

DOI: 10.24850/j-tyca-2021-02-10

Artículos

## **Ajuste de la distribución de valores extremos de dos componentes (TCEV) por medio de máxima verosimilitud**

### **Fit of the two-component extreme value (TCEV) distribution through of maximum likelihood**

Daniel Francisco Campos-Aranda<sup>1</sup>

<sup>1</sup>Profesor jubilado de la Universidad Autónoma de San Luis Potosí, México, campos\_aranda@hotmail.com

Autor para correspondencia: Daniel Francisco Campos-Aranda, campos\_aranda@hotmail.com

#### **Resumen**

En muchas cuencas medianas y grandes de nuestro país y del mundo, su registro anual de crecientes está integrado por eventos que fueron generados por fenómenos físicamente diferentes. Por ejemplo, muchas crecientes se originan con tormentas locales y una porción pequeña se debe a lluvias ciclónicas de amplia cobertura y duración, que generan crecientes extraordinarias. Para este tipo de registros se propone la

distribución TCEV (*two-component extreme value*) con cuatro parámetros de ajuste, la cual tiene una base teórica que permite una interpretación aproximada para dos mecanismos de generación de las crecientes y que además es capaz de reproducir la variabilidad real del coeficiente de asimetría. En este trabajo se detalla su génesis y el método de ajuste por máxima verosimilitud, según dos versiones numéricas: (1) sustitución sucesiva y (2) maximización de la función objetivo. Se procesaron seis registros de crecientes, cuya amplitud varió de 31 a 72 datos, con tres a seis valores dispersos o crecientes que se apartan de la tendencia general. Se contrastan las predicciones del modelo TCEV, ajustado con cada método numérico, contra las obtenidas con las distribuciones de aplicación establecida bajo precepto (LP3, GVE y LOG) y la Wakeby. Aceptando el error estándar de ajuste como criterio de selección, se deduce que la distribución TCEV es la mejor opción en dos de los seis registros procesados. Por último, se formulan las conclusiones, las cuales sugieren la aplicación sistemática de la distribución TCEV con ambos métodos numéricos en los registros que presentan dos poblaciones mezcladas.

**Palabras clave:** distribución TCEV, proceso de Poisson, máxima verosimilitud, algoritmo de Rosenbrock, error estándar de ajuste, homogeneidad y estacionariedad, contraste de predicciones.

## Abstract

The annual record of floods, in many medium and large basins of our country and the world, is made up of events generated by physically different phenomena. For example, many floods originate from local

storms and a small portion is generated by cyclonic rains of wide coverage and duration, which generate extraordinary floods. The TCEV distribution (two-component extreme value) with four fitting parameters has been proposed for this type of records. TCEV has a theoretical basis that allows an approximate interpretation for two mechanisms of generation of the floods and it is also capable of reproducing the real variability of the asymmetry coefficient. This paper details its genesis and the fitting method by maximum likelihood, according to two numerical versions: (1) successive substitution and (2) objective function maximization. Six registries of floods were processed, the amplitude of which varied from 31 to 72 data, with three to six outliers or floods values that depart from the general trend. The predictions of the TCEV model, fitted with both numerical methods, are compared with those obtained using the standard application distributions (LP3, GVE and LOG) and the Wakeby distribution. Accepting the standard error of fit as a selection criterion, it follows that the TCEV distribution is the best option in two of the six processed records. Lastly, as conclusion, the systematic application of the TCEV distribution is suggested, using both numerical methods, in records that present two mixed populations.

**Keywords:** TCEV distribution, Poisson process, maximum likelihood, Rosenbrock algorithm, standard error of fit, homogeneity and stationarity, prediction contrast.

Recibido: 26/09/2019

Aceptado: 28/06/2020

## Introducción

### Generalidades del AFC

El *análisis de frecuencia de crecientes* (AFC) es una de las estimaciones básicas de la hidrología superficial, ya que permite la planeación, diseño y revisión de todas las obras hidráulicas requeridas por la sociedad, sean éstas de aprovechamiento, como los embalses; o de protección, como los diques, encauzamientos, presas de control, puentes, obras de drenaje urbano, etcétera. El AFC intenta definir la relación entre los gastos máximos de las crecientes anuales y su probabilidad de ser excedidos. Seleccionada tal probabilidad o riesgo para la *creciente de diseño*, su valor debe estimarse con la mayor precisión posible, ya que un error por defecto genera un aumento del riesgo adoptado y un error por exceso, origina un aumento del costo del proyecto (Botto, Ganora, Laio, & Claps, 2014).

El procedimiento más exacto para realizar un AFC consiste en representar el registro disponible de gastos máximos anuales por una función de distribución de probabilidades (FDP), y con base en tal modelo probabilístico realizar las inferencias buscadas o *predicciones*. Esta técnica puede conducir a estimaciones inexactas debido principalmente a los

cuatro factores siguientes: (1) errores de medición en los datos; (2) amplitud finita del registro disponible; (3) presencia de mecanismos diferentes de generación de las crecientes, y (4) elección de la FDP (Merz & Blöschl, 2008).

Francés-García (1995) indica que históricamente la evolución de los AFC ha pasado por tres etapas: (I) búsqueda y ajuste de una mejor FDP a los gastos máximos anuales, hasta la década de 1960; (II) mejora de los métodos de estimación de los parámetros de ajuste de las FDP, durante el decenio de 1970, y (III) aumento en la exactitud de las estimaciones, desde la década de 1980, mediante el incremento de los registros de gastos máximos, a través del análisis regional y/o la incorporación de la información histórica de crecientes.

Una cuarta etapa evolutiva de los AFC inicia en el siglo XXI y se refiere al procesamiento de registros que no son estacionarios, debido a que presentan tendencias, y/o aumento o disminución de la variabilidad, como consecuencia de cambios físicos en las cuencas o del cambio climático (Khaliq, Ouarda, Ondo, Gachon, & Bobée, 2006; Katz, 2013; López-de-la-Cruz & Francés, 2014; Prosdocimi, Kjeldsen & Svensson, 2014; Campos-Aranda, 2018).

## **AFC con base en FDP mixtas**

Desde la mitad del siglo pasado cobró interés la presencia de dos propiedades estadísticas observadas en los registros de gastos máximos anuales. La primera se llamó "efecto pata de perro" (*dog leg effect*), designada así por Potter (1958). Se observa al dibujar en el papel de probabilidad los datos y encontrar un cambio brusco de pendiente debido a la presencia de valores dispersos (*outliers*). La segunda propiedad se denominó "fenómeno de separación", estudiada y documentada por Matalas, Slack y Wallis (1975), quienes dibujaron en las abscisas los valores promedio de los coeficientes de asimetría o de sesgo, y en las ordenadas sus respectivas desviaciones estándar de los registros sintéticos generados con las FDP más comunes. Al llevar a tal gráfica los puntos de los registros históricos disponibles de gastos máximos anuales en cada una de las 14 regiones hidrológicamente homogéneas de EUA se observó que las FDP quedaban por debajo de los puntos reales; esto indicaba que tales FDP no podían representar la variabilidad muestral que tienen las crecientes.

Detectadas por los hidrólogos las deficiencias citadas de las FDP utilizadas hacia el final de la década de 1970, se propuso su solución a través de los *modelos probabilísticos mixtos*, los cuales toman en cuenta los diferentes orígenes en las crecientes de un registro. Los modelos mixtos se aplican bajo dos enfoques diferentes: (a) crecientes independientes y separables, y (b) crecientes independientes que están mezcladas. El ejemplo clásico del primer enfoque es el modelo sugerido por Waylen y Woo (1982), aplicable cuando las crecientes originadas por fusión de la nieve y lluvia son plenamente identificables por su fecha de ocurrencia en el registro disponible de gastos máximos anuales. Las crecientes de fusión de nieve originan gastos bajos y ocurren en

primavera-verano, y las de lluvia generan gastos mayores y suceden en el invierno de octubre a febrero. Para el segundo enfoque se han propuesto las mezclas de FDP, por ejemplo: (1) la denominada Gumbel doble, sugerida por González-Villarreal (1970); (2) la distribución Wakeby, propuesta por Houghton (1978), que intenta modelar la porción izquierda y derecha de manera separada; (3) el modelo TCEV, desarrollado por Rossi, Fiorentino y Versace (1984); y (4) la función Gumbel mixta aplicada recientemente por Molina-Aguilar, Gutiérrez-López y Aparicio-Mijares (2018). Rulfova, Buishand, Roth y Kysely (2016) propusieron la distribución TCGEV (*two-component generalized extreme value*) de dos componentes de valores extremos generalizados, como un modelo más flexible para el análisis de precipitaciones máximas anuales de seis horas de duración.

La distribución TCEV se describe con detalle en este artículo debido a que no ha sido aplicada en México de manera sistemática, ni bajo precepto, ni como una opción de los modelos probabilísticos sugeridos en registros de crecientes con poblaciones mezcladas.

## Objetivo

En este trabajo se expone con detalle el origen teórico de la distribución TCEV (*two-component extreme value*) con cuatro parámetros de ajuste, que conduce a un modelo probabilístico adecuado a registros de

crecientes anuales que proceden de dos mecanismos físicamente diferentes, pero que están mezclados. El método de ajuste de la distribución TCEV, por medio de máxima verosimilitud, se describe en sus dos versiones: (1) método numérico de sustituciones sucesivas y (2) método numérico de optimización. Se procesan, con ambos métodos, seis registros de crecientes tomados de la literatura especializada, los cuales varían de 31 a 72 datos, y tienen entre tres y seis valores dispersos. Enseguida, se contrastan seis predicciones de periodos de retorno de 10, 25, 50, 100, 500 y 1 000 años, cada una con seis valores procedentes del ajuste de las distribuciones TCEV, Log-Pearson tipo III, General de Valores Extremos, Logística Generalizada y Wakeby. Por último, se formulan las conclusiones de estudio.

## **Otros métodos de ajuste y aplicaciones de la distribución TCEV**

Francés (1998) amplió el método de máxima verosimilitud para el uso conjunto de información sistemática e histórica de crecientes. En cambio, Beran, Hosking y Arnell (1986) exponen tres ecuaciones para cuantificar los momentos de probabilidad pesada (MPP) e indican que tales expresiones deben resolverse de manera iterativa, por lo cual resultan más complicadas que el método de máxima verosimilitud. El procedimiento del método de los MPP ha sido expuesto por Singh (1998),

así como el basado en el principio de máxima entropía, sugerido originalmente por Fiorentino, Arora y Singh (1987).

Por otra parte, Fiorentino, Versace y Rossi (1985) fueron los primeros en indicar que la distribución TCEV mostró un buen ajuste en registros de precipitación máxima diaria anual del sur de Italia, que fueron procesados como series de duración parcial o de magnitudes superiores a un valor umbral. Posteriormente, Cannarozzo, D'Asaro y Ferro (1995) aplicaron la distribución TCEV de manera regional en Sicilia, Italia, usando registros de lluvia y de crecientes. Ferro y Porto (2006) usan el modelo TCEV con el enfoque jerárquico regional en un AFC para Sicilia, Italia. Tal enfoque regional fue propuesto originalmente por Fiorentino, Gabriele, Rossi y Versace (1987). Boni, Parodi y Rudari (2006), así como Aronica y Candela (2007) emplean la distribución TCEV en los análisis regionales de lluvias que han descrito. Escalante-Sandoval y Reyes-Chávez (2004) realizan un análisis bivariado de crecientes, empleando como distribuciones marginales la TCEV.

## **Descripción de la teoría y datos**

### **FDP de los modelos mixtos**

Waylen y Woo (1982) llaman  $R$  a la variable aleatoria que representa a las crecientes anuales originadas por la precipitación, que son las más grandes, y  $S$  a las generadas por la fusión de la nieve, que son las básicas o menores;  $X_i$  es una nueva variable aleatoria que es el máximo de ambas;  $i$  es el año, y  $n$  es la amplitud del registro disponible. Como se acepta la hipótesis de que  $R$  y  $S$  son independientes, su FDP es igual al producto de ambas, y cada una se estima por separado con los  $n$  datos correspondientes tomados del registro. Entonces, la función de probabilidad de no excedencia de  $X_i$  es (Francés-García, 1995):

$$X_i = \max(R_i, S_i) \quad (1)$$

$$F_X(x) = P(X \leq x) = P(R \leq x, S \leq x) = P(R \leq x) \cdot P(S \leq x) = F_R(x) \cdot F_S(x) \quad (2)$$

En el caso más simple,  $F_R(x)$  y  $F_S(x)$  son FDP Gumbel, con dos parámetros de ajuste, ubicación y escala, cada una. La aplicación de la distribución Gumbel doble de cinco parámetros de ajuste se ha resuelto de manera práctica a través de optimización numérica no restringida (Gómez, Aparicio, & Patiño, 2010); su FDP es (González-Villarreal, 1970):

$$F_X(x) = P(X \leq x) = F_R(x) \cdot [p + (1 - p) \cdot F_S(x)] \quad (3)$$

En la ecuación anterior,  $p$  es la probabilidad de tener eventos ordinarios. En la distribución Wakeby (Houghton, 1978), la nueva variable aleatoria está definida por la suma, ya que la creciente máxima anual se produce por la combinación en el tiempo de los dos mecanismos generadores y entonces su FDP se expresa en forma explícita como:

$$x = -a \cdot [1 - F_X(x)]^b + c \cdot [1 - F_X(x)]^d + e \quad (4)$$

La estimación de sus cinco parámetros de ajuste se ha resuelto por medio del método de los momentos L (Hosking & Wallis, 1997) y mediante optimización numérica restringida (Campos-Aranda, 2001). Por último, la distribución Gumbel mixta ha sido ajustada con base en algoritmos modernos de búsqueda (Molina-Aguilar *et al.*, 2018); su FDP es similar a la Ecuación (3), ésta es:

$$F_X(x) = P(X \leq x) = p \cdot F_R(x) + (1 - p) \cdot F_S(x) \quad (5)$$

## Génesis de la distribución TCEV

### Proceso compuesto de Poisson

Rossi *et al.* (1984) establecieron que un principio teórico es necesario que respalde a una cierta FDP, para que pueda ser considerada y aceptada como un modelo de origen o fuente (*parent distribution*) de las crecientes de una cierta región geográfica. Debido a las limitaciones que tienen los registros disponibles de crecientes, además de sus particularidades estadísticas, la FDP por seleccionar debe contar con una estructura probabilística que simule la procedencia real de las crecientes de la zona bajo estudio.

Al respecto, Rossi *et al.* (1984) indican que la hipótesis básica de la teoría de valores extremos es muy restrictiva, al considerar que las crecientes anuales provienen del valor máximo de la una serie amplia de variables aleatorias independientes e idénticamente distribuidas (*iid*). Un planteamiento más flexible y de mayor similitud con la realidad física considera que las crecientes anuales ( $X$ ) son el máximo de una secuencia de  $k$  números aleatorios con distribución de Poisson, que son variables aleatorias *iid* no negativas  $Z_i$ , con  $i = 1, 2, \dots, k$ , las cuales son también independientes de  $k$ . Lo anterior equivale a modelar  $X$  como el máximo de un proceso compuesto de Poisson, cuya FDP es (Todorovic & Zelenhasic, 1970):

$$F_X(x) = P[X \leq x] = \exp\{-\lambda[1 - F_Z(x)]\} \quad \text{para } x \geq 0 \quad (6)$$

en donde  $\lambda = E[k]$  es el parámetro del proceso de Poisson. Si  $Z$  se adopta como una variable aleatoria exponencial, se tiene que:

$$F_Z(z) = 1 - \exp(-x/\theta) \text{ para } z \geq 0 \quad (7)$$

siendo  $\theta = E[Z]$ . Sustituyendo la Ecuación (7) en la Ecuación (6) se obtiene:

$$F_X(x) = \exp[-\lambda \cdot \exp(-x/\theta)] \text{ para } x \geq 0 \quad (8)$$

La expresión anterior se puede transformar en la FDP Gumbel o tipo I de valores extremos, haciendo su parámetro de ubicación  $\varepsilon$  igual a  $\theta$  por el logaritmo natural de  $\lambda$ , obteniéndose:

$$F_X(x) = \exp\left\{-\exp\left[-\frac{x-\varepsilon}{\theta}\right]\right\} \text{ para } -\infty < x < \infty \quad (9)$$

siendo:

$$\varepsilon = \theta \cdot \ln \lambda \quad (10)$$

Y:

$$\lambda = \exp(\varepsilon/\theta) \quad (11)$$

La Ecuación (8) tiene una discontinuidad en  $x = 0$ , que vale  $\exp(-\lambda)$ , es decir, exhibe una componente delta, mientras que la Ecuación (9) se extiende hacia los valores negativos de  $x$ . El valor de  $\exp(-\lambda)$  es cercano a cero, excepto quizás en climas áridos por la ocurrencia de valores nulos (Rossi *et al.*, 1984).

### Desarrollo del modelo mixto

Un enfoque posible para tomar en cuenta en los AFC la presencia de valores dispersos (*outliers*) y series de una asimetría elevada consiste en aceptar que el registro de crecientes proviene de dos mecanismos diferentes de generación: uno de ellos origina crecientes de baja magnitud, pero frecuentes, y el otro genera crecientes extraordinarias esporádicas. Entonces, asumiendo que existen dos secuencias de crecientes independientes de variables *iid*,  $Z_{1i}$ ,  $i = 1, 2, \dots, K_1$  y  $Z_{2j}$ ,  $j = 1, 2, \dots, K_2$ , cada una definida por un proceso compuesto de Poisson con parámetros  $\lambda_1 = E[K_1]$  y  $\lambda_2 = E[K_2]$ , respectivamente, con  $\lambda_1 > \lambda_2$ . El número total  $K = K_1 + K_2$  de crecientes independientes en un año será también un proceso compuesto de Poisson con parámetro  $\lambda = \lambda_1 + \lambda_2$ , mientras que la magnitud  $Z$  de las crecientes anuales estará definida por la mezcla de FDP, es decir (Rossi *et al.*, 1984):

$$f(z) = p \cdot F_{Z_1}(z) + (1 - p) \cdot F_{Z_2}(z) \text{ para } z \geq 0 \quad (12)$$

en donde  $p = \lambda_1/\lambda$  es la proporción de  $Z_1$  en la mezcla (crecientes básicas u ordinarias) y  $(1 - p)$  la de las crecientes extraordinarias con mucha mayor variabilidad. Sustituyendo la Ecuación (6) en la Ecuación (12) se obtiene:

$$F_X(x) = \exp\{-\lambda_1[1 - F_{Z_1}(x)] - \lambda_2[1 - F_{Z_2}(x)]\} \text{ para } x \geq 0 \quad (13)$$

Designando con  $X_1$  y  $X_2$  a los máximos anuales de  $Z_1$  y  $Z_2$  se tiene que:

$$F_X(x) = F_{X_1}(x) \cdot F_{X_2}(x) \quad (14)$$

Por último, si  $Z_1$  y  $Z_2$  son variables aleatorias exponenciales, como en la Ecuación (8), la Ecuación (14) se transforma en la distribución de valores extremos de dos componentes (TCEV):

$$F_X(x) = \exp[-\lambda_1 \cdot \exp(-x/\theta_1) - \lambda_2 \cdot \exp(-x/\theta_2)] \text{ para } x \geq 0 \quad (15)$$

Sus cuatro parámetros de ajuste caracterizan las crecientes básicas y las extraordinarias con el número medio de crecientes independientes por año ( $\lambda_1 > 0$ ,  $\lambda_2 \geq 0$ ) y la amplitud media del gasto máximo anual ( $\theta_2 \geq \theta_1 > 0$ ). Fiorentino *et al.* (1985) destacan que la Ecuación (15) tiene una probabilidad finita cercana a cero, cuyo valor es  $\exp(-\lambda)$ , y que en

ausencia de la componente de valores extraordinarios, la distribución TCEV se reduce a la distribución Gumbel. La función de densidad de probabilidad de  $X$ , según la ecuación anterior es:

$$f_X(x) = F_X(x) \cdot \psi(x) \quad (16)$$

en la cual:

$$\psi(x) = (\lambda_1/\theta_1) \cdot \exp(-x/\theta_1) + (\lambda_2/\theta_2) \cdot \exp(-x/\theta_2) \quad (17)$$

Generalizando la Ecuación (11) y aplicándola a la Ecuación (15), se obtiene que el modelo TCEV equivale al producto de dos distribuciones Gumbel (Rossi *et al.*, 1984; Metcalfe, 1997), ésta es:

$$F_X(x) = \exp\{-\exp[-(x - \varepsilon_1)/\theta_1]\} \cdot \exp\{-\exp[-(x - \varepsilon_2)/\theta_2]\} \quad (18)$$

En la ecuación anterior,  $\varepsilon_1$  y  $\varepsilon_2$  son los parámetros de ubicación y  $\theta_1$  y  $\theta_2$  los de escala de cada distribución Gumbel.

### Probabilidad de valores dispersos

Beran *et al.* (1986) exponen y analizan diversos aspectos teóricos complementarios relativos a la distribución TCEV. Encuentran que al igual que la distribución Wakeby, el modelo TCEV tiene extremo derecho más denso que la mayoría de las FDP utilizadas en hidrología. También demuestran que la distribución TCEV es capaz de mostrar el fenómeno de separación al utilizar registros amplios de Inglaterra.

Beran *et al.* (1986) sugieren que la proporción de valores dispersos de un registro es de interés y puede ser utilizada para validar la aceptación de la distribución TCEV. Esta proporción denominada  $q$  es función de  $\theta$  y  $\lambda$ , según las expresiones siguientes:

$$\theta = \theta_2/\theta_1 \quad (19)$$

$$\lambda = \lambda_2/\lambda_1^{1/\theta} \quad (20)$$

$$q = -\frac{\lambda}{\theta} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \lambda^j \cdot \Gamma\left(\frac{j+1}{\theta}\right) \quad (21)$$

Beran *et al.* (1986) exponen una gráfica para estimar  $q$ , cuyas rectas varían de 0.01 a 0.90, con  $\theta \cdot \ln \lambda$  en las abscisas y  $\theta$  en las ordenadas. También indican que en la Ecuación (21) la serie converge rápidamente para  $q < 0.90$  y que cuando se empleó la distribución TCEV con 2 334 estaciones-año de 57 registros de Inglaterra se obtuvo un valor disperso por cada 33 datos máximos anuales; es decir,  $q = 3.03 \%$ ; en cambio, Rossi *et al.* (1984) encuentran en los ríos italianos un valor

disperso por cada siete datos anuales, es decir,  $q = 14.3 \%$ . Para estimar el valor de la función Gamma se puede emplear la aproximación de Stirling (Davis, 1972):

$$\Gamma(\omega) \cong e^{-\omega} \cdot \omega^{\omega-1/2} \cdot \sqrt{2\pi} \left( 1 + \frac{1}{12\omega} + \frac{1}{288\omega^2} - \frac{139}{51840\omega^3} - \frac{571}{2488320\omega^4} + \dots \right) \quad (22)$$

## Ajuste por máxima verosimilitud

### 1. Método Numérico de Sustitución Sucesiva

Rossi *et al.* (1984) exponen el logaritmo natural de la función de máxima verosimilitud y lo designan por  $L_{fmv}$ ; su expresión es:

$$L_{fmv} = \sum_{i=1}^n \ln f_X(x_i) = \sum_{i=1}^n \ln F_X(x_i) + \sum_{i=1}^n \ln \psi(x_i) \quad (23)$$

Igualando a cero las derivadas parciales de  $L_{fmv}$  con respecto a los cuatro parámetros de ajuste y realizando varias operaciones algebraicas, Rossi *et al.* (1984) obtienen las cuatro ecuaciones siguientes, las cuales se resuelven mediante una técnica numérica de sustitución sucesiva:

$$\lambda_j = \lambda_j \frac{\sum_{i=1}^n \frac{\exp(-x_i/\theta_j)}{\psi(x_i)}}{\theta_j \sum_{i=1}^n \exp(-x_i/\theta_j)} \text{ para } j = 1, 2 \quad (24)$$

$$\theta_j = \frac{\sum_{i=1}^n \frac{x_i \cdot \exp(-x_i/\theta_j)}{\psi(x_i)}}{\sum_{i=1}^n x_i \cdot \exp\left(-\frac{x_i}{\theta_j}\right) + \sum_{i=1}^n \frac{\exp(-x_i/\theta_j)}{\psi(x_i)}} \text{ para } j = 1, 2 \quad (25)$$

Para asegurar una convergencia rápida conviene iniciar con unos valores de  $\lambda$  y  $\theta$  lo más aproximados posible. Para la búsqueda de los valores iniciales, primero se dibujan en el papel Gumbel-Powell (Chow, 1964) los datos del registro de crecientes, asignándoles una posición gráfica o probabilidad de no excedencia por medio de la fórmula de Weibull (Benson, 1962):

$$P(X < x) = \frac{m}{n+1} \quad (26)$$

donde  $m$  es el número de orden del dato cuando éstos han sido ordenados de menor a mayor;  $n$  es el número de observaciones, crecientes o datos. Enseguida se identifica la serie básica y se representa por una línea recta o distribución Gumbel (Ecuación (9)), definiendo un punto  $F1, X1$  en su inicio y otro hacia el final de los datos, que se designa  $F2, X2$ . Después se traza otra recta o modelo Gumbel con mayor pendiente para representar los valores extremos o crecientes extraordinarias, usando como punto inicial  $F2, X2$ , y definiendo un punto final en  $F3, X3$  hacia el último de los

datos. Estas tres parejas de valores son llevados a las fórmulas siguientes, que proceden de la Ecuación (9) (Campos-Aranda, 2002), para definir los valores iniciales  $\lambda_j$  y  $\theta_j$ :

$$\theta_1 = \frac{(X_2 - X_1)}{\{-\ln[-\ln(F_2)]\}} - \{-\ln[-\ln(F_1)]\} \quad (27)$$

$$\varepsilon_1 = X_1 - \theta_1 \{-\ln[-\ln(F_1)]\} \quad (28)$$

$$\lambda_1 = \exp(\varepsilon_1 / \theta_1) \quad (29)$$

$$\theta_2 = \frac{(X_3 - X_2)}{\{-\ln[-\ln(F_3)]\}} - \{-\ln[-\ln(F_2)]\} \quad (30)$$

$$\varepsilon_2 = X_2 - \theta_2 \{-\ln[-\ln(F_2)]\} \quad (31)$$

$$\lambda_2 = \exp(\varepsilon_2 / \theta_2) \quad (32)$$

Con base en los valores iniciales, se aplican por primera vez la Ecuación (24) y Ecuación (25), y se evalúa la función logarítmica de máxima verosimilitud ( $L_{fmv}$ ), con la Ecuación (23), y con la Ecuación (15) y Ecuación (17) auxiliares. También se calcula el error estándar de ajuste con la Ecuación (33).

Enseguida, los nuevos valores de  $\lambda_j$  y  $\theta_j$  se convierten en los iniciales y se vuelven a aplicar la Ecuación (24) y Ecuación (25). Si el valor de  $L_{fmv}$

disminuyó, se repite el proceso; si aumentó, se suspende la sustitución sucesiva. Este proceso se programó en lenguaje *Basic* y se llamó a tal código *TCEVMV*.

Al concluir el proceso, se tienen los valores óptimos de los parámetros de ajuste y con ellos se aplica de manera repetida la Ecuación (15) para obtener parejas de valores de  $x$  y  $F_X(x)$ , a fin de construir el modelo TCEV en el papel Gumbel-Powell, y obtener las predicciones buscadas asociadas con periodos de retorno ( $Tr$ ) o intervalos promedio de recurrencia de 10, 25, 50, 100, 500 y 1 000 años. El  $Tr$  es el recíproco de la probabilidad de excedencia y entonces los  $Tr$  citados corresponden a probabilidades de no excedencia [ $F_X(x)$ ] de 0.90, 0.96, 0.98, 0.99, 0.998 y 0.999, respectivamente.

## 2. Método de maximización de la función objetivo

Metcalfe (1997) propuso maximizar el logaritmo de la función de máxima verosimilitud (Ecuación (23)), para el ajuste de la distribución TCEV; por ello, Campos-Aranda (2002) usó el valor negativo de tal expresión ( $L_{fmv}$ ), como *función objetivo* a minimizar por medio de un algoritmo numérico de múltiples variables no restringidas (Rosenbrock, 1960), a través del código disponible en *Fortran* (Kuester & Mize, 1973), el cual fue traducido al lenguaje *Basic* por facilidad de manejo de datos e impresiones (Campos-Aranda, 2003).

Las pruebas numéricas demostraron que el uso de la Ecuación (18) en la Ecuación (23) es más conveniente que la aplicación de la Ecuación (15) debido a una mayor estabilidad numérica y una convergencia más rápida hacia el mínimo buscado. Este algoritmo numérico se programó en lenguaje *Basic* y se designó a tal código *TCEVROS*.

Para el arranque del algoritmo de Rosenbrock se definen unos valores iniciales de las variables a optimizar o parámetros de ajuste de la distribución TCEV; con base en la Ecuación (27), Ecuación (28), Ecuación (29), Ecuación (30) y Ecuación (31). Con tales valores iniciales ( $\varepsilon_1$ ,  $\theta_1$ ,  $\varepsilon_2$  y  $\theta_2$ ) se calcula la primera magnitud de la función objetivo (Ecuación (23), auxiliándose de la Ecuación (17)) y se comienza la búsqueda del mínimo por medio del algoritmo de Rosenbrock.

Al concluir el proceso, se definen las etapas y evaluaciones de la función objetivo (FO) y se obtienen los valores óptimos de los parámetros de ajuste, y con ellos se aplica de manera repetida la Ecuación (18) para obtener parejas de valores de  $x$  y  $F_x(x)$ , así como las predicciones, como se hizo en el método numérico de sustitución sucesiva.

## **Error estándar de ajuste**

Con fines de comparación cuantitativa del ajuste logrado con la distribución TCEV y el método de máxima verosimilitud frente a otros

modelos probabilísticos, se propuso estimar el llamado *error estándar de ajuste (EEA)*, definido como (Kite, 1977):

$$EEA = \left[ \frac{\sum_{i=1}^n (Q_{oi} - Q_{ci})^2}{n - np} \right]^{1/2} \quad (33)$$

siendo  $n$  el número total de datos;  $Q_{oi}$ , el gasto máximo anual observado ordenado de menor a mayor; y  $Q_{ci}$ , el gasto máximo calculado con la distribución TCEV (Ecuación (15) o Ecuación (18)), para la misma probabilidad de no excedencia asignada al gasto observado mediante la fórmula de Weibull (Ecuación (26)). Por último,  $np$  es el número de parámetros de ajuste; en este caso, cuatro:  $\lambda_1$  o  $\varepsilon_1$ ;  $\theta_1$ ,  $\lambda_2$  o  $\varepsilon_2$ ,  $\theta_2$ .

Como la Ecuación (15) y la Ecuación (18) no tienen solución inversa, el planteamiento fue utilizar el método numérico de bisección con tolerancia de error igual a 0.0001 entre la probabilidad calculada (Ecuación (26)) y la obtenida con la Ecuación (15) o con la Ecuación (18), utilizando valores límites inferior de 0.001 de  $Q_{oi}$  y superior de tres veces  $Q_{oi}$ . La evaluación del *EEA* se intentó al terminar el cálculo de los valores iniciales de los parámetros de ajuste (Ecuación (27), Ecuación (28), Ecuación (29), Ecuación (30), Ecuación (31) y Ecuación (32)). Algunas veces el método numérico de bisección no funcionó, por ejemplo, cuando la distribución TCEV no coincide con los datos; para tales casos se suprime el uso de la subrutina del *EEA*. El otro cálculo del *EEA* se realizó al terminar el método numérico de sustitución o el algoritmo de Rosenbrock, y por lo general concluyó de modo satisfactorio.

## Registros de crecientes por procesar

Se procesaron seis series de gastos máximos anuales ( $\text{m}^3/\text{s}$ ) tomadas de la literatura especializada, las cuales se exponen en orden progresivo de amplitud ( $n$ ) en la Tabla 1. Su procedencia es la siguiente: (1) tomado de Haan, Barfield y Hayes (1994) de la localidad llamada Beargrass Creek con 31 valores; (2) el registro de 37 valores de la estación Santa Cruz, lo han expuesto Molina-Aguilar *et al.* (2018); (3) Francés-García (1995) presentó el del río Turia en la estación E-25 con 41 datos; (4) el registro de 53 datos en la estación Huites procede de Campos-Aranda (1999); (5) Gómez *et al.* (2010) mostraron el registro de 58 datos en la estación La Cuña; por último, (6) Kite (1991) presentó los datos del río St. Mary con 72 datos.

**Tabla 1.** Gastos máximos anuales ( $\text{m}^3/\text{s}$ ) en los seis registros procesados de las estaciones hidrométricas: Beargrass Creek, Santa Cruz, E-25 río Turia, Huites, La Cuña y río St. Mary.

Núm.	Registro número								
	1	2	3	4		5		6	
1	51.3	2 142.0	139	2 085	1 530	784.0	595.2	565	824
2	22.4	1 023.4	90	2 531	8 000	736.8	110.2	294	292
3	23.8	837.6	63	14 376	5 496	510.0	523.9	303	345

4	49.6	1 161.2	2 300	2580	3 385	461.0	1 636.3	569	442
5	25.4	1 062.0	26	1 499	1 374	411.0	1 168.0	232	360
6	60.0	784.2	260	1 165	1 245	326.0	295.0	405	371
7	34.5	1 086.3	117	1 127	2 299	349.8	212.8	228	544
8	36.5	487.8	76	3 215	1 345	130.4	367.4	232	552
9	21.7	677.0	514	10 000	11 350	690.0	144.6	394	651
10	44.5	807.0	84	3 229	2 509	266.0	78.4	238	190
11	35.1	553.0	90	677	2 006	199.0	261.9	524	202
12	30.0	1 252.0	3 700	1 266	1 180	690.0	196.3	368	405
13	42.2	369.5	88	1 025	-	340.6	46.8	464	583
14	25.0	293.0	155	955	-	249.6	313.8	411	725
15	37.4	1 157.2	199	4 780	-	350.0	319.6	368	232
16	93.4	762.2	60	696	-	317.0	621.1	487	974
17	68.0	1074.0	58	593	-	732.6	824.5	394	456
18	27.6	1 280.0	79	3 010	-	265.1	-	337	289
19	26.0	1 002.0	150	1 908	-	743.6	-	385	348
20	111.0	3 680.0	918	15 000	-	463.9	-	351	564
21	32.6	861.0	90	1 396	-	1 474.9	-	518	479
22	24.7	888.8	133	1 620	-	323.0	-	365	303
23	20.2	1 166.4	25	2 702	-	160.4	-	515	603
24	41.1	950.0	150	1 319	-	763.8	-	280	514
25	20.0	7000.0	136	1 944	-	578.0	-	289	377
26	147.2	484.0	35	2 420	-	191.8	-	255	318
27	60.9	920.6	43	2 506	-	2 440.0	-	334	342
28	33.1	812.0	34	1 534	-	238.4	-	456	593
29	58.9	3 332.4	40	1 508	-	622.1	-	479	378

30	35.4	898.0	238	1 558	-	1 374.0	-	334	255
31	64.3	2 790.0	37	2 200	-	439.7	-	394	292
32	-	620.0	49	2 225	-	280.2	-	348	-
33	-	1 495.0	32	7 960	-	267.2	-	428	-
34	-	836.0	42	4 001	-	287.3	-	337	-
35	-	940.0	34	1 067	-	280.7	-	311	-
36	-	3 080.0	117	3 233	-	156.5	-	453	-
37	-	1 550.0	64	1 119	-	455.5	-	328	-
38	-	-	48	6 178	-	501.2	-	564	-
39	-	-	48	4 443	-	385.0	-	527	-
40	-	-	42	1 474	-	698.2	-	510	-
41	-	-	144	2 508	-	184.7	-	371	-

Para las tres estaciones de aforo de México se citan enseguida su clave, ubicación y periodo de registro: (1) Santa Cruz: 10040, río San Lorenzo de la Región Hidrológica No. 10 (Sinaloa), 1944-1980; (2) Huites: 10037, río Fuerte de la Región Hidrológica No. 10 (Sinaloa), 1941-1993; y (3) La Cuña: 12054, río Verde de la Región Hidrológica No. 12-3 (río Santiago), 1947-2004.

## Verificación de la homogeneidad de los registros

Para que los resultados del AFC sean confiables, los datos a utilizar deben proceder de un *proceso aleatorio estacionario*, lo cual implica que no haya cambiado en el tiempo. Entonces los registros de crecientes deben estar integrados por datos independientes, que estén libres de componentes determinísticas, para que tal registro sea *homogéneo*.

A fin de comprobar lo anterior, se aplicaron siete pruebas estadísticas, una general, el Prueba de Von Neumann y seis específicas: dos de persistencia (Anderson y Sneyers), dos tendencia (Kendall y Spearman), una para el cambio en la media (Cramer) y la última para buscar inconsistencia en la dispersión (Bartlett). Estas pruebas se pueden consultar en WMO (1971), y Machiwal y Jha (2008). Todas las pruebas citadas se aplicaron con un nivel de significancia ( $\alpha$ ) del 5 % y seis de ellas indican que los registros seleccionados son homogéneos. La prueba de Bartlett detecta pérdida de homogeneidad por exceso de variabilidad debido a la presencia de los valores dispersos.

Como complemento se aplicó la prueba de Wald-Wolfowitz, no paramétrica, que ha sido utilizada por Bobée y Ashkar (1991), y por Rao y Hamed (2000) para probar *independencia y estacionariedad* en registros de gastos máximos anuales ( $x_i$ ). Con base en esta prueba, también se aprobó la calidad estadística de los registros por procesar.

## Resultados y su análisis

## Ajustes según método numérico de sustitución

En la Tabla 2 y Tabla 3 se muestran los resultados del ajuste de la distribución TCEV a los seis registros procesados con base en la Ecuación (24) y Ecuación (25). Se observa que el número de iteraciones varió de 1 a 15 y que únicamente en el registro 3 del río Turia no se pudo evaluar el *EEA* inicial. La valoración de la calidad del ajuste logrado con este método será estimada durante el contraste de sus predicciones.

**Tabla 2.** Resultados del ajuste de la distribución TCEV según el método numérico de sustitución en las tres estaciones hidrométricas indicadas.

<b>Estación</b>	Beargrass Creek	Santa Cruz	Río Turia
<b>Datos en</b>	Haan <i>et al.</i> (1994)	Molina-Aguilar <i>et al.</i> (2018)	Francés-García (1995)
<b>Número de datos</b>	31	37	41
<b>Mínimo, máximo</b>	20.0, 147.2	293.0, 7 000.0	25.0, 3 700.0
<b>F1, X1</b>	0.20, 24.0	0.022, 300.0	0.10, 35.0
<b>F2, X2</b>	0.83, 62.0	0.830, 1600.0	0.90, 500.0
<b>F3, X3</b>	0.97, 148.0	0.974, 7000.0	0.98, 1750.0
<b>(<math>\lambda_1</math>) inicial</b>	6.282	7.662	2.904

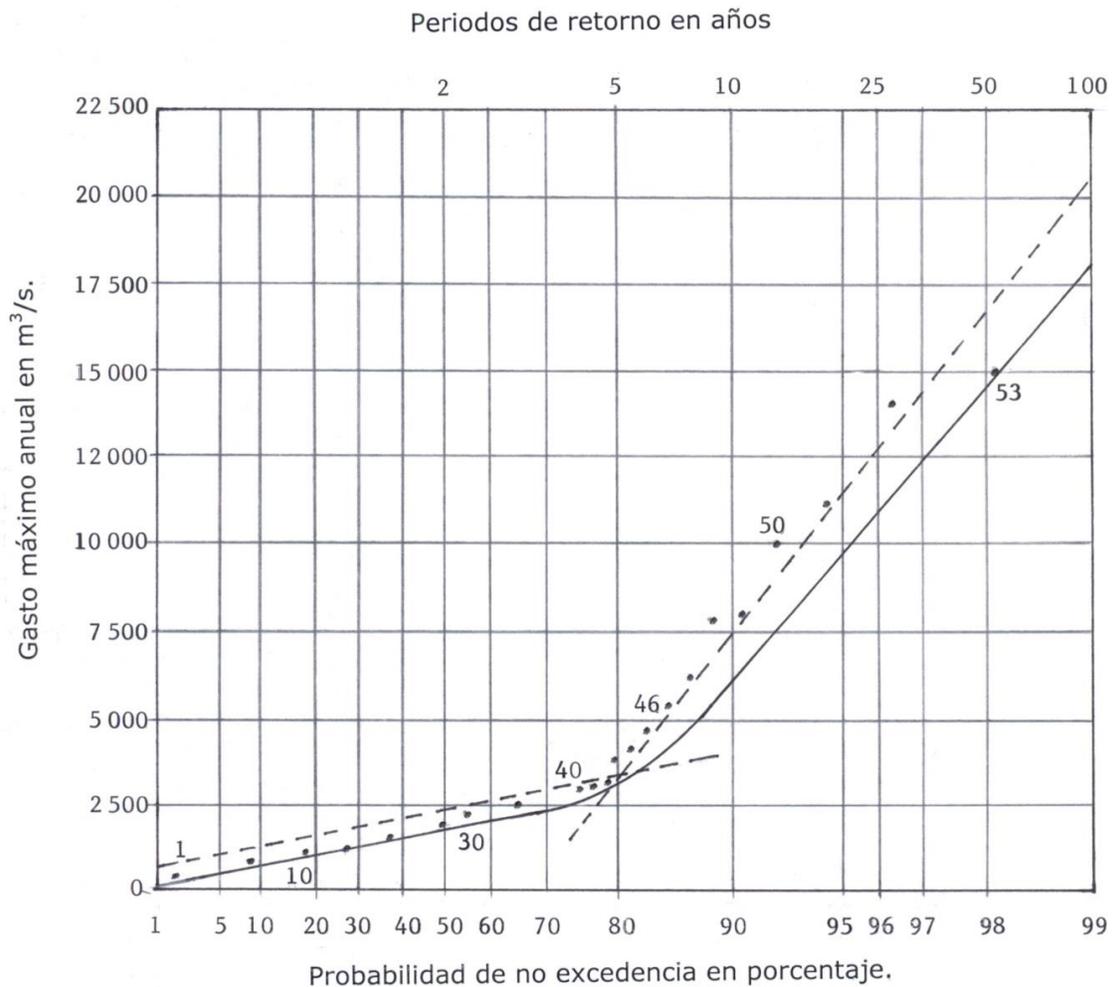
$(\theta_1)_{\text{inicial}}$	17.6	430.5	150.8
$(\lambda_2)_{\text{inicial}}$	0.688	0.333	0.204
$(\theta_2)_{\text{inicial}}$	47.5	2 760.3	756.9
$\text{FO}(\ln L)_{\text{inicial}}$	-139.490	-293.149	-278.073
$\text{EEA}_{\text{inicial}} (\text{m}^3/\text{s})$	11.6	375.3	-
<b>Número de iteraciones:</b>	4	1	15
$(\lambda_1)_{\text{final}}$	12.120	6.558	5.030
$(\theta_1)_{\text{final}}$	15.8	337.0	38.0
$(\lambda_2)_{\text{final}}$	0.784	0.238	0.138
$(\theta_2)_{\text{final}}$	46.4	2180.2	1401.5
$\text{FO}(\ln L)_{\text{final}}$	-137.440	-290.976	-243.446
$\text{EEA}_{\text{final}} (\text{m}^3/\text{s})$	6.9	328.3	246.8

**Tabla 3.** Resultados del ajuste de la distribución TCEV según método numérico de sustitución en las tres estaciones hidrométricas indicadas.

<b>Estación</b>	Huites	La Cuña	St. Mary's River
<b>Datos en</b>	Campos-Aranda (1999)	Gómez <i>et al.</i> (2010)	Kite (1991)
<b>Número de datos</b>	53	58	72
<b>Mínimo, máximo</b>	593.0, 15 000.0	46.8, 2 440.0	190.0, 974.0
<b><math>F_1, X_1</math></b>	0.10, 1 000.0	0.018, 50.0	0.05, 220.0
<b><math>F_2, X_2</math></b>	0.80, 3 500.0	0.905, 860.0	0.95, 660.0

<b><math>F3, X3</math></b>	0.98, 16 700.0	0.984, 2 475.0	0.99, 1 040.0
<b><math>(\lambda_1)_{inicial}</math></b>	5.857	5.047	22.894
<b><math>(\theta_1)_{inicial}</math></b>	1 071.1	219.2	108.2
<b><math>(\lambda_2)_{inicial}</math></b>	0.422	0.263	0.870
<b><math>(\theta_2)_{inicial}</math></b>	5 495.4	886.0	233.1
<b><math>FO(\ln L)_{inicial}</math></b>	-473.167	-412.894	-452.755
<b><math>EEA_{inicial} (m^3/s)</math></b>	790.1	116.4	33.9
<b>No. iteraciones:</b>	5	5	1
<b><math>(\lambda_1)_{final}</math></b>	6.450	4.945	21.770
<b><math>(\theta_1)_{final}</math></b>	655.4	160.9	100.1
<b><math>(\lambda_2)_{final}</math></b>	0.358	0.253	0.706
<b><math>(\theta_2)_{final}</math></b>	5088.6	734.5	181.1
<b><math>FO(\ln L)_{final}</math></b>	-466.344	-408.736	-452.673
<b><math>EEA_{final} (m^3/s)</math></b>	478.7	83.1	30.3

En la Figura 1 se tienen dibujados, en el papel de probabilidad Gumbel-Powell, los datos del registro en la estación Huites, en intervalos de cinco valores del 1 al 40, y de ahí en adelante de 1 en 1. Con líneas punteadas se muestran las rectas que representan a las dos poblaciones y con línea continua la distribución TCEV calculada con el método numérico de sustitución sucesiva. Se observa que el modelo TCEV sigue de manera aproximada a las crecientes básicas hasta la número 42, pero se aparta de las extraordinarias desde la número 43 y coincide con la última, que es el dato 53.



**Figura 1.** Ajuste con el método de sustitución sucesiva de la distribución TCEV a las crecientes de la estación Huites, en el papel de probabilidad Gumbel-Powell.

## Ajustes según método numérico de optimización

En los registros de Santa Cruz y La Cuña, los valores iniciales definidos por los puntos  $F1, X1$  a  $F3, X3$  no permitieron el arranque del algoritmo de Rosenbrock y, por ello, se modificaron ligeramente, como se muestra en la Tabla 4 y Tabla 5, de resultados del ajuste de la función TCEV a los seis registros procesados. Se observa que algunas veces este método mejoró el ajuste (menores  $L_{fmv}$  y  $EEA$ ), como en el primero y el último de los registros procesados, y en otras ocasiones no superó los resultados del método numérico de sustitución, como fue el caso de los registros del río Turia y de Huites. Por lo anterior, se recomienda que de manera normal se apliquen ambos métodos de ajuste por máxima verosimilitud.

**Tabla 4.** Resultados del ajuste de la distribución TCEV según el método numérico de optimización en las tres estaciones hidrométricas indicadas.

Estación	Beargrass Creek	Santa Cruz	Est. 25 Río Turia
<b><math>F1, X1</math></b>	0.20, 24.0	0.10, 550.0	0.10, 35.0
<b><math>F2, X2</math></b>	0.83, 62.0	0.78, 1 300.0	0.90, 500.0
<b><math>F3, X3</math></b>	0.97, 148.0	0.95, 3 830	0.98, 1 750.0
<b><math>(\epsilon_1)_{inicial}</math></b>	32.387	830.945	160.738
<b><math>(\theta_1)_{inicial}</math></b>	17.624	336.852	150.759
<b><math>(\epsilon_2)_{inicial}</math></b>	-17.785	-932.923	-1 203.201
<b><math>(\theta_2)_{inicial}</math></b>	47.848	1 603.572	756.855
<b><math>FO(\ln L)_{inicial}</math></b>	-139.490	-291.499	-278.073

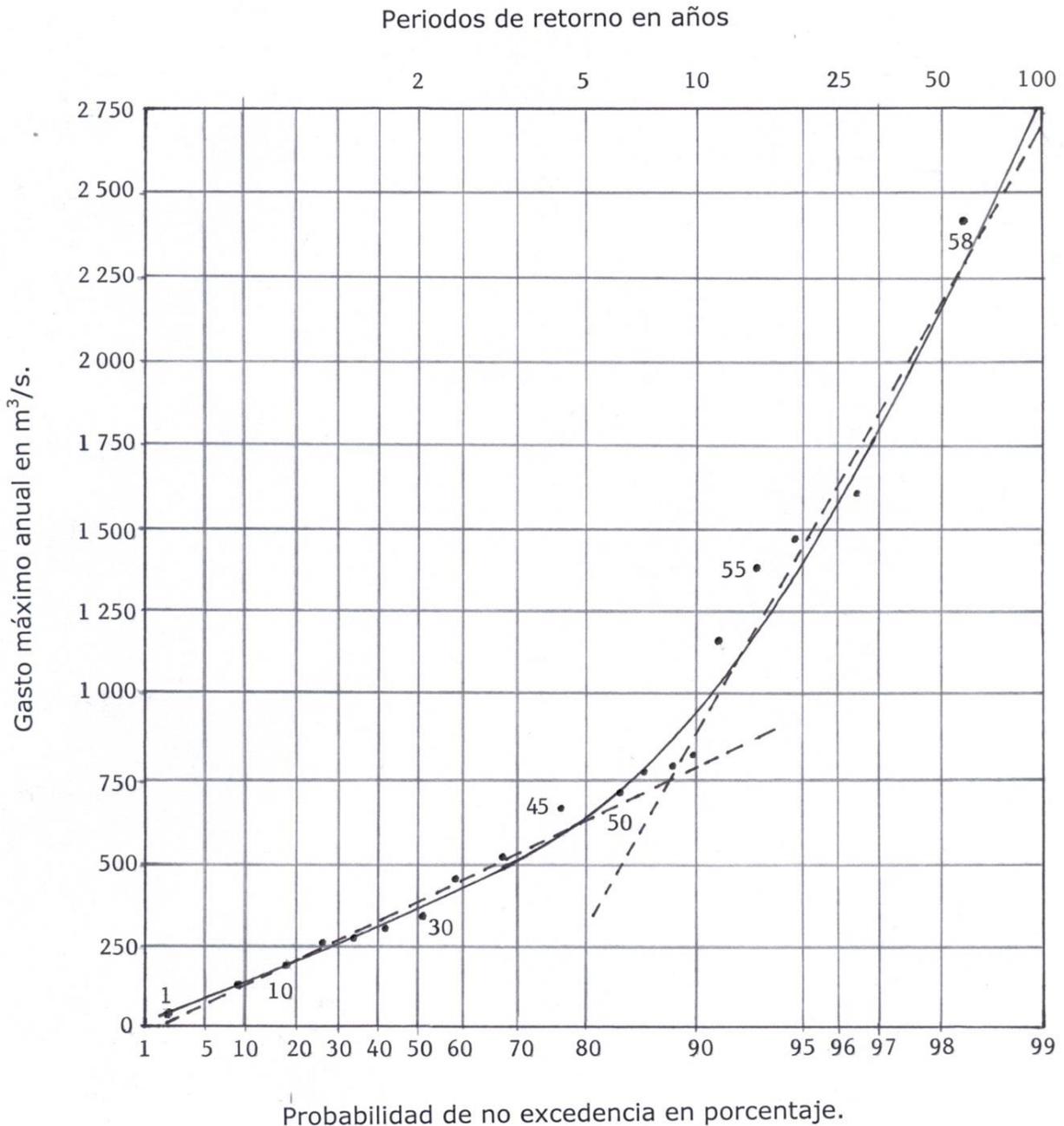
<b><math>EEA_{inicial}</math> (m<sup>3</sup>/s)</b>	11.6	432.9	-
<b>Número de etapas</b>	6	8	17
<b>Número de evaluaciones</b>	72	118	232
<b><math>(\epsilon_1)_{final}</math></b>	27.918	729.148	60.849
<b><math>(\theta_1)_{final}</math></b>	8.133	277.394	36.710
<b><math>(\epsilon_2)_{final}</math></b>	-19.230	-1 274.057	-1 913.153
<b><math>(\theta_2)_{final}</math></b>	43.013	1 719.063	1 133.334
<b>FO(ln L)<sub>final</sub></b>	-134.291	-289.773	-243.634
<b><math>EEA_{final}</math> (m<sup>3</sup>/s)</b>	5.9	397.9	267.0

**Tabla 5.** Resultados del ajuste de la distribución TCEV según método numérico de optimización en las tres estaciones hidrométricas indicadas.

<b>Estación</b>	Huites	La Cuña	St. Mary's River
<b><math>F1, X1</math></b>	0.10, 1 000.0	0.05, 80.0	0.01, 185.0
<b><math>F2, X2</math></b>	0.80, 3 500.0	0.88, 750.0	0.90, 580.0
<b><math>F3, X3</math></b>	0.98, 16 700.0	0.97, 1 850.0	0.99, 1 010.0
<b><math>(\epsilon_1)_{inicial}</math></b>	1 893.361	313.058	344.690
<b><math>(\theta_1)_{inicial}</math></b>	1 071.135	212.414	104.565
<b><math>(\epsilon_2)_{inicial}</math></b>	-4742.809	-827.543	168.192
<b><math>(\theta_2)_{inicial}</math></b>	5 495.425	766.904	182.996
<b>FO(ln L)<sub>inicial</sub></b>	-473.167	-411.328	-455.004
<b><math>EEA_{inicial}</math> (m<sup>3</sup>/s)</b>	790.1	93.9	50.4

<b>Número de etapas</b>	12	6	8
<b>Número de evaluaciones</b>	184	127	98
<b><math>(\epsilon_1)_{final}</math></b>	1 445.455	280.490	315.582
<b><math>(\theta_1)_{final}</math></b>	639.558	162.207	96.423
<b><math>(\epsilon_2)_{final}</math></b>	-6 417.104	-1 157.494	165.375
<b><math>(\theta_2)_{final}</math></b>	5 969.293	852.125	145.344
<b><math>FO(\ln L)_{final}</math></b>	-466.337	-408.863	-451.528
<b><math>EEA_{final} (m^3/s)</math></b>	540.6	54.4	22.5

En la Figura 2 se tienen dibujados, en el papel de probabilidad Gumbel-Powell, los datos del registro en la estación La Cuña en intervalos de cinco valores del 1 al 50, y de ahí en adelante de 1 en 1. Con líneas punteadas se muestran las rectas que representan a las dos poblaciones y con línea continua la distribución TCEV calculada con el método de maximización de la función objetivo. Se observa que el modelo TCEV representa de forma excelente a todas las crecientes.



**Figura 2.** Ajuste con el método de maximización de la función objetivo de la distribución TCEV a las crecientes de la estación La Cuña, en el papel de probabilidad Gumbel-Powell.

## Contrastes de las predicciones de la distribución TCEV

Un primer contraste se realizó entre las predicciones obtenidas con cada uno de los dos métodos de ajuste de la distribución TCEV aplicados. Se deduce que ambos métodos conducen a predicciones similares en todos los registros, destacando el último por su similitud y el de la estación Huites por sus diferencias. Lo anterior ratifica la sugerencia de ajustar el modelo TCEV con ambos métodos expuestos.

El segundo contraste de predicciones se llevó a cabo utilizando las tres distribuciones cuya aplicación ha sido sugerida bajo precepto. La Log-Pearson tipo III (LP3) en EUA; la General de Valores Extremos (GVE) en Inglaterra, de 1975 a 2000; y la Logística Generalizada (LOG), que actualmente se aplica en Inglaterra (Shaw, Beven, Chappell, & Lamb, 2011). Además, se aplicó la distribución Wakeby (Houghton, 1978), que ha demostrado gran versatilidad y habilidad descriptiva, al ser una función mixta.

La distribución LP3 se ajustó con el método de momentos en el dominio logarítmico (WRC, 1977) y en dominio real (Bobée, 1975), adoptando la de menor *EEA*; en cambio, las funciones GVE, LOG y Wakeby se aplicaron con el método de los momentos L (Stedinger, Vogel & Foufoula-Georgiou, 1993; Hosking & Wallis, 1997; Campos-Aranda,

2018). En la Tabla 6 se exponen las predicciones estimadas con cada uno de los cuatro modelos probabilísticos citados.

**Tabla 6.** Predicciones de los periodos de retorno indicados, obtenidas con cinco distribuciones de probabilidad, en los seis registros procesados.

Estación(*) FDP (**)	EEA (m <sup>3</sup> /s)	Periodos de retorno, en años					
		10	25	50	100	500	1 000
<b>Beargrass Creek (nvd = 3; q = 9.7 %)</b>							
TCEV (ns)	6.9	102	139	171	203	277	309
TCEV (no)	5.9	78	118	149	179	248	278
LP3 (dl)	3.4	77	111	143	184	320	404
GVE (mL)	7.7	76	107	138	177	313	398
LOG (mL)	4.9	74	105	136	178	334	440
WAK (mL)	7.2	79	110	138	170	266	319
<b>Santa Cruz (nvd = 6; q = 16.2 %)</b>							
TCEV (ns)	328.3	2 080	3 845	5 375	6 898	10 417	11 929
TCEV (no)	397.9	2 620	4 225	5 435	6 635	9 408	10 600
LP3 (mdl)	255.6	2 510	3 943	5 445	7 440	14 947	20 026
GVE (mL)	499.8	2 348	3 677	5 127	7 133	15 296	21 234
LOG (mL)	340.5	2 290	3 571	5 003	7 033	15 694	22 265
WAK (mL)	502.2	2 348	3 725	5 205	7 213	15 084	20 616
<b>Est. 25, río Turia (nvd = 4; q = 9.8 %)</b>							
TCEV (ns)	246.8	381	1 708	2 693	3 671	5 932	6 905

TCEV (no)	267.0	638	1 712	2 510	3 300	5 129	5 915
LP3 (mdl)	207.0	408	1 093	2 323	4 975	30 201	66 688
GVE (mL)	441.0	453	958	1 666	2 883	10 223	17 612
LOG (mL)	416.9	355	750	1 307	2 269	8 144	14 119
WAK (mL)	501.0	369	793	1 385	2 395	8 404	14 381
<b>Huites (<math>nvd = 6</math>; <math>q = 11.3 \%</math>)</b>							
TCEV (ns)	478.7	6 240	11 050	14 630	18 200	26 387	29 916
TCEV (no)	540.6	7 025	12 680	16 870	21 045	30 675	34 814
LP3 (mdl)	949.9	6 290	10 492	15 061	21 303	45 851	63 053
GVE (mL)	1007.1	5 948	9 614	13 600	19 091	41 299	57 363
LOG (mL)	984.2	5 786	9 322	13 263	18 830	42 460	60 302
WAK (mL)	893.1	6 303	10 070	13 834	18 626	35 412	46 091
<b>La Cuña (<math>nvd = 5</math>; <math>q = 8.6 \%</math>)</b>							
TCEV (ns)	83.1	847	1 359	1 858	2 369	3 554	4 064
TCEV (no)	54.4	926	1 575	2 168	2 762	4 137	4 729
LP3 (mdr)	74.8	991	1 389	1 719	2 078	3 025	3 486
GVE (mL)	97.1	941	1 373	1 784	2 290	3 964	4 979
LOG (mL)	61.1	913	1 338	1 764	2 313	4 304	5 614
WAK (mL)	95.2	958	1 400	1 799	2 265	3 668	4 446
<b>St. Mary's River (<math>nvd = 4</math>; <math>q = 5.6 \%</math>)</b>							
TCEV (ns)	30.3	568	680	769	862	1 099	1 211
TCEV (no)	22.3	595	707	795	888	1 122	1 233
LP3 (mdl)	16.8	600	712	797	883	1 093	1 188
GVE (mL)	22.1	603	713	796	881	1 079	1 167

LOG (mL)	15.5	589	706	807	919	1 240	1 410
WAK (mL)	23.3	605	715	794	868	1 025	1 087

\**nvd* = número de valores dispersos.

*n* = número de datos del registro.

$q = (nvd/n) \cdot 100$ .

\*\* *Método de ajuste:*

ns = numérico de sustitución.

no = numérico de optimización.

mdl = momentos en el dominio logarítmico.

mdr = momentos en el dominio real.

mL = momentos L.

En términos generales, las distribuciones LP3, GVE, LOG y WAK conducen a predicciones mayores en todos los registros procesados en los periodos de retorno (*Tr*) de 500 y 1 000 años. Adoptando el criterio de selección del menor *EEA*, en los tres primeros registros la función LP3 sería la opción a tomar, pero ello conduce a predicciones elevadas en los *Tr* de 500 y 1 000 años. Lo anterior es notable en el río Turia y en Santa Cruz, y podría ser aceptable en el registro 1. Lo contrario ocurre en los registros de Huites y de La Cuña, en los cuales la mejor opción es la distribución TCEV, en el primero a través del método numérico de sustitución y en el segundo con el de optimización (ver Figura 1 y Figura 2). Por último, en el registro de St. Mary's River, la distribución con el menor *EEA* es la Logística Generalizada, cuyas predicciones son las más grandes, pero del mismo orden de magnitud que las obtenidas con los otros modelos probabilísticos.

Conviene destacar que las predicciones de la distribución TCEV en la estación Huites (Figura 1), prácticamente coinciden con las obtenidas con el enfoque global por Campos-Aranda (1999).

Otra apreciación que se detecta en la Tabla 5 es la siguiente: cuando el porcentaje de valores dispersos ( $q$ ) es cercano al 10 % o lo excede; las distribuciones que se aplican bajo precepto y la Wakeby pueden conducir a predicciones muy elevadas en los  $Tr$  superiores a los 100 años, sobre todo si tales valores dispersos son muy grandes comparados con las crecientes ordinarias, como en el caso de los registros del río Turia, Huites y Santa Cruz. En tales registros, el modelo TCEV es una opción que siempre hay que emplear, pues su origen teórico justifica su aplicación en muestras de crecientes que proceden de dos mecanismos físicos diferentes de formación, pero que están mezcladas y por ello muestran valores dispersos (*outliers*).

## Conclusiones

*Primera:* la distribución TCEV constituye una opción muy importante para modelar crecientes cuando el registro disponible consiste de eventos asociados con dos procesos hidrometeorológicos diferentes; es decir, cuando existen gastos que se apartan bastante de la tendencia general de los datos, al dibujarlos en el papel Gumbel-Powell.

*Segunda:* de acuerdo con los resultados de la Tabla 2, Tabla 3, Tabla 4 y Tabla 5, el ajuste de la distribución TCEV mediante el método de máxima verosimilitud, según el proceso numérico de sustitución sucesiva y de optimización mediante el algoritmo de Rosenbrock, son procedimientos simples que por lo general convergen y se complementan, ya que unas veces el primero aporta mejores resultados (menores  $L_{fmv}$  y  $EEA$ ) y otras el segundo.

*Tercera:* el contraste de las predicciones con seis periodos de retorno de la distribución TCEV (mostrado en la Tabla 6) permite observar que sus dos métodos de ajuste por máxima verosimilitud conducen a predicciones similares en todos los registros. Lo anterior ratifica la sugerencia de ajustar el modelo TCEV con ambos procesos expuestos.

*Cuarta:* las predicciones de las distribuciones que se aplican bajo precepto (LP3, GVE, y LOG) y del modelo probabilístico Wakeby pueden servir para ratificar o acotar las estimaciones obtenidas con la función TCEV, y ayudar en la selección de las crecientes de diseño, al tomar en cuenta la génesis de las crecientes, el  $EEA$  y las capacidades predictivas de cada distribución.

### **Agradecimientos**

Se agradecen los comentarios y correcciones sugeridas por los dos árbitros anónimos A y B, que permitieron mejorar la redacción del texto, incorporar los datos procesados y corregir omisiones descriptivas de los cálculos realizados.

## Referencias

- Aronica, G. T., & Candela, A. (2007). Derivation of flood frequency curves in poorly gauged Mediterranean catchments using a simple stochastic hydrological rainfall-model. *Journal of Hydrology*, 347, 132-142.
- Benson, M. A. (1962). Plotting positions and economics of engineering planning. *Journal of Hydraulics Division*, 88(6), 57-71.
- Beran, M., Hosking, J. R. M., & Arnell, N. (1986). Comment on "Two-component extreme value distribution for flood frequency analysis, by Rossi *et al.*, (1984)". *Water Resources Research*, 22(2), 263-266.
- Bobée, B. (1975). The Log-Pearson type 3 distribution and its application to Hydrology. *Water Resources Research*, 11(5), 681-689.
- Bobée, B., & Ashkar, F. (1991). *The Gamma Family and derived distributions applied in hydrology*. Littleton, USA: Water Resources Publications.
- Boni, G., Parodi, A., & Rudari, R. (2006). Extreme rainfall events: Learning from raingauge time series. *Journal of Hydrology*, 327(3-4), 304-314.
- Botto, A., Ganora, D., Laio, F., & Claps, P. (2014). Uncertainty compliant design flood estimation. *Water Resources Research*, 50(5), 4242-4253.
- Campos-Aranda, D. F. (2018). Ajuste con momentos L de las distribuciones GVE, LOG y PAG no estacionarias en su parámetro de

- ubicación, aplicado a datos hidrológicos extremos. *Agrociencia*, 52(2), 169-189.
- Campos-Aranda, D. F. (2003). Capítulo 9. Optimización numérica. En: *Introducción a los métodos numéricos: software en Basic y aplicaciones en hidrología superficial* (pp. 172-211). San Luis Potosí, México: Editorial Universitaria Potosina.
- Campos-Aranda, D. F. (2002). *Ajuste de la distribución TCEV por medio de optimización numérica no restringida*. En: XVII Congreso Nacional de Hidráulica. Monterrey, N. L. *Avances en Hidráulica*, 9, 527-532.
- Campos-Aranda, D. F. (2001). Contraste de dos procedimientos de ajuste de la distribución Wakeby en modelación probabilística de crecientes. *Agrociencia*, 35(4), 429-439.
- Campos-Aranda, D. F. (1999). Hacia el enfoque global en el análisis de frecuencia de crecientes. *Ingeniería Hidráulica en México*, 14(1), 23-42.
- Cannarozzo, M., D'Asaro, F., & Ferro, V. (1995). Regional rainfall and flood frequency analysis for Sicily using the two component extreme value distribution. *Hydrological Sciences Journal*, 40(1), 19-40.
- Chow, V. T. (1964). Statistical and probability analysis of hydrologic data. Section 8-I: Frequency Analysis. In: Chow, V. T. (ed.). *Handbook of applied hydrology* (pp. 8.1-8.42). New York, USA: McGraw-Hill Book Co.
- Davis, P. J. (1972). Chapter 6. Gamma Function and related functions. In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of*

- mathematical functions* (pp. 253-296). New Work, USA: Dover Publications.
- Escalante-Sandoval, C. A., & Reyes-Chávez, L. (2004). *Análisis bivariado de gastos máximos anuales con distribuciones marginales TCEV*. En: XVIII Congreso Nacional de Hidráulica. San Luis Potosí, S.L.P., México. *Avances en Hidráulica*, 11, 523-529.
- Ferro, V., & Porto, P. (2006). Flood frequency analysis for Sicily, Italy. *Journal of Hydrologic Engineering*, 11(2), 110-122.
- Fiorentino, M., Versace, P., & Rossi, F. (1985). Regional flood frequency estimation using the two-component extreme value distribution. *Hydrological Sciences Journal*, 30(1), 51-64.
- Fiorentino, M., Arora, K., & Singh, V. P. (1987). The two-component extreme value distribution for flood frequency analysis: Derivation of a new estimation method. *Stochastic Hydrology and Hydraulics*, 1(3), 199-208.
- Fiorentino, M., Gabriele, S., Rossi, F., & Versace, P. (1987). Hierarchical approach for regional flood frequency analysis. In: Singh, V. P. (ed.). *Regional flood frequency analysis* (pp. 35-49). Dordrecht, The Netherlands: Reidel Publishing Company.
- Francés-García, F. (1995). *Utilización de la información histórica en el análisis regional de las avenidas* (nomografía No. 27). Barcelona, España: Centro Internacional de Métodos Numéricos en Ingeniería.
- Francés, F. (1998). Using the TCEV distribution function with systematic and non-systematic data in a regional flood frequency analysis. *Stochastic Hydrology and Hydraulics*, 12(4), 267-283.

- Gómez, J. F., Aparicio, M., & Patiño, C. (2010). *Manual de análisis de frecuencias en hidrología*. Jiutepec, México: Instituto Mexicano de Tecnología del Agua.
- González-Villarreal, F. J. (1970). *Contribución al análisis de frecuencias de valores extremos de los gastos máximos en un río (No. 277)*., México, DF, México: Instituto de Ingeniería de la Universidad Nacional Autónoma de México.
- Haan, C. T., Barfield, B. J., & Hayes, J. C. (1994). Chapter 2. Hydrologic frequency analysis. *Design hydrology and sedimentology for small catchments*. (pp. 5-36). San Diego, USA: Academic Press.
- Hosking, J. R., & Wallis, J. R. (1997). Appendix: *L*-moments for some specific distributions. In: *Regional Frequency Analysis. An approach based on L-moments* (pp. 191-209). Cambridge UK: Cambridge University Press.
- Houghton, J. C. (1978). Birth of a parent: The Wakeby distribution for modeling flood flows. *Water Resources Research*, 14(6), 1105-1109.
- Katz, R. W. (2013). Chapter 2. Statistical Methods for Nonstationary Extremes. In: Aghakouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (eds.). *Extremes in a changing climate* (pp. 15-37). Dordrecht, The Netherlands: Springer.
- Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J. C., Gachon, P., & Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, 329(3-4), 534-552.

- 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.
- Kite, G. W. (1991). Chapter 4. Frequency analysis. In: *Hydrologic Applications: Computer programs for water resources engineering*. (pp. 72-98). Littleton, USA: Water Resources Publications.
- Kuester, J. L., & Mize, J. H. (1973). Chapter 9. Multivariable unconstrained methods. In: *Optimization techniques with Fortran* (pp. 297-365). New York, USA: McGraw-Hill Book Co.
- López-de-la-Cruz, J., & Francés, F. (2014). La variabilidad climática de baja frecuencia en la modelación no estacionaria de los regímenes de las crecidas en las regiones hidrológicas Sinaloa y Presidio-San Pedro. *Tecnología y ciencias del agua*, 5(4), 79-101.
- Machiwal, D., & Jha, M. K. (2008). Comparative evaluation of statistical tests for time series analysis: Applications to hydrological time series. *Hydrological Sciences Journal*, 53(2), 353-366.
- Matalas, N. C., Slack, J. R., & Wallis, J. R. (1975). Regional skew in search of a parent. *Water Resources Research*, 11(6), 815-826.
- Merz, R., & Blöschl, G. (2008). Flood Frequency Hydrology: 1. Temporal, spatial and causal expansion of information. *Water Resources Research*, 44(8), 1-17.
- Metcalfe, A. V. (1997). Chapter 4. Extreme value and related distributions. In: *Statistics in Civil Engineering* (pp. 81-115). London, UK: Arnold Publishers.

- Molina-Aguilar, J. P., Gutiérrez-López, M. A., & Aparicio-Mijares, F. J. (2018). Búsqueda armónica para optimizar la función Gumbel mixta univariada. *Tecnología y ciencias del agua*, 9(5), 280-322.
- Potter, W. D. (1958). Upper and lower frequency curves for peak rates of runoff. *EOS. Transactions of AGU*, 39(1), 100-105.
- Prosdocimi, I., Kjeldsen, T. R., & Svensson, C. (2014). Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. *Natural Hazards and Earth System Sciences*, 14(5), 1125-1144.
- Rao, A. R., & Hamed, K. H. (2000). Theme 1.8. Tests on hydrologic data. In: *Flood Frequency Analysis* (pp. 12-21). Boca Raton, USA: CRC Press.
- Rosenbrock, H. H. (1960). An automatic method of finding the greatest or least value of a function. *Computer Journal*, 3(3), 175-184.
- Rossi, F., Fiorentino, M., & Versace, P. (1984). Two-component extreme value distribution for flood frequency analysis. *Water Resources Research*, 20(7), 847-856.
- Rulfova, Z., Buishand, A., Roht, M., & Kysely, J. (2016). A two-component generalized extreme value distribution for precipitation frequency analysis. *Journal of Hydrology*, 534, 659-668.
- Shaw, E. M., Beven, K. J., Chappell, N. A., & Lamb, R. (2011). Chapter 13: Estimating floods and low flows in the UK. In: *Hydrology in Practice* (4<sup>th</sup> ed.) (pp. 322-350). London, UK: Spon Press.
- Singh, V. P. (1998). Chapter 22. Two-component extreme value distribution. In: *Entropy-based parameter estimation in hydrology*

- (Vol. 10). (pp. 347-362). Dordrecht, The Netherlands: Kluwer Academic Publishers, Water Science and Technology Library.
- Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Chapter 18: Frequency Analysis of Extreme Events. In: Maidment, D. R. (ed.). *Handbook of Hydrology* (pp. 18.1-18.66). New York, USA: McGraw-Hill, Inc.
- Todorovic, P., & Zelenhasic, E. (1970). A stochastic model for flood analysis. *Water Resources Research*, 6(6), 1641-1648.
- WRC, Water Resources Council. (1977). *Guidelines for determining flood flow Frequency* (Bulletin # 17A). Washington, DC, USA: Hydrology Committee, Water Resources Council.
- WMO, World Meteorological Organization. (1971). Annexed III. In: *Climatic Change* (pp. 58-71) (Technical Note No. 79). Geneva, Switzerland: Secretariat of the World Meteorological Organization.
- Waylen, P., & Woo, M. K. (1982). Prediction of annual floods generated by mixed processes. *Water Resources Research*, 18(4), 1283-1286.