DOI: 10.24850/j-tyca-14-01-05

Articles

Comparative Frequency Analysis with L moments between the Kappa distribution and six of general application

Análisis de frecuencias comparativo con momentos L entre la distribución Kappa y seis de aplicación generalizada

Daniel Francisco Campos-Aranda[1], ORCID: https://orcid.org/0000-0002-9876-3967

[1]Retired professor of the Autonomous University of San Luis Potosi, San Luis Potosí, Mexico, campos_aranda@hotmail.com

Corresponding author: Daniel Francisco Campos-Aranda, campos_aranda@hotmail.com

Abstract

Frequency analysis (FA) allows estimating magnitudes for the maximum annual data of floods and daily rains associated with low exceedance probabilities. Such estimates or *predictions* allow the hydrological design of hydraulic works of exploitation or protection. The FA comprehends five stages: (1) verification of the randomness of the data; (2) adoption of a *probability distribution function* (PDF); (3) fitting the PDF; (4) evaluation of the goodness of fit, and (5) selection of the results. In this study, the theoretical bases of the Kappa distribution of four parameters (*u*, *a*, *k*, *h*), obtained through the L-moment method, which is described in detail, are exposed. The Kappa distribution, when its second shape parameter *h* takes values of -1, 0, and 1, reproduces the Generalized Logistics (GLO), Generalized Extreme Value (GEV), and Generalized Pareto (GPA) distributions. Three *joint* records of annual floods of peak flow (*Qp*) and runoff volume (*Vol*) were processed, and two yearly records of *Qp* and three of maximum annual daily precipitation. Five distributions were used to fit each of the 11 processed records: Kappa, the distribution reproduced according to the value of *h* (GLO, GEV o GPA); Log-Pearson type III; Log-Normal, and Wakeby. The statistical quality of each fit was quantified with the standard error of fit and the mean absolute error. The Conclusions highlight the similarity of the results (errors and predictions) in the 11 records processed and suggest the systematic application of the Kappa distribution, to complement those prescribed by government agencies and those of generalized use.

Keywords: Kappa distribution, L moments and ratios, standard error of fit, mean absolute error, *Q*-*Q* graphics, predictions.

Resumen

El *análisis de frecuencias* (AF) permite estimar magnitudes de los datos máximos anuales de crecientes y lluvias diarias asociadas con bajas probabilidades de ser excedidas. Tales estimaciones o *predicciones* permiten el diseño hidrológico de las obras hidráulicas de aprovechamiento o protección. El AF comprende cinco etapas: (1) verificación de la aleatoriedad de los datos; (2) adopción de una *función de distribución de probabilidades* (FDP); (3) ajuste de la FDP; (4) evaluación del ajuste logrado, y (5) selección de los resultados. En este estudio se exponen las bases teóricas de la distribución Kappa de cuatro parámetros de ajuste (u, a, k, h) obtenidos a través del método de los momentos L, que se describe con detalle. La distribución Kappa, cuando su segundo parámetro de forma h toma valores de -1, 0 y 1 reproduce a las distribuciones Logística Generalizada (LOG), General de Valores Extremos (GVE) y Pareto Generalizada (PAG). Se procesaron tres registros *conjuntos* de crecientes anuales de gasto pico (*Qp*) y volumen escurrido (*Vol*), dos registros anuales de *Qp* y tres de precipitación máxima diaria anual. A cada uno de los 11 registros procesados se les ajustaron cinco distribuciones: Kappa, la que reproduce según el valor de h (LOG, GVE o PAG); Log-Pearson tipo III; Log-Normal, y Wakeby. La calidad estadística de cada ajuste se cuantificó con el error estándar de ajuste y el error absoluto medio. Las conclusiones destacan la similitud de los resultados (errores y predicciones), en los 11 registros procesados y sugieren la aplicación sistemática de la distribución Kappa, para complementar las de aplicación bajo precepto y las de uso generalizado.

Palabras clave: distribución Kappa, momentos y cocientes L, error estándar de ajuste, error absoluto medio, gráficos *Q*-*Q*, predicciones.

Received: 12/22/2020

Accepted: 09/18/2021

Introduction

Frequency Analysis (FA) is a procedure that seeks to define the magnitudes of a random variable associated with low probabilities of being exceeded; Such estimates are called *Predictions,* and the hydrological dimensioning of all hydraulic works, such as reservoirs, retaining dams, bridges, urban drainage, etc., is based on them. The most common random variables processed with FA are the so-called *extreme hydrological data*, such as floods, levels in rivers and seas, winds, and precipitation. The maximum storms that occur in the basins originate from floods, whose basic annual data are the peak flow and the runoff volume. Regarding storms and due to the scarcity of pluviographs, it is common to process the data from the pluviometers through the maximum annual daily rainfall (Rao & Hamed, 2000; Meylan, Favre, & Musy, 2012; Stedinger, 2017; Johnson & Sharma, 2017).

FA consists of the following five stages: (1) verification of the randomness of the data or available sample; (2) adoption of a probabilistic model or *probability distribution function* (PDF), with which the desired predictions are made; (3) application of a method for estimating the fit parameters of the tested PDF; (4) calculation of the indicators of the quality of the adjustment achieved, between the PDF and the sample and (5) selection of the results (Hosking & Wallis, 1997).

This study's first stage was carried out based on the Wald–Wolfowitz Test. For the second stage, the following seven PDFs were applied and contrasted: Kappa of four fit parameters, the three distributions with applications suggested under precept or norm: Log–Pearson type III (LP3), General Extreme Values (GVE, by its acronym in Spanish) and Generalized Logistics (LOG, idem) and three frequently used: the Log–Normal (LN3), the Generalized Pareto (PAG, idem) and the Wakeby (WAK), all with three fit parameters, except the last, with five (Gómez, Aparicio, & Patiño, 2010; Nguyen, El-Outayek, Lim, & Nguyen, 2017; Campos-Aranda, 2019).

Regarding stage three, the method of moments L was adopted, which has been established as a simple, accurate, and robust procedure for estimating the fit parameters of the PDF used in hydrological FA (Hosking & Wallis, 1997). Another advantage of the L–moments method has been highlighted by Kjeldsen, Ahn, and Prosdocimi (2017), about the use of the L–quotient diagram, to select an appropriate PDF for the available data. Finally, about the fourth stage of FA, the two most widely used indicators were applied: the standard error of fit and the mean absolute error (Kite, 1977; Willmott & Matsuura, 2005; Chai & Draxler, 2014).

The FAs have two implicit challenges, the first seek distributions that best represent the records of extreme hydrological data, which are becoming more extensive, and the second, outside the scope of this work, is related to the processing of non–random data: dependent or non–stationary (Khaliq, Ouarda, Ondo, Gachon, & Bobée, 2006; Meylan *et al*., 2012; Katz, 2013).

It is important to note that the use of PDFs with four or more fit parameters, such as the Kappa and Wakeby models, has proliferated due to their versatility and ability to represent data samples where the PDF origin is unknown (Singh & Deng, 2003; Asquith, 2011).

On the other hand, Hosking and Wallis (1993) were the first to use the Kappa distribution to generate synthetic data to find the best PDF in a regional FA of floods. Parida (1999) was the first to use the Kappa distribution in the FA of the rain that occurred during the monsoon in India in the months from June to September.

The basic *objectives* of this study were the following three: (1) to present a summary of the theory behind the Kappa distribution; (2) to describe in detail the method of moments L for the estimation of its four fit parameters and (3) perform a goodness-of-fit contrast and predictions, between the Kappa distribution and six of general application, three of an

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

application under precept (LP3, GVE, LOG) and three commonly used (LN3, PAG, and WAK).

Methods and materials

The genesis of the Kappa distribution

Hosking (1994) explained the development of the Kappa distribution of 4 fit parameters, starting from the following two transformations, which follow three PDFs currently used in the FA of floods, these are:

$$X = u + \frac{\alpha\left(1 - e^{-k \cdot Y}\right)}{k} \quad \text{when } k \neq 0 \qquad (1)$$

$$X = u + \alpha \cdot Y \quad \text{when } k = 0 \qquad (2)$$

where X and Y are real random variables and u, a and k are real fit parameters of location, scale, and shape. If Y comes from an exponential distribution, with a PDF of the form:

$$F(y) = 1 - e^{-y} \quad \text{with } y \geq 0 \tag{3}$$

then X will have a Generalized Pareto (PAG) PDF with three fit parameters. If Y comes from a Gumbel distribution (extreme values type I or double exponential), with a PDF of the form:

$$F(y) = \exp(-e^{-y}) \tag{4}$$

then X will have a Generalized extreme values (GVE) PDF with three Fit parameters. Finally, if Y comes from a Logistics distribution, with FDP of the form:

$$F(y) = \frac{1}{1+e^{-y}} \tag{5}$$

then X will have a Generalized Logistics (LOG) PDF with three adjustment parameters. On the other hand, equations (3), (4), and (5) can be obtained as special cases with $h = 1$, $h = 0$, and $h = -1$ from the following PDFs:

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

$$F(y) = (1 - h \cdot e^{-y})^{1/h} \quad \text{for } y \geq \ln h \text{ if } h \neq 0 \tag{6}$$

$$F(y) = \exp(-e^{-y}) \quad \text{when } h = 0 \tag{7}$$

h is a second form parameter. The Kappa distribution of four fit parameters (*u*,*α*,*k*,*h*), which includes the PAG, GVE, and LOG distributions as special cases, is obtained for the random variable *X* by applying transformations 1 and 2 to the random variable *Y*, where PDF is defined by equations (6) and (7). The PDF of the Kappa distribution is as follows (Hosking, 1994):

$$F(x, \theta) = \left\{1 - h[1 - k(x - u)/\alpha]_{+}^{1/k}\right\}^{1/h} \quad \text{when } k \neq 0, h \neq 0 \tag{8}$$

In this equation, the + sign of the rectangular parentheses indicates positivity for its internal expression (Dupuis & Winchester, 2001). Coles (2001) clarifies, for the GVE distribution, that any combination of fit parameters (θ) that violates the positivity condition above implies that at least one of the observed points (x) is beyond the final points of the distribution and then the likelihood function is zero, and its logarithmic version is $-\infty$.

The above equation includes distributions of two parameters obtained as expressions or limit forms when *k* or *h* tend to zero. Under such a condition, the probability density functions [$f(x)$] and quantiles or inverse solution of the Kappa distribution are (Hosking, 1994; Asquith, 2011):

$$f(x) = \frac{1}{\alpha}[1 - k(x-u)/\alpha]^{\left(\frac{1}{k}\right)-1}[F(x)]^{1-k} \quad (9)$$

$$x(F) = u + \frac{\alpha}{k}\left[1 - \left(\frac{1-F^h}{h}\right)^k\right] \quad (10)$$

In both expressions, the fit parameters: u, α, k and h are real numbers; furthermore, $\alpha > 0$. In Equation (10), F is the probability of non–exceedance, and the ratio $(1–F^h)/h$ defines the limits of the variable x, already known in the LOG, GVE, and PAG distributions (Hosking & Wallis, 1997; Stedinger, 2017).

Hosking (1994) establishes four conditions that define the space of the fit parameters (Figure 1) to ensure the existence of moments L (conditions *a* and *b*) and the uniqueness of the four fit parameters (conditions *c* and *d*), given the first four moments L, such conditions are (a) $k > -1$; (b) if $h < 0$, then $h \cdot k > -1$; (c) $h > -1$ and (d) $k + 0.725 \cdot h > -1$.

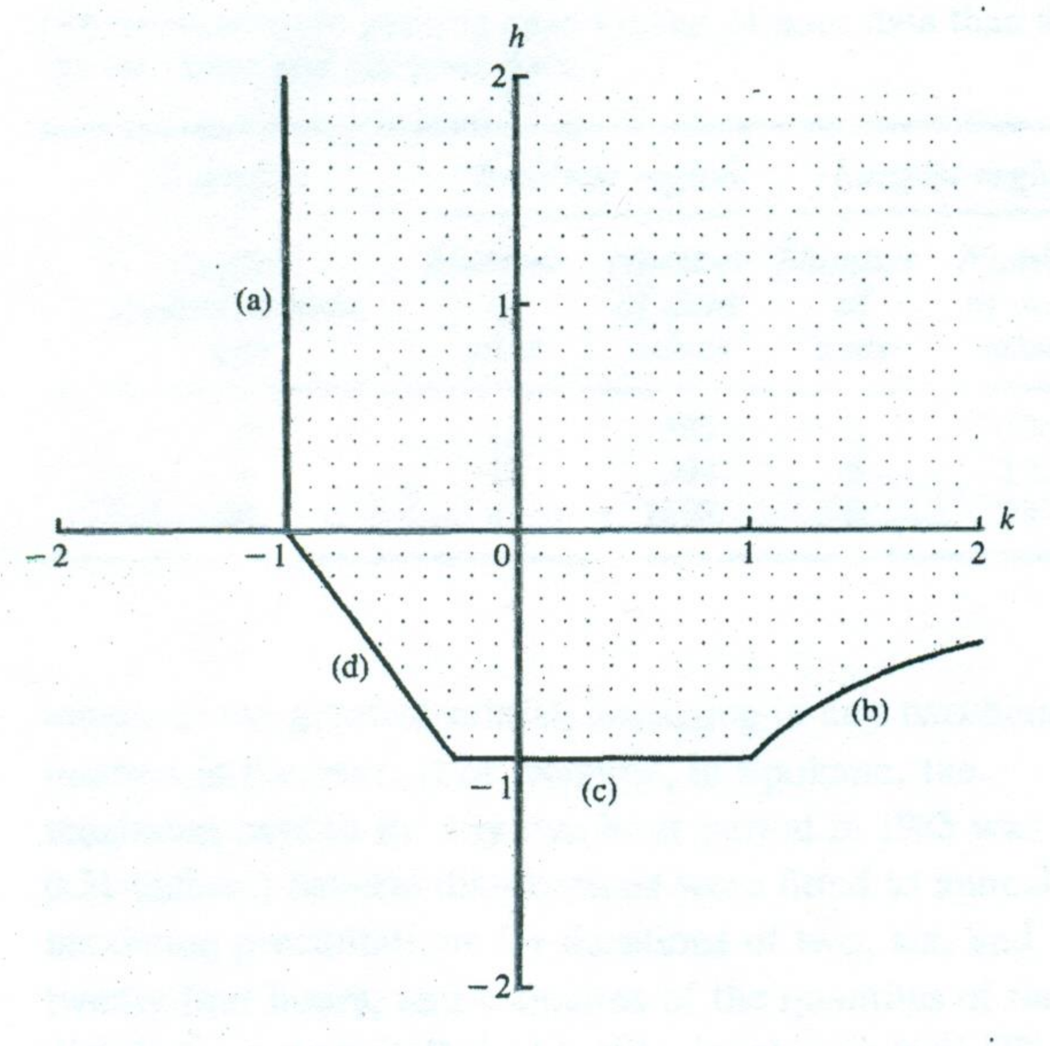

Figure 1. Space (area of points) of the fit parameters *k* and *h* of the Kappa distribution ensures the occurrence of the first four moments L and the uniqueness of its fit parameters (Hosking, 1994).

Sample moments and L ratios

The L moments are linear combinations of the weighted probability moments (b_r) developed by Greenwood, Landwehr, Matalas, and Wallis (1979); therefore, they are robust to the scattered values of the sample. Its calculation begins by ordering the series of annual hydrological data (x_i) from smallest to largest ($x_1 \leq x_2 \leq \cdots \leq x_n$), and then probability–weighted moments are obtained using the following expression (Hosking & Wallis, 1997; Rao & Hamed, 2000; Asquith, 2011; Stedinger, 2017):

$$b_r = \frac{1}{n}\sum_{i=r+1}^{n} \frac{(i-1)(i-2)\cdots(i-r)}{(n-1)(n-2)\cdots(n-r)} x_i \quad (11)$$

In the previous expression, the order number r varies from 0 to 3, and n is the data number of the annual series. It follows that b_0 is equal to the arithmetic mean. The moments L of the sample (l) and their respective quotients (t) of similarity with the coefficients of variation, asymmetry, and kurtosis are:

$$l_1 = b_0 \quad (12)$$

$$l_2 = 2 \cdot b_1 - b_0 \quad (13)$$

$$l_3 = 6 \cdot b_2 - 6 \cdot b_1 + b_0 \quad (14)$$

$$l_4 = 20 \cdot b_3 - 30 \cdot b_2 + 12 \cdot b_1 - b_0 \quad (15)$$

$$t_2 = l_2 / l_1 \quad (16)$$

$$t_3 = l_3 / l_2 \quad (17)$$

$$t_4 = l_4 / l_2 \quad (18)$$

The moment ratio diagram L has t_3 on the abscissa axis and t_4 on the ordinate axis. The PDFs of three fit parameters are curved lines, and the PDFs of two fit parameters are points.

In Figure 2, the domain area of the Kappa distribution is shown in the moment ratio diagram L; the points L, G, and E correspond to the Logistic, Gumbel, and Exponential distributions; its lower extreme curve is the limit of the values of t_4 as a function of t_3 (Hosking & Wallis, 1997).

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

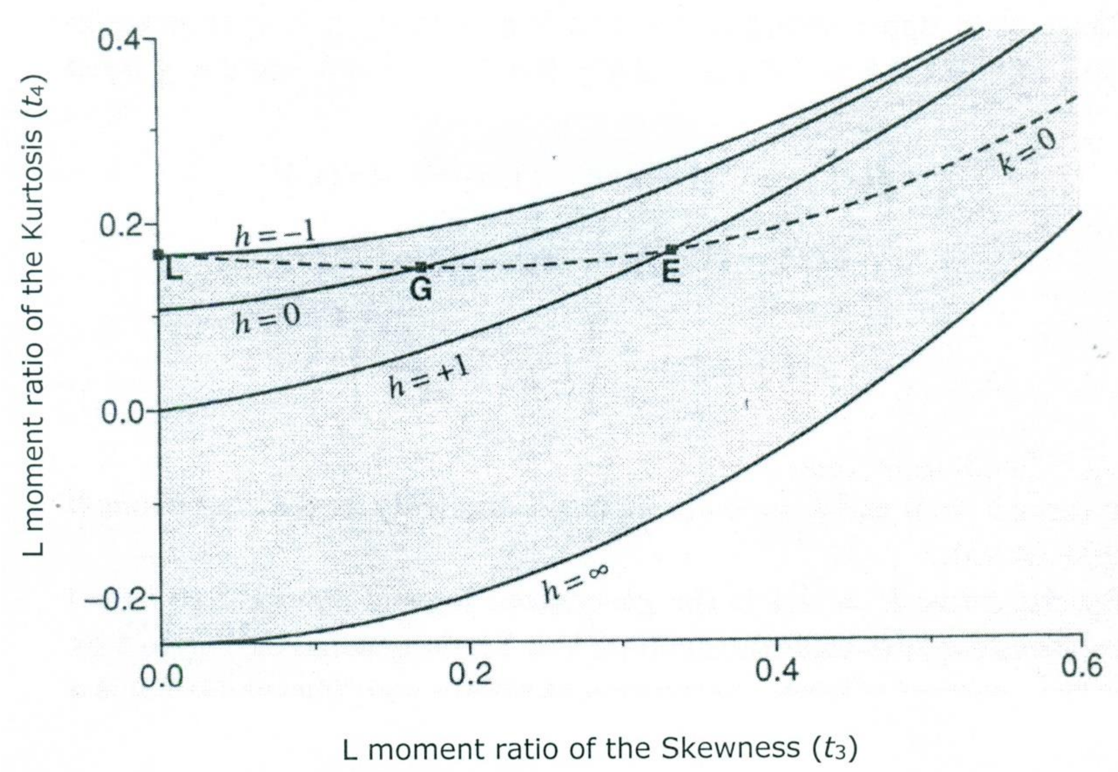

Figure 2. The Kappa distribution domain area is in the L-moment ratio diagram (Hosking & Wallis, 1997).

Moments and ratios L of the Kappa distribution

Hosking and Wallis (1997) expose the equations of the first two moments L, and of the ratios $t3$ and t_4, of the Kappa distribution; these are:

$$l_i = u + \alpha(1 - g_1)/k \tag{19}$$

$$l_2 = \alpha(g_1 - g_2)/k \tag{20}$$

$$t_3 = (-g_1 + 3g_2 - 2g_3)/(g_1 - g_2) \tag{21}$$

$$t_4 = (-g_1 + 6g_2 - 10g_3 + 5g_4)/(g_1 - g_2) \tag{22}$$

in which:

$$g_r = \frac{r\Gamma(1+k)\cdot\Gamma(-k-r/h)}{(-h)^{1+k}\Gamma(1-r/h)} \quad \text{when } h < 0 \tag{23}$$

$$g_r = \frac{r\Gamma(1+k)\cdot\Gamma(r/h)}{h^{1+k}\Gamma(1+k+r/h)} \quad \text{when } h > 0 \tag{24}$$

For the estimation of the Gamma function $\Gamma(\omega)$, the Stirling formula (Davis, 1972) was used:

$$\Gamma(\omega) \cong e^{-\omega} \cdot \omega^{\omega - \frac{1}{2}} \cdot (2\pi)^{1/2} \cdot F1 \tag{25}$$

being:

$$F1 = \left(1 + \frac{1}{12\cdot\omega} + \frac{1}{288\cdot\omega^2} - \frac{139}{51840\cdot\omega^3} - \frac{571}{2488320\cdot\omega^4} + \cdots\right)$$

Estimation of fit parameters

Kjeldsen *et al.* (2017) suggest the following procedure: defining a fixed and approximate value of h, based on the ratios t_3 and t_4 of the data record, in the moment ratio diagram L. When points are obtained above the curve of the LOG distribution implies that $h<-1$, and then h will take its limit value of -1 (condition *c* in Figure 1), as proposed by Hosking and Wallis (1993). The foregoing is equivalent to adjusting the LOG distribution, which proceeds when t3 and t4 define points above its curve in the moment ratio L (Campos–Aranda, 2013).

To define the value of the form parameter h, when it is located between the curves of the LOG and GVE distributions, with a negative value, or between the curve of the GVE and the PAG, with a positive value, the respective t_4 of each curve is calculated for the data t_3 of the sample, with the following equations (Hosking & Wallis, 1997):

$$t_4^{\text{LOG}} = 0.16667 + 0.83333 \cdot t_3^2 \qquad (26)$$

$$t_4^{\text{GVE}} = 0.10701 + 0.1109 \cdot t_3 + 0.84838 \cdot t_3^2 - 0.06669 \cdot t_3^3 + SUM1$$

$$SUM1 = 0.00567 \cdot t_3^4 - 0.04208 \cdot t_3^5 + 0.03763 \cdot t_3^6 \qquad (27)$$

$$t_4^{\text{PAG}} = 0.20196 \cdot t_3 + 0.95924 \cdot t_3^2 - 0.20096 \cdot t_3^3 + 0.04061 \cdot t_3^4 \qquad (28)$$

As already stated, if t_4 is more significant than t_4^{LOG}, then h=−1. If t_4 is greater than t_4^{GVE}, h will be negative, and its value is obtained by linear interpolation, with the expression:

$$h = -(t_4 - t_4^{\text{GVE}})/(t_4^{\text{LOG}} - t_4^{\text{GVE}}) \qquad (29)$$

When t_4 is less than t_4^{GVE}, h will be positive, and its value is obtained by linear interpolation with the equation:

$$h = (t_4^{\text{GVE}} - t_4)/(t_4^{\text{GVE}} - t_4^{\text{PAG}}) \qquad (30)$$

The previous expression is also used in cases where t_4 is less than t_4^{PAG} and therefore, h will be greater than the unit.

Next, the k form parameter is estimated by trial and error as a function of t_3 with Equation (21), using equations (23) or (24), depending on whether h is negative or positive. This numerical calculation was solved with the bisection method (Campos–Aranda, 2003), adopting a tolerance of 0.00001 for the difference between the t_3 of the sample and that estimated with Equation (21). The beginning of the bisection method must respect the limits imposed on k in Figure 1; therefore, when $h>0$, k took initial values of −1 and 2. But when $h<0$, its initial negative value was defined with the expression $k = -1-0.725\cdot(h)$ and its initial positive value with the equation $k = -(1/h)$. Estimated the value of k, based on Equation (20) and the value of l_2, a is obtained, and finally, with Equation (19) and the magnitude l_1, the value of u is defined.

Other methods of fitting the Kappa distribution

Dupuis and Winchester (2001) have commented that the estimation of the four fit parameters of the Kappa distribution through the L moments method is not always feasible or calculable. Therefore they expose the maximum likelihood method as a viable option. These authors emphasize that the method of L moments presents computational difficulties when h is negative, especially if it is less than −1. For such situations ($h<-1$), Park and Park (2002) have proposed the method of maximum likelihood with penalties.

Singh and Deng (2003) propose the estimation method designated POME (*principle of maximum entropy*) and compare it against the existing ones of ordinary moments, L moments, and maximum likelihood. They conclude that the POME and maximum likelihood methods are identical, and the ordinary moments method is the least reliable.

Distributions used in the contrast

Logically, the LOG, GVE, and PAG distributions are the first three of the contrast because they are the ones that encompass and reproduce the Kappa model. Later, the two most widely used were included, the Log-Normal with three fit parameters (LN3) and the Wakeby with five fit parameters (Houghton, 1978). These five distributions were applied with the L moments method (Hosking & Wallis, 1997). The Log–Pearson type III (LP3) distribution was fitted with its classic method, that of moments, in the logarithmic domain (WRC, 1977) and in the real domain (Bobée, 1975), adopting the best fit; that is, with minor fit errors.

Three of the six distributions mentioned have been proposals for application by norm or precept. The first was the LP3, in the U.S.A, from the beginning of the seventies. Afterward, the Natural Environment Research Council (NERC, 1975) adopted the GVE for England until the end of the last century, when it changed to the LOG (Shaw, Beven, Chappel, & Lamb, 2011). The three widely used distributions have been LN3, PAG, and Wakeby.

Q–*Q* diagnostic graph

Nguyen *et al.* (2017) have suggested two assessments to select the optimal PDF for a record of extreme hydrological data: (1) descriptive ability and (2) predictive ability. The first refers to the accuracy with which the PDF being tested reproduces the sample data, and the second is logically associated with the variability of its predictions about the dispersion of the sample predictions. There are three techniques to test descriptive ability: (1) diagnostic charts, (2) statistical tests, and (3) goodness-of-fit indices.

Empirical versus estimated *probability P*–*P* and observed *quantity* versus estimated *Q*–*Q* diagnostic plots have become popular (Coles, 2001; Wilks, 2011) and provide a simple and effective way to compare the results of a contrasted PDF. For a sample of data x_i sorted from smallest to largest, an empirical probability (p) is assigned to them, for example, with the Cunnane formula, which according to Stedinger (2017), leads to unbiased values in most of the PDFs used in hydrology, this is:

$$p = \frac{m-0.40}{n+0.20} \tag{31}$$

m is the data order number, and n is the data number. For each datum x_i, its probability is obtained with the equation of the tested PDF. For the case of the Kappa distribution, with Equation (8). The P–P graph is defined with the following abscissa and ordinate points:

$$\left[\frac{m-0.40}{n+0.20}, F(x_i)\right] \quad \text{for } i = 1,2,\ldots,n \tag{32}$$

The Q–Q graph uses Equation (10) or the inverse solution of the PDF Kappa to define points on the ordinates and is made up of the following points:

$$\left[x_i, x\left(\frac{m-0.40}{n+0.20}\right)\right] \quad \text{for } i = 1,2,\ldots,n \tag{33}$$

The disadvantage of diagnostic graphs lies in the subjective assessment made when comparing various PDFs since a numerical value is not available (Nguyen *et al.*, 2017). Campos–Aranda (2019) visualizes the *Q*–*Q* graph as more useful to observe overestimated predictions (as it is located above the 45° line) or underestimated (as it is located below).

Standard error of fit (*EEA*)

Goodness-of-fit indices have the advantage of being easy to calculate and commonly involve the difference between the observed values x_i and the estimated values $\hat{x}$ with the PDF being tested. The *EEA*, by its acronym in Spanish, is the most common (Chai & Draxler, 2014); it was established in the mid-1970s (Kite, 1977) and has been applied in Mexico using Weibull's empirical formula (Benson, 1962); now it will be applied using Equation (31). The expression of the *EEA* is:

$$EEA = \left[\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{(n - np)}\right]^{1/2} \quad (34)$$

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

x_i is the n observed data sorted from smallest to largest, $\hat{x}$ The estimates for the probability are estimated with Equation (31) and the contrasted PDF. *np* is the number of fit parameters of the FDP, with four for the Kappa, five for the Wakeby, and three for the rest of the contrasted ones. The *EEA* has the units of the variable x_i.

Mean Absolute Error (*EAM*)

Its advantages lie in having the units of the variable, just like the *EEA*, and preventing the impact of the dispersed values from being squared and, therefore, *EEA* ≥ *EAM* by its acronym in Spanish (Willmott & Matsuura, 2005). Its expression is (Nguyen *et al.*, 2017):

$$EAM = \frac{\sum_{i=1}^{n} |x_i - \hat{x}_i|}{n - np} \tag{35}$$

Processed hydrological data records

Aldama, Ramirez, Aparicio, Mejia–Zermeño and Ortega–Gil (2006) expose the *joint* records of peak flow (*Qp*) in m^3/s and volume (*Vol*) in millions of m^3 (Mm^3) per year of the input floods at 15 important reservoirs and one in the project. Of such records, the two with the greatest amplitude were selected. The first ones correspond to the 64-entry data to the La Boquilla dam on the Conchos river tributary of the Bravo river, which defines Hydrological Region No. 24-1. The following records to be processed are the 61 entry data to the Adolfo Ruiz Cortines dam (*Mocuzari*), on the Mayo river, in Sonora, Mexico.

Gomez *et al.* (2010) expose the *joint* records of *Qp* and *Vol* with 55 data from the La Cuña gauging station, located in the Rio Verde Hydrological Region No. 12–3 (Rio Santiago), Mexico. These flood records were also processed.

The following two records of annual floods were taken from the contrast made by Campos–Aranda (2013) and belong to samples that define their location in the L ratio diagram below the PAG distribution curve, for which their *h* value will be greater than the unity. One record contains 53 floods at the Luis Donaldo Colosio Dam (Huites) entrance on the Fuerte river in Sinaloa, Mexico. The other 67 floods at the entrance to the El Cuchillo dam, on the San Juan river, in the state of Nuevo Leon, Mexico. This record was previously processed by Campos-Aranda (1998). Due to the difficulty involved in obtaining these records, they are shown in Table 1.

Table 1. Records of the annual floods (m^3/s) of the hydrometric stations Huites in Sinaloa, Mexico, and El Cuchillo in the state of Nuevo Leon, Mexico.

No.	Huites (1941-1993)			El Cuchillo (1927-1993)				No.
1	2085	1908	1119	1817.0	3358.0	603.0	396.0	19
2	2531	15000	6178	54.4	526.4	348.2	210.0	20
3	14376	1396	4443	34.2	1393.9	1274.0	366.9	21
4	2580	1620	1474	994.0	1173.1	726.8	183.9	22
5	1499	2702	2508	332.5	381.8	5540.0	115.2	23
6	1165	1319	1530	162.8	469.0	464.5	963.7	24
7	1127	1944	8000	2736.5	2084.4	470.6	688.6	25
8	3215	2420	5496	393.5	167.4	655.4	8315.1	26
9	10000	2506	3385	602.8	2511.5	649.0	505.3	27
10	3229	1534	1374	1307.0	302.7	454.6	392.6	28
11	677	1508	1245	139.8	384.5	3355.5	161.5	29
12	1266	1558	2299	6758.5	125.0	1935.0	153.9	30
13	1025	2200	1345	720.4	900.0	643.7	250.9	31
14	955	2225	11350	404.0	1927.0	1081.0	–	32

15	4780	7960	2509	1194.0	177.0	2000.0	–	33
16	696	4001	2006	817.2	1677.6	3500.0	–	34
17	593	1067	1182	675.5	492.0	355.0	–	35
18	3010	3233	–	1584.8	584.5	225.9	–	36

Finally, three records of annual maximum daily precipitation (*PMD*) from a pluviometric station in each geographical area of San Luis Potosi were processed. Mexquitic was processed from the Altiplano (n=72), from the Middle Zone, San Francisco (n=50), and from the Huasteca region, Xilitla (n=51). These records were analyzed by Campos-Aranda (2019) to obtain their optimal PDFs and integrated based on the CONAGUA monthly Excel file provided to the author; therefore, they are reproduced in Table 2.

Table 2. Records of annual *PMD* (millimeters) in the three pluviometric stations indicated in the state of San Luis Potosi, Mexico.

No.	Mexquitic (1943-2014)			San Francisco (1961-2013)		Xilitla (1964-2014)		No.
1	48.3	42.0	51.0	15.0	135.0	146.5	138.0	27
2	60.0	39.0	50.0	25.0	45.0	159.0	200.0	28
3	57.0	66.0	25.0	12.0	42.4	207.4	182.0	29
4	47.0	48.0	12.0	12.0	43.0	224.5	330.0	30

5	35.5	55.0	28.0	84.0	36.0	224.7	148.0	31
6	56.0	54.0	30.0	24.0	32.0	163.0	222.0	32
7	40.0	57.0	62.5	111.5	64.5	193.5	113.0	33
8	25.0	64.0	41.5	42.0	49.0	184.0	187.0	34
9	83.0	68.0	53.2	36.0	39.0	128.0	163.0	35
10	84.0	40.0	35.8	53.0	64.0	147.0	136.0	36
11	40.0	55.2	50.0	21.0	44.0	299.0	92.0	37
12	31.0	40.5	30.0	57.0	44.0	111.0	160.0	38
13	52.0	55.3	41.0	42.0	54.0	200.0	173.0	39
14	59.0	51.0	69.0	43.0	40.0	103.0	203.0	40
15	47.0	27.2	75.0	22.0	45.5	420.0	131.0	41
16	40.0	74.0	53.0	78.0	42.0	226.0	118.0	42
17	45.0	44.2	53.0	57.0	27.0	220.0	98.5	43
18	61.0	46.0	53.5	42.0	30.0	180.0	320.0	44
19	21.0	45.3	107.0	30.0	75.0	160.0	247.0	45
20	30.5	73.0	41.0	85.0	42.0	232.0	144.0	46
21	26.0	31.7	–	78.0	40.0	140.0	162.0	47
22	51.5	36.0	–	41.0	87.0	137.0	122.0	48
23	32.0	59.0	–	22.0	28.0	126.7	93.5	49
24	28.0	46.0	–	23.0	51.0	200.0	213.0	50

25	62.0	20.0	–	38.0	–	280.0	147.0	51
26	65.0	15.5	–	43.4	–	184.0	–	52

Wald-Wolfowitz test

This nonparametric test was developed by Abraham Wald and Jacob Wolfowitz based on the work of Richard L. Anderson on the serial correlation coefficient. It has been used by Bobée and Ashkar (1991), Rao and Hamed (2000), and Meylan *et al.* (2012) to test *independence* and *stationarity* in records of maximum annual expenditures (X_i). Due to the above, it was proposed to apply the test to the records of annual *Qp*, *Vol,* and *PMD*, which must be samples of random values. The Wald–Wolfowitz test statistic is:

$$R = \sum_{i=1}^{n-1} x_i \cdot x_{i+1} + x_n \cdot x_1 \qquad (36)$$

When the size (n) of the series or sample (x_i) is not small, and its data are independent, R comes from a Normal distribution with mean and variance, given by the following expressions:

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

$$E[R] = \bar{R} = \frac{S_1^2 - S_2}{n-1} \quad (37)$$

$$Var[R] = \frac{S_2^2 - S_4}{n-1} + \frac{S_1^4 - 4 \cdot S_1^2 \cdot S_2 + 4 \cdot S_1 \cdot S_3 + S_2^2 - 2 \cdot S_4}{(n-1)(n-2)} - \bar{R}^2 \quad (38)$$

in which:

$$S_w = \sum_{i=1}^{n} x_i^w \quad (39)$$

Finally, U is calculated with the equation:

$$U = \frac{R - \bar{R}}{\sqrt{Var[R]}} \quad (40)$$

The value of U follows a Normal distribution (0,1) and can be used to test the independence of the series data with a level of significance α, commonly 5%. In a two-tailed test, the standardized normal variable is $Z_{\alpha/2} \cong 1.96$; then, when the absolute value of U is less than 1.96, the series will be made up of independent values (random sample).

Procedure flow chart

Figure 3 shows schematically the process followed to contrast the Kappa distribution.

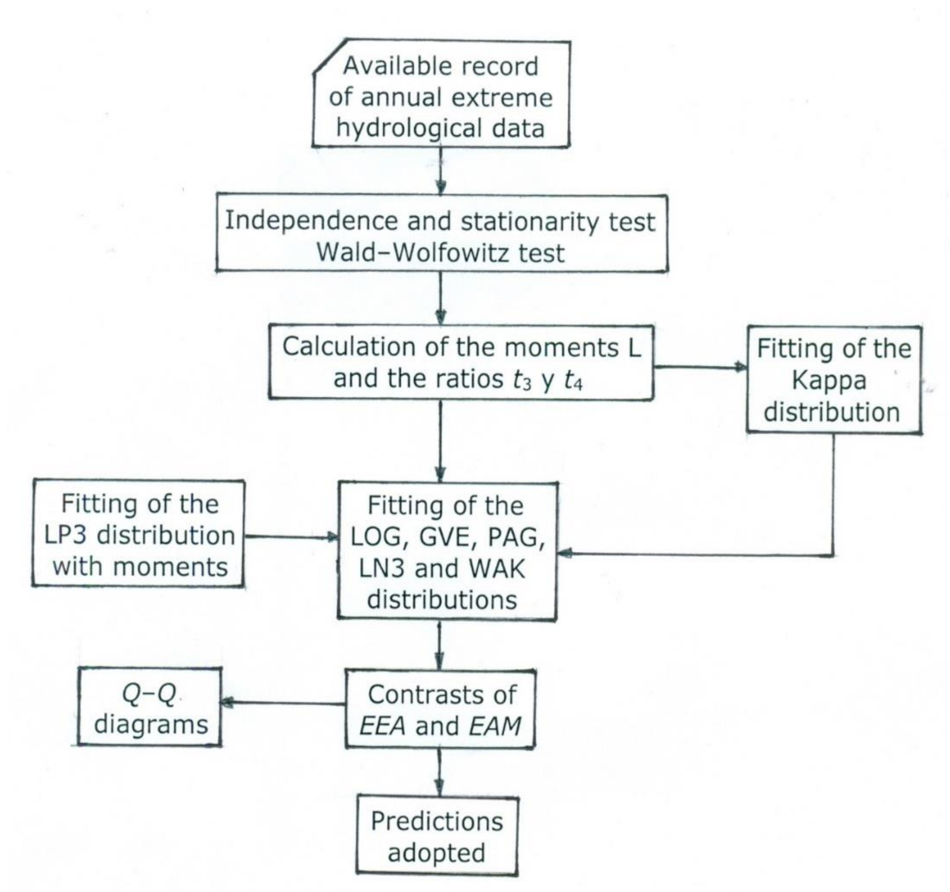

Figure 3. A flowchart that exposes the process followed for contrasting the Kappa distribution with the six of general application.

Results and their discussion

Test of randomness and moments and ratios L

In the fourth column of Table 3, the values of the *U* statistic (Equation (40)) are shown, defining that the 11 records processed are random. The rest of the columns show the magnitudes of the moments and ratios L of each record (equations (12) to (18)).

Table 3. General data and values of moments and ratios L of the 11 records processed.

No.	**Registro:**	***A*** ***H***	***U***	l_1	l_2	t_3	t_4
1	La Boquilla *Qp*	21003	−0.084	858.707	373.7957	0.30456	0.15338
2	La Boquilla *Vol*	21003	−0.251	292.896	128.3310	0.25226	0.12746
3	Mocúzari *Qp*	10719	0.147	1477.266	547.3816	0.34733	0.26482

4	Mocúzari *Vol*	10719	1.189	158.709	66.0485	0.26164	0.16069
5	La Cuña *Qp*	19097	0.284	499.875	205.4937	0.38716	0.25523
6	La Cuña *Vol*	19097	0.213	154.839	74.4854	0.43326	0.28467
7	Huites *Qp*	26020	−0.135	3176.464	1453.9170	0.50858	0.31947
8	El Cuchillo *Qp*	8794	−0.821	1139.560	651.0005	0.51895	0.32702
9	Mexquitic *PMD*	1749	1.307	47.794	9.5213	0.05475	0.15491
10	S.Francisco *PMD*	1066	−0.565	46.726	12.7573	0.23966	0.24363
11	Xilitla *PMD*	630	−0.278	181.163	34.6449	0.21794	0.18487

Symbology:

A = watershed area, in km^2, in the first eight records.

H = altitude, in masl, in the last three records.

U = statistical of the Wald–Wolfowitz test.

L_1 = first order moment L, in m^3/s or Mm^3.

L_2 = second order L moment.

T_3 = ratio of moments L of asymmetry.

T_4 = ratio of moments L of kurtosis.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Fit parameters of the Kappa distribution

Based on the values in Table 3, equations (19) to (30), and the procedure described for estimating the four fit parameters of the Kappa distribution, the results shown in Table 4 were obtained, in which the number of iterations performed with the bisection method is also indicated, to estimate the first parameter of form *k*. In the fourth column, the first contrast PDF of the Kappa distribution is indicated, selected from LOG, GVE, or PAG, according to the proximity of the value of *h* to −1, 0, 1, respectively. It is observed in the Xilitla record that, although *h* is negative, it is closer to zero than to −1.

Table 4. Results of the fit with L moments of the Kappa distribution in the 11 processed records of extreme hydrological data.

No.	Record:	ITE	PDF	Fit parameters:			
				h	***k***	***a***	***u***
1	La Boquilla *Qp*	22	PAG	0.8864	0.0546	789.107	132.918
2	La Boquilla *Vol*	18	PAG	0.7652	0.1076	276.658	65.128
3	Mocúzari *Qp*	16	LOG	−0.8928	−0.3431	453.746	1166.899
4	Mocúzari *Vol*	14	GVE	0.4140	−0.0348	105.492	75.685
5	La Cuña *Qp*	16	GVE	0.2888	−0.2685	228.381	258.746

6	La Cuña *Vol*	13	GVE	0.4287	−0.3155	78.311	60.399
7	Huites *Qp*	14	PAG	1.1640	−0.3337	1652.386	613.021
8	El Cuchillo *Qp*	17	PAG	1.2424	−0.3440	730.866	−33.717
9	Mexquitic *PMD*	12	LOG	−0.7337	−0.0215	10.363	46.070
10	San Fco. *PMD*	16	LOG	−1.0000	−0.2404	11.590	41.914
11	Xilitla *PMD*	17	GVE	−0.3970	−0.1523	38.453	159.906

Symbology:

ITE = number of iterations of the bisection method.

PDF = PDF close to the Kappa distribution, according to the value of *h*.

Observations regarding the fit achieved

Based on the results (errors and predictions) concentrated in Table 5, it can be observed in eight of the 11 records processed that the minimum *EEA* and *EAM*, indicated in italics, coincide with defining the best fit. In records 2, 5, and 11, the lowest *EEA* and *EMAE* were obtained for different PDFs. The selection (shaded line) between both distributions was based on contrasting the predictions for the return periods greater than 100 years, adopting the intermediate or more representative ones, according to the author's criteria. Such a selection may seem subjective, but due to the similarity of all the predictions, such an adoption judgment is not entirely wrong.

Table 5. The contrast of goodness-of-fit indicators and predictions between the Kappa distributions, the one that represents (LOG, GVE, or PAG) and three of general use (LP3, LN3, and WAK) in the 11 hydrological data records processed.

NR	PDF	*SE of fits*	*MAE*	Return periods, in years						
				25	50	100	500	1000	5000	10000
1	KAP	79.5	62.9	2461	2912	3346	4292	4674	5508	5845
1	PAG	*76.9*	*58.1*	2474	2925	3355	4282	4652	5448	5765
1	LP3	86.1	62.8	2491	3015	3544	4784	5318	6547	7069
1	LN3	99.4	72.8	2462	3033	3650	5279	6074	8163	9174
1	WAK	78.3	59.0	2477	2938	3382	4351	4744	5601	5947
2	KAP	*19.8*	12.5	817	948	1070	1319	1414	1608	1682
2	PAG	22.2	12.9	816	933	1035	1225	1290	1412	1453
2	LP3	21.4	13.4	849	997	1135	1416	1521	1731	1808
2	LN3	23.3	17.7	818	985	1159	1598	1804	2324	2567
2	WAK	20.7	*12.0*	790	923	1067	1522	1808	2860	3571
3	KAP	153.2	87.7	3782	4873	6244	10991	13988	24420	31004
3	LOG	155.3	86.6	3765	4853	6223	10988	14010	24585	31302
3	LP3	148.9	94.6	3917	4878	5956	8978	10536	14855	17058
3	LN3	152.8	99.5	3910	4858	5911	8807	10270	14240	16219

3	WAK	*152.0*	*82.1*	3780	4902	6303	11051	13982	23917	30052
4	KAP	*8.2*	*5.8*	434	517	602	808	900	1122	1221
4	GVE	9.1	6.5	431	526	629	911	1053	1441	1636
4	LP3	11.4	8.5	423	505	588	783	868	1067	1152
4	LN3	8.3	6.0	431	520	613	849	960	1244	1377
4	WAK	8.9	6.0	435	515	595	773	847	1015	1086
5	KAP	59.1	38.2	1419	1835	2335	3920	4844	7786	9489
5	GVE	*58.7*	37.5	1406	1838	2372	4160	5254	8926	11172
5	LP3	72.1	42.2	1408	1762	2154	3228	3767	5219	5938
5	LN3	59.2	37.8	1441	1839	2293	3592	4270	6167	7139
5	WAK	59.1	*36.6*	1442	1844	2304	3646	4367	6469	7599
6	KAP	17.7	8.8	495	664	873	1576	2007	3459	4350
6	GVE	18.6	9.7	486	662	888	1702	2234	4156	5411
6	LP3	*12.0*	*6.9*	517	694	907	1574	1952	3101	3736
6	LN3	16.7	8.1	506	670	863	1439	1752	2661	3142
6	WAK	18.0	9.0	495	664	872	1561	1978	3358	4193
7	KAP	789.8	390.9	10173	13939	18689	35052	45301	80588	102652
7	PAG	799.5	393.9	10068	13834	18629	35436	46133	83632	107487
7	LP3	949.8	400.7	10492	15061	21301	45851	63053	129752	175998
7	LN3	*784.3*	*376.8*	10283	14164	18971	34551	43592	71547	87183
7	WAK	815.4	410.2	10070	13834	18626	35412	46091	83504	107304
8	KAP	237.0	101.1	4282	6010	8205	15865	20718	37637	48342
8	PAG	248.1	102.1	4218	5948	8178	16170	21357	39946	52007

8	LP3	*217.7*	*95.9*	4403	6209	8467	15903	20281	34031	41838
8	LN3	227.6	97.1	4333	6115	8338	15630	19901	33229	40742
8	WAK	253.5	106.3	4212	5946	8187	16245	21495	40376	52675
9	KAP	1.7	1.3	80	88	96	115	123	143	152
9	LOG	1.7	1.2	80	88	96	117	127	150	160
9	LP3	2.2	1.8	82	89	97	113	120	135	141
9	LN3	2.1	1.3	79	86	92	104	109	120	124
9	WAK	*1.5*	*1.0*	80	89	98	120	129	153	163
10	KAP	3.5	2.9	97	117	139	208	247	367	435
10	LOG	3.5	2.9	97	116	139	208	247	366	433
10	LP3	3.9	3.2	98	113	128	166	183	224	242
10	LN3	3.7	3.1	98	114	131	172	191	239	261
10	WAK	*3.3*	*2.7*	99	119	141	198	227	304	342
11	KAP	6.7	4.7	318	365	416	558	630	831	934
11	GVE	7.0	*4.6*	319	361	406	518	570	702	764
11	LP3	7.0	4.8	319	362	407	522	577	718	786
11	LN3	7.5	4.9	318	359	400	501	547	660	712
11	WAK	*6.4*	4.8	318	366	418	556	625	808	899

Symbology:

NR = registration number, according to Table 3 or Table 4.

PDF = probability distribution function tested.

According to data, the SE of fits = standard error of fit, in m3/s, Mm3, or mm.

According to data, MAE = means an absolute error in m3/s, Mm3 ó mm.

Figure 4 and Figure 5 show the Q–Q diagnostic graphs of the two best fits achieved with the Kappa distribution obtained for records 2 and 4, respectively. Both records correspond to the annual runoff volume of floods in Mm^3. In Figure 4, up to data number 59, the Kappa model reproduces the sample data exactly, and from then on, it overestimates three values and underestimates the last two relatives of the scattered data. On the other hand, in Figure 5, only one lack of accuracy was detected in the last two data, but it is not very severe.

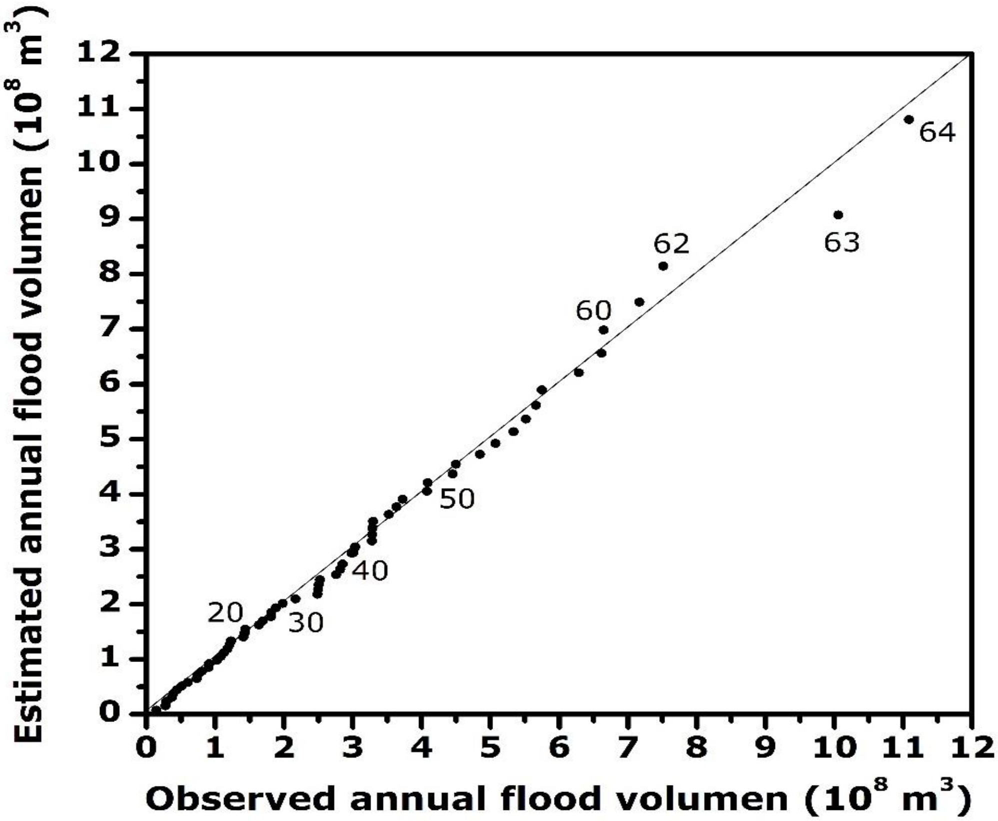

Figure 4. *Q*–*Q* graph of the Record 2 obtained with the Kappa distribution.

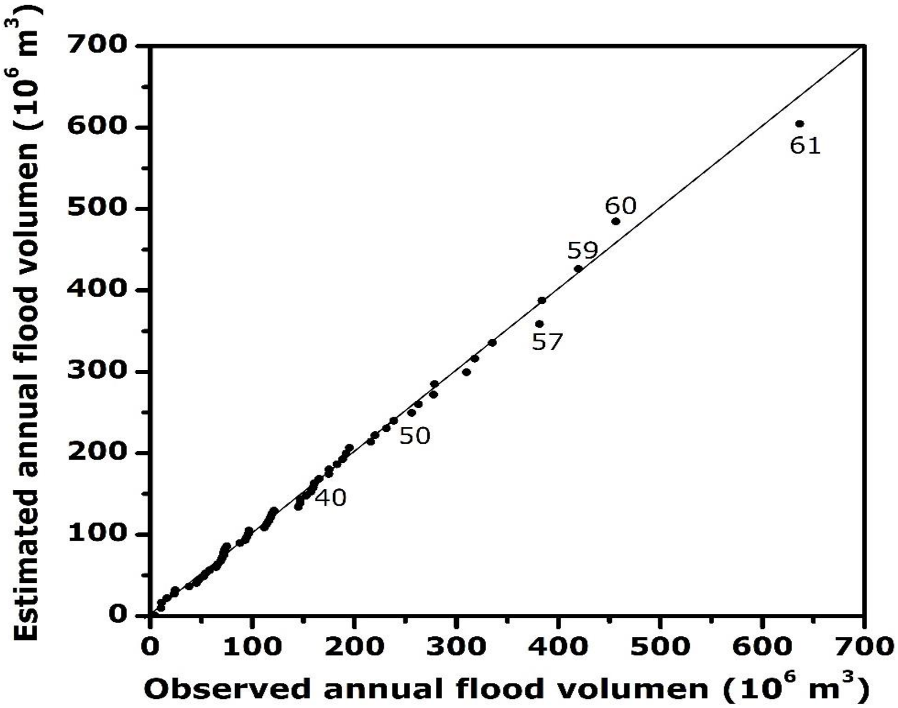

Figure 5. *Q*–*Q* graph of the Record 4 obtained with the Kappa distribution.

Observations regarding the predictions

In the 11 records processed, the predictions of the different PDFs corresponding to return periods (*Tr*) of less than 100 years are quite similar. Generally, the FDP adopted for having a lower value of the *SE of fits* and the *MAE* leads to representative predictions in the high return periods (*Tr*>500 years). The two conditions mentioned above result in confidence for all predictions and those adopted.

The contrast of results (errors and predictions) between the Kappa distributions and the one that includes or reproduces indicates the great similarity of values. The above is notable in records 1, 2, 3, 7, and 10. In records 2 and 4, the Kappa distribution was the best option, and the Wakeby function led to the best fit in records 3 and 5 of peak flow. And 9, 10, and 11 of the annual *PMD*. Concerning the above, Campos-Aranda (2019) suggests it be applied under precept in the annual *PMD* records.

Hydrological uncertainties

The FA implies several hydrological weaknesses or uncertainties, which can be summarized in the following four. The first refers to the representativeness, in the future, of the available hydrological record. The second is associated with the extrapolation that must be done to estimate return periods greater than 100 years. The third lies in the FA itself, as it has to select a PDF to make the predictions. Finally, due to excess flexibility, hydrological uncertainty is considered to be the tendency of the distributions of more than three adjustment parameters to model the available record too accurately and thus lose prediction capacity. Such is the case of the distributions: Kappa, TCEV, mixed Gumbel, and Wakeby.

Kjeldsen, Lamb, and Blazkova (2014) explain the general aspects of the uncertainty in the FA, and Botto, Ganora, Laio, and Claps (2014) developed a procedure based on the minimum cost-benefit criterion, which allows correcting the hydrological uncertainties of the Afs. Botto, Ganora, Claps, and Laio (2017) exposed the simple operational version of such a process.

Conclusions

Respecting the limits of the shape parameters (*h* and *k*), shown in Figure 1, for the Kappa distribution ensures the existence of its four moments L and the uniqueness of its fit parameters (*u*, *a*, *k*, *h*). In addition, the L moments method (equations (19) to (30)) procedure to obtain the fit parameters is simple and free of computational complications.

In the 11 records processed, the contrast between the results (errors and predictions) of the Kappa distribution and the one it encompasses (LOG, GVE, or PAG) showed great similarity. The foregoing implies that the Kappa model with four fit parameters reproduces the results of the aforementioned distributions and, perhaps, its adjustment is closer to the real data, as its parameter h varies between −1 and a value greater than unity, as stated for registers 7 and 8.

The Kappa distribution, due to its wide area of values for the L ratios of skewness (t_3) and kurtosis (t_4) that it encompasses or reproduces, shown in Figure 2, is a probabilistic model that will lead to an excellent fit to the data or available sample of maximum annual hydrological values and therefore should be systematically included in their frequency analyses.

Although the Kappa distribution was applied locally to 11 specific records in this work, various authors have highlighted that their best performance occurs in the *regional* FAs. The foregoing opens several options for its contrast in such studies.

In addition, the observations deduced from Table 5 of results (errors and predictions) allow us to suggest the application of the Kappa distribution on a routine basis to *complement* those of application under precept (LP3, GVE, and LOG), as well as those of widespread use (LN3, PAG, and WAK); especially in the selection of the predictions to be adopted in the three extreme return periods of 1 000, 5 000 and 10 000 years.

Acknowledgments

The comments and corrections suggested by the anonymous referees A, B, and D are appreciated; the former aided in improving the structure and writing of the article, and the second ones allowed to include relevant aspects of frequency analysis not initially addressed, such as challenges and hydrological uncertainty.

References

Aldama, A. A., Ramírez, A. I., Aparicio, J., Mejía-Zermeño, R., & Ortega-Gil, G. E. (2006). *Seguridad hidrológica de las presas en México.* Jiutepec, México: Instituto Mexicano de Tecnología del Agua.

Asquith, W. H. (2011). Chapter 9. L-moments of four and more parameter univariate distributions. In: *Distributional analysis with L-moment statistics using the R environment for statistical computing* (pp. 203-233). Lubbock, USA: author's edition.

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. (1975). The Log-Pearson type 3 distribution and its application to Hydrology. *Water Resources Research*, 11(5), 681-689. DOI: 10.1029/WR011i005p00681

Bobée, B., & Ashkar, F. (1991). Chapter 1: Data requirements for hydrologic frequency analysis. In: *The Gamma Family and derived distributions applied in Hydrology* (pp. 1-12). Littleton, USA: Water Resources Publications.

Botto, A., Ganora, D., Laio, F., & Claps, P. (2014). Uncertainty compliant design flood estimation. *Water Resources Research*, 50(5), 4242-4253. DOI: 10.1002/2013WR014981

Botto, A., Ganora, D., Claps, P., & Laio, F. (2017). Technical note: Design flood under hydrological uncertainty. *Hydrology and Earth System Sciences*, 21(7), 3353-3358. DOI: 10.5194/hess-21-3353-2017

Campos-Aranda, D. F. (1998). *Revisión de la creciente de diseño de la presa El Cuchillo, en el estado de Nuevo León, México (Enfoque probabilístico global)*. Tomo 1 (pp. 349-357). XVIII Congreso Latinoamericano de Hidráulica, del 13 al 16 de octubre, Oaxaca, Oaxaca, México.

Campos-Aranda, D. F. (2003). Capítulo 3: Solución de ecuaciones no lineales. En: *Introducción a los métodos numéricos: Software en Basic y aplicaciones en hidrología superficial* (pp. 44-65). San Luis Potosí, México: Editorial Universitaria Potosina.

Campos-Aranda, D. F. (2013). Contraste de la distribución logística generalizada en 31 registros históricos de eventos máximos anuales. *Ingeniería, Investigación y Tecnología*, 14(1), 113-123.

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

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

Coles, S. (2001). Theme 2.6.7: Model diagnostics and Chapter 3: Classical Extreme Value Theory and Models. In: *An introduction to statistical modeling of extreme values* (pp. 36-44, 45-73) London, UK: Springer-Verlag.

Davis, P. J. (1972). Gamma Function and related functions. Chapter 6 (pp 253-296). In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of Mathematical Functions*. New York, USA: Dover Publications.

Dupuis, D. J., & Winchester, C. (2001). More on the four-parameter Kappa distribution. *Journal of Statistical Computation and Simulation*, 71(2), 99-113.

Gómez, J. F., Aparicio, M., & Patiño, C. (2010). Capítulo 6: Análisis de frecuencias bivariado para la estimación de avenidas de diseño. En: *Manual de análisis de frecuencias en hidrología* (pp. 106-127). Jiutepec, México: Instituto Mexicano de Tecnología del Agua.

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. DOI: 10.1029/WR015i005p01049

Hosking, J. R. M. (1994). The four-parameter Kappa distribution. *IBM Journal of Research and Development*, 38(3), 251-258.

Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, 29(2), 271-281.

Hosking, J. R. M., & 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.

Johnson, F., & Sharma, A. (2017). Design Rainfall. Chapter 125. In: Singh, V. P. (ed.). *Handbook of Applied Hydrology* (2nd ed.) (pp. 125.1-125.13). New York, USA: McGraw-Hill Education.

Katz, R. W. (2013). Statistical methods for nonstationary extremes (Chapter 2). 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. DOI: 10.1016/j.jhydrol.2006.03.004

Kite, G. W. (1977). Chapter 12: Comparison of frequency distributions. In: *Frequency and risk analyses in hydrology* (pp. 156-168). Fort Collins, USA: Water Resources Publications.

Kjeldsen, T. R., Lamb, R., & Blazkova, S. D. (2014). Chapter 8: Uncertainty in flood frequency analysis. In: Beven, K., & Hall, J. (eds.). *Applied Uncertainty Analysis for Flood Risk Management* (pp. 153-197). London, UK: Imperial College Press.

Kjeldsen, T. R., Ahn, H., & Prosdocimi, L. (2017). On the use de a four-parameter kappa distribution in regional frequency analysis. *Hydrological Sciences Journal*, 62(9), 1354-1363. DOI: 10.1080/02626667.2017.1335400

Meylan, P., Favre, A. C. & Musy, A. (2012). Chapter 3: Selecting and checking data series and Theme 9.3: Frequency analysis in a non-stationarity context. In: *Predictive hydrology. A frequency analysis approach* (pp. 29-70, 176-179). Boca Raton, USA: CRC Press.

NERC, Natural Environment Research Council. (1975). Volume I: Hydrological Studies, chapter 2: Statistical flood frequency analysis. In: *Flood Studies Report* (pp. 107-251). London, UK: Natural Environment Research Council.

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

Parida, B. P. (1999). Modeling of Indian summer monsoon rainfall using a four-parameter Kappa distribution. *International Journal of Climatology*, 19(12), 1389-1398.

Park, J. S., & Park, B. J. (2002). Maximum likelihood estimation of the four-parameter Kappa distribution using the penalty method. *Computer and Geosciences*, 28(1), 65-68.

Rao, A. R., & Hamed, K. H. (2000). Theme 1.8: Tests on hydrologic data and Chapters 7, 8 and 9. In: *Flood Frequency Analysis* (pp. 12-21, 207-321). Boca Raton, USA: CRC Press.

Shaw, E. M., Beven, K. J., Chappel, N. A., & Lamb, R. (2011). Chapter 13: Estimating floods and low flows in the UK. In: *Hydrology in practice* (4th ed.) (pp. 322-350). London, UK: Spon Press.

Singh, V. P., & Deng, Z. Q. (2003). Entropy-based parameter estimation for Kappa distribution. *Journal of Hydrologic Engineering*, 8(2), 81-92. DOI: 10.1061/(ASCE)1084-0699(2003)8:2(81)

Stedinger, J. R. (2017). Flood Frequency Analysis. Chapter 76. In: Singh, V. P. (ed.). *Handbook of Applied Hydrology* (2nd ed.) (pp. 76.1-76.8). New York, USA: McGraw-Hill Education.

WRC, Water Resources Council. (1977). *Guidelines for determining flood flow frequency.* Bulletin # 17A of the Hydrology Committee. Revised edition. Washington, DC, USA: Water Resources Council.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82. DOI: 10.3354/cr030079

Wilks, D. S. (2011). Theme 4.5: Qualitative assessments of the goodness fit. In: *Statistical methods in the atmospheric sciences* (3rd ed.) (pp. 112-116). San Diego, USA: Academic Press (Elsevier).