírica; su designación proviene de sus autores Capéraà, Fougères y Genest (1997). Su expresión es:

$$\lambda_U^{CFG} = 2 - 2 \exp \left\{ \frac{1}{n} \sum_{i=1}^n \ln \left[\sqrt{\ln \frac{1}{U_i} \cdot \ln \frac{1}{V_i}} / \ln \left(\frac{1}{\max(U_i, V_i)^2} \right) \right] \right\} \quad (12)$$

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, como es el caso de los cuatro estimadores expuestos por AghaKouchak, Sellars y Sorooshian (2013).

Cómulas de Arquímedes trivariadas

Cómulas de Arquímedes simétricas

Chen y Guo (2019) indican para variables aleatorias multivariadas, mayores de dos ($d \geq 3$) y correlacionadas, que la familia de Cópulas de Arquímedes se divide en *simétricas* y *asimétricas*. Las primeras son fáciles de construir y tienen un único parámetro de asociación (θ), lo cual obliga a que todos los pares de variables muestren la misma estructura y grado de dependencia (Zhang & Singh, 2019).

Para las dos familias de Cópulas de Arquímedes expuestas y sus versiones *multivariadas simétricas* ($d \geq 3$) se indica su rango de θ y las funciones generadoras $\varphi(s)$, y su primera y segunda derivadas $\varphi'(s)$,

$\varphi''(s)$, donde s es la variable aleatoria en el intervalo de 0 a 1 (Grimaldi & Serinaldi, 2006a; Xu, Yin, Guo, Liu, & Hong, 2016; Chen & Guo, 2019; Zhang & Singh, 2019).

(3) Familia de Cópula multivariada de Frank

El rango de θ es $(0, +\infty)$ y el valor de $\theta = 1$ indica la condición de independencia en u_k :

$$C(u_1, u_2, \dots, u_d) = -\frac{1}{\theta} \ln \left[1 + \frac{\prod_{k=1}^d (e^{-\theta u_k} - 1)}{(e^{-\theta} - 1)^{d-1}} \right] \quad (13)$$

$$\varphi(s) = -\ln \left(\frac{e^{-\theta s} - 1}{e^{-\theta} - 1} \right) \quad (14)$$

$$\varphi'(s) = \frac{\theta}{1 - e^{\theta s}} \quad (15)$$

$$\varphi''(s) = \frac{\theta^2}{e^{\theta s} - 2 + e^{-\theta s}} \quad (16)$$

(4) Familia de Cópula multivariada de Gumbel–Hougaard

El rango de θ es $(1, +\infty)$ y el límite de $\theta = 1$ corresponde a la condición de independencia en u_k :

$$C(u_1, u_2, \dots, u_d) = \exp \left\{ - \left[\sum_{k=1}^d (-\ln u_k) \right]^\theta \right\} \quad (17)$$

$$\varphi(s) = [-\ln(s)]^\theta \quad (18)$$

$$\varphi'(s) = -\frac{\theta}{s} [-\ln(s)]^{\theta-1} \quad (19)$$

$$\varphi''(s) = \frac{\theta}{s^2} \{(\theta - 1)[- \ln(s)]^{\theta-2} + [- \ln(s)]^{\theta-1}\} \quad (20)$$

Cóputas de Arquímedes asimétricas

Para modelar diferentes estructuras de dependencia en variables aleatorias multivariadas, Chen y Guo (2019) recurren al enfoque de Grimaldi y Serinaldi (2006b) de aplicar Cóputas de Arquímedes *anidadas* (*nested*, en inglés). Con tal enfoque, las Cóputas de Arquímedes trivariadas *asimétricas* de dos parámetros de asociación (θ_1 y θ_2) más comunes tienen la fórmula general: $C(u, v, w) = C_{\theta_1}(w, C_{\theta_2}(u, v))$ y se exponen las dos siguientes (Grimaldi & Serinaldi, 2006b; Xu *et al.*, 2016; Zhang & Singh, 2019; Chen & Guo, 2019):

(5) Familia de Cóputa asimétrica de Frank trivariada

Con $\theta_2 \geq \theta_1 \in [0, \infty)$ y $\tau_{12}, \tau_{13}, \tau_{23} \in [0, 1]$ para tres variables aleatorias con dependencia positiva:

$$C(u, v, w) = -\frac{1}{\theta_1} \ln \left\{ 1 - F_1^{-1} \left(1 - \left[1 - F_2^{-1} \left(1 - e^{-\theta_2 u} \right) \left(1 - e^{-\theta_2 v} \right) \right]^{\theta_1 / \theta_2} \right) \left(1 - e^{-\theta_1 w} \right) \right\} \quad (21)$$

siendo:

$$F_1 = 1 - e^{-\theta_1}$$

$$F_2 = 1 - e^{-\theta_2}$$

(6) Familia de Cópula asimétrica de Gumbel–Hougaard trivariada

Con $\theta_2 \geq \theta_1 \in [1, \infty)$ y $\tau_{12}, \tau_{13}, \tau_{23} \in [0, 1]$ para tres variables aleatorias con dependencia positiva:

$$C(u, v, w) = \exp \left\{ - \left[((-\ln u)^{\theta_2} + (-\ln v)^{\theta_2})^{\theta_1/\theta_2} + (-\ln w)^{\theta_1} \right]^{1/\theta_1} \right\} \quad (22)$$

Estimación de probabilidades empíricas

Las probabilidades de no excedencia *empíricas* univariadas y trivariadas se estimaron con base en la fórmula de Gringorten (Ecuación (23)), la cual ha sido sugerida por varios autores para los análisis de frecuencias bivariados, y por Zhang y Singh (2007) para los trivariados. Tal ecuación es:

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

siendo i el número del dato cuando se ordenan de menor a mayor, y n su número total o de años de los registros de gastos máximos anuales.

En los análisis bivariados se puede trabajar gráficamente en el plano bidimensional, como lo ha expuesto Campos-Aranda (2023).

Para el caso de las probabilidades trivariadas se trabajó en el espacio tridimensional con los gastos máximos anuales QX y QY en el plano x, y , y el QZ en el eje perpendicular (z). El proceso numérico comienza guardando los registros históricos de gasto máximo (QX, QY y QZ) anuales en archivos QXh, QYh y QZh ; además se ordenaron en forma progresiva de magnitud en archivos QXo, QYo y QZo . Enseguida se procesa cada dato anual para comparar el valor histórico contra el ordenado, y se contabilizan las veces que el segundo resultó menor o igual, y se designa NQX, NQY y NQZ . Lo anterior equivale a cambiar el dato original de cada terna de valores anuales históricos por su número de orden o *rango*.

Después se compara cada terna histórica de rangos contra todas las demás y se contabilizan las veces en que los tres rangos (condición AND) resultan menores; tal cantidad se denomina $NQXYZ$; es decir, el número de ocurrencias de combinaciones de qx, qy y qz menores en el espacio tridimensional. Por último, se aplica la fórmula de posición gráfica de Gringorten. Para el caso trivariado es la siguiente:

$$F_e(x, y, z) = P(QX \leq qx, QY \leq qy, QZ \leq qz) = \frac{NQXYZ_i - 0.44}{n + 0.12} \quad (24)$$

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):

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

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

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

Búsqueda de las distribuciones marginales óptimas

La búsqueda de las distribuciones *marginales* idóneas tomó en cuenta las características estadísticas de los datos hidrológicos por procesar de la Tabla 1. 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 tres mejores distribuciones por su menor cercanía a las *cinco curvas* que se muestran en tal gráfico.

Tabla 1. Gastos máximos anuales (m^3/s) disponibles y estimados en las estaciones hidrométricas indicadas de la cuenca del río Tempoal en el periodo común de 1960 a 2002.

Núm.	Año mes	Tempoal	El Cardón	Los Hules	Terrerillos	NXYZ
1	1960 NOV	1277.0	1080.0	320.8	247.1	2
2	1961 JUN	852.9	303.5	434.5	525.0	6
3	1962 JUN	739.2	246.7	457.5	529.3	5
4	1963 JUL	1800.0	481.0	947.4	895.9	18
5	1964 DIC	748.0	122.6	258.0	397.1	1
6	1965 AGO	792.7	202.0	283.9	659.4	5
7	1966 JUN	1778.0	287.0	742.2	1121.7	16
8	1967 SEP	2245.0	854.2	1009.4	1153.0	25
9	1968 SEP	1145.0	476.0	1096.0	611.2	11
10	1969 SEP	1948.0	555.8	825.0	2224.2	28
11	1970 SEP	1418.0	339.9	800.0	1049.3	16
12	1971 OCT	1630.0	720.4	1064.0	1488.5	22
13	1972 JUL	989.0	185.8	450.0	529.0	6
14	1973 JUN	1668.0	387.0	749.0	1740.0	20
15	1974 SEP	4950.0	1198.3	1950.0	3187.8	37
16	1975 SEP	4040.0	1204.2	2470.0	2085.0	33
17	1976 OCT	1275.0	185.0	472.0	792.3	6
18	1977 OCT	514.0	179.1	559.0	162.9	1
19	1978 SEP	3725.0	1390.0	2874.0	2152.3	34
20	1979 SEP	1655.9	667.0	1082.0	514.2	9
21	1980 SEP	1162.0	357.0	583.2	994.1	14

Núm.	Año mes	Tempoal	El Cardón	Los Hules	Terrerillos	NXYZ
22	1981 AGO	2020.0	733.9	1650.3	(1151.3)	24
23	1982 SEP	539.6	133.1	268.8	491.4	2
24	1983 JUL	868.0	269.8	544.0	743.5	8
25	1984 SEP	4030.0	572.0	2834.9	2981.0	30
26	1985 JUL	1882.0	457.0	938.4	1487.7	22
27	1986 JUN	476.0	130.0	308.0	434.0	1
28	1987 JUL	1765.0	346.8	1440.0	2635.0	19
29	1988 SEP	3265.0	356.0	4350.0	3710.0	21
30	1989 SEP	649.0	306.0	644.0	2100.0	5
31	1990 AGO	1611.0	141.8	(3463.7)	204.5	1
32	1991 OCT	3532.0	1248.0	(1072.2)	2860.0	36
33	1992 OCT	2291.0	790.0	762.8	1607.5	29
34	1993 SEP	6120.0	865.5	1684.1	3422.5	37
35	1994 SEP	1133.0	412.0	723.8	1237.9	15
36	1995 AGO	741.9	381.6	440.9	474.0	3
37	1996 AGO	683.0	218.0	804.0	507.6	4
38	1997 OCT	905.0	85.7	428.4	362.5	1
39	1998 SEP	1266.9	(271.9)	204.3	994.4	13
40	1999 OCT	2693.7	602.9	630.9	3328.3	30
41	2000 JUN	641.2	(185.5)	84.9	753.4	4
42	2001 SEP	1847.9	498.3	278.5	1512.2	24
43	2002 SEP	926.4	134.0	496.7	822.2	5
Estadístico <i>U</i>		0.686	0.460	0.538	0.993	-

Estas distribuciones tienen tres parámetros de ajuste y son los siguientes: 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 a la distribución Log-Pearson tipo III (LP3). Este proceso ha sido descrito por Campos-Aranda (2023).

Además, se aplicaron las distribuciones Kappa y Wakeby, las cuales han mostrado gran versatilidad y universalidad para representar series de datos hidrológicos extremos debido a que tienen 4 y 5 parámetros de ajuste (Hosking, 1994; Hosking & Wallis, 1997; Kjeldsen, Ahn, & Prosdoci, 2017).

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 de las crecientes anuales a través del método de momentos L, según procedimientos expuestos por Hosking y Wallis (1997), y Stedinger (2017).

Errores de ajuste

El primer criterio aplicado para la selección de la mejor FDP a unos datos o serie disponible fueron los llamados *errores de ajuste* (Kite, 1977; Willmott & Matsuura, 2005; Chai & Draxler, 2014). Este criterio 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) 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) se buscan para la misma probabilidad de no excedencia, asignada a los datos por la fórmula empírica de Gringorten (Ecuación (23)).

Estimación del parámetro de dependencia θ

El método más simple para estimar el parámetro θ de las *FC* bivariadas (Ecuación (1)) es por tanteos, igualando el cociente tau de Kendall y en las *FC* trivariadas simétricas (ecuaciones (13) y (17)), buscando que los estadísticos de los errores de ajuste (ecuaciones (25) a (26)) sean mínimos.

Estimación de los parámetros de dependencia θ_1 y θ_2

La búsqueda del valor mínimo de la Ecuación (25) o error medio estándar para el ajuste de las *FC* trivariadas asimétricas definidas por las ecuaciones (21) y (22) se realizó con base en el algoritmo Complex de múltiples variables restringidas o acotadas para encontrar los valores óptimos de θ_1 y θ_2 , cumpliendo con la condición $\theta_2 > \theta_1$.

El *algoritmo Complex* es una técnica de exploración local (Box, 1965), que se guía exclusivamente por lo que encuentra a su paso; sus antecedentes, una descripción breve de su proceso operativo y su código OPTIM en lenguaje *Basic* se pueden consultar en Campos-Aranda (2003). En Bunday (1985) se tiene otra descripción y código de este método de búsqueda.

Las designaciones principales en el código OPTIM son NX y NY, que definen el número de variables de decisión y dependientes, función de las primeras; para el caso analizado, dos (θ_1, θ_2) y una ($\theta_2 > \theta_1$). Una ventaja importante del código OPTIM radica en permitir un fácil acceso de los límites ($L = lower, U = upper$), nombres y valores iniciales de las variables, en la subrutina citada, por medio de las designaciones siguientes: XL(I), XU(I), XN\$(I), X(I), YL(J), YU(J), YN\$(J) y Y(J). Para el caso estudiado, I varía de 1 a 2 y J = 1.

En todas las variables de decisión se usó como límite inferior 0.001 y como límite superior 10 y 20 en la FC de Frank para θ_1 y θ_2 , y de 5 y 15 en la FC de Gumbel-Hougaard. La única variable dependiente se definió por el cociente de θ_2 entre θ_1 , con límite inferior de 1 y superior de 5; valor que fue adoptado de manera arbitraria.

La función objetivo se denomina FO en el código OPTIM y se define al final del programa; corresponde lógicamente a la Ecuación (25), con nombre FO\$=EME de error medio estándar. Para los criterios de convergencia de las desviaciones absoluta y relativa de la FO se utilizaron los valores siguientes: 0.0002 y 0.00001.

Ratificación de la función Cópula seleccionada

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 trivariadas (Ecuación (24)). Yue (2000) indica una forma simple y práctica de representar a las probabilidades conjuntas empíricas y teóricas, que 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 $\alpha = 5 \%$, que es:

$$D_n = \frac{1.358}{\sqrt{n}} \quad (28)$$

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

Periodos de retorno trivariados

Tipos OR y AND

El primer *periodo de retorno trivariado* del evento (X, Y, Z) se define bajo la condición OR, lo cual indica que los límites x, y o z , o los tres *pueden* ser excedidos, y entonces la ecuación clásica del periodo de retorno o inverso de la probabilidad de excedencia será (Genest & Chebana, 2017; Zhang & Singh, 2019):

$$T_{XYZ} = \frac{1}{P(X>x \text{ or } Y>y \text{ or } Z>z)} = \frac{1}{1-F_{XYZ}(x,y,z)} = \frac{1}{1-C[F_X(x),F_Y(y),F_Z(z)]} \quad (29)$$

en la cual $C[F_X(x), F_Y(y), F_Z(z)] = C(u, v, w)$ es la *FC* seleccionada o probada.

El segundo *periodo de retorno trivariado* del evento (X, Y, Z) está asociado al caso en que los tres límites *son* excedidos ($X > x, Y > y, Z > z$) o condición AND; su ecuación es la (30) siguiente (Zhang & Singh, 2019):

$$T'_{XYZ} = \frac{1}{P(X>x \text{ and } Y>y \text{ and } Z>z)} = \frac{1}{F'_{XYZ}(x,y,z)} = \frac{1}{1-u-v-w+C(u,v)+C(u,w)+C(v,w)-C(u,v,w)} \quad (30)$$

Para la aplicación de la Ecuación (30) anterior, se observa que se requieren las tres *FC* bivariadas y la trivariada.

Secundario o de tipo Kendall

Salvadori y De Michele (2004) introducen con detalle el concepto del *periodo de retorno bivariado secundario* (ζ), designado así para enfatizar que el periodo de retorno conjunto T_{XY} es el primario, de donde procede al utilizar las *isolíneas* que define la *FC* aplicada, cuya expresión es:

$$L_s = [(u, v) \in \mathbb{I}^2: C(u, v) = s] \quad (31)$$

Siendo:

s = variable aleatoria unitaria $0 < s \leq 1$

C = *FC* probada

Después se define una región $B_C(s)$ en el espacio unitario (\mathbb{I}^2) sobre la isolínea, debajo de ella y a la izquierda, que será:

$$B_C(s) = \{(u, v) \in \mathbb{I}^2: C(u, v) \leq s\} \quad (32)$$

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

$$K_C(s) = s - \frac{\varphi(s)}{\varphi'(s)} \quad (33)$$

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

$$\zeta = \frac{1}{1-K_C(s)} \quad (34)$$

cuyo denominador es la probabilidad de excedencia (*survival function*), la cual corresponde a eventos probablemente destructivos o peligrosos.

La distribución paramétrica de Kendall (Ecuación (33)), para las Cópulas de Arquímedes *trivariadas* simétricas es la siguiente (Barbe,

Genest, Ghoudi, & Rémillard, 1996; Grimaldi & Serinaldi, 2006a; Zhang & Singh, 2019):

$$K_C(s) = P[C(u, v, w) \leq s] = s - \frac{\varphi(s)}{\varphi'(s)} - \frac{\varphi^2(s) \cdot \varphi''(s)}{2[\varphi'(s)]^3} \quad (35)$$

Sustituyendo las ecuaciones (18) a (20) en la (35) se obtiene una expresión para la distribución de Kendall de la *FC* de Gumbel-Hougaard trivariada simétrica, que será utilizada posteriormente.

Gräler *et al.* (2013) extienden la desigualdad $T_{XYZ} \leq T'_{XYZ}$, la cual indica que el periodo de retorno tipo OR es siempre menor que el tipo AND e indican que T_{KEN} está intermedio entre los dos citados. T_{KEN} se obtiene con la Ecuación (34). Entonces se tiene:

$$T_{XYZ} \leq T_{KEN} \leq T'_{XYZ} \quad (36)$$

Salvadori, De Michele y Durante (2011) destacan que la estimación de los periodos de retorno y sus eventos de diseño en los análisis de frecuencias *multivariados* son un problema difícil. Para resolverlo establecen un marco teórico basado en las *FC* y en la distribución de Kendall, el cual aplican mediante simulación numérica.

Registros de crecientes anuales por procesar

El río Tempoal es uno de los afluentes importantes del río Moctezuma, que junto con el río Tampaón forman el río Pánuco de la Región

Hidrológica No. 26 de México. El río Tempoal tiene cinco estaciones hidrométricas: El Cardón, Los Hules, Terrerillos, Platón Sánchez y Tempoal, cuyas áreas de cuenca son 609, 1 269, 1 493, 4 700 y 5 275 km², respectivamente. En la Figura 2 se muestra la ubicación y morfología del río Tempoal (tomada de Campos-Aranda, 2015).

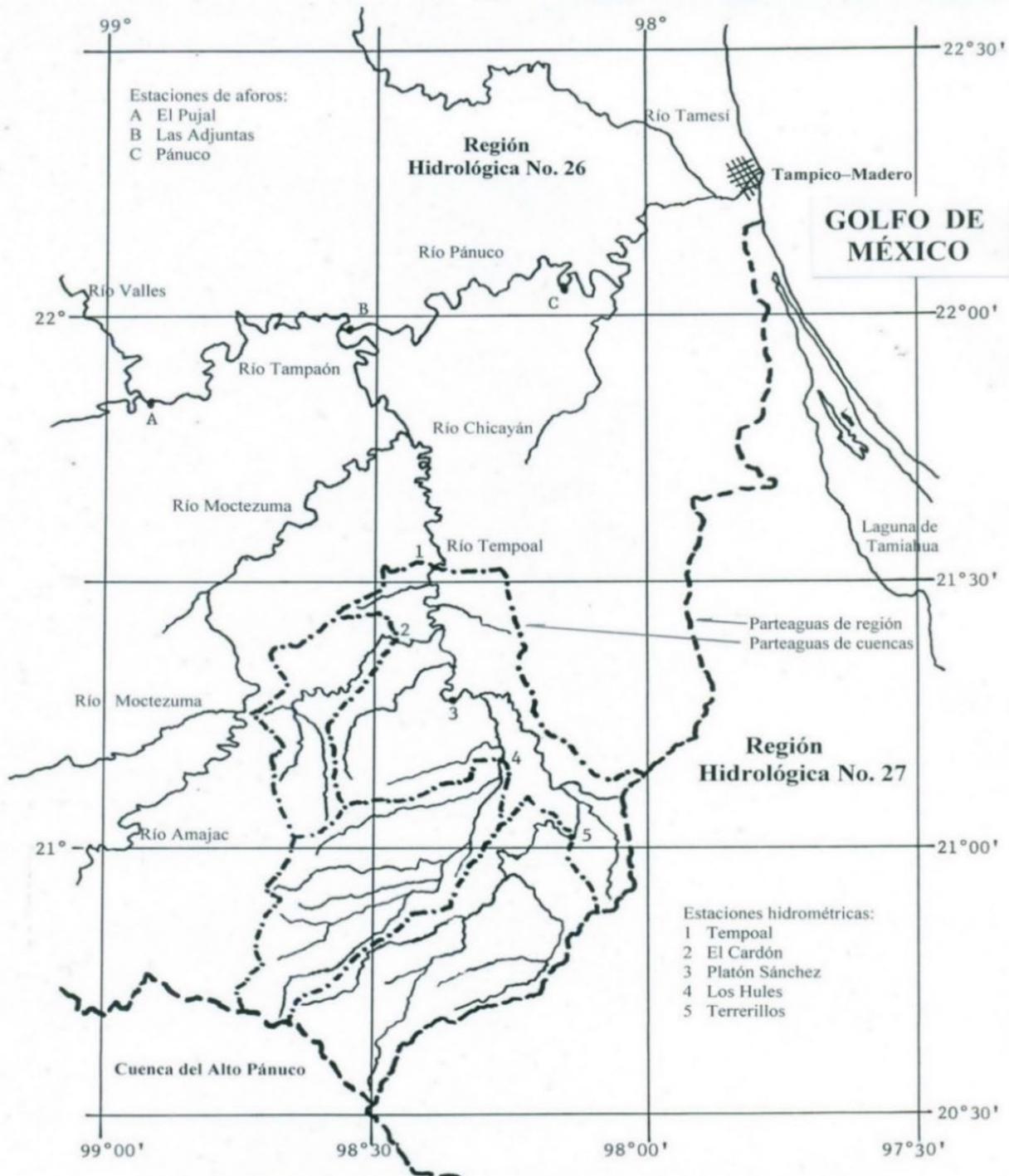


Figura 2. Localización geográfica y esquema de morfología del río Temporal de la Región Hidrológica No. 26 (Pánuco), México.

Campos-Aranda (2015) expuso los registros disponibles de gasto máximo anual en m^3/s en el periodo de 1960 a 2002 en las cinco estaciones hidrométricas del río Tempoal, procedentes del sistema BANDAS (IMTA, 2003) del archivo *Datos Hidrométricos Mensuales*, con nueve años por página. Tales registros están incompletos, con dos años faltantes en 1998 y 2000 en El Cardón; dos más en Los Hules en 1990 y 1991; uno en Terrerillos en 1981, y 18 años faltantes en la estación de aforos Platón Sánchez en el periodo de 1960 a 1977.

Con base en el algoritmo de Beale y Little (1975) se estimaron de manera simultánea los datos faltantes y el periodo sin registro de la estación Platón Sánchez. Los gastos anuales estimados por Campos-Aranda (2015) se exponen entre paréntesis. En la columna final de la Tabla 1 se tiene el número de ocurrencias de la Ecuación (24) para la terna Tempoal-El Cardón-Terrerillos, que se define posteriormente.

En la columna 3 de la Tabla 1 se muestran los datos disponibles y sus meses de ocurrencia se tienen en la columna 2 para las crecientes anuales (QX) de la estación Tempoal, que es la base. En cambio, en las columnas 4, 5 y 6 se exponen las crecientes anuales de las estaciones auxiliares *simultáneas por mes*, como se explica en el inciso siguiente.

Simultaneidad de los eventos por analizar

A principios de este siglo, cuando iniciaron los análisis de frecuencias de crecientes *trivariados*, por ejemplo, los de Zhang y Singh (2007), las variables procesadas, gasto máximo (Q), volumen escurrido (V) y duración total (D) procedían del hidrograma de la creciente anual. Lo

anterior implica, como ya se indicó, que los gastos máximos anuales de las estaciones auxiliares deben de corresponder con el mismo evento registrado en la estación base, es decir, se debe cumplir una *simultaneidad de eventos*.

Para la estación base Tempoal, el número de meses de la columna 2 de la Tabla 1 establece una *época húmeda* perfectamente definida de junio a octubre, ya que solo ocurren dos crecientes fuera de ella en noviembre de 1960 y en diciembre de 1964. El mes con más ocurrencias es septiembre con 18 eventos; en cambio, en los meses restantes acontecen entre 5 y 7 eventos. La existencia de una época húmeda de cinco meses claramente definida permite verificar la *simultaneidad de crecientes* por meses.

En la Tabla 1 se indican sombreadas las crecientes que no son las máximas anuales; por lo tanto, establecen una falta de *simultaneidad* mensual con las de la estación Tempoal. Son 13 en la estación El Cardón, 6 en Los Hules y 10 en Terrerillos.

Respecto a los valores extremos máximos (*outliers*) del registro conjunto (1960-2002) en Tempoal, se observan cuatro valores superiores a los 4 000 m³/s en los años 1974, 1975, 1984 y 1993.

Resultados y su discusión

Verificación de la aleatoriedad

Con base en el test de Wald-Wolfowitz (Bobée & Ashkar, 1991; Rao & Hamed, 2000; Meylan *et al.*, 2012) se probó la independencia y estacionariedad de los cuatro registros de gasto máximo de la Tabla 1.

Los valores de estadístico U de tal prueba se tienen en el último reglón de la Tabla 1; como $U < 1.96$ se deduce que los registros son *aleatorios*.

Distribución marginal en la estación Tempoal

En la Tabla 2 se muestran los resultados del ajuste de las tres FDP idóneas según distancia absoluta ponderada en el diagrama de cocientes L (Hosking & Wallis, 1997), y de los modelos Kappa y Wakeby. Se observa, respecto a las predicciones, una excelente similitud, excepto para la distribución Log-normal que conduce a valores elevados.

Tabla 2. Errores de ajuste y predicciones (m^3/s) de las tres FDP idóneas y dos de uso generalizado en el registro de gastos máximos anuales (1960-2002) de la estación hidrométrica *Tempoal*, México.

FDP	EEA	EAM	Periodos de retorno en años					
			25	50	100	500	1 000	5 000
PAG	152.0	108.1	4 720	5 694	6 695	9 103	10 183	12 792
PE3	151.9	108.1	4 704	5 670	6 656	9 023	10 073	12 581
LN3	183.0	121.3	4 689	5 830	7 097	10 588	12 354	17 148
Kappa	162.7	112.0	4 720	5 766	6 880	9 755	11 130	14 680
Wakeby	158.9	114.6	4 724	5 727	6 765	9 325	10 493	13 373
PAG mod	153.5	111.1	4 652	5 512	6 356	8 256	9 049	10 836

La mejor distribución, la Pareto Generalizada (PAG), establece errores de ajuste bajos según se observa en la Tabla 2. Sin embargo, para una probabilidad de no excedencia (p) del 1 % define un valor de 496.8

m³/s, con lo cual el valor más bajo del registro, que es 476.0 m³/s, genera una probabilidad negativa.

Para corregir lo anterior se aplicó la versión modificada del método de momentos L expuesta por Rao y Hamed (2000), y Campos-Aranda (2014), que conduce a un valor de 458.0 m³/s para una $p = 1 \%$, las predicciones mostradas en el renglón final de la Tabla 2 y parámetros de ubicación (u_1), escala (a_1) y forma (k_1) siguientes: 444.2926, 1364.267, 0.026739; su ecuación es:

$$F(x) = 1 - \left[1 - \frac{k_1(x-u_1)}{a_1} \right]^{1/k_1} \quad (37)$$

Distribución marginal en la estación El Cardón

De manera similar, en la Tabla 3 se exponen los resultados de los ajustes en el registro de crecientes anuales de la estación El Cardón. Nuevamente, la mejor FDP, la PAG, define un valor de 93.9 m³/s para una $p = 1 \%$, que es superior al mínimo registrado de 85.7 m³/s.

Tabla 3. Errores de ajuste y predicciones (m^3/s) de las tres FDP idóneas y dos de uso generalizado en el registro de gastos máximos anuales (1960–2002) de la estación hidrométrica *El Cardón*, México.

FDP	EEA	EAM	Periodos de retorno en años					
			25	50	100	500	1 000	5 000
PAG	52.3	28.0	1 274	1 488	1 689	2 111	2 275	2 619
PE3	60.6	31.6	1 267	1 506	1 747	2 313	2 561	3 145
LP3	65.7	42.5	1 194	1 394	1 590	2 024	2 204	2 603
Kappa	44.0	29.5	1 266	1 442	1 594	1 869	1 961	2 128
Wakeby	55.1	30.0	1 274	1 493	1 701	2 145	2 321	2 696
PAGmod	45.1	28.3	1 246	1 425	1 585	1 891	1 999	2 207

Para corregir lo anterior se aplicó la versión modificada del método de momentos L expuesta por Rao y Hamed (2000), y Campos-Aranda (2014), que conduce a un valor de $79.5 \text{ m}^3/\text{s}$ para una $p = 1 \%$, las predicciones mostradas en el renglón final de la Tabla 3 y parámetros de ubicación (u_2), escala (a_2) y forma (k_2) siguientes: 74.78195, 471.314, 0.168322 con la ecuación siguiente:

$$F(y) = 1 - \left[1 - \frac{k_2(y-u_2)}{a_2} \right]^{1/k_2} \quad (38)$$

La FDP Pearson tipo III también conduce a un valor inferior a $85.7 \text{ m}^3/\text{s}$ para $p = 1 \%$ de $73.8 \text{ m}^3/\text{s}$, pero sus errores de ajuste son elevados, al igual que sus predicciones, por ello no se adoptó.

Distribución marginal en la estación Terrerillos

Por último, con base en un procedimiento semejante al descrito para los dos registros anteriores, se obtuvieron los resultados mostrados en la Tabla 4. Se observa que la distribución Kappa conduce a los errores de ajuste más bajos, pero todas sus predicciones se consideran reducidas y además, para una $p = 1\%$ define un valor de $212.0 \text{ m}^3/\text{s}$, que es superior al valor más bajo del registro que fue de $162.9 \text{ m}^3/\text{s}$.

Tabla 4. Errores de ajuste y predicciones (m^3/s) de las tres FDP idóneas y dos de uso generalizado en el registro de gastos máximos anuales (1960-2002) de la estación hidrométrica *Terrerillos*, México.

FDP	EEA	EAM	Periodos de retorno en años					
			25	50	100	500	1 000	5 000
PAG	174.9	105.8	3 603	4 173	4 695	5 741	6 129	6 906
LP3	197.2	143.3	3 352	3 871	4 358	5 370	5 759	6 566
PE3	211.7	132.7	3 586	4 256	4 928	6 498	7 182	8 789
Kappa	113.6	95.7	3 524	3 861	4 105	4 441	4 523	4 636
Wakeby	209.5	130.2	3 585	4 241	4 895	6 405	7 053	8 553
PAGmod	150.2	107.2	3 530	4 018	4 443	5 230	5 499	5 997

Por otra parte, la distribución PAG también define un valor superior para $p = 1\%$ con $180.9 \text{ m}^3/\text{s}$. Nuevamente, para corregir lo anterior, se aplicó la versión modificada del método de momentos L expuesta por Rao y Hamed (2000), y Campos-Aranda (2014), que conduce a un valor de

144.1 m³/s para una $p = 1 \%$, las predicciones mostradas en el renglón final de la Tabla 4 y parámetros de ubicación (u_3), escala (a_3) y forma (k_3) siguientes 129.786, 1430.494, 0.199064, con la ecuación siguiente:

$$F(z) = 1 - \left[1 - \frac{k_3(z-u_3)}{a_3} \right]^{1/k_3} \quad (39)$$

Predicciones para contraste en la estación Tempoal

Para el contraste de las *predicciones* obtenidas con la *FC* en los análisis de frecuencias trivariados en la estación hidrométrica Tempoal con las cercanas que mostraron *dependencia regional*, primero se estimaron las del periodo completo de datos en la estación Tempoal, que abarcó de 1954 a 2006 ($n = 49$).

Los seis datos por incluir fueron las crecientes anuales del lapso de 1954 a 1959, que son 2 110.0, 6 000.0, 4 424.0, 449.0, 4 100.0 y 1 507.6 m³/s. Estas crecientes establecen un nuevo valor mínimo de 449.0 m³/s, pero lo más relevante es que en un lapso de seis años ocurre una creciente semejante a la máxima de todo el registro de 43 años de 6 120 m³/s y dos más superiores a 4 000 m³/s.

Lo anterior seguramente dará origen a predicciones más grandes, al haber mayor número de eventos extremos. Las distancias absolutas ponderadas definen como FDP idóneas las tres primeras que se exponen en la Tabla 5 de resultados.

Tabla 5. Errores de ajuste y predicciones (m^3/s) de las tres FDP idóneas y dos de uso generalizado en el registro de gastos máximos anuales (1954-2002) de la estación hidrométrica *Tempoal*, México.

FDP	EEA	EAM	Periodos de retorno en años					
			25	50	100	500	1 000	5 000
PE3	253.4	148.4	5 262	6 339	7 436	10 056	11 215	13 976
PAG	246.5	145.5	5 282	6 336	7 394	9 866	10 937	13 438
LP3	261.6	195.4	4 967	5 885	6 807	8 962	9 895	12 063
Kappa	221.3	165.0	5 256	6 158	7 003	8 758	9 433	10 835
Wakeby	253.1	152.3	5 281	6 340	7 404	9 897	10 981	13 519
PAGmod	236.1	146.1	5 253	6 260	7 256	9 524	10 481	12 663

Las dos primeras distribuciones idóneas conducen a errores de ajuste bajos, principalmente del error absoluto medio, pero definen valores con una probabilidad de no excedencia del 1 % mayores que el mínimo del registro ($449.0 \text{ m}^3/\text{s}$); por ello, no son aceptables.

La versión modificada del método de momentos L para el ajuste de la distribución PAG expuesta por Rao y Hamed (2000), y Campos-Aranda (2014) conduce a un valor de $433.0 \text{ m}^3/\text{s}$ para una $p = 1 \%$ y las predicciones mostradas en el renglón final de la Tabla 5, que serán las utilizadas en los contrastes.

Como se anticipó, al aumentar los seis años de registro, las crecientes de los periodos de retorno de 100 a 5 000 años de la Tabla 5 son superiores a las de la Tabla 2 de un 14.2 % a un 16.9 %.

Cocientes tau de Kendall

Con base en la Ecuación (7) se calcularon los cocientes tau de Kendall mostrados en la Tabla 6 para los datos de las cinco estaciones hidrométricas del sistema del río Tempoal (Tabla 1) analizadas por *duplas*. El cociente tau de Kendall más bajo entre las estaciones Los Hules y Terrerillos conduce a un valor de 3.61 en la Ecuación (8), por lo cual las estaciones no son independientes.

Tabla 6. Cocientes tau de Kendall bivariados para las crecientes anuales (1960-2002) de las cuatro estaciones hidrométricas del sistema del río Tempoal, México.

Estaciones hidrométricas	El Cardón	Los Hules	Terrerillos
Tempoal	0.6013	0.5127	0.5991
El Cardón	1	0.4640	0.4839
Los Hules		1	0.4219

Para la estación Tempoal, sus mejores duplas de mayor correlación o *dependencia regional* se forman con El Cardón y Terrerillos; por lógica, la terna por analizar es Tempoal-El Cardón-Terrerillos. Conviene observar que ambas correlaciones son similares en magnitud y de signo positivo. Por lo anterior, se puede intuir que los ajustes de las *FC* trivariadas simétrica y asimétrica serán semejantes.

Dependencia observada por duplas

Con base en la Ecuación (22) y los datos de la Tabla 1 se calcularon los valores de la dependencia observada en la cola derecha, expuestos en la Tabla 7. Se observa que los valores mayores de λ_U^{CFG} ocurren entre el registro de crecientes de las estaciones Tempoal con Terrerillos y con El Cardón; por lo tanto, se asemeja al comportamiento de los cocientes tau de Kendall.

Tabla 7. Valores de la dependencia en la cola derecha (λ_U^{CFG}) en las crecientes anuales (1960-2002) de las cuatro estaciones hidrométricas del sistema del río Tempoal, México.

Estaciones hidrométricas	El Cardón	Los Hules	Terrerillos
Tempoal	0.6451	0.5942	0.6693
El Cardón	1	0.5500	0.5186
Los Hules		1	0.5619

Al buscar en la Tabla 4 —expuesta por Campos-Aranda (2023)— valores aproximados a los definidos en la Tabla 6 y Tabla 7 para el tau de Kendall y la dependencia en la cola derecha (λ_U^{CFG}), se encontró que la *FC* de Gumbel-Hougaard parece reproducir tales magnitudes, por lo anterior, será tal *FC* con la que se probará modelar las duplas Tempoal-El Cardón, Tempoal-Terrerillos y El Cardón-Terrerillos, así como la terna Tempoal-El Cardón-Terrerillos.

Ajuste de las FC en los análisis bivariados

Como ya se indicó, para la aplicación de la Ecuación (30), relativa al periodo de retorno trivariado de tipo AND, se requieren las FC bivariadas $C(u,v)$, $C(u,w)$ y $C(v,w)$, las cuales se adoptan con base en el proceso siguiente. En la última columna de la Tabla 8 se muestran los valores del estimador de la dependencia observada (Ecuación (22)) para las duplas indicadas.

Tabla 8. Indicadores estadísticos del ajuste de las funciones Cópula bivariadas entre las crecientes anuales de las estaciones hidrométricas indicadas del sistema del río Tempoal, México.

FC	θ	EME	EAM	DP	DN	MDP	MDN	$(\lambda_U^{CFG}) \lambda_U$
Estaciones Tempoal-El Cardón								(0.6451)
Frank	7.964	0.0316	0.0237	17	26	0.0782	-0.0575	0.0000
G-H	2.5083	0.0306	0.0231	20	23	0.0685	-0.0569	0.6817
Estaciones Tempoal-Terrerillos								(0.6693)
Frank	7.905	0.0372	0.0283	19	24	0.0766	-0.0891	0.0000
G-H	2.4945	0.0381	0.0306	22	21	0.0758	-0.0831	0.6797
Estaciones El Cardón-Terrerillos								(0.5186)
Frank	5.451	0.0255	0.0194	27	16	0.0493	-0.0652	0.0000
G-H	1.9378	0.0296	0.0222	25	18	0.0641	-0.0743	0.5700

Significado de los nuevos acrónimos:

DP, DN: número de diferencias positivas y negativas.

MDP, MDN: máxima diferencia positiva y negativa.

Se observa que en los tres casos, la *FC* de Gumbel-Hougaard (G-H) aporta una dependencia en la cola extrema derecha (Ecuación (9)), ligeramente mayor que la observada; por lo anterior, se concluye que la selección de tal *FC* fue acertada.

Selección y ratificación de la *FC* trivariada

Funciones Cópula simétricas

A los datos *conjuntos* anuales de gasto máximo de las estaciones Tempoal, El Cardón y Terrerillos, tomados de la Tabla 1, se les ajustaron las *FC* trivariadas ($d = 3$) de Frank y Gumbel-Hougaard, definidas por las ecuaciones (13) y (17). Tal ajuste se realizó por tanteos del valor de su parámetro de asociación (θ), buscando los menores errores de ajuste según las ecuaciones (25) a (27). Tales cálculos se realizaron con base en un programa de cómputo en *Basic*, desarrollado ex profeso. Los resultados obtenidos se muestran en la Tabla 9.

Tabla 9. Indicadores estadísticos del ajuste de las funciones Cópula trivariadas *simétricas* indicadas en las ternas de crecientes anuales de las estaciones Tempoal-El Cardón-Terrerillos, México.

<i>FC</i>	θ	<i>EME</i>	<i>EAM</i>	<i>DP</i>	<i>DN</i>	<i>MDP</i>	<i>MDN</i>
Frank	7.995	0.0277	0.0213	20	23	0.0614	-0.0692
G-H	2.795	0.0294	0.0225	19	24	0.0492	-0.0792

Significado de los nuevos acrónimos:

DP, *DN*: número de diferencias positivas y negativas.

MDP, *MDN*: máxima diferencia positiva y negativa.

Por otra parte, la Ecuación (28) define a $D_n = 0.2071$ y como la diferencia máxima absoluta de las *FC* de Frank y de Gumbel-Hougaard simétricas en la Tabla 9 es 0.0792, el test de Kolmogorov-Smirnov permite la adopción de cualquiera de ellas. Los coeficientes de correlación (r_{xy}) entre las probabilidades empíricas (Ecuación (24)) y las teóricas, estimadas con las *FC* de Frank y de G-H *simétricas* fueron 0.9953 y 0.9944; por lo tanto, ambas *FC* definen buenos ajustes.

Funciones Cópulas asimétricas

La aplicación de las *FC* asimétricas trivariadas, con dos parámetros de asociación (θ_1, θ_2), a los datos de la Tabla 1 para la terna procesada se realizó con base en el algoritmo Complex de múltiples variables acotadas. Los valores iniciales en la *FC* de Frank 2 y 8; en la de Gumbel-Hougaard, 1.5 y 4. Los valores óptimos encontrados de θ_1 y θ_2 , y sus indicadores de ajuste se han concentrado en la Tabla 10.

Tabla 10. Indicadores estadísticos del ajuste de las funciones Cópula trivariadas *asimétricas* en las ternas de crecientes anuales de las estaciones Tempoal-El Cardón-Terrerillos, México.

<i>FC</i>	θ_1	θ_2	<i>EME</i>	<i>EAM</i>	<i>DP</i>	<i>DN</i>	<i>MDP</i>	<i>MDN</i>
Frank	7.1294	10.4163	0.0275	0.0208	19	24	0.0611	-0.0659
G-H	2.3475	3.8288	0.0290	0.0227	16	27	0.0573	-0.0773

Significado de los nuevos acrónimos:

θ_1, θ_2 : parámetros de asociación de la *FC* asimétrica.

Nuevamente, la *FC* de Frank define el mejor ajuste. Como ya se indicó, la Ecuación (28) define a $D_n = 0.2071$ - y como la diferencia máxima absoluta de las *FC* de Frank y de Gumbel-Hougaard asimétricas en la Tabla 10 es 0.0773, el test de Kolmogorov-Smirnov ratifica la adopción de alguna de ellas. Los coeficientes de correlación (r_{xy}) entre las probabilidades empíricas (Ecuación (24)) y las teóricas, estimadas con las *FC* de Frank y de G-H *asimétricas* resultaron de 0.9955 y 0.9945; por lo tanto, ambas *FC* muestran un ajuste excelente.

Adopción de una *FC* trivariada

El resultado de la Tabla 8 —de ser la *FC* de Gumbel-Hougaard la adoptada, debido a la reproducción que hace del valor de λ_U^{CFG} para las tres *duplas* que se establecen y que se analizaron— orienta su selección para el caso trivariado.

Tal selección no se considera inapropiada, ya que según se observa en la Tabla 9 y Tabla 10, tal *FC* de Gumbel-Hougaard muestra ajustes bastante similares a los de la *FC* de Frank simétrica y asimétrica. Lo anterior quedó verificado con base en los coeficientes de correlación (r_{xy}) entre las probabilidades empíricas y las teóricas trivariadas de ambas *FC*, que resultaron prácticamente iguales.

Periodos de retorno de las crecientes de diseño

Suponiendo que en las cercanías de aguas debajo de la estación hidrométrica *Tempoal* se van a construir diques para protección de

planicies de inundación con fines agrícolas e industriales y un puente para cruzarlo, entonces se requiere estimar *eventos de diseño* con periodos de retorno conjuntos o trivariados de 50, 100, 500 y 1 000 años. Por lo anterior, es necesario estimar gastos o *crecientes de diseño* (QX) en la estación base *Tempoal* con los cuatro periodos de retorno conjuntos T_{KEN} citados.

Estimación de las crecientes de diseño trivariadas

Desigualdad de los periodos de retorno trivariados

Definidos los cuatro periodos de retorno conjuntos de diseño, se aplica su probabilidad de no excedencia (u, v, w) respectiva de 0.98, 0.99, 0.998 y 0.999 en la Ecuación (5) con los valores de θ mostrados en la Tabla 8. Además, se aplica la Ecuación (17) con $\theta = 2.795$ y la Ecuación (22) con $\theta_1 = 2.3475$ y $\theta_2 = 3.8288$, para obtener las probabilidades que requieren las ecuaciones (29) y (30) de los periodos de retorno de tipo OR y AND. Para la estimación del T_{KEN} se aplicó la Ecuación (35).

Los resultados de la Tabla 11 muestran enorme similitud en los periodos de retorno trivariados de tipo OR y AND, de las FC simétrica y asimétrica. Por lo anterior, se puede utilizar el T_{KEN} para obtener los eventos de diseño conjuntos o crecientes anuales en la estación base Tempoal. Los resultados de la columna 3 de la Tabla 11 permiten *verificar* la Ecuación (36) de la desigualdad de los Tr trivariados.

Tabla 11. Periodos de retorno conjuntos de tipo OR, AND y *secundario* estimados con las *FC* trivariadas de Gumbel-Hougaard simétrica y asimétrica, para las ternas de crecientes anuales de las estaciones Tempoal-El Cardón-Terrerillos, México.

<i>Tr</i> (años)	Tipo de <i>Tr</i> con la <i>FC</i> simétrica			<i>FC</i> asimétrica	
	OR	Secundario	AND	OR	AND
50	33.9	94.2	119.0	33.8	118.0
100	67.7	189.0	240.0	67.5	238.1
500	337.6	947.5	1 208.8	336.9	1 199.1
1 000	675.1	1 895.7	2 419.6	673.6	2 399.8

Estimación de los eventos de diseño

Con base en la Ecuación (35) de la distribución de Kendall trivariada, establecida para la *FC* de Gumbel-Hougaard simétrica, se buscó por tanteos el periodo de retorno univariado (*Tr*) y su respectiva probabilidad de no excedencia (*s*), que definen un periodo de retorno *secundario* igual al conjunto o trivariado de diseño. Encontrado tal valor de la variable unitaria (*s*), se obtienen con la solución inversa de la distribución marginal (Ecuación (37)), las variables *QX*, respectivas. En la Tabla 12 se exponen los resultados.

Tabla 12. Eventos de diseño obtenidos con el periodo de retorno secundario que iguala al univariado de las *ternas de crecientes conjuntas* de las estaciones del sistema del río Tempoal, México.

<i>Tr</i> (años) univariado	Tempoal-El Cardón-Terrerillos	
	Prob. (s) del <i>Tr</i> secundario	Valor del <i>QX</i> (m³/s)
50	0.962075	4 719
100	0.980916	5 569
500	0.996164	7 497
1 000	0.998081	8 303

Las predicciones trivariadas de la Tabla 12 resultan inferiores a las predicciones univariadas de la estación base Tempoal mostradas en la Tabla 5, como se muestra en la Figura 3. Los valores de diseño de la Tabla 12, son inferiores un 24.6, 23.3, 21.3 y 20.8 %, respectivamente, en relación con las predicciones de la Tabla 5.

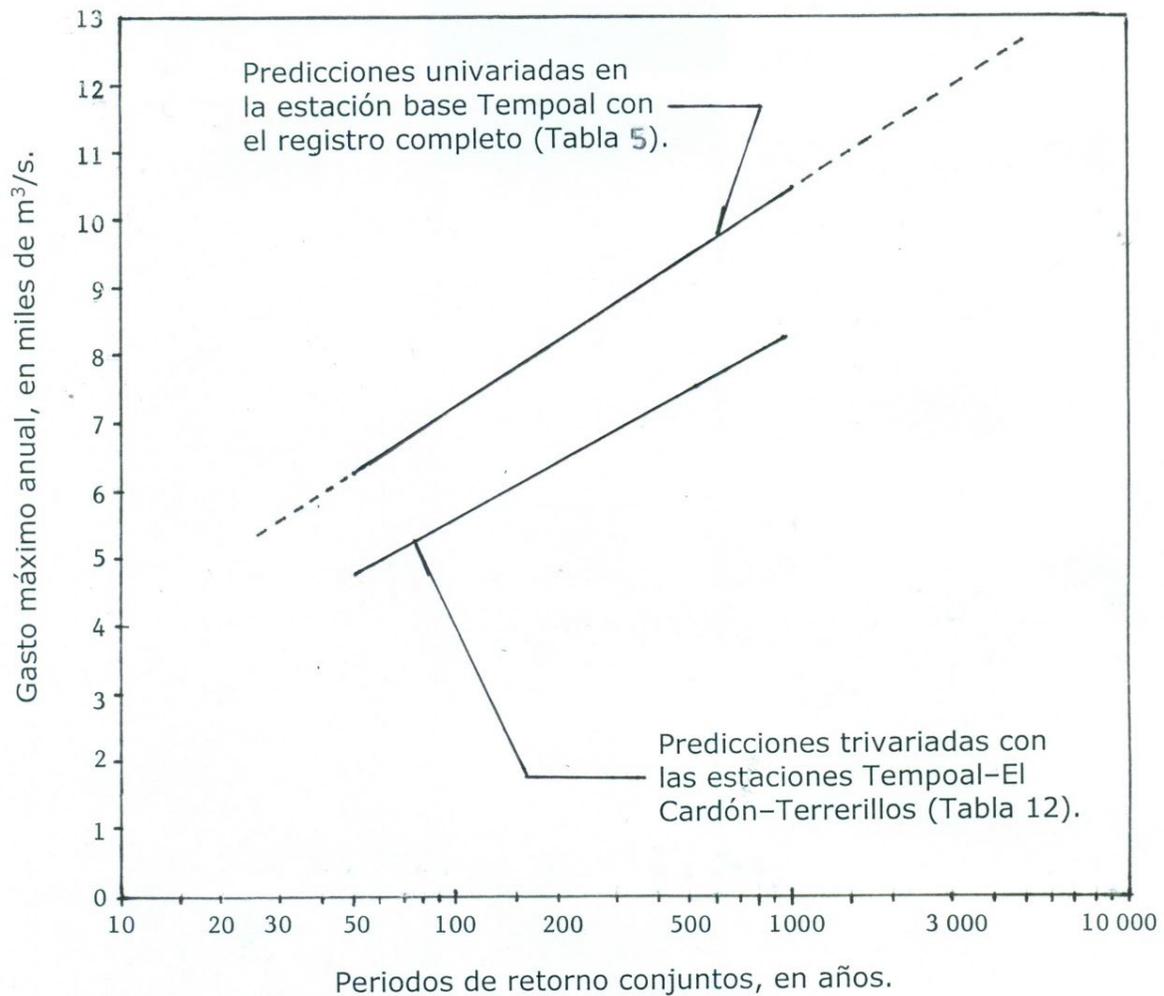


Figura 3. Gráfica de predicciones de diseño obtenidas con los enfoques univariado y trivariado para las crecientes anuales de la estación base Tempoal, México.

Otro estudio similar

Campos-Aranda (2022) expone una aplicación numérica en la cual las predicciones obtenidas con el análisis de frecuencias *bivariado* de crecientes con dependencia regional resultaron ligeramente mayores que

las estimadas con el registro completo de la estación de aforos base y, por ello, son las adoptadas. En estos casos, el análisis de frecuencias de crecientes bivariado se considera un éxito debido a que conduce a predicciones más severas o críticas.

Enfoque futuro de aplicación de las FC

Los análisis de frecuencias de crecientes bivariados y trivariados en una estación base o de interés para un proyecto, con dependencia regional, es decir, con registros cercanos de estaciones hidrométricas auxiliares que muestran correlación con el de la estación base, deberán evolucionar para procesar FDP marginales de series o registros *no estacionarios*. Hacia tal eventualidad, ya se han comenzado sugerir diversos enfoques, como el de Bender, Wahl y Jensen (2014), y el de Chebana y Ouarda (2021).

Conclusiones

Los análisis de frecuencias de crecientes *trivariados* de las variables gasto máximo en la estación base (QX) y en las estaciones auxiliares (QY y QZ), que exhiben una correlación o dependencia regional, y tienen igual amplitud de registro, permitirán un contraste de la estimación univariada de *crecientes de diseño* del registro completo de la estación base contra las obtenidas con la *FC* asociada con un periodo de retorno conjunto.

El uso de las *funciones Cópula (FC)* en los análisis de frecuencias trivariados permite construir la distribución conjunta con base en las funciones marginales. Por lo anterior, las distribuciones de probabilidad

idóneas de QX , QY y QZ se definen con la exactitud máxima posible, y pueden ser diferentes y de cualquier tipo.

La estimación del periodo de retorno trivariado de tipo AND requiere de las distribuciones bivariadas, en el caso estudiado de las duplas de variables $QX-QY$, $QX-QZ$ y $QY-QZ$. Por lo anterior, primero se buscan FC que reproduzcan la dependencia observada (λ_U^{CFG}) y muestren un buen ajuste con las variables conjuntas citadas.

En la aplicación numérica descrita para un lapso común de 43 crecientes anuales registradas en el sistema del río Tempoal, de la Región Hidrológica No. 26 (Pánuco), México, se emplearon como estación base Tempoal, y como auxiliares El Cardón y Terrerillos, por ser las que mostraron mayor correlación. Se aplicaron dos familias de FC : Frank y Gumbel-Hougaard.

Para las ternas de datos anuales de QX , QY y QZ se aplicaron FC de Arquímedes trivariadas simétricas, con un parámetro de asociación (θ) y trivariadas asimétricas, con dos parámetros de asociación (θ_1 , θ_2) de las familias citadas. Finalmente, se estimaron periodos de retorno conjuntos de tipo OR, AND y Kendall. Estos últimos permiten obtener los eventos de diseño de QX mostrados en la Tabla 12.

Este tipo de análisis de frecuencias de crecientes con *dependencia regional* conduce en algunos casos a predicciones mayores que las estimaciones univariadas realizadas con el registro completo de la estación base. En otras ocasiones, como el caso expuesto, sus predicciones son menores y entonces permiten verificar la tendencia de la relación periodo de retorno contra gasto de diseño, como se mostró en la Figura 3.

Los análisis de frecuencias de crecientes trivariados con *dependencia regional* descritos tienen gran sencillez y no presentan complicaciones de cómputo cuando se realizan con base en las *FC*.

Agradecimientos

Se agradecen los comentarios y correcciones sugeridas por los árbitros anónimos C, D, E y del editor temático, los cuales permitieron volver más explícito el trabajo, al cambiar el orden de algunos temas o cálculos, así como reforzar conceptos teóricos relacionados con las funciones Cópula y plantear opciones para investigaciones futuras. También se incluyó la referencia de un estudio bivariado similar por aplicar la dependencia o correlación regional.

Referencias

- AghaKouchak, A., Sellars, S., & Sorooshian, S. (2013). Chapter 6. Methods of tail dependence estimation. In: AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (eds.). *Extremes in a changing climate* (pp. 163-179). Dordrecht, The Netherlands: Springer.
- Aldama, A. A., 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.
- Barbe, P., Genest, C., Ghoudi, K., & Rémillard, B. (1996). On Kendall's Process. *Journal of Multivariate Analysis*, 58(2), 197-229.

- Beale, E. M. L. & Little, R. J. A. (1975). Missing values in multivariate analysis. *Journal of Royal Statistical Society Series B*, 37(1), 129-145. DOI: 10.1111/j.2517-6161.1975.tb01037.x
- Bender, J., Wahl, T., & Jensen, J. (2014). Multivariate design in the presence of non-stationarity. *Journal of Hydrology*, 514(June), 123-130. DOI: 10.1016/j.jhydrol.2014.04.017
- Bobée, B. (1975). The Log-Pearson type 3 distribution and its application to Hydrology. *Water Resources Research*, 11(5), 681-689. DOI: 10.1029/WR011i005p00681
- Bobée, B., & Ashkar, F. (1991). Chapter 1. Data requirements for hydrologic frequency analysis. In: *The Gamma Family and derived distributions applied in hydrology* (pp. 1-12). Littleton, USA: Water Resources Publications.
- Box, M. J. (1965). A new method of constrained optimization and a comparison with other methods. *Computer Journal*, 8(1), 42-52.
- Bunday, B. D. (1985). Theme 6.2. The complex method. In: *Basic optimisation methods* (pp. 98-106). London, England: Edward Arnold publishers, Ltd.
- Campos-Aranda, D. F. (2003). Capítulo 7. Integración numérica y Capítulo 9. Optimización numérica. En: *Introducción a los métodos numéricos: software en Basic y aplicaciones en hidrología superficial* (pp. 137-153, 172-211). San Luis Potosí, México: Editorial Universitaria Potosina.
- Campos-Aranda, D. F. (2014). Predicción de crecientes usando la distribución Pareto Generalizada ajustada con tres métodos simples. *Tlálloc*, 65, octubre-diciembre, 7-26.

- Campos-Aranda, D. F. (2015). Estimación simultánea de datos hidrológicos anuales faltantes en múltiples sitios. *Ingeniería. Investigación y Tecnología*, 16(2), 295-306.
- Campos-Aranda, D. F. (2022). Análisis de frecuencias de crecientes bivariado con dependencia regional y funciones Cópula. *Aqua-LAC*, 14(2), 47-61. DOI: 10.29104/phi-aqualac/2022-v14-2-11
- Campos-Aranda, D. F. (2023). Selección y aplicación de funciones Cópula con dependencia en su extremo derecho al análisis de frecuencias conjunto (Q,V) de crecientes anuales. *Tecnología y ciencias del agua*, 14(5), 120-188. DOI: 10.24850/j-tyca-14-05-03
- Capéraà, P., Fougères, A. L., & Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3), 567-577. DOI: 10.1093/biomet/84.3.567
- 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
- Chebana, F., & Ouarda, T. B. M. J. (2021). Multivariate non-stationary hydrological frequency analysis. *Journal of Hydrology*, 593(February), 125907. DOI: 10.1016/j.jhydrol.2020.125907
- Chen, L., & Guo, S. (2019). Chapter 2. Copula theory and Chapter 3. Copula-based flood frequency analysis. In: *Copulas and its application in hydrology and water resources* (pp. 13-38, 39-71). Gateway East, Singapore: Springer.

- Chowdhary, H., & Singh, V. P. (2019). Chapter 11. Multivariate frequency distributions in hydrology. In: Teegavarapu, R. S. V., Salas, J. D., & Stedinger, J. R. (eds.). *Statistical analysis of hydrologic variables* (pp. 407-489). Reston, USA: American Society of Civil Engineers.
- Davis, P. J., & Polonsky, I. (1972). Chapter 25. Numerical interpolation, differentiation and integration. In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of mathematical functions* (9th print.) (pp. 875-926). New York, USA: Dover Publications.
- Dupuis, D. J. (2007). Using Copulas in hydrology: Benefits, cautions, and issues. *Journal of Hydrologic Engineering*, 12(4), 381-393. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(381)
- Escalante-Sandoval, C. A., & Raynal-Villaseñor, J. A. (1994). A trivariate extreme value distribution applied to flood frequency analysis. *Journal of Research of the National Institute of Standards and Technology*, 99(4), 369-375.
- Escalante-Sandoval, C., & Raynal-Villaseñor, J. (2008). Trivariate generalized extreme value distribution in flood frequency analysis. *Hydrological Sciences Journal*, 53(3), 550-567.
- 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.
- Gräler, B., van den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., & Verhoest, N. E. C. (2013). Multivariate return periods in hydrology: A critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, 17(4), 1281-1296. DOI: 10.5194/hess-17-1281-2013
- Grimaldi, S., & Serinaldi, F. (2006a). Design hyetograph analysis with 3-copula function. *Hydrological Sciences Journal*, 51(2), 223-238. DOI: 10.1623/hysj.51.2.223
- Grimaldi, S., & Serinaldi, F. (2006b). Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources*, 29(8), 1155-1167. DOI: 10.1016/j.advwatres.2005.09.005
- Hosking, J. R. 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, England: Cambridge University Press.

- IMTA, Instituto Mexicano de Tecnología del Agua. (2003). *Banco Nacional de Datos de Aguas Superficiales (Bandas)* (8 CD). Jiutepec, México: Comisión Nacional del Agua, Secretaría de Medio Ambiente y Recursos Naturales, Instituto Mexicano de Tecnología del Agua.
- 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
- Meylan, P., Favre, A. C., & Musy, A. (2012). Chapter 3. Selecting and checking data series and Theme 9.2. Multivariate Frequency Analysis using Copulas. In: *Predictive hydrology. A frequency analysis approach* (pp. 29-70, 164-176). Boca Raton, USA: CRC Press.
- Nelsen, R. B. (2006). Chapter 4. Archimedean Copulas. In: *An introduction to Copulas* (2nd ed.) (pp. 109-155). New York, USA: Springer Series in Statistics.
- Nieves, A., & Domínguez, F. C. (1998). Secciones 6.2 y 6.3. Cuadratura de Gauss e integrales múltiples. *Métodos numéricos. Aplicados a la ingeniería* (pp. 416-434). México, DF, México: Compañía Editorial Continental.

- Poulin, A., Huard, D., Favre, A. C., & Pugin, S. (2007). Importance of tail dependence in bivariate frequency analysis. *Journal of Hydrologic Engineering*, 12(4), 394-403. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(394)
- Rao, A. R., & Hamed, K. H. (2000). Theme 1.8. Tests on hydrologic data and Theme 8.3: The Generalized Pareto distribution. *Flood frequency analysis* (pp. 12-21, 271-290). Boca Raton, USA: CRC Press.
- Requena, A. I., Mediero, L., & Garrote, L. (2013). A bivariate return period based on copulas for hydrologic dam design: Accounting for reservoir routing in risk estimation. *Hydrology and Earth System Sciences*, 17(8), 3023-3038. DOI: 10.5194/hess-17-3023-2013
- Salvadori, G., & De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40(W12511), 1-17. DOI: 10.1029/2004WR003133
- Salvadori, G., & De Michele, C. (2007). On the use of Copulas in Hydrology: Theory and Practice. *Journal of Hydrologic Engineering*, 12(4), 369-380. DOI: 10.1061/(ASCE)1084-0699(2007)12:4(369)
- Salvadori, G., De Michele, C., Kottegoda, N. T., & Rosso, R. (2007). Chapter 3. Bivariate analysis via Copulas; Appendix B: Dependence, and Appendix C: Families of Copulas. *Extremes in nature. An approach using Copulas* (pp. 131-175, 219-232, 233-269). Dordrecht, The Netherlands: Springer.

- Salvadori, G., De Michele, C., & Durante, F. (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences*, 15(11), 3293-3305. DOI: 10.5194/hess-15-3293-2011
- 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* (9th print.) (pp. 997-1010). New York, USA: Dover Publications.
- WRC, Water Resources Council. (1977). *Guidelines for determining flood flow frequency* (revised edition). Bulletin #17A of the Hydrology Committee. Washington, DC, USA: Water Resources Council.
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82. DOI: 10.3354/cr030079
- Xu, C., Yin, J., Guo, S., Liu, Z., & Hong, X. (2016). Deriving design flood hydrograph based on conditional distribution: A case study of Danjiangkou reservoir in Hanjiang basin. *Mathematical Problems in Engineering*, 2016(4319646), 1-16. DOI: 10.1155/2016/4319646
- 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. DOI: 10.1016/S0022-1694(99)00168-7

- Yue, S. (2000). Joint probability distribution of annual maximum storm peaks and amounts as represented by daily rainfalls. *Hydrological Sciences Journal*, 45(2), 315-326. DOI: 10.1080/02626660009492327
- Yue, S., & Rasmussen, P. (2002). Bivariate frequency analysis: Discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14), 2881-2898. DOI: 10.1002/hyp.1185
- Zhang, L., & Singh, V. P. (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 and Chapter 4. Symmetric Archimedean Copulas. *Copulas and their applications in water resources engineering*. pp. (pp. 62-122, 123-171). Cambridge, United Kingdom: Cambridge University Press.