

DOI: 10.24850/j-tyca-15-2-5

Notas

Contraste de Análisis de Frecuencias entre las distribuciones Beta–kappa y Beta–Pareto con tres de aplicación generalizada

Contrast of Frequency Analysis between Beta–kappa and Beta–Pareto distributions with three of widespread application

Daniel Francisco Campos Aranda¹, ORCID: <https://orcid.org/0000-0002-9876-3967>

¹Profesor Jubilado de la Universidad Autónoma de San Luis Potosí, San Luis Potosí, México, campos_aranda@hotmail.com

Autor para correspondencia: Daniel Francisco Campos Aranda, campos_aranda@hotmail.com

Resumen

El diseño hidrológico de varias obras hidráulicas o la revisión de las construidas, se basa en las *Crecientes de Diseño*, que son gastos máximos del río, asociados a bajas probabilidades de excedencia o *predicciones*. Su



estimación más confiable se realiza a través del *Análisis de Frecuencias*. Proceso estadístico que consiste en representar el registro de gastos máximos anuales, con una *función de distribución de probabilidades* (FDP) o modelo probabilístico, utilizado para realizar las predicciones buscadas. En este estudio de contraste se proponen las FDP Beta-kappa y Beta-Pareto y se consideraron FDP de uso generalizado, las tres siguientes: la Log-Pearson tipo III, la General de Valores Extremos y la Logística Generalizada. Por lo anterior, se expone, para las dos primeras FDP, un resumen de su teoría y su método de ajuste por máxima verosimilitud. Se procesan once registros de datos hidrológicos extremos anuales y se contrastan los ajustes con dos índices: el error estándar de ajuste y el error absoluto medio. La selección de las predicciones en los siete periodos de retorno (Tr) estudiados, se basó en los valores menores de los errores de ajuste y en la búsqueda de predicciones *representativas*, en los $Tr \geq 500$ años. Las Conclusiones sugieren la inclusión de las distribuciones Beta-kappa y Beta-Pareto en los análisis de frecuencias debido a su versatilidad y facilidad de ajuste.

Palabras clave: distribución Beta-kappa, distribución Beta-Pareto, ajuste por máxima verosimilitud, error estándar de ajuste, error absoluto medio, gráficos Q-Q, predicciones.

Abstract

The hydrological design of several hydraulic works or the revision of the constructed ones is based on the *Design Floods*, which are maximum flows of the river, associated with low probabilities of exceedance or *predictions*.



Its most reliable estimate is made through *Frequency Analysis*. Statistical process that consists of representing the record of maximum annual flows, with a *probability distribution function* (PDF) or probabilistic model, used to make the desired predictions. In this contrast study, the Beta-kappa and Beta-Pareto FDPs are proposed, and the following three were considered to be widely used FDPs: Log-Pearson type III, General Extreme Values, and Generalized Logistics. Therefore, it is exposed, for the first two FDP, a summary of his theory and his method of fit for maximum likelihood is presented. Eleven annual extreme hydrological data records are processed and the fits are contrasted with two indices: the standard error of fit and the mean absolute error. The selection of the predictions in the seven return periods (Tr) studied was based on the lower values of the fit errors and on the search for *representative* predictions, in the $Tr \geq 500$ years. The Conclusions suggest the inclusion of the Beta-kappa and Beta-Pareto distributions in the frequency analysis due to their versatility and fit facility.

Keywords: Beta-kappa distribution, Beta-Pareto distribution, maximum likelihood fit, standard error of fit, mean absolute error, Q-Q graphics, predictions.

Recibido: 10/06/2021

Aceptado: 26/07/2022

Introducción

Etapas del Análisis de Frecuencias

El dimensionamiento hidrológico de las obras hidráulicas, como son: diques de protección, canalizaciones y puentes, así como las diversas estructuras del drenaje urbano, se basa en las llamadas *Crecientes de Diseño* (CD). La estimación hidrológica de las CD más exacta, se realiza a través del *Análisis de Frecuencias* (AF). Procedimiento estadístico que consiste en interpretar o caracterizar el registro disponible de eventos hidrológicos, por ejemplo, crecientes o lluvias máximas, en términos de sus probabilidades de ocurrencia futuras (Bobée & Ashkar, 1991).

El AF involucra las cinco etapas siguientes: (1) integración y verificación de la aleatoriedad del registro o muestra disponible; (2) selección de la *función de distribución de probabilidades* (FDP) o modelo probabilístico que representará a los datos y permitirá las estimaciones o *predicciones* asociadas a bajas probabilidades de excedencia; (3) ajuste de las diversas FDP probadas, es decir, obtención de sus parámetros de ajuste con los variados métodos disponibles; (4) evaluación de la calidad estadística del ajuste logrado entre los datos y la FDP, por medio de gráficos e índices de diagnóstico y (5) selección de los resultados (Kite,

1977; Bobée & Ashkar, 1991; Rao & Hamed, 2000; Meylan, Favre & Musy, 2012; Stedinger, 2017; Teegavarapu, Salas & Stedinger, 2019).

En este estudio de contraste, en la etapa uno se seleccionaron cuatro registros de gastos pico y volumen *conjuntos* de crecientes anuales y tres registros de lluvias máximas diarias anuales. En total se procesaron once series de datos hidrológicos extremos y se verificó su aleatoriedad con el test de Wald–Wolfowitz. En la etapa dos, se aborda el *objetivo* del estudio al seleccionar a las FDP Beta–kappa (BEK) y Beta–Pareto (BEP) para contrastarlas contra tres de aplicación generalizada, que fueron: la Log–Pearson tipo III (LP3), la General de Valores Extremos (GVE) y la Logística Generalizada (LOG). Todas las FDP citadas tienen tres parámetros de ajuste.

En la etapa tres, las FDP BEK y BEP se ajustaron con el método de máxima verosimilitud propuesto por Mielke y Johnson (1974). La distribución LP3 se ajustó con su método clásico de momentos en el dominio logarítmico (WRC, 1977) y los modelos GVE y LOG con el método de momentos L (Hosking & Wallis, 1997).

Para la etapa cuatro se calcularon dos índices (*EEA* y *EAM*), el error estándar de ajuste (Kite, 1977; Chai & Draxler, 2014) y el error absoluto medio (Willmott & Matsuura, 2005). Finalmente, para la etapa cinco de selección de resultados, se tomaron en cuenta los valores del *EEA* y *EAM*, así como los valores obtenidos para las *predicciones*.

Antecedentes de las distribuciones BEK y BEP



Strupczewski, Markiewicz, Kochanek y Singh (2008) indican que existen muy pocas referencias sobre el uso de FDP, con dos parámetros de forma para modelar eventos hidrológicos extremos y citan las dos siguientes. El sistema de distribuciones del estadístico francés Halphen propuesto en 1941, tiene una frontera inferior en cero y dos parámetros de forma, pero debido a su complejidad matemática quedó en el olvido. Por otra parte, las distribuciones designadas por Mielke y Johnson (1974) BEK y BEP son modelos con dos parámetros de forma y uno de escala.

Wilks (1993) en un trabajo pionero de contraste de FDP de tres parámetros de ajuste, con datos máximos de precipitación, procesados como series anuales y de magnitudes superiores a un valor umbral, encontró que la distribución BEK describe bastante bien a las series anuales y el modelo BEP se ajusta mejor a las series de duración parcial. Campos–Aranda (1998) expuso varias aplicaciones de la distribución BEP. Mason, Waylen, Mimmack, Rajaratnam y Harrison (1999) emplean las FDP BEK y BEP en un estudio de detección de cambios ocurridos en lluvias extremas.

Öztekin (2007) contrasta los modelos BEK y BEP contra la distribución Wakeby, encontrado que esta última conduce a mejores ajustes o similares, en registros de lluvia máxima. Murshed, Kim y Park (2011) exponen para la distribución BEK la estimación de sus parámetros de ajuste mediante los métodos de momentos y momentos L. Por último, Nguyen, El Outayek, Lim y Nguyen (2017) incluyen a las FDP BEK y BEP



en su estudio de contraste de lluvias máximas anuales, basado en las habilidades descriptivas y predictivas de las FDP.

Objetivos

De este estudio de contraste fueron los tres siguientes: (1) exponer un resumen de la teoría de las distribuciones BEK y BEP; (2) describir con detalle su método de ajuste por máxima verosimilitud, a través de sus ecuaciones iterativas para la estimación de sus tres parámetros de ajuste y (3) realizar un contraste de bondad de ajuste y predicciones, entre las distribuciones BEK y BEP y las tres de aplicación generalizada (LP3, GVE y LOG).

En términos generales, la técnica estadística del AF tiene suposiciones debatibles en cada una de sus cinco etapas. Desde la representatividad en el futuro del registro disponible, hasta la adopción de resultados, con base en gráficos e índices de diagnóstico; pasando por la selección de varias FDP para realizar las predicciones buscadas. Es la etapa dos del AF, la que abre posibilidades para probar nuevos modelos probabilísticos, ya que como se sabe, ninguna FDP es mejor y su *idoneidad* depende de cada registro procesado.

Métodos y materiales

Ecuaciones de las distribuciones BEK y BEP

Mielke y Johnson (1974) exponen la FDP y la función de densidad de probabilidad (*fdp*) de la variable aleatoria Beta generalizada de segunda clase (x):

$$F(x) = \frac{(x/\beta)^{\gamma} {}_2F_1[\alpha+1, \gamma/\theta; 1+\gamma/\theta; -(x/\beta)^{\theta}]}{\left(\frac{\gamma}{\theta}\right) B(\gamma/\theta, \alpha+1-\gamma/\theta)} \quad x \geq 0 \quad (1)$$

$$F(x) = 0 \quad x \geq 0 \quad (2)$$

$$f(x) = \frac{(x/\beta)^{\gamma-1} [1+(x/\beta)^{\theta}]^{-(\alpha+1)}}{(\beta/\theta) B(\gamma/\theta, \alpha+1-\gamma/\theta)} \quad x > 0 \quad (3)$$

$$f(x) = 0 \quad x \geq 0 \quad (4)$$

en las cuales, $a > 0$, $\beta > 0$, $\theta > 0$ y $0 < \gamma < \theta(a+1)$. El numerador de la ecuación 1 es la serie hipergeométrica Gaussiana siguiente (Oberhettinger, 1972):



$${}_1F_2(a, b; c; z) = \sum_{n=0}^{\infty} \frac{\Gamma(n+a)\Gamma(n+b)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(n+c)} \frac{z^n}{n!} \quad (5)$$

y la función del denominador es:

$$B(\delta, \epsilon) = \Gamma(\delta)\Gamma(\epsilon)/\Gamma(\delta + \epsilon) \quad (6)$$

Mielke y Johnson (1974) indican que los cálculos asociados con las ecuaciones 1 y 3, no son simples, pero estableciendo dos restricciones al parámetro γ se obtienen distribuciones con ventajas computacionales importantes. La primera restricción es $\gamma = a\theta$ y conduce a la distribución Beta- κ (BEK), designada así por su similitud con la distribución Kappa de Mielke (1973), cuyas ecuaciones de su FDP y fdp son:

$$F(x) = \{(x/\beta)^\theta / [1 + (x/\beta)^\theta]\}^\alpha \quad x \geq 0 \quad (7)$$

$$f(x) = (\alpha\theta/\beta)(x/\beta)^{\alpha\theta-1} [1 + (x/\beta)^\theta]^{-(\alpha+1)} \quad x > 0 \quad (8)$$

con $a > 0$, $\beta > 0$ y $\theta > 0$. β es el parámetro de escala y a y θ los de forma. La función de cuantiles, designando a $F(x) = p$, es la siguiente:

$$x(p) = \beta [p^{1/\alpha} / (1 - p^{1/\alpha})]^{1/\theta} \quad (9)$$

La segunda restricción es $\gamma=\theta$, la cual define la distribución Beta- P (BEP), designada así por su semejanza con el modelo probabilístico tipo Pareto, sus FDP y fdp son:

$$F(x) = 1 - \left[1 + \left(\frac{x}{\beta}\right)^\theta\right]^{-\alpha} \quad x \geq 0 \quad (10)$$

$$f(x) = (\alpha\theta/\beta)(x/\beta)^{\theta-1} \left[1 + (x/\beta)^\theta\right]^{-(\alpha+1)} \quad x > 0 \quad (11)$$

con $a>0$, $\beta>0$ y $\theta>0$. Nuevamente, β es el parámetro de escala y a y θ los de forma. La función de cuantiles, designando a $F(x)=p$, es la siguiente:

$$x(p) = \beta \left[(1-p)^{-1/\alpha} - 1 \right]^{1/\theta} \quad (12)$$

Ajuste por Máxima Verosimilitud de BEK y BEP

Mielke y Johnson (1974) exponen los procedimientos de cálculo de los parámetros de ajuste (a , β , θ) con el método de máxima verosimilitud,



según tres ecuaciones de aplicación iterativa (j), partiendo de valores iniciales de β_0 y θ_0 . Para la distribución BEK tales expresiones son:

$$\alpha_j = n \left\{ \sum_{i=1}^n \ln \left[1 + (x_i/\beta_{j-1})^{-\theta_{j-1}} \right] \right\}^{-1} \quad (13)$$

$$\beta_j = \frac{1}{n} \left(1 + \frac{1}{\alpha_j} \right) \beta_{j-1} \sum_{i=1}^n \left[1 + (x_i/\beta_{j-1})^{-\theta_{j-1}} \right]^{-1} \quad (14)$$

$$\theta_j = n \left\{ \sum_{i=1}^n \frac{[(x_i/\beta_j)^{\theta_{j-1}} - \alpha_j] \ln(x_i/\beta_j)}{1 + (x_i/\beta_j)^{\theta_{j-1}}} \right\}^{-1} \quad (15)$$

Para la distribución BEP sus ecuaciones iterativas son:

$$\alpha_j = n \left\{ \sum_{i=1}^n \ln \left[1 + (x_i/\beta_{j-1})^{\theta_{j-1}} \right] \right\}^{-1} \quad (16)$$

$$\beta_j = \frac{1}{n} (1 + \alpha_j) \beta_{j-1} \sum_{i=1}^n \left[1 + (x_i/\beta_{j-1})^{-\theta_{j-1}} \right]^{-1} \quad (17)$$

$$\theta_j = n \left\{ \sum_{i=1}^n \frac{[\alpha_j (x_i/\beta_j)^{\theta_{j-1}} - 1] \ln(x_i/\beta_j)}{1 + (x_i/\beta_j)^{\theta_{j-1}}} \right\}^{-1} \quad (18)$$

Distribuciones del contraste y su ajuste

La distribución LP3 se ajustó con su método clásico, el de momentos en el dominio logarítmico (WRC, 1977). En cambio, las distribuciones GVE y LOG se ajustaron con base en el método de los momentos L, que ha demostrado ser eficiente y robusto, incluso en muestras pequeñas, según ecuaciones expuestas por Hosking y Wallis (1997).

Campos-Aranda (2002) expuso seis métodos de ajuste de la distribución LP3, acotando su aplicabilidad con base en el cociente de estimaciones (CE), definido como:

$$CE = \frac{X_{100} - X_{50}}{X_{50} - X_{25}} \quad (19)$$

en el cual, X_{Tr} es la *predicción* del periodo de retorno (Tr) indicado en años, igual al recíproco de la probabilidad de excedencia (q), cuando se procesan series o muestras de eventos hidrológicos máximos anuales. Cuando el $CE \leq 1.00$ los métodos de ajuste idóneos son la mezcla de momentos y el de momentos en el dominio real y cuando $CE \geq 1.30$ los métodos más convenientes son el de máxima verosimilitud y el de máxima entropía. Cuando CE fluctúa entre los límites citados, el método de momentos en el dominio logarítmico es aceptable y generalmente conduce a un buen ajuste.

Por otra parte, la distribución General de Valores Extremos (GVE), nace o proviene del Teorema de Fisher–Tippet–Gnedenko, que establece como el modelo probabilístico de los valores máximos muestreados anualmente a la GVE y sus tres casos particulares, según el valor de su parámetro de forma. Sin embargo, como se indica al final del inciso de los Objetivos, se siguen proponiendo y usando otras distribuciones, debido a la naturaleza aleatoria de los registros de datos hidrológicos extremos, para buscar su distribución *idónea*, con base en los indicadores estadísticos del ajuste logrado.

Gráfico de diagnóstico Q–Q

Nguyen *et al.* (2017) han sugerido dos evaluaciones para seleccionar la FDP óptima a un registro de datos hidrológicos extremos: (1) la habilidad descriptiva y (2) la habilidad predictiva. La primera se refiere a la exactitud con la cual la FDP que se prueba, reproduce a los datos de la muestra y la segunda está asociada lógicamente, con la variabilidad de sus predicciones, en relación con la dispersión que tienen las predicciones de la muestra. Para probar la habilidad descriptiva se tienen tres técnicas: (1) los gráficos de diagnóstico; (2) las pruebas estadísticas y (3) los índices de bondad de ajuste (Meylan *et al.*, 2012).

Los gráficos de diagnóstico $P-P$ y $Q-Q$ de *probabilidad* empírica contra calculada y de *cantidad* observada contra estimada se han popularizado (Coles, 2001; Wilks, 2011) y constituyen una manera simple y efectiva de comparar los resultados de una FDP contrastada. Para una muestra de datos x_i ordenados de menor a mayor, se les asigna una probabilidad empírica (p), por ejemplo, con la fórmula de Cunnane, que de acuerdo con Stedinger (2017) conduce a valores insesgados en la mayoría de las FDP utilizadas en hidrología, esta es:

$$p = \frac{m-0.40}{n+0.20} \quad (20)$$

en la cual, m es el número de orden del dato y n su número total. Para cada dato x_i se obtiene su probabilidad con la ecuación de la FDP probada. Para el caso de las distribuciones BEK y BEP, con las expresiones 7 y 10. El gráfico de $P-P$ se define con los puntos de abscisas y ordenadas siguientes:

$$\left[\frac{m-0.40}{n+0.20}, F(x_i) \right] \quad \text{para } i = 1, 2, \dots, n \quad (21)$$

El gráfico de $Q-Q$ utiliza las ecuaciones 9 y 12 o soluciones inversas de las FDP BEK y BEP, para definir los puntos de las ordenadas y está integrado por los puntos siguientes:

$$\left[x_i, x \left(\frac{m-0.40}{n+0.20} \right) \right] \quad \text{para } i = 1, 2, \dots, n \quad (22)$$

La desventaja de los gráficos de diagnóstico radica en la apreciación subjetiva que se hace al comparar diversas FDP, ya que no se dispone de un valor numérico (Nguyen *et al.*, 2017). Campos-Aranda (2019) visualiza más útil la gráfica de $Q-Q$, para detectar predicciones sobreestimadas (por quedar arriba de la recta a 45°) o subestimadas (por estar debajo).

Error Estándar de Ajuste (*EEA*)

Los índices de bondad de ajuste tienen la ventaja de ser de cálculo fácil y comúnmente involucran a la diferencia entre los valores observados x_i y los estimados \hat{x} con la FDP que se contrasta. El *EEA* es el más común (Chai & Draxler, 2014), se estableció a mediados de los años setenta (Kite, 1977) y se ha aplicado en México haciendo uso de la fórmula empírica de Weibull (Benson, 1962); ahora se aplicará utilizando la ecuación 20 de Cunnane. La expresión del *EEA* es:

$$EEA = \left[\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{(n-np)} \right]^{\frac{1}{2}} \quad (23)$$



x_i son los n datos observados ordenados de menor a mayor, \hat{x}_i los estimados, para la probabilidad estimada con la ecuación 20 y la FDP que se contrasta. np es el número de parámetros de ajuste de la FDP, con tres para las aplicadas en este estudio.

Error Absoluto Medio (*EAM*)

Sus ventajas radican en tener las unidades de la variable, al igual que el *EEA*, y evitar que el impacto de los valores dispersos sea elevado al cuadrado y por ello $EEA \geq EAM$ (Willmott & Matsuura, 2005). Su expresión es (Nguyen *et al.*, 2017):

$$EAM = \frac{\sum_{i=1}^n |x_i - \hat{x}_i|}{n - np} \quad (24)$$

Registros de datos hidrológicos procesados



Aldama, Ramírez, Aparicio, Mejía–Zermeño y Ortega–Gil, (2006) exponen los registros *conjuntos* de gasto pico (Qp) en m^3/s y volumen (Vol) en millones de m^3 (Mm^3) anuales de las crecientes de entrada a 15 embalses importantes de México y uno en proyecto. De tales registros conjuntos, se seleccionaron los tres que se consideran *complicados* para su análisis probabilístico, ya que incluyen valores dispersos (*outliers*) y tienen grandes diferencias entre sus valores bajos y los máximos. Los primeros, corresponden a los 43 datos de ingreso a la presa El Infiernillo en el río Balsas, entre los estados de Michoacán y Guerrero, tiene un área de cuenca de 108000 km^2 . Los segundos registros por procesar son los 52 datos de ingreso a la presa Huites en el río Fuerte, en el estado de Sinaloa, con un área de cuenca de 26020 km^2 . Los terceros fueron los 37 datos de ingreso a la presa Guamúchil, sobre río Mocerito, también el estado de Sinaloa y con un área de cuenca de 1630 km^2 .

Por otra parte, Domínguez y Arganis (2012) exponen los registros conjuntos de Qp y Vol con 47 datos de entrada a la presa Malpaso, en el río Grijalva, en el estado de Chiapas, México, con un área de cuenca de 34800 km^2 . Tales registros crecientes también fueron procesados.

Por último, se analizaron tres registros de precipitación máxima diaria (PMD) anual de una estación pluviométrica de cada zona geográfica del estado de San Luis Potosí, México. Del Altiplano se procesó la estación Los Filtros ($n=66$), ubicada en la ciudad de San Luis Potosí; de la Zona Media la localizada en la ciudad de Río Verde ($n=52$) y de la región Huasteca la del poblado de Tanquián de Escobedo ($n=52$). Las altitudes de las estaciones pluviométricas citadas son: 1904, 987 y 87 en m.s.n.m. Estos registros fueron analizados por Campos–Aranda (2019) para



obtener sus FDP óptimas e integrados con base en el archivo mensual en Excel de la CONAGUA, proporcionado al autor; por ello, se reproducen en la Tabla 1.

Tabla 1. Registros de *PMD* anual (milímetros) en las tres estaciones pluviométricas indicadas del estado de San Luis Potosí, México.

No.	Los Filtros (1949-2014)		Río Verde (1961-2013*)		Tanquián (1961-2014**)		No.
1	15.9	66.5	46.4	28.2	107.0	120.0	34
2	20.6	26.0	52.2	61.8	55.5	62.0	35
3	50.9	31.5	33.4	37.8	90.3	64.0	36
4	40.5	46.5	27.0	55.6	171.0	132.0	37
5	63.6	44.0	39.2	39.2	81.5	80.0	38
6	41.9	41.0	79.0	44.0	176.0	72.0	39
7	60.0	55.0	43.1	74.1	176.5	82.0	40
8	35.9	21.5	40.5	51.7	71.5	88.0	41
9	48.6	29.8	57.7	34.0	78.0	185.0	42
10	63.0	41.5	63.7	41.5	109.0	105.0	43
11	35.5	25.4	81.5	43.5	84.0	67.0	44
12	40.0	59.0	52.5	32.3	87.0	201.0	45

13	63.2	33.5	48.5	126.3	111.0	89.0	46
14	39.4	46.5	51.3	58.5	99.5	200.9	47
15	27.2	51.0	117.5	53.7	166.5	90.0	48
16	59.0	40.0	57.3	99.1	113.5	110.0	49
17	32.0	35.5	61.8	81.4	148.0	68.0	50
18	30.0	45.5	83.4	31.8	117.0	89.0	51
19	40.2	25.9	71.7	63.2	85.0	190.0	52
20	31.5	20.7	35.0	-	73.0	-	53
21	31.5	37.5	87.0	-	103.0	-	54
22	52.0	40.2	86.3	-	79.0	-	55
23	52.3	111.0	37.1	-	81.5	-	56
24	31.3	43.3	30.2	-	113.5	-	57
25	35.0	76.9	87.1	-	94.0	-	58
26	28.5	42.8	46.7	-	54.0	-	59
27	57.2	46.1	79.9	-	86.0	-	60
28	58.0	42.5	51.0	-	98.0	-	61
29	42.9	45.3	61.0	-	63.0	-	62
30	26.4	44.5	92.3	-	200.8	-	63
31	65.5	26.0	97.4	-	106.0	-	64
32	22.0	59.1	38.7	-	370.0	-	65

33	51.2	44.1	43.9	-	84.0	-	66
----	------	------	------	---	------	---	----

* un año faltante

** dos años faltantes

Test de Wald–Wolfowitz

Esta prueba no paramétrica ha sido utilizada por Bobée y Ashkar (1991), Rao y Hamed (2000) y Meylan *et al.* (2012) para probar *independencia y estacionariedad* en registros de gastos máximos anuales (X_i). Por lo anterior, se propuso aplicar la prueba a los registros de Qp , Vol y PMD anuales, que deben ser muestras de valores *aleatorios*. A. Wald y J. Wolfowitz basándose en el trabajo de R. L. Anderson sobre el coeficiente de correlación serial desarrollaron tal prueba, cuyo estadístico es:

$$R = \sum_{i=1}^{n-1} x_i \cdot x_{i+1} + x_n \cdot x_1 \quad (25)$$

Cuando el tamaño (n) de la serie o muestra (x_i) no es pequeño y sus datos son independientes, R procede de una distribución Normal con media y varianza, dadas por las expresiones siguientes:

$$E[R] = \bar{R} = \frac{S_1^2 - S_2}{n-1} \quad (26)$$

$$Var[R] = \frac{S_2^2 - S_4}{n-1} + \frac{S_1^4 - 4 \cdot S_1^2 \cdot S_2 + 4 \cdot S_1 \cdot S_3 + S_2^2 - 2 \cdot S_4}{(n-1)(n-2)} - \bar{R}^2 \quad (27)$$

en las cuales:

$$S_w = \sum_{i=1}^n x_i^w \quad (28)$$

Finalmente, se calcula U , con la ecuación:

$$U = \frac{R - \bar{R}}{\sqrt{Var[R]}} \quad (29)$$

El valor de U sigue una distribución Normal (0,1) y se puede usar para probar la independencia de los datos de la serie con un nivel de significancia α , comúnmente del 5%. En una prueba de dos colas la variable normal estandarizada es $Z_{\alpha/2} \cong 1.96$; entonces, cuando el valor absoluto de U es menor que 1.96 la serie estará integrada por valores independientes (*muestra aleatoria*).

Resultados y su Discusión



Prueba de aleatoriedad y cocientes de momentos L

En la tercera columna de la Tabla 2, se exponen los valores de la estadística U (ecuación 29), definiendo que los 11 registros procesados son aleatorios. En el resto de columnas se muestran las magnitudes de la media aritmética y los cocientes de momentos L de cada registro.

Los registros originales de la estación hidrométrica Guamúchil expuestos por Aldama *et al.* (2006), incluyen al final cuatro valores del Qp y Vol máximos extremos, obtenidos en la presa Eustaquio Buelna con operación inversa de su tránsito de crecientes, los cuales originan que tales registros no sean aleatorios ($U = 4.156$ y $U = 2.597$). Eliminando esos cuatro datos de Qp y Vol se obtienen los valores de U mostrados en la Tabla 2.

Tabla 2. Datos generales y cocientes de momentos L de los 11 registros procesados.

No.	Registro en:	U	\bar{X}	t_3	t_4
1	El Infiernillo Qp	-0.707	5499.512	0.49875	0.36082
2	El Infiernillo Vol	-0.022	2244.791	0.37866	0.16834



3	Huites Qp	-0.090	3305.135	0.49313	0.30438
4	Huites Vol	-0.602	841.769	0.30773	0.17585
5	Guamúchil Qp	-1.418	1610.854	0.70597	0.51098
6	Guamúchil Vol	1.043	38.747	0.57062	0.28975
7	Malpaso Qp	-0.666	2153.234	0.40782	0.25020
8	Malpaso Vol	0.558	1583.168	0.29264	0.13777
9	Los Filtros PMD	-0.616	43.005	0.13516	0.16115
10	Río Verde PMD	0.179	58.442	0.20926	0.09728
11	Tanquián PMD	-0.746	112.086	0.36652	0.22984

Simbología:

U = estadístico del Test de Wald-Wolfowitz

\bar{X} = media aritmética, en m^3/s , Mm^3 o milímetros

t_3 = cociente de momentos L de asimetría

t_4 = cociente de momentos L de curtosis

Ajuste de la distribución BEK

Tomando en cuenta que los parámetros de ajuste de las distribuciones Beta se obtienen por sustituciones sucesivas (j) o iteraciones, primero se ajustaron los modelos considerados de uso generalizado; es decir, las



distribuciones LP3, GVE y LOG a cada uno de los 11 registros procesados, para darles tal acepción.

Con lo anterior se dispone de los errores de ajuste (*EEA* y *EAM*) y de las predicciones relativas a los siete periodos de retorno analizados (25, 50, 100, 500, 1000, 5000 y 10000 años), obtenidas con las tres distribuciones citadas. Los valores de los errores de ajuste se utilizaron como magnitudes *a no superar* y para el caso de las predicciones, en los periodos de retorno menores a 100 años, *como valores a igualar*, con el ajuste por iteraciones de las distribuciones Beta.

Los parámetros de ajuste de la distribución BEK se estimaron con base en las ecuaciones 13 a 15, usando como valores iniciales de β_0 y θ_0 la media aritmética (Tabla 2) y un valor de 5.0, respectivamente. En cada iteración (j), se evaluaron los errores estándar de ajuste (ecuación 23) y absoluto medio (expresión 24), haciendo uso de las fórmulas 9 y 20, para obtener los valores estimados \hat{x}_i . La comparación entre errores obtenidos con la BEK y sus predicciones, permitió definir el número de iteraciones realizadas, las cuales se exponen en la Tabla 3, así como los parámetros de ajuste obtenidos. El número máximo de iteraciones se limitó a 500.

Tabla 3. Número de iteraciones (j) y valores de los parámetros de ajuste (β , a , θ) de la distribución BEK en los 11 registros procesados.

No.	Registro en:	j	β	a	θ
1	El Infiernillo <i>Qp</i>	274	2619.301	2.576858	2.538683
2	El Infiernillo <i>Vol</i>	3	2463.009	0.517784	2.692160

3	Huites <i>Qp</i>	27	1488.418	1.888118	2.045599
4	Huites <i>Vol</i>	3	923.356	0.553249	3.152180
5	Guamúchil <i>Qp</i>	37	304.088	2.013669	1.384388
6	Guamúchil <i>Vol</i>	4	43.366	0.397784	1.791456
7	Malpaso <i>Qp</i>	500	701.759	8.699795	2.733231
8	Malpaso <i>Vol</i>	2	1936.030	0.461082	3.178971
9	Los Filtros <i>PMD</i>	1	42.399	0.862399	5.390516
10	Río Verde <i>PMD</i>	1	57.438	0.800002	4.660408
11	Tanquián <i>PMD</i>	200	50.479	7.188239	3.506981

Ajuste de la distribución BEP

Un procedimiento idéntico al descrito se siguió para el ajuste de la BEP, pero ahora se utilizaron las ecuaciones 16 a 18, para sus parámetros de ajuste y las expresiones 12 y 20, para sus valores estimados \hat{x}_i y así poder evaluar los errores de ajuste (ecuaciones 23 y 24) y sus predicciones (fórmula 12). Los resultados de muestran en la Tabla 4.

Tabla 4. Número de iteraciones (j) y valores de los parámetros de ajuste (β , α , θ) de la distribución BEP en los 11 registros procesados.



No.	Registro en:	j	β	α	θ
1	El Infiernillo <i>Qp</i>	70	2919.052	0.360741	5.423902
2	El Infiernillo <i>Vol</i>	2	1572.835	1.001999	2.413874
3	Huites <i>Qp</i>	3	1920.472	0.762506	2.641226
4	Huites <i>Vol</i>	216	803.068	1.317579	2.335993
5	Guamúchil <i>Qp</i>	6	430.447	0.676817	1.973478
6	Guamúchil <i>Vol</i>	4	15.870	0.983551	1.417081
7	Malpaso <i>Qp</i>	9	1408.943	0.513508	5.187900
8	Malpaso <i>Vol</i>	150	1972.335	1.914897	1.882710
9	Los Filtros <i>PMD</i>	1	42.774	1.178656	4.616619
10	Río Verde <i>PMD</i>	2	55.730	1.081484	4.527451
11	Tanquián <i>PMD</i>	16	76.011	0.345564	8.324678

Estrategia de selección y resultados

El planteamiento general para la selección de las predicciones adoptadas, en cada registro procesado, se basó en los valores mínimos de los errores *EEA* y *EAM*, así como en las magnitudes de las predicciones extremas de

periodos de retorno (Tr) 1000, 5000 y 10000 años, para buscar unas estimaciones *representativas* de las cinco obtenidas en cada Tr .

Exclusivamente para el *registro 3* (Huites Qp), la distribución LP3 condujo a los menores errores de ajuste y también a las predicciones adoptadas, dada la extraordinaria similitud que mostraron todas las estimaciones.

Es importante destacar que la distribución LP3, condujo a los menores errores de ajuste en los *registros 4, 5, 6 y 7*. En el primero y el último, tal distribución no fue seleccionada por conducir a predicciones muy bajas; por el contrario, en los registros 5 y 6. Para estos cuatro registros sus CE fueron: 1.015, 1.831, 1.619 y 1.226; evidentemente, en los registros 5 y 6 no resulta aplicable el método de ajuste por momentos en el dominio logarítmico. En el resto de registros procesados, el CE no resultó menor de 1.00.

En el registro 1, la distribución BEP reporta los errores de ajuste más bajos, pero todas sus predicciones se consideran elevadas. Por lo anterior, se adoptó la distribución GVE, por tener valores mucho menores de los errores de ajuste que el modelo LOG. *En el registro 2*, las distribuciones LP3 y GVE condujeron a los menores errores de ajuste; se adopta la primera por sus predicciones más severas.

En el registro 4, la distribución GVE aportó los siguientes valores mínimos de los errores de ajuste, pero sus predicciones fueron reducidas y por ello, se adoptó la distribución BEK con errores bajos y predicciones representativas, incluso similares a las de la LOG.

En los registros 5, 6 y 7, la distribución BEK adoptada, reporta los siguientes valores más bajos de los errores de ajuste y sus predicciones se aceptan representativas, debido a la gran semejanza que muestran con las de los modelos GVE y LOG.

En los registros 8, 9 y 10, se adoptaron las distribuciones GVE, BEP y GVE, que reportaron los errores de ajuste mínimos. Respecto a sus predicciones, pueden ser consideradas levemente escasas las de la GVE y algo elevadas las de la BEP, al compararlas con las del modelo LOG.

En el registro 11, las distribuciones LP3 y BEP condujeron a los menores errores de ajuste; se adopta la segunda por sus predicciones más severas.

Tabla 5. Contraste de indicadores de bondad de ajuste y predicciones entre las distribuciones BEK y BEP y las tres de uso generalizado (LP3, GVE y LOG), en los 11 registros de datos hidrológicos procesados.

NR	FDP	EEA	EAM	Periodos de retorno, en años						
				25	50	100	500	1000	5000	10000
1	BEK	1575	622.3	13364	17659	23267	43955	57770	108934	143049
1	BEP	1072	436.1	15125	21556	30719	69927	99654	226845	323281
1	LP3	1078	458.2	15257	20645	27643	53011	69619	129550	168642
1	GVE	1727	652.3	14499	19961	27423	57155	78386	163182	223718
1	LOG	1829	707.3	14095	19509	27104	58949	82743	182964	257909
2	BEK	416.0	276.8	6235	8158	10612	19379	25084	45626	59015
2	BEP	552.8	352.7	5851	7861	10514	20521	27343	53207	70868

2	LP3	468.2	233.3	7055	9394	12240	21356	26650	43240	52705
2	GVE	407.2	257.2	6542	8552	11017	19156	24074	40360	50200
2	LOG	460.5	299.0	6373	8456	11138	20835	27199	50330	65532
3	BEK	990.8	482.7	9648	13644	19219	42339	59437	130555	183116
3	BEP	1023.0	505.6	9443	13367	18884	42024	59292	131850	186018
3	LP3	938.4	411.3	10856	15381	21438	44405	59970	117987	156804
3	GVE	974.8	453.0	9994	14008	19464	41019	56297	116801	159676
3	LOG	1056.0	492.8	9696	13680	19246	42424	59637	131640	185161
4	BEK	88.3	61.1	2086	2624	3284	5491	6845	11409	14211
4	BEP	94.3	59.8	2198	2799	3539	6025	7559	12772	16003
4	LP3	64.5	45.4	2109	2501	2899	3854	4278	5287	5732
4	GVE	86.0	58.1	2155	2672	3265	5005	5949	8732	10242
4	LOG	101.3	70.8	2111	2670	3354	5622	6999	11593	14387
5	BEK	1552	620.8	5044	8416	13961	44842	74022	236881	390495
5	BEP	1639	640.5	4771	8039	13522	45140	75850	253085	425248
5	LP3	1152	493.1	7779	14739	27483	112419	204067	803624	1444814
5	GVE	1693	716.1	5864	9679	15832	48841	79098	241647	390542
5	LOG	1820	729.2	5422	8962	14714	46057	75171	234269	382083
6	BEK	18.1	11.5	150	226	336	831	1224	3009	4429
6	BEP	23.3	12.3	156	259	429	1369	2253	7152	11761
6	LP3	16.6	11.2	199	346	584	1833	2933	8432	13126
6	GVE	22.3	13.4	148	224	334	825	1208	2913	4246
6	LOG	23.7	13.9	142	217	326	829	1235	3103	4611
7	BEK	213.8	106.4	4986	6453	8332	15039	19381	34957	45005

7	BEP	285.4	152.3	4715	6118	7936	14521	18837	34465	44707
7	LP3	185.3	93.6	5127	6550	8294	14040	17498	28887	35735
7	GVE	315.6	139.0	5013	6442	8243	14475	18403	32031	40620
7	LOG	346.7	160.2	4896	6357	8279	15478	20356	38726	51186
8	BEK	279.9	163.8	4093	5143	6428	10708	13324	22114	27498
8	BEP	280.2	151.4	4317	5423	6728	10826	13206	20810	25270
8	LP3	277.7	141.2	4482	5573	6767	9968	11545	15709	17733
8	GVE	226.4	140.3	4253	5265	6406	9676	11407	16383	19019
8	LOG	262.4	165.7	4168	5272	6610	10965	13567	22086	27182
9	BEK	2.5	1.6	74	85	97	131	149	200	228
9	BEP	2.5	1.5	76	87	99	134	152	205	232
9	LP3	2.9	1.7	74	82	90	107	114	131	139
9	GVE	3.7	1.5	74	81	88	104	110	123	129
9	LOG	3.3	1.5	74	83	93	121	135	172	191
10	BEK	4.4	2.9	108	126	147	208	241	341	395
10	BEP	4.1	3.2	106	123	142	198	228	317	366
10	LP3	3.4	2.2	110	126	142	184	204	255	279
10	GVE	3.0	2.2	110	125	141	181	199	244	265
10	LOG	3.8	2.9	109	127	147	208	241	339	392
11	BEK	13.0	7.2	220	269	329	521	635	1005	1224
11	BEP	12.2	6.4	233	296	377	659	839	1468	1868
11	LP3	11.9	6.7	231	283	344	535	644	985	1180
11	GVE	15.6	7.6	228	280	344	551	673	1071	1307
11	LOG	16.3	7.9	223	278	348	598	759	1337	1712

Simbología:

NR = número de registro, según Tablas 3 ó 4

FDP = función de distribución de probabilidades probada

EEA = error estándar de ajuste, en m^3/s , Mm^3 ó mm , según datos

EAM = error absoluto medio, en m^3/s , Mm^3 ó mm , según datos

En las Figuras 1 y 2 se muestran los gráficos de diagnóstico $Q-Q$ de los dos mejores ajustes logrados con las distribuciones BEK en el registro 7 y BEP en el registro 9.

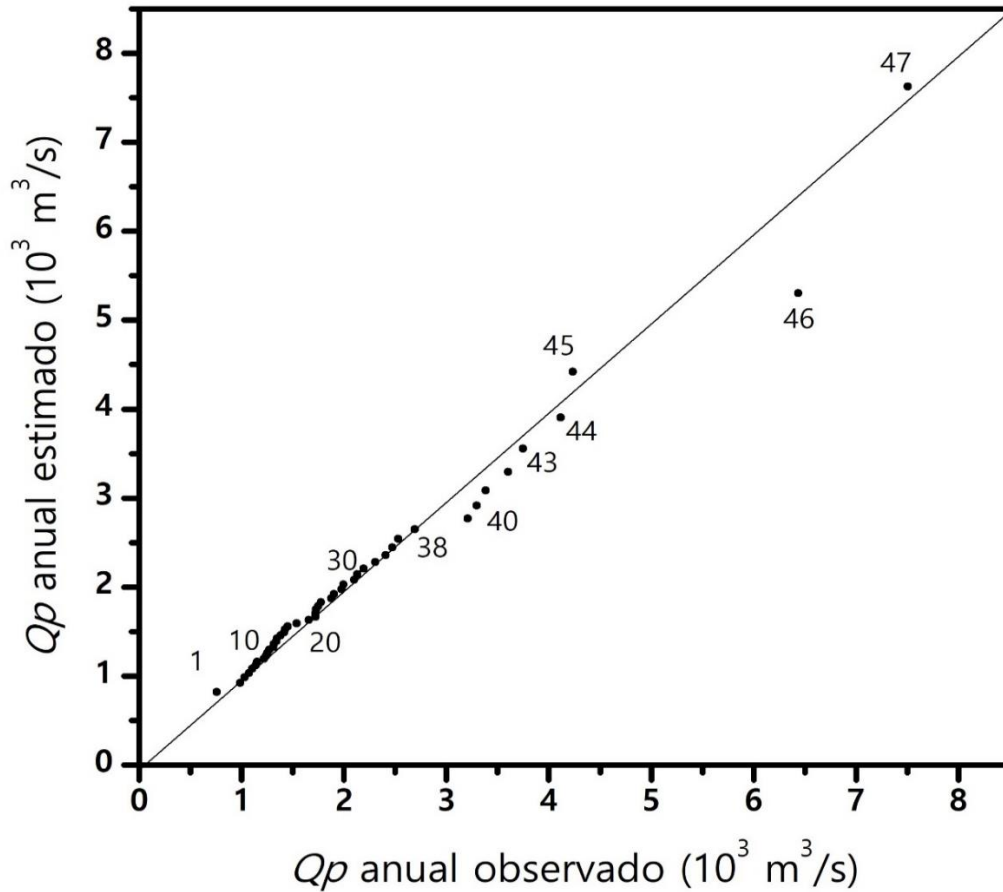


Figura 1. Gráfico Q–Q del registro 7 de Q_p anuales en la presa Malpaso obtenido con la distribución BEK.

En la Figura 1, hasta el dato número 38 el modelo BEK reproduce de manera exacta los datos de la muestra y a partir de ahí, subestima los seis siguientes valores y el dato 46; por último, sobreestima ligeramente el 45 y el último. En cambio, en la Figura 2, únicamente se detecta falta de exactitud en los seis últimos datos, pero no severa.

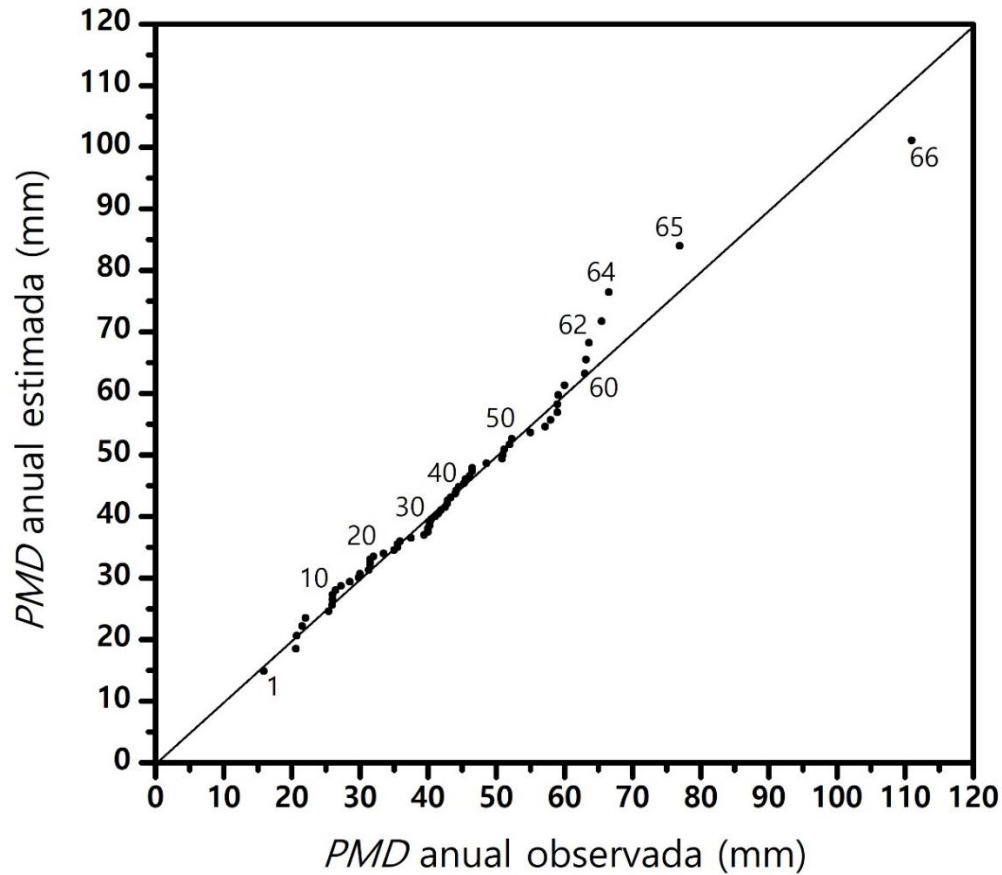


Figura 2. Gráfico Q-Q del registro 9 de *PMD* anual en la estación Los Filtros obtenido con la distribución BEP.

Conclusiones

Las distribuciones BEK y BEP contrastadas tienen las dos ventajas siguientes: (1) muestran gran versatilidad para representar registros de eventos hidrológicos extremos debido a su cola derecha densa y al uso de dos parámetros de forma y (2) su método de ajuste por máxima verosimilitud es eficiente, simple y sin complicaciones de cómputo. Por lo anterior, se recomienda su inclusión rutinaria en los análisis de frecuencias de eventos hidrológicos extremos.

En los once registros procesados, las predicciones de los tres primeros periodos de retorno ($Tr < 100$ años), son bastante similares. En general, la FDP adoptada por tener menor valor del EEA y EAM , conduce a predicciones representativas en los cuatro últimos periodos de retorno altos ($Tr \geq 500$ años). Las dos condiciones anteriores originan confianza en todas las *predicciones* calculadas y en las adoptadas.

Las observaciones deducidas de la Tabla 5 de resultados (errores y predicciones), permiten sugerir la aplicación de las distribuciones BEK y BEP de manera rutinaria para *complementar* a las de aplicación de uso generalizado (LP3, GVE y LOG); sobre todo en la selección de las predicciones por adoptar en los tres periodos de retorno extremos de 1000, 5000 y 10000 años. Lo anterior, fue verificado con la distribución BEK que se adoptó en dos registros de Qp y dos de Vol (del 4 al 7) de los ocho procesados y el modelo BEP se adoptó en dos registros de PMD (9 y 11) de los tres analizados.

Se debe destacar que las Conclusiones anteriores están basadas en los resultados de los 11 registros procesados y por ello, pueden dar la impresión de que las distribuciones BEK y BEP son mejores que las de

aplicación generalizada, pero no es así; sino que únicamente deben ser consideradas opciones prácticas y factibles para los análisis de frecuencias de datos hidrológicos extremos.

Agradecimientos

Se agradecen las observaciones y correcciones sugeridas por los árbitros anónimos C, D, E y el Editor. Tales correcciones permitieron depurar el texto, para adecuarlo al contexto hidrológico mexicano y ayudaron a justificar la propuesta de aplicación rutinaria de las distribuciones Beta-kappa y Beta-Pareto, en los análisis de frecuencias de datos extremos: crecientes (Q_p, Vol) y lluvias máximas diarias.

Referencias

- Aldama, A. A., Ramírez, A. I., Aparicio, J., Mejía-Zermeño, R. & Ortega-Gil, G. E. (2006). *Seguridad Hidrológica de las Presas en México*. Jiutepec, Morelos: Instituto Mexicano de Tecnología del Agua. 188 páginas.
- Benson, M. A. (1962). Plotting positions and economics of engineering planning. *Journal of Hydraulics Division*, 88(6), 57–71. DOI: 10.1061/jYCEAj.0001293
- Bobée, B. & Ashkar, F. (1991). *The Gamma Family and derived distributions applied in Hydrology*. Chapter 1: Data requirements for hydrologic frequency analysis (pp. 1–12). Littleton, Colorado, U.S.A.: Water Resources Publications. 203 p.

- Campos–Aranda, D. F. (1998). Distribución de probabilidades β - p : Descripción y aplicación en hidrología superficial. *XV Congreso Nacional de Hidráulica (AMH)*, páginas 965–971. 13 al 16 de octubre. Oaxaca, Oaxaca, México.
- Campos–Aranda, D. F. (2002). Contraste de seis métodos de ajuste de la distribución Log–Pearson tipo III en 31 registros históricos de eventos máximos anuales. *Ingeniería Hidráulica en México*, XVII(2), 77–97.
- Campos–Aranda, D. F. (2019). Mejores FDP en 19 series amplias de PMD anual del estado de San Luis Potosí, México. *Tecnología y ciencias del agua*, 10(5), 34–74. DOI: 10.24850/j-tyca-2019-05-02
- Coles, S. (2001). *An introduction to Statistical Modeling of Extreme Values*. Theme 2.6.7: Model diagnostics (pp. 36–44). London, England: Springer–Verlag. 208 p.
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250. DOI:10.5194/gmd-7-1247-2014
- Domínguez M., R. & Arganis J., M. L. (2012). Validation of method to estimate design discharge flow for dam spillways with large regulating capacity. *Hydrological Sciences Journal*, 57(3), 460–478. DOI:10.1080/02626667.2012.665993
- Hosking, J. R. M. & J. R. Wallis. (1997). *Regional Frequency Analysis. An Approach Based on L-moments*. Appendix: L-moments for some

specific distributions, pp. 191–209. Cambridge, England: Cambridge University Press. 224 p.

Kite, G. W. (1977). *Frequency and Risk Analyses in Hydrology*. Chapter 12: Comparison of frequency distributions (pp. 156–168). Fort Collins, Colorado, U.S.A.: Water Resources Publications. 224 p.

Mason, S. J., Waylen, P. R., Mimmack, G. M., Rajaratnam, B., & Harrison, J.M. (1999). Changes in extreme rainfall events in South Africa. *Climatic Change*, 41(2), 249–257. DOI: 10.1023/A:1005450924499

Meylan, P., Favre, A. C. & Musy, A. (2012). *Predictive Hydrology. A Frequency Analysis Approach*. Chapter 3: Selecting and checking data series (pp. 29–70). Boca Raton, Florida, U.S.A.: CRC Press. 212 p.

Mielke, P. W. (1973): Another family of distributions for describing and analyzing precipitation data. *Journal of Applied Meteorology*, 12(2), 275–280. DOI: 10.1175/1520-0450(1973)012<0275:AFODFD>2.0.CO;2

Mielke, P. W., & Johnson, E. S. (1974). Some generalized Beta distributions of the second kind having desirable application features in hydrology and meteorology. *Water Resources Research*, 10(2), 223–226. DOI: 10.1029/WR010i002p00223

Murshed, M. S., Kim, S., & Pak, J. S. (2011). Beta- κ distribution and its application to hydrologic events. *Stochastic Environmental Research and Risk Assessment*, 25(7), 897–911. DOI 10.1007/s00477-011-0494-4

- Nguyen, T. H., El Outayek, S., Lim, S. H., & Nguyen, T. V. T. (2017). A systematic approach to selecting the best probability models for annual maximum rainfalls – A case study using data in Ontario (Canada). *Journal of Hydrology*, 553, 49–58. DOI: 10.1016/j.jhydrol.2017.07.052
- Oberhettinger, F. (1972). Hypergeometric Functions. Chapter 15 (pp 555–566). In M. Abramowitz & I. A. Stegun (Eds.). *Handbook of Mathematical Functions*. New York, U.S.A.: Dover Publications. Ninth printing. 1046 p.
- Öztekin, T. (2007). Wakeby distribution for representing annual extreme and partial duration rainfall series. *Meteorological Applications*, 14(4), 381–387. DOI:10.1002/met.37
- Rao, A. R. & Hamed, K. H. (2000). *Flood Frequency Analysis*. Theme 1.8: Tests on hydrologic data (pp. 12–21) and Chapters 7, 8 and 9 (pp. 207–321). Boca Raton, Florida, U.S.A.: CRC Press. 350 p.
- Stedinger, J. R. (2017). Flood Frequency Analysis. Chapter 76 (pp. 76.1–76.8). In V. P. Singh (Ed.). *Handbook of Applied Hydrology*. New York, U.S.A.: McGraw–Hill Education. Second Edition.
- Strupczewski, W. G., Markiewicz, I., Kochanek, K., & Singh, V. P. (2008). Short walk into two–shape–parameter flood frequency distributions. *Hydrology and Hydraulics*, V. P Singh (ed.), pp. 669–716. Highlands Ranch, Colorado, U.S.A.: Water Resources Publications. 1080 p.
- Teegavarapu, R. S. V., Salas, J. D., & Stedinger, J. R. (2019). *Statistical Analysis of Hydrologic Variables*. Chapter 1: Introduction (pp. 1–4). Reston, Virginia, U.S.A.: American Society of Civil Engineers. 548 p.

- WRC (1977). *Guidelines for Determining Flood Flow Frequency*. Bulletin # 17A of the Hydrology Committee. Washington, D. C., U.S.A.: Water Resources Council. Revised edition.
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82. DOI: 10.3354/cr030079
- Wilks, D. S. (1993). Comparison of three-parameter probability distributions for representing annual extreme and partial duration precipitation series. *Water Resources Research*, 29(10), 3543–3549. DOI: 10.1029/93WR01710
- Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Theme 4.5: Qualitative assessments of the goodness fit (pp. 112–116). San Diego, U.S.A.: Academic Press (Elsevier). Third edition. 676 p.