

Ajuste de las distribuciones GVE, LOG y PAG con momentos L depurados (1, 0)

• Daniel Francisco Campos-Aranda •

Profesor Jubilado de la Universidad Autónoma de San Luis Potosí, México

*Autor de correspondencia

Resumen

Campos-Aranda, D. F. (julio-agosto, 2015). *Nota técnica. Ajuste de las distribuciones gve, log y pag con momentos L depurados (1, 0). Tecnología y Ciencias del Agua*, 6(4), 153-167.

Los momentos estadísticos han sido utilizados para caracterizar las distribuciones de probabilidad, así como las muestras de datos observados. En este trabajo se describe someramente la teoría de los momentos L y de los llamados momentos L depurados (1, 0), que son capaces de reducir la influencia del valor más bajo de la muestra, para mejorar el ajuste y obtener predicciones extremas más confiables. Se citan las ecuaciones, recientemente expuestas en la literatura estadística, que permiten estimar los parámetros de ubicación (μ), escala (σ) y forma (k) de las tres funciones de distribución de probabilidades más utilizadas en los análisis de frecuencias de datos hidrológicos extremos, que son la General de Valores Extremos (GVE), la Logística Generalizada (LOG) y la Pareto Generalizada (PAG). Estas tres distribuciones se ajustaron con los métodos de momentos L y momentos L depurados (1, 0), a los 21 registros de gasto máximo anual disponibles en la Región Hidrológica núm. 10 (Sinaloa). Se evaluó la calidad de cada ajuste a través del error estándar. El análisis de los resultados indica que la distribución PAG conduce a los menores errores de ajuste en trece registros con el método de los momentos L depurados (1, 0) y en el resto con el de momentos L . Las conclusiones sugieren la aplicación sistemática de los tres modelos probabilísticos utilizados y del método de los momentos L depurados (1, 0), como una versión avanzada del procedimiento de los momentos L , actualmente de uso universal.

Palabras clave: momentos L , momentos L depurados (1, 0), distribuciones de probabilidad GVE, LOG y PAG, error estándar de ajuste, Región Hidrológica núm. 10 (Sinaloa).

Abstract

Campos-Aranda, D. F. (July-August, 2015). *Technical Note. Fitting of GPA, GLO and GEV Distributions with Trimmed L-moments (1, 0). Water Technology and Sciences (in Spanish)*, 6(4), 153-167.

Statistical moments have been used to characterize probability distributions and samples of observed data. This work briefly describes the theory of L-moments and trimmed L-moments (1,0), which can reduce the influence of the lowest value in a sample, in order to improve the fit and obtain more reliable extreme predictions. Recent equations found in the statistical literature that estimate the location, scale and shape of the probability distribution functions are cited, which are often used to analyze the frequencies of extreme hydrological data. These include the General Extreme Values (GEV), Generalized Logistic (GLO) and Generalized Pareto (GPA) equations. These three distributions were fitted with the methods of L-moments and trimmed L-moments (1,0) to 21 annual maximum flow registries from Hydrological Region No. 10 (Sinaloa). The quality of each fit was evaluated based on the standard error. The analysis of the results indicate that the GPA distribution provides the smallest fitting errors for 13 registries using the trimmed L-moments (1,0) and for the rest of the registries using L-moments. The conclusions suggest that the three probabilistic models studied can be applied with the trimmed L-moments (1,0) as an advanced version of the L-moments procedure which is universally used.

Keywords: L-moments, trimmed L-moments (1, 0), probability distributions GEV, GLO and GPA, standard error of fit, Hydrological Region No. 10 (Sinaloa).

Recibido: 05/03/15
Aceptado: 11/05/15

Introducción

Los momentos estadísticos han sido usados de modo tradicional para caracterizar las *funciones de distribución de probabilidades* (FDP), o bien las series de datos observados. Los momentos L , de creación más reciente, son combinaciones lineales de los valores esperados de los estadísticos de orden (Asquith, 2011), comparables a los momentos convencionales, pero tienen menores varianzas muestrales y son más robustos a la influencia de los valores dispersos o extremos (*outliers*). De manera análoga a los momentos estadísticos, los momentos L de orden uno a cuatro, también caracterizan, respectivamente, localización, escala, asimetría y curtosis.

De acuerdo con Karvanen (2006), el concepto de los momentos L se originó hacia finales de los años sesenta, con varios resultados desvinculados sobre combinaciones lineales de los estadísticos de orden, que culminan con el trabajo de Greenwood, Landwehr, Matalas y Wallis (1979). Hosking (1990) unifica la teoría de los momentos L y proporciona guías para sus aplicaciones prácticas, las cuales se han realizado en las disciplinas de la hidrología, meteorología, control de calidad e ingeniería civil. Karvanen (2006) también cita avances recientes en los aspectos teóricos, por ejemplo: (1) Mudholkar y Hutson (1998) rempazan los valores esperados de los estadísticos de orden por funcionales que inducen estimadores rápidos como la mediana, el estimador Gastwirth y la media truncada; estos análogos, llamados momentos LQ , siempre existen y son simples de evaluar; (2) Elamir y Seheult (2003) introducen una extensión natural de los momentos L , llamados *momentos depurados*; se designan momentos TL de *Trimmed*, que significa depurado, truncado o cortado; (3) también Elamir y Seheult (2004) derivan la estructura exacta de la varianza de muestreo de los momentos L .

Por otra parte, el análisis de frecuencias de datos hidrológicos, crecientes, lluvias máximas y sequías, busca principalmente valores de éstos, asociados con bajas probabilidades de exceedencia. Tales magnitudes, llamadas *predicciones*,

se obtienen con base en el ajuste de una FDP a la muestra de datos disponible. Entre las FDP que más se han utilizado en estos estudios y que incluso su aplicación ha sido establecida bajo precepto, se tienen los modelos probabilísticos: General de Valores Extremos (GVE), Logística Generalizada (LOG) y Pareto Generalizada (PAG). Estas FDP han sido adoptadas en los análisis de frecuencias debido a su extremo derecho grueso o denso (*heavy tail*), que es la porción donde se estiman las predicciones (El Adlouni, Bobée, & Ouarda, 2008).

El objetivo de este trabajo consiste en exponer con detalle los aspectos teóricos asociados con los llamados *momentos L depurados*, en especial de aquellos que eliminan el valor más bajo de la serie o muestra de datos, y su aplicación en el ajuste de las distribuciones GVE, LOG y PAG. Lo anterior se complementa con un contraste numérico en los 21 registros disponibles de crecientes anuales de la Región Hidrológica núm. 10 (Sinaloa), que no tienen régimen de escurrimientos afectado por embalses y que cuentan con más de 20 datos, según la información actualizada hasta el año 2011, proporcionada por la Comisión Nacional del Agua (Conagua).

Resumen de la teoría operativa

Definición poblacional de los momentos L

Sea X una variable aleatoria y sea $X_{j:n}$ uno de sus estadísticos de orden, es decir, una variable aleatoria distribuida como el j -ésimo elemento más pequeño de una muestra aleatoria de tamaño n , tomada de la distribución de X , entonces, los momentos L de X son las cantidades siguientes (Hosking, 1990):

$$\lambda_1 = E(X_{1:n}) \quad (1)$$

$$\lambda_2 = \frac{1}{2}E(X_{2:n} - X_{1:n}) \quad (2)$$

$$\lambda_3 = \frac{1}{3}E(X_{3:n} - 2X_{2:n} + X_{1:n}) \quad (3)$$

$$\lambda_4 = \frac{1}{4}E(X_{4:n} - 3X_{3:n} + 3X_{2:n} - X_{1:n}) \quad (4)$$

cuya expresión general es:

$$\lambda_r = \frac{1}{r} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} E(X_{r-j:r}) \quad r = 1, 2, \dots \quad (5)$$

En la ecuación (5), el segundo paréntesis es un cociente de factoriales que define el número de combinaciones posibles de los m términos, tomando q en cada arreglo; su expresión general es (Bhattacharai, 2004; Asquith, 2011):

$$\binom{m}{q} = \frac{m!}{q! (m-q)!} \quad \text{para } q \leq m \quad (6)$$

Por convención $0! = 1$. Recordar que la función factorial Gamma es $\Gamma(m+1) = m!$, la cual se puede estimar con la fórmula de Stirling (Davis, 1965), que será expuesta posteriormente. Por otra parte, los valores esperados de los estadísticos de orden de la ecuación (5) se estimarán con la expresión (Hosking & Wallis, 1997; Kottegoda & Rosso, 2008; Asquith, 2011):

$$E(X_{r:n}) = \frac{n!}{(r-1)! (n-r)!} \int_0^1 x(F) \cdot F^{r-1} \cdot (1-F)^{n-r} dF \quad (7)$$

siendo $x(F)$ la función de cuantiles o solución inversa de la función de distribución de probabilidades acumuladas $F(x) = F$.

Momentos L depurados poblacionales y de la muestra

Como ya se indicó brevemente, los momentos L tienen usos similares a los momentos convencionales, además existen siempre que la media de la distribución exista, aunque algunos momentos de orden superior puedan no existir y son relativamente robustos a los efectos de los valores extremos. Hosking (2007) indica que estas dos ventajas son insuficientes en el manejo de ciertos datos con muchos valores extremos y de FDP que involucran colas muy gruesas (*very heavy tails*), ya que el primer momento puede no existir. Los momentos L depurados son generaliza-

ciones que no requieren que exista la media de la FDP de soporte. Se definen de manera general como (Elamir & Seheult, 2003; Hosking, 2007):

$$\lambda_r^{(s,t)} = \frac{1}{r} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} E(X_{r+s-j:r+s+t}) \quad r = 1, 2, \dots \quad (8)$$

siendo s y t enteros positivos. El término depurado (*trimmed*), que también puede ser truncado o cortado, es apropiado debido a que la definición de $\lambda_r^{(s,t)}$ no involucra los valores esperados de los estadísticos de orden de los s datos o valores más pequeños ni de los t mayores de la muestra de tamaño $r + s + t$; es decir, les asigna peso cero. Los cocientes de momentos L depurados $\tau_r^{(s,t)} = \lambda_r^{(s,t)} / \lambda_2^{(s,t)}$ son medidas adimensionales de la forma de la distribución. Hosking (2007) señala que una cuestión para decidir es el grado de depuración o de corte apropiado y al respecto indica que la inspección de los datos de la muestra puede ayudar en tal selección. De manera reciente, Elamir (2010) encontró que la cantidad de depuración conduce a diferentes varianzas de los estimados; entonces, una selección óptima será la suma mínima de los errores absolutos entre los datos y los estimados, con la FDP ajustada con tales momentos L depurados.

A partir de una muestra ordenada $(x_1 \leq x_2 \leq \dots \leq x_n)$, los momentos L depurados $\lambda_r^{(s,t)}$ pueden ser estimados de manera no sesgada con la expresión siguiente (Elamir & Seheult, 2003; Hosking, 2007):

$$l_r^{(s,t)} = \frac{1}{r} \sum_{i=s+1}^{n-t} \left[\frac{\sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \binom{i-1}{r+s-j-1} \binom{n-i}{t+j}}{\binom{n}{r+s+t}} \right] \cdot x_i \quad (9)$$

Momentos L depurados (1, 0) poblacionales y de la muestra

Ahmad, Shabri y Zakaria (2011b) indican que los momentos L depurados fueron introducidos por Elamir y Seheult (2003) como una opción para

tomar en cuenta los valores extremos; son fáciles de calcular y más robustos que los momentos L a la presencia de tales datos atípicos o dispersos (*outliers*). También señalan que la mayoría de los trabajos de aplicación de los momentos L depurados se han centrado en casos simétricos de truncado, por lo general suprimiendo el valor más bajo y el más grande de la muestra conceptual ($s = t = 1$); por ejemplo, Abdul-Moniem y Selim (2009), y Bílková (2014), proponen emplear los momentos L depurados (1, 0), en los cuales sólo el valor más bajo de la muestra es eliminado, buscando con ello un mejor ajuste de la FDP a los valores extremos superiores. De acuerdo con la ecuación (8) y considerando $s = 1$ y $t = 0$, se definen los siguientes cuatro primeros momentos L depurados (1, 0), según Ahmad *et al.* (2011b):

$$\lambda_1^{(1,0)} = E(X_{2,2}) \quad (10)$$

$$\lambda_2^{(1,0)} = \frac{1}{2}E(X_{3,3} - X_{2,3}) \quad (11)$$

$$\lambda_3^{(1,0)} = \frac{1}{3}E(X_{4,4} - 2X_{3,4} + X_{2,4}) \quad (12)$$

$$\lambda_4^{(1,0)} = \frac{1}{4}E(X_{5,5} - 3X_{4,5} + 3X_{3,5} - X_{2,5}) \quad (13)$$

Los cocientes poblacionales de momentos L depurados de asimetría y curtosis son, respectivamente:

$$\tau_3^{(1,0)} = \frac{\lambda_3^{(1,0)}}{\lambda_2^{(1,0)}} \quad (14)$$

$$\tau_4^{(1,0)} = \frac{\lambda_4^{(1,0)}}{\lambda_2^{(1,0)}} \quad (15)$$

Las estimaciones no sesgadas de los momentos depurados (1, 0) se basan en una muestra (x_i) de datos ordenados en forma progresiva y se obtienen a partir de la ecuación (8) modificada:

$$l_r^{(1,0)} = \frac{1}{r} \sum_{i=2}^n \left[\frac{\sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \cdot \binom{i-1}{r-j} \cdot \binom{n-i}{j}}{\binom{n}{r+1}} \right] \cdot x_i \quad (16)$$

Abdul-Moniem y Selim (2009) aplican los métodos de los momentos L y de los depurados simétricamente ($s = t = 1$) al ajuste de la distribución Pareto Generalizada; manteniendo conocido el parámetro de forma (k), estiman los parámetros de ubicación (u) y de escala (a) en muestras sintéticas de tamaños 10(10)40, con diez mil replicaciones cada una, en las cuales $u = 2$ y $a = 0.20(0.20)1.40$. Encontraron que el error cuadrático medio (ECM) entre los parámetros estimados y los reales o asignados decrece conforme n se incrementa, además creció conforme aumentó el parámetro de escala, pero tal error es siempre menor en las estimaciones realizadas con los momentos L depurados.

Ahmad *et al.* (2011b) también utilizan la distribución Pareto Generalizada y a través de simulación aleatoria encuentran que los momentos L depurados (1, 0) reducen la influencia probable del dato más pequeño de la muestra en la estimación de los eventos de alto periodo de retorno. Sus resultados con datos reales muestran que, en algunos casos, los momentos L depurados (1, 0) son una mejor opción que los momentos L , en especial en la estimación de predicciones de baja probabilidad de exceedencia.

Ahmad, Shabri y Zakaria (2013) contrastan las distribuciones GVE, LOG y PAG ajustadas por momentos L y momentos depurados (1, 0) y (1,1), en 12 registros de gastos máximos anuales de la porción noreste de Malasia; encuentran que los momentos depurados (1, 0) conducen a mejores ajustes en la mayoría de los registros procesados.

Ajuste con momentos L de las distribuciones GVE, LOG y PAG

De acuerdo con Hosking y Wallis (1997), según su tabla 5.1, estas tres funciones de distribución

de probabilidad (FDP), cuando su parámetro de forma $k < 0$, tienen sus colas derechas más gruesas o densas de entre todas las que son utilizadas de manera regular en los análisis de frecuencias de datos hidrológicos; además, ellas tienen una frontera superior cuando $k > 0$ y cuando $k = 0$ definen distribuciones de dos parámetros: la Gumbel, la logística y la exponencial, respectivamente.

El método de obtención de sus parámetros de ubicación, escala y forma (u , a y k) de las distribuciones GVE, LOG y PAG con base en los momentos L se puede consultar, respectivamente, en Stedinger, Vogel y Foufoula-Georgiou (1993), Hosking y Wallis (1997), y Rao y Hamed (2000). A continuación se citan para cada una de estas tres FDP su fórmula matemática o $F(x)$, su solución inversa o función de cuantiles $x(F)$ y las ecuaciones de cálculo necesarias para estimar sus tres parámetros de ajuste.

Distribución General de Valores Extremos

Intervalo de x : $u + a/k \leq x < \infty$ si $k < 0$; $-\infty < x < \infty$ si $k = 0$ y $-\infty < x \leq u + a/k$ si $k > 0$.

$$F(x) = e^{-e^{-y}} \quad (17)$$

siendo y la variable reducida:

$$y = -\frac{1}{k} \ln \left[1 - k \left(\frac{x-u}{a} \right) \right] \quad \text{para } k \neq 0 \quad (18)$$

$$y = \frac{x-u}{a} \quad \text{para } k = 0 \quad (19)$$

$$x(F) = u + \frac{a}{k} \left[1 - (-\ln F)^k \right] \quad \text{para } k \neq 0 \quad (20)$$

$$x(F) = u - a \cdot \ln(-\ln F) \quad \text{para } k = 0 \quad (21)$$

$$c = \frac{2}{3+t_3} - 0.63093 \quad (22)$$

$$k \approx 7.8590 \cdot c + 2.9554 \cdot c^2 \quad (23)$$

$$a = \frac{l_2 \cdot k}{(1-2^{-k}) \cdot \Gamma(1+k)} \quad (24)$$

$$u = l_1 - \frac{a}{k} [1 - \Gamma(1+k)] \quad (25)$$

La función factorial Gamma fue estimada con la fórmula de Stirling (Davis, 1965), que para valores grandes del argumento ε es bastante aproximada con un error cercano a cero, y está dada por:

$$\Gamma(\varepsilon) \approx e^{-\varepsilon} \cdot \varepsilon^{\varepsilon-1/2} \cdot \sqrt{2\pi} \cdot \left(1 + \frac{1}{12 \cdot \varepsilon} + \frac{1}{288 \cdot \varepsilon^2} - \frac{139}{51840 \cdot \varepsilon^3} - \frac{571}{2488320 \cdot \varepsilon^4} + \dots \right) \quad (26)$$

Distribución Logística Generalizada

Intervalo de x , idéntico al de la GVE:

$$F(x) = \frac{1}{1 + e^{-y}} \quad (27)$$

siendo y igual a la ecuación (18):

$$x(F) = u + \frac{a}{k} \left\{ 1 - \left[\frac{(1-F)}{F} \right]^k \right\} \quad \text{para } k \neq 0 \quad (28)$$

$$x(F) = u - a \cdot \ln \left[\frac{(1-F)}{F} \right] \quad \text{para } k = 0 \quad (29)$$

$$k = -t_3 \quad (30)$$

$$a = \frac{l_2 \cdot \sin(\pi \cdot k)}{\pi \cdot k} \quad (31)$$

$$u = l_1 - a \left(\frac{1}{k} - \frac{\pi}{\sin(\pi \cdot k)} \right) \quad (32)$$

Distribución Pareto Generalizada

Intervalo de x : $u \leq x < \infty$ si $k \leq 0$; $u \leq x \leq u + a/k$ si $k > 0$.

$$F(x) = 1 - e^{-y} \quad (33)$$

siendo y igual a la ecuación (18):

$$x(F) = u + \frac{a}{k} \left[1 - (1-F)^k \right] \quad \text{para } k \neq 0 \quad (34)$$

$$x(F) = u - a \cdot \ln(1-F) \quad \text{para } k = 0 \quad (35)$$

$$k = \frac{(1-3t_3)}{(1+t_3)} \quad (36)$$

$$a = l_2(1+k) \cdot (2+k) \quad (37)$$

$$u = l_1 - l_2(2+k) \quad (38)$$

Ajuste con momentos L depurados (1, 0) de las distribuciones GVE, LOG y PAG

Ahmad, Shabri y Zakaria (2011a), y Ahmad *et al.* (2013) obtienen las expresiones para estimar los tres parámetros de ajuste de estas distribuciones, basados en los momentos L depurados (1, 0), éstas son:

Distribución General de Valores Extremos

$$k = 0.4813 - 2.1364 \cdot t_3^{(1,0)} + 0.6994 \cdot (t_3^{(1,0)})^2 - 0.1991 \cdot (t_3^{(1,0)})^3 \quad (39)$$

$$a = \frac{-2(6)^k \cdot l_2^{(1,0)} \cdot [1/\Gamma(k)]}{3(2^k - 3^k)} \quad (40)$$

$$u = l_1^{(1,0)} - \frac{a}{k} + \frac{a}{2^k \cdot [1/\Gamma(k)]} \quad (41)$$

La ecuación (39) fue obtenida con base en la expresión de t_3 en función k expuesta por Ahmad *et al.* (2013), y es válida en el intervalo para k de -0.70 a 0.50. El recíproco de la función Gamma fue estimado con la siguiente fórmula del producto infinito de Euler (Davis, 1965), la cual fue evaluada para $n = 50$; con ello, ya no cambian sus cinco cifras decimales:

$$1/\Gamma(z) = z \cdot e^{\gamma z} \prod_{n=1}^{\infty} \left[\left(1 + \frac{z}{n} \right) \cdot e^{-z/n} \right] \quad (42)$$

siendo $\gamma = 0.5772156649$ la constante de Euler con el valor aproximado indicado. La función $1/\Gamma(z)$ es continua y pasa por cero en los puntos $z = 0, -1, -2, -3, \dots$

Distribución Logística Generalizada

$$k = \frac{4 - 27 \cdot t_3^{(1,0)}}{20} \quad (43)$$

$$a = \frac{4 \cdot k \cdot l_2^{(1,0)}}{6 \cdot \Gamma(2-k) \cdot \Gamma(1+k) - 3 \cdot \Gamma(3-k) \cdot \Gamma(1+k)} \quad (44)$$

$$u = l_1^{(1,0)} - \frac{a}{k} + \frac{a \cdot \Gamma(2-k) \cdot \Gamma(1+k)}{k} \quad (45)$$

Distribución Pareto Generalizada

$$k = \frac{4 - 12 \cdot t_3^{(1,0)}}{4 + 3 \cdot t_3^{(1,0)}} \quad (46)$$

$$a = \frac{l_2^{(1,0)}(1+k) \cdot (2+k) \cdot (3+k)}{3} \quad (47)$$

$$u = l_1^{(1,0)} - a \left[\frac{3+k}{(1+k) \cdot (2+k)} \right] \quad (48)$$

Indicador cuantitativo del ajuste

Coles (2001) indica que la motivación para ajustar una FDP a los datos es poder obtener

conclusiones respecto a la población de la cual probablemente proceden los valores disponibles. Como la confiabilidad de tales conclusiones depende de la exactitud de la FDP ajustada, entonces es necesario verificar que tal modelo probabilístico es capaz de reproducir los datos que se emplearon para estimarlo (Flowers-Cano, Flowers, & Rivera-Trejo, 2014).

Desde mediados de la década de los años setenta se estableció el *error estándar de ajuste* (EEA) como una medida cuantitativa que estima la calidad estadística del ajuste y que además permite la comparación objetiva entre las diversas distribuciones de probabilidad que se prueban o ajustan a una muestra, ya que tiene las unidades de los datos. Su expresión es la siguiente (Kite, 1977):

$$EEA = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n - np}} \quad (49)$$

en donde n es el número de datos de la muestra; x_i , los datos ordenados de menor a mayor; $\hat{x}_i$, los valores estimados con la función de cuantiles o solución inversa, utilizando las ternas de parámetros de ajuste (u, a, k) que se prueban para una probabilidad de no excedencia $P(X < x)$, estimada con la fórmula de Weibull (Benson, 1962):

$$P(X < x) = \frac{m}{n+1} \quad (50)$$

en la cual m es el número de orden del dato, con 1 para el menor y n para el mayor. Por último, np es el número de parámetros de ajuste; para las distribuciones GVE, LOG y PAG es tres.

Indicadores cualitativos del ajuste

Recientemente se han popularizado dos *gráficos de diagnóstico*: el de *Probabilidades* y el de *Cantidades* (Coles, 2001; Wilks, 2011); el primero emplea en el eje de las abscisas la probabilidad empírica estimada con la ecuación (50) y en las ordenadas la probabilidad que define el modelo

para cada dato disponible ordenados en forma progresiva de magnitud $(x_1 \leq x_2 \leq \dots \leq x_n)$, la cual se estima con las ecuaciones (17), (27) y (33) para las distribuciones GVE, LOG y PAG, respectivamente. En el gráfico de Cantidades se indican en las abscisas el valor del dato observado $(x_1 \leq x_2 \leq \dots \leq x_n)$ y en las ordenadas el valor estimado con la solución inversa del modelo; es decir, con las ecuaciones (20), (28) y (34) para las funciones GVE, LOG y PAG, para la probabilidad empírica estimada con la ecuación (50). En resumen, las coordenadas x y y de cada gráfico son (Coles, 2001):

Probabilidades

$$\left[\left(\frac{i}{n+1}, F(x_i) \right) : i = 1, 2, \dots, n \right] \quad (51)$$

Cantidades

$$\left[\left(x_i, x \left(\frac{i}{n+1} \right) \right) : i = 1, 2, \dots, n \right] \quad (52)$$

Descripción de la aplicación numérica

Registros hidrométricos procesados

La aplicación numérica que se describe utilizó los 21 registros de crecientes máximas anuales disponibles en la Región Hidrológica núm. 10 (Sinaloa) en estaciones hidrométricas cuyo régimen de escurrimiento no está afectado por embalses y cuentan con más de 20 años de datos en el sistema BANDAS (IMTA, 2002). Los registros de las estaciones hidrométricas Huites y Guamúchil abarcan hasta el año en que la construcción del embalse respectivo afectó su régimen; los datos de la estación San Francisco proceden del *Boletín Hidrológico* 36 (SRH, 1975). Esta información hidrométrica se actualizó en las oficinas de la Comisión Nacional del Agua (Conagua) y entonces 12 registros fueron ampliados con datos que abarcan, como máximo, hasta el año 2011, los cuales fueron Santa Cruz, Jaina, Palo Dulce, Ixpalino, Chinipas, Tamazula, Acatitán, Choix, Badiraguato, Zopilote, Chico

Ruiz y El Bledal. En las primeras cinco columnas del cuadro 1 se exponen las características generales de los registros que serán procesados, expuestos en orden decreciente de tamaño de cuenca; el registro más corto tiene 23 años y los más amplios 60 años, con un valor mediano de 33 años. Campos-Aranda (2014a) presenta un mapa con la localización de las 21 estaciones hidrométricas procesadas y sus respectivas cuencas de drenaje.

Pruebas estadísticas aplicadas a los registros

Para verificar la calidad estadística de los registros hidrométricos que serán procesados, se aplicó una prueba general y seis tests específicos que buscan persistencia (Anderson y Sneyers), tendencia (Kendall y Spearman), exceso o déficit de variabilidad (Bartlett) y cambio en

la media (Cramer). La prueba general fue la de Von Neumann, que detecta aleatoriedad contra componentes determinísticas no especificadas. Todas estas pruebas se pueden consultar en WMO (1971), y Machiwal y Jha (2008, 2012).

La prueba de Von Neumann encontró sólo dos registros no homogéneos: Choix y Pericos, ambos con persistencia detectada por los tests de Anderson y de Sneyers. Tres registros (Santa Cruz, Jaina y Acatitán) mostraron tendencia descendente y uno (Badiraguato) ascendente ligera, de acuerdo con los tests de Kendall y Spearman. Al tomar en cuenta que los registros que mostraron tendencia descendente tienen años faltantes, principalmente en su lapso final de 1995 al 2011, se consideraron susceptibles de procesamiento probabilístico, ya que tal componente determinística puede estar siendo inducida por la falta de continuidad.

Cuadro 1. Características generales y momentos L depurados (1, 0) de los 21 registros de crecientes máximas anuales disponibles en las estaciones hidrométricas indicadas de la Región Hidrológica núm. 10 (Sinaloa).

Núm.	Estación hidrométrica	Área de cuenca (km ²)	Años de registro (n)	$I_1 = I_1^{(0,0)}$	$I_1^{(1,0)}$	$I_2^{(1,0)}$	$I_3^{(1,0)}$	$t_3^{(1,0)}$
1	Huites	26 057	1942-1992 (51)	3328.333	4854.961	1706.992	802.426	0.4701
2	San Francisco	17 531	1941-1973 (33)	1724.636	2444.118	769.869	312.918	0.4065
3	Santa Cruz	8 919	1944-2011 (56)	972.500	1464.255	525.671	247.989	0.4718
4	Jaina	8 179	1942-2011 (60)	973.850	1441.582	519.757	259.900	0.5000
5	Palo Dulce	6 439	1958-1986 (26)	1083.423	1552.551	551.855	320.997	0.5817
6	Ixpalino	6 166	1953-2009 (46)	1191.891	1663.188	510.515	251.830	0.4933
7	La Huerta	6 149	1970-1999 (28)	945.107	1272.204	258.448	10.619	0.0411
8	Chinipas	5 098	1965-2008 (33)	981.030	1310.680	308.185	81.948	0.2659
9	Tamazula	2 241	1963-1999 (33)	593.909	772.023	181.468	78.403	0.4321
10	Naranjo	2 064	1939-1984 (45)	633.311	965.053	350.734	136.605	0.3895
11	Acatitán	1 884	1955-2008 (49)	797.878	1242.170	462.272	182.957	0.3958
12	Guamúchil	1 645	1940-1971 (32)	702.344	1003.141	319.356	141.162	0.4420
13	Choix	1 403	1956-2005 (41)	334.927	469.731	143.144	66.263	0.4629
14	Badiraguato	1 018	1960-1999 (40)	1033.350	1650.228	763.269	464.662	0.6088
15	El Quelite	835	1961-2001 (33)	479.091	697.343	216.313	71.132	0.3288
16	Zopilote	666	1939-2010 (60)	337.117	493.580	143.607	29.706	0.2069
17	Chico Ruiz	391	1977-2006 (23)	206.343	296.257	79.715	12.380	0.1553
18	El Bledal	371	1938-1994 (57)	288.895	411.405	126.451	53.259	0.4212
19	Pericos	270	1961-1992 (30)	250.800	336.685	81.431	22.299	0.2773
20	La Tina	254	1960-1983 (24)	104.958	165.297	67.297	36.277	0.5391
21	Bamícori	223	1951-1983 (33)	189.182	280.225	94.765	33.181	0.3501

Momentos L depurados (1, 0)

Con base en la ecuación (16) se calcularon estos momentos, los cuales se han concentrado en las columnas 6, 7 y 8 del cuadro 1; en su columna final se tiene el cociente de asimetría, evaluado según la ecuación (14).

Errores estándar de ajuste

Con base en los momentos L y sus cocientes (Hosking & Wallis, 1997; Campos-Aranda, 2014a), se obtuvieron los tres parámetros de ajuste de las distribuciones GVE, LOG y PAG, por medio de las ecuaciones (22) a (26), (30) a (32) y (36) a (38), respectivamente. Después, haciendo uso de sus soluciones inversas,

ecuaciones (20), (28) y (34), se realizaron las estimaciones $\hat{x}_i$ o predicciones necesarias para aplicar la ecuación (49) del error estándar de ajuste (EEA). Los resultados se han concentrado en el cuadro 2, en las columnas indicadas con "moL".

De manera similar, pero utilizando los valores de los momentos L depurados (1, 0) del cuadro 1, se estimaron los tres parámetros de ajuste de las funciones GVE, LOG y PAG, a través de las ecuaciones (39) a (42), (43) a (45) y (46) a (48). En seguida, a partir de sus soluciones inversas, expresiones (20), (28) y (34), se obtuvieron los valores de $\hat{x}_i$ para obtener los errores estándar de ajuste, los cuales se tienen en el cuadro 2, en las columnas con designación "moLD".

Cuadro 2. Errores estándar de ajuste (m^3/s) obtenidos con los métodos de momentos L y momentos L depurados (1, 0), en las estaciones hidrométricas indicadas de la Región Hidrológica núm. 10 (Sinaloa).

Núm.	Estación hidrométrica	Error estándar de ajuste (EEA)						Error relativo del máximo EEA (%)
		GVE		LOG		PAG		
		moL	moLD	moL	moLD	moL	moLD	
1	Huites	979	925	1061	976	834	(822)	29.1
2	San Francisco	377	351	419	385	302	(295)	42.0
3	Santa Cruz	388	389	402	398	393	(378)	6.3
4	Jaina	349	352	369	362	337	(335)	10.1
5	Palo Dulce	745	763	767	765	(716)	743	7.1
6	Ixpalino	346	354	366	364	(335)	337	9.3
7	La Huerta	96	134	124	211	(53)	54	298.1
8	Chinipas	101	102	118	123	90	[90]	36.7
9	Tamazula	144	145	148	148	146	(142)	4.2
10	Naranjo	156	148	173	164	127	(124)	39.5
11	Acatitán	189	191	210	207	176	(174)	20.7
12	Guamúchil	235	236	247	243	(225)	226	9.8
13	Choix	101	102	106	105	99	(98)	8.2
14	Badiraguato	838	836	888	829	808	(801)	10.9
15	El Quelite	87	86	98	98	74	[74]	32.4
16	Zopilote	51	56	63	79	28	(27)	192.6
17	Chico Ruiz	28	33	35	44	[17]	17	158.8
18	El Bledal	64	65	67	67	69	(64)	7.8
19	Pericos	29	28	33	34	(24)	25	41.7
20	La Tina	90	92	92	93	(87)	90	6.9
21	Bamícori	52	49	57	54	43	(41)	39.0

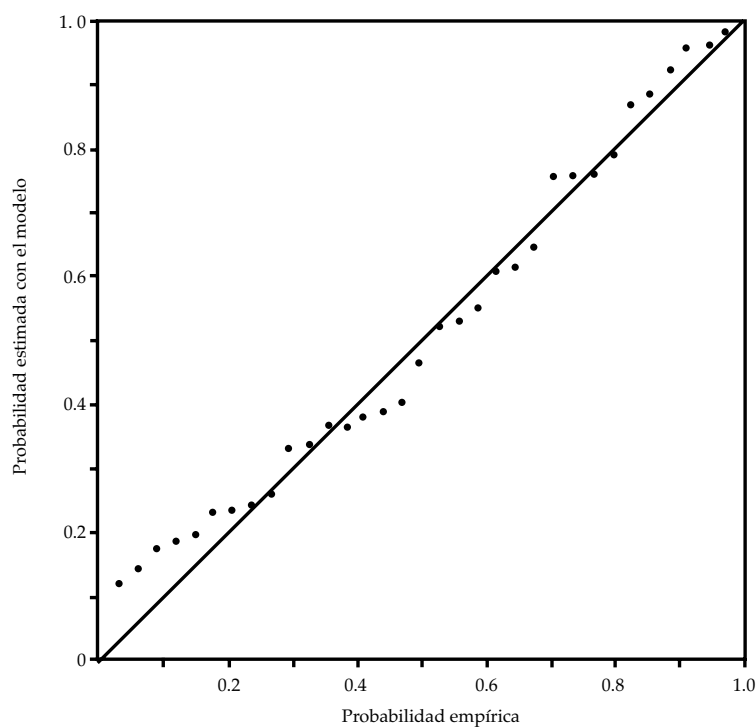


Figura 1. Gráfico de probabilidad para el registro de crecientes anuales de la estación hidrométrica San Francisco, a través del ajuste de la distribución GVE con un EEA = 351 m³/s.

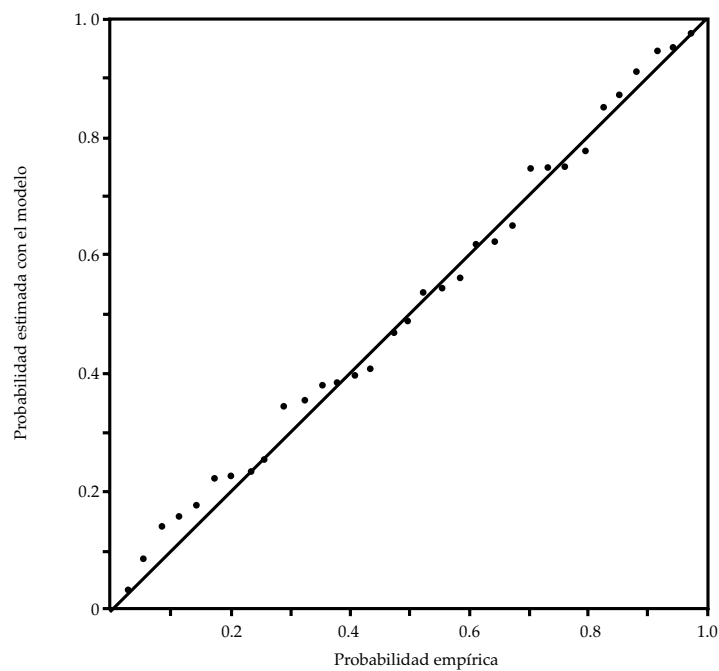


Figura 2. Gráfico de probabilidad para el registro de crecientes anuales de la estación hidrométrica San Francisco, a través del ajuste de la distribución PAG con un EEA = 295 m³/s.

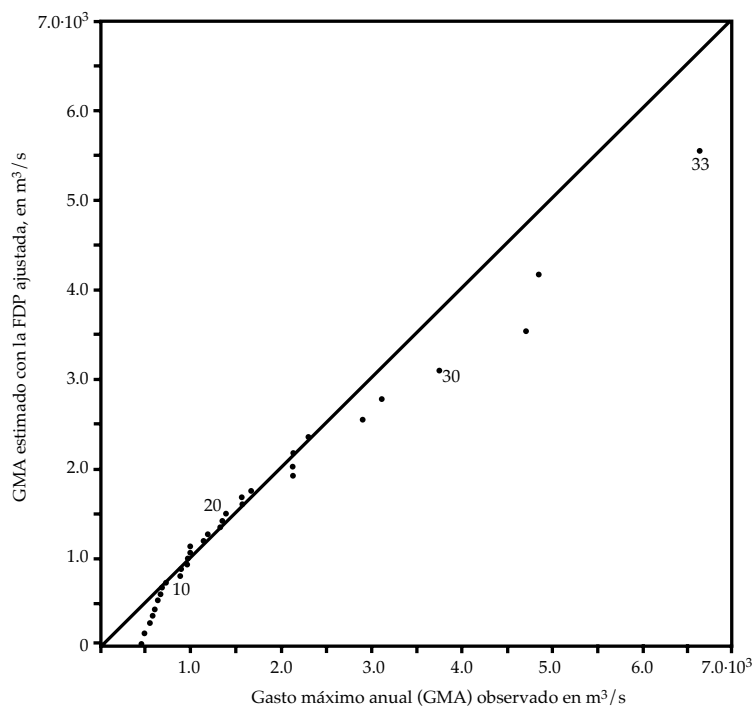


Figura 3. Gráfico de cantidades para el registro de crecientes anuales de la estación hidrométrica San Francisco, a través del ajuste de la distribución LOG con un EEA = 385 m³/s.

Gráficos de diagnóstico

En las figuras 1 y 2 se presentan los gráficos de probabilidad del registro de la estación hidrométrica San Francisco, con $n = 33$, para las distribuciones GVE y PAG. Se observa, en general, un mejor ajuste del modelo PAG con todos los datos de la muestra, sobre todo para los valores bajos, en los que la función GVE presenta la mayor dispersión.

Respecto a los gráficos de cantidades, se exponen los de las distribuciones LOG y PAG en las figuras 3 y 4, para el mismo registro en la estación de aforos San Francisco, destacando el ajuste excelente que tiene la función PAG en los valores bajos y cómo ambas funciones no pueden reproducir los valores altos del registro a partir del dato 28.

Análisis de los resultados y predicciones de diseño

En el cuadro 2 se han indicado entre paréntesis circular los valores mínimos del error estándar de ajuste; se observa que en todos los registros, la distribución PAG condujo a los valores menores, por ello es la seleccionada para exponer sus predicciones, las cuales se han concentrado en el cuadro 3. Campos-Aranda (2014b) expone varios métodos simples para el ajuste de la función PAG.

En tres registros, los de las estaciones Chinipas, El Quelite y Chico Ruiz, ocurren valores iguales del EEA y para tales casos se decidió adoptar las predicciones más altas del método de momentos L o de los momentos L depurados (1, 0), lo cual se indica con paréntesis rectan-

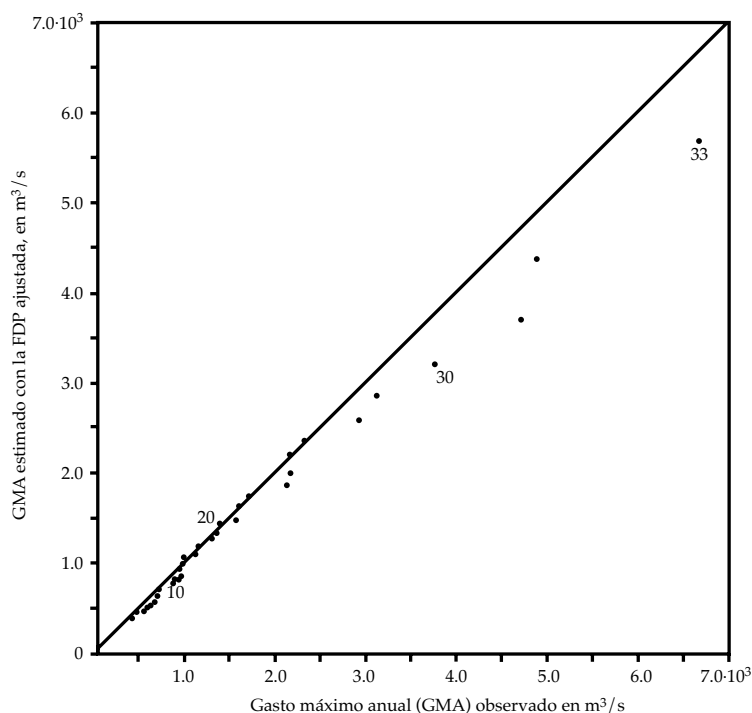


Figura 4. Gráfico de cantidades para el registro de crecientes anuales de la estación hidrométrica San Francisco, a través del ajuste de la distribución PAG con un EEA = 295 m³/s.

gular en el cuadro 2. Conviene destacar que en dos primeros registros citados, el ajuste de la distribución PAG con momentos L depurados (1, 0), condujo a predicciones más altas en los periodos de retorno elevados.

También se debe destacar que el ajuste por momentos L depurados (1, 0) de la función PAG llevó a un valor menor del EEA en más de la mitad (catorce) de los 21 registros procesados. Lo anterior demuestra que tal método en efecto otorga mayor ponderación a los datos de mayor magnitud a través de un mejor ajuste de la FDP, lo que conduce a predicciones más confiables.

Con el método de momentos L depurados (1, 0), se obtienen menores EEA apreciables en el ajuste de la distribución GVE en ciertos registros, como Huites, San Francisco, Naranjo, Badiraguato y Bamícori. En relación con la función LOG, lo anterior ocurre en los mismos registros y en Jaina y Guamúchil.

Un contraste individual entre las predicciones mostradas en el cuadro 3 y las adoptadas

por Campos-Aranda (2014a), procedentes de las distribuciones GVE, LOG y Log-Pearson tipo III, demostró una similitud alta en todas las estaciones hidrométricas hasta los periodos de retorno de 100 años, pero también mostró que la función PAG aporta, en algunos registros, predicciones menores en los periodos de retorno elevados (>100 años). En la última columna del cuadro 2 se tiene el llamado error relativo del máximo EEA, calculado con la expresión siguiente:

$$ER = \frac{EEA_M - EEA_m}{EEA_m} 100 \quad (53)$$

en la cual EEA_M y EEA_m son los errores estándar de ajuste máximo y mínimo de los seis valores calculados. Destacan los ER de las estaciones de aforos La Huerta, Zopilote y Chico Ruiz, cuyos registros carecen de valores extremos, ya que el parámetro de forma (k), estimado con momentos L de la distribución PAG, resultó

Cuadro 3. Predicciones (m^3/s) obtenidas con la distribución Pareto Generalizada en los 21 registros de crecientes anuales de las estaciones hidrométricas indicadas de la Región Hidrológica núm. 10 (Sinaloa).

Núm.	Estación hidrométrica	Método de ajuste	Periodos de retorno en años						
			5	10	25	50	100	500	1 000
1	Huites	moLD	4 480	6 760	10 612	14 326	18 909	34 121	43 336
2	San Francisco	moLD	2 433	3 500	5 116	6 515	8 086	12 525	14 840
3	Santa Cruz	moLD	1 346	2 047	3 235	4 383	5 804	10 538	13 417
4	Jaina	moLD	1 278	1 951	3 146	4 355	5 910	11 464	15 058
5	Palo Dulce	moL	1 301	1 973	3 251	4 635	6 526	14 025	19 350
6	Ixpalino	moL	1 608	2 307	3 421	4 433	5 622	9 236	11 256
7	La Huerta	moL	1 521	1 764	1 935	2 002	2 041	2 079	2 085
8	Chinipas	moLD	1 442	1 853	2 328	2 642	2 921	3 455	3 645
9	Tamazula	moLD	754	1 003	1 398	1 754	2 171	3 428	4 125
10	Naranjo	moLD	979	1 467	2 184	2 786	3 445	5 226	6 115
11	Acatitán	moLD	1 251	1 894	2 848	3 659	4 555	7 017	8 267
12	Guamúchil	moL	996	1 438	2 111	2 696	3 357	5 237	6 225
13	Choix	moLD	442	634	955	1 261	1 635	2 853	3 580
14	Badiraguato	moLD	1 154	1 966	3 668	5 691	8 689	22 446	33 516
15	El Quelite	moLD	748	1 047	1 440	1 734	2 026	2 696	2 981
16	Zopilote	moLD	579	759	943	1 049	1 133	1 269	1 309
17	Chico Ruiz	moL	348	446	540	593	633	694	710
18	El Bledal	moLD	404	578	849	1 089	1 366	2 177	2 616
19	Pericos	moL	374	480	599	676	742	865	906
20	La Tina	moL	150	239	392	539	722	1 335	1 709
21	Bamícori	moLD	296	428	608	749	893	1 245	1 403

positivo en las tres, y en la función GVE también es positivo en la primera y cercano a cero en las restantes. En estos tres registros, el ajuste de las distribuciones GVE y LOG a través del método de los momentos L depurados (1, 0) conduce a EEA mayores que los del método de momentos L , lo cual indica que no es conveniente eliminar su dato más bajo.

Conclusiones

Los resultados numéricos de este trabajo establecen a la distribución Pareto Generalizada como un modelo probabilístico que conviene probar en los análisis de frecuencias de crecientes y de otros datos hidrológicos extremos, ya que conduce a errores estándar de ajuste (EEA) bajos.

En diez de los 21 registros procesados en la aplicación numérica descrita, los EEA obtenidos con las distribuciones GVE y LOG están dentro del 10% del error relativo (ecuación (47)), como se observa en la columna final del cuadro 2; por lo anterior, su aplicación debe seguir siendo una norma o precepto.

Los resultados concentrados en el cuadro 2 relativos al EEA demuestran que en determinados registros de crecientes anuales, más de la mitad de los 21 procesados, el método de los momentos L depurados (1, 0) es una buena opción para abatirlo y lograr mejores ajustes con cada una de las tres funciones de probabilidad aplicadas (GVE, LOG y PAG); por ello se sugiere utilizarlo sistemáticamente, como una versión avanzada del procedimiento de los momentos L , el cual ya es de uso universal en los análisis de frecuencias hidrológicas.

Agradecimientos

Se agradece a los ingenieros de la Conagua, José Luis Juárez Rubio de la Dirección Local San Luis Potosí y Alejandro García Ruiz de la GASIR, el haber proporcionado al autor la información hidrométrica actualizada de crecientes anuales en 19 estaciones de aforos de la Región Hidrológica núm. 10 (Sinaloa).

Referencias

- Abdul-Moniem, I. B., & Selim, Y. M. (2009). TL-Moments and L-Moments Estimation for the Generalized Pareto Distribution. *Applied Mathematical Sciences*, 3(1), 43-52.
- Ahmad, U. N., Shabri, A., & Zakaria, Z. A. (2011a). Flood Frequency Analysis of Annual Maximum Stream Flows Using L-Moments and TL-Moments Approach. *Applied Mathematical Sciences*, 5(5), 243-253.
- Ahmad, U. N., Shabri, A., & Zakaria, Z. A. (2011b). Trimmed L-Moments (1, 0) for the Generalized Pareto Distribution. *Hydrological Sciences Journal*, 56(6), 1053-1060.
- Ahmad, U. N., Shabri, A., & Zakaria, Z. A. (2013). An Analysis of Annual Maximum Stream Flows in Terengganu, Malaysia using TL-Moments Approach. *Theoretical and Applied Climatology*, 111(3-4), 649-663.
- Asquith, W. H. (2011). *Distributional Analysis with L-moment Statistics using the R Environment for Statistical Computing*. Chapter 3: Order Statistics (pp. 47-59) and Chapter 6: L-Moments (pp. 87-122). Texas, USA: Author Edition (ISBN-13: 978-1463508418).
- Benson, M. A. (1962). Plotting Positions and Economics of Engineering Planning. *Journal of Hydraulics Division*, 88(6), 57-71.
- Bhattarai, K. P. (2004). Partial L-Moments for the Analysis of Censored Flood Samples. *Hydrological Sciences Journal*, 49(5), 855-868.
- Bílková, D. (2014). Trimmed L-Moments: Analogy of Classical L-Moments. *American Journal of Mathematics and Statistics*, 4(2), 80-106.
- Campos-Aranda, D. F. (2014a). Análisis regional de frecuencia de crecientes en la Región Hidrológica núm. 10 (Sinaloa), México. 2: contraste de predicciones locales y regionales. *Agrociencia*, 48(3), 255-270.
- Campos-Aranda, D. F. (octubre-diciembre, 2014b). Predicción de crecientes usando la distribución Pareto Generalizada ajustada con tres métodos simples. *Revista digital Tláloc*, 65.
- Coles, S. (2001). Theme 2.6.7: Model Diagnostics (pp. 36-44). In *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.
- Davis, P. J. (1965). Gamma Function and Related Functions. Chapter 6 (pp. 253-296). In M. Abramowitz & I. A. Stegun (Eds.). *Handbook of Mathematical Functions*. New York: Dover Publications.
- El Adlouni, S., Bobée, B., & Ouarda, T. B. M. J. (2008). On the Tails of Extreme Event Distributions in hydrology. *Journal of Hydrology*, 355(1-4), 16-33.
- Elamir, E. A. H. (2010). Optimal Choices for Trimming in Trimmed L-Moment Method. *Applied Mathematical Sciences*, 4(58), 2881-2890.
- Elamir, E. A. H., & Seheult, A. H. (2003). Trimmed L-Moments. *Computational Statistics & Data Analysis*, 43(3), 299-314.
- Elamir, E. A. H., & Seheult, A. H. (2004). Exact Variance Structure of Sample L-Moments. *Journal of Statistical Planning and Inference*, 124(2), 337-359.
- Flowers-Cano, R. S., Flowers, R. F., & Rivera-Trejo, F. (2014). Evaluación de criterios de selección de modelos probabilísticos: validación con series de valores máximos simulados. *Tecnología y Ciencias del Agua*, 5(5), 189-197.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form. *Water Resources Research*, 15(5), 1049-1054.
- Hosking, J. R. M. (1990). L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society, B*, 52(1), 105-124.
- Hosking, J. R. M. (2007). Some Theory and Practical Uses of Trimmed L-Moments. *Journal of Statistical Planning and Inference*, 137(9), 3024-3039.
- Hosking, J. R. M., & Wallis, J. R. (1997). Chapter 2: L-Moments (pp. 14-43); Chapter 5: Choice Of A Frequency Distribution (pp. 73-86); Appendix: L-Moments For Some Specific Distributions (pp. 191-209). In *Regional Frequency Analysis. An Approach Based on L-moments*. Cambridge, England: Cambridge University Press.
- IMTA (2002). *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.
- Karvanen, J. (2006). Estimation of Quantile Mixtures Via L-Moments and Trimmed L-Moments. *Computational Statistics & Data Analysis*, 51(2), 947-959.
- Kite, G. W. (1977). Chapter 12: Comparison of Frequency Distributions (pp. 156-168). In *Frequency and Risk Analyses in Hydrology*. Colorado, USA: Water Resources Publications.
- Kottegoda, N. T., & Rosso, R. (2008). Chapter 3: Random Variables and their Properties (pp. 83-164). In *Applied Statistics for Civil and Environmental Engineers*. Chichester, United Kingdom: Blackwell Publishing.

- Machiwal, D., & Jha, M. K. (2008). Comparative Evaluation of Statistical Tests for Time Series Analysis: Application to Hydrological Time Series. *Hydrological Sciences Journal*, 53(2), 353-366.
- Machiwal, D., & Jha, M. K. (2012). Chapter 4: Methods for Time Series Analysis (pp. 51-84). In *Hydrologic Time Series Analysis: Theory and Practice*. Dordrecht, The Netherlands: Springer.
- Mudholkar, G. S., & Hutson, A. D. (1998). LQ-Moments: Analogs of L-Moments. *Journal of Statistical Planning and Inference*, 71(1-9), 191-208.
- Rao, A. R., & Hamed, K. H. (2000). Chapter 9: The Logistic Distribution (pp. 291-321). In *Flood Frequency Analysis*. Boca Raton, USA: CRC Press.
- SRH (1975). *Boletín Hidrológico núm. 36. Región Hidrológica núm. 10 (Sinaloa)*. Tomos I y VI. México, DF: Secretaría de Recursos Hidráulicos, Dirección de Hidrología.
- Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Chapter 18: Frequency Analysis of Extreme Events (pp. 18.1-18.66). In D. R. Maidment (Ed.). *Handbook of Hydrology*. New York: McGraw-Hill.
- Wilks, D. S. (2011). Theme 4.5: Qualitative Assessments of the Goodness Fit (pp. 112-116). In *Statistical Methods in the Atmospheric Sciences* (3rd edition). San Diego, USA: Academic Press (Elsevier).
- WMO (1971). Annexed III: Standard Tests of Significance to Be Recommended in Routine Analysis of Climatic Fluctuations (pp. 58-71). Technical Note No. 79. In *Climatic Change*. Geneva: World Meteorological Organization.

Dirección institucional del autor

Dr. Daniel Francisco Campos Aranda

Profesor jubilado de la Universidad Autónoma de San Luis Potosí
Genaro Codina 240, Colonia Jardines del Estadio
78280 San Luis Potosí, San Luis Potosí, México
campos_aranda@hotmail.com