DOI: 10.24850/j-tyca-2022-04-01

Articles

Estimation of annual runoff volume in hydrological Region No. 10 (Sinaloa), Mexico, through regional frequencies analysis

Estimación del volumen escurrido anual en la Región Hidrológica No. 10 (Sinaloa), México, mediante análisis de frecuencias regional

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

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

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

Abstract

One of the objectives of hydrological studies is the estimation of the annual runoff volume (*VEA*, for its name in Spanish) of watersheds without gauging. Such data is basic in the design of reservoirs. In this study, the estimation is provided through *regional frequency analysis*, based on the *Runoff Index* method. This approach was applied in the Hydrological Region No. 10 (Sinaloa), Mexico, for which 22 *VEA* records were processed, with amplitudes that varied from 24 to 57 years. Through the Discordance Test, three records were eliminated. Regional hydrological homogeneity was verified with the remaining 19 records and three statistical techniques: seasonality indices, linearity of ordinary moments, and linear regression between probability-weighted moments. The tests of loss of homogeneity of each record led to the elimination of four, due to their ascending linear trend. The 15 remaining records were processed in a dimensionless way, by dividing their data over their annual mean runoff volume (*VEMA*) and they were concatenated with the stations-years method, forming a series of 539 data. Three probabilistic models were applied to obtain the *regional growth curve*: Pearson type III, Log-Normal, and Potential Transformation; choosing the one with the smallest standard error of fit. The scaling of the regional Log-Normal distribution was achieved with a logarithmic relationship between the *VEMA* and the basin area of the processed records. In the Conclusions section, the established procedure is detailed and the importance of the results to generate synthetic sequences of *VEA* is highlighted (which are equally probable to occur). Therefore, the systematic application of the Runoff Index procedure as a method of estimating the *VEA* in other regions of the country is recommended.

Keywords: Annual runoff volume, L moments, discordance test, seasonality indices, regional hydrological homogeneity, linear trend, standard error of fit, synthetic sequences.

Resumen

Un objetivo de los estudios hidrológicos es la estimación del volumen escurrido anual (*VEA*) de cuencas sin aforos. Datos que son básicos en el diseño de embalses. En este estudio, tal problema se resuelve por medio del *análisis de frecuencias regional*, con base en el método del *índice de escurrimientos*. Este enfoque se aplicó en la Región Hidrológica No. 10 (Sinaloa), México, procesando 22 registros de *VEA*, con amplitudes que variaron de 24 a 57 años. Por medio de la prueba de discordancia se eliminaron tres registros. La homogeneidad hidrológica regional se verificó con los 19 registros restantes y tres técnicas estadísticas: índices de estacionalidad, linealidad de los momentos ordinarios y regresión lineal entre momentos de probabilidad ponderada. Las pruebas de pérdida de homogeneidad de cada registro condujeron a eliminar cuatro, por mostrar tendencia lineal ascendente. Los 15 registros remanentes se procesaron de forma adimensional, al dividir sus datos entre su volumen escurrido medio anual (*VEMA*) y se concatenaron con el método de las estaciones-años, formando una serie de 539 datos. Para obtener la *curva de crecimiento regional* se aplicaron tres modelos probabilísticos: Pearson tipo III, Log-Normal y Transformación Potencial; adoptando el de menor error estándar de ajuste. El escalamiento de la distribución Log-Normal regional se logró con una relación logarítmica entre el *VEMA* y el área de cuenca de los registros procesados. Las Conclusiones detallan el procedimiento establecido y destacan la importancia de los resultados para generar secuencias sintéticas de *VEA*, igualmente probables de ocurrir. Por ello, se recomienda la aplicación sistemática del método del Índice de escurrimiento en otras regiones del país, para contar con tal método de estimación del *VEA*.

Palabras clave: volumen escurrido anual, momentos L, prueba de discordancia, índices de estacionalidad, homogeneidad hidrológica regional, tendencia lineal, error estándar de ajuste, secuencias sintéticas.

Received: 13/04/2020

Accepted: 11/05/2021

Introduction

Reservoir sizing

Surface Hydrology, about the estimation of hydraulic resources, seeks to answer the following two questions: how much water is available? and what are the extreme values? Both issues involve estimating the available volumes of water and its extreme magnitudes (floods and droughts), as well as their occurrence or behavior over time.

The classic hydraulic works for utilization are the *reservoirs*, which can be divided into large and small; the former is multipurpose and generally store water from the wet years and use it in the dry years, the latter store water from the rainy seasons and release it in the next dry season. The hydrological design of reservoirs, whether large or small, requires the estimation of the annual and monthly runoff regime at the project site, for their sizing. Furthermore, designing floods are necessary to guarantee their hydrological safety (Carr & Underhill, 1974; Zsuffa & Gálai, 1987; Skertchly-Molina, 1989).

Due to the low density of hydrometric stations, the runoff regime estimation at the reservoir project site was generally carried out before the regional approach, based on the concept of similarity of basins. Then, a hydrological model of the rain-runoff process, calibrated in a gauged basin, was applied to another *similar* but without hydrometric data (Cannarozzo, Noto, Viola, & La-Loggia, 2009).

Currently, the *regional frequencies analysis* allows the use of all the hydrometric information available in a geographic area, through statistical methods or techniques which help define a *homogeneous* region or zone from a hydrological point of view, and then it is feasible and reliable to establish a common dimensionless probability distribution function (PDF), about a mean or median *index value*. Subsequently, in each homogeneous region, an empirical relationship is obtained between the index value and various morphological and climatic characteristics of the basin easily calculable, making it possible to generate synthetic samples of the annual runoff volume (*VEA*, by its acronym in Spanish) at any site in the studied region (Cannarozzo *et al*., 2009). Blöschl, Sivapalan, Wagener, Viglione, and Savenije (2013) indicate that the hypothesis of this method establishes that the annual mean runoff volume can vary between sites in a homogeneous region, within which the PDF is identical and defined by the dimensionless growth curve.

On the other hand, Skertchly-Molina (1989) has presented a thorough classification of the methods that allow dimensioning the reservoirs. Such techniques can be divided into two types: (1) those which process historical runoff records and (2) those which use synthetic records that are equally likely to occur. Regarding the convenience of one or the other, McMahon and Mein (1986) point out that the historical record of runoff is the least probable or feasible method that will repeat exactly in the future. In addition, they indicate that having a way to view the sensitivity of the reservoir´s necessary capacity estimate is very useful to make more reliable decisions. This is possible by making use of synthetic series of annual runoff (Campos-Aranda, 2011).

OPEN ACCESS

Check for updates

Objective

The present study focuses on the *regional frequency analysis* of the annual runoff volume (*VEA*, by its acronym in Spanish), in millions of cubic meters (Mm^3), from 22 hydrometric stations of the Hydrological Region No. 10 (Sinaloa), Mexico. The results allow the generation of synthetic sequences, of any size, equally likely to occur from the *VEA*, at the sites of the hydrometric stations used, and at any location without gauging data within the hydrological region analyzed. The *Runoff Index* method was used, similar to the well–known Flood Index. We worked with 19 records, after applying the Disagreement Test and the regional hydrological homogeneity was verified based on three statistical techniques. Next, the regional PDF was defined, through the standard error of fit, and finally, a logarithmic equation that relates the annual mean runoff volume and the basin area was obtained.

Methods and materials

Runoff Index Method

Hosking and Wallis (1997) mention that the techniques integrating the so-called *index flood method* can be applied to any type of data. This method was proposed by Dalrymple (1960) to process flood data, also called maximum flows (hence its name) allows estimating a summary of statistical parameters from the grouping of several different samples or records (Rahman, Haddad, & Eslamian, 2014; Ouarda, 2017).

When there are records of annual runoff volumes at *NS* sites and each location j, the number of data is n_i of volumes V_i^j. Then, we have that $V_j(F)$ with $0<F<1$, is the quantile function of the PDF of each site j. The basic hypothesis of the runoff index method establishes that the sites j integrate a *homogeneous region*, which implies that the PDFs of the *NS* sites are identical, except for the scaling factor, which is precisely the *Runoff Index* (IE^j). The equation of the method is:

$$V^j(F) = IE^j \cdot q(F) \text{ with } j = 1,2,\ldots,NS \quad (1)$$

where $q(F)$ is the *regional growth curve*, that is, a dimensionless quantile function common to all sites, and $q(F)$ is a regional PDF or common function of the values V_i^j/IE^j. In general, the IE^j is equal to the mean annual runoff volume (*$VEMA^j$*, by its acronym in Spanish) of the data from each site, but the median can be used (Hosking & Wallis, 1997).

In the IE^j method, the following four hypotheses are accepted: (1) the runoff from each site are independent, (2) the runoff from different sites is independent, and (3) the PDFs at different sites are identical except for the scaling factor and (4) the mathematical form of the regional growth curve is perfectly specified (Hosking & Wallis, 1997).

The regional frequencies analysis, including the IE^j, has two fundamental applications, within the homogeneous region; the first is carried out in sites with short or unreliable records and the second in localities without hydrometric data, using the empirical equations obtained for the relationship between the $VEMA^j$ and the basin area, or other morphological characteristics of the basin, or the main channel.

Moments and L ratios

The L moments (λ_r) are linear combinations of the *weighted probability moments* (β_r) developed by Greenwood, Landwehr, Matalas, and Wallis (1979), which are statistical parameters of the ordered data. The L moments have proven to be an efficient and robust system for the fit of PDFs currently used in hydrology or established under the precept. They are calculated with the following equations (Hosking & Wallis, 1997: Rao & Hamed, 2000; Stedinger, 2017):

$$\lambda_1 = \beta_0 \tag{2}$$

$$\lambda_2 = 2 \cdot \beta_1 - \beta_0 \tag{3}$$

$$\lambda_3 = 6 \cdot \beta_2 - 6 \cdot \beta_1 + \beta_0 \tag{4}$$

$$\lambda_4 = 20 \cdot \beta_3 - 30 \cdot \beta_2 + 12 \cdot \beta_1 - \beta_0 \tag{5}$$

In such a system, the quotients (τ) of L moments are defined, starting with L–Cv which is analogous to this coefficient, and then those of similarity with the asymmetry and kurtosis coefficients, these are:

$$\tau_2 = \lambda_2/\lambda_1 \tag{6}$$

$$\tau_3 = \lambda_3/\lambda_2 \tag{7}$$

$$\tau_4 = \lambda_4/\lambda_2 \tag{8}$$

In a sample of size n, with its elements X_i exposed in ascending order ($X_1 \leq X_2 \leq \cdots \leq X_n$) the unbiased estimators of β_r are obtained with the following general expression:

$$\beta_r = \frac{1}{n}\sum_{i=r+1}^{n}\frac{(i-1)\cdot(i-2)\cdots(i-r)}{(n-1)\cdot(n-2)\cdots(n-r)}X_i \text{ con } r = 0,1,\ldots,4 \quad (9)$$

Discordancy test

In any regional frequencies analysis, at least the following two verifications must be carried out: (1) each sample or series of data from a site must be reviewed to look for erroneous data, that is, values that are too large or extremely low, as well as repeated, which could originate in the transcription and (2) the samples should be compared with each other and with the closest ones, to verify a change in magnitude, for example as the size of the basin grows or its location varies from one area to another from the analyzed region (Hosking & Wallis, 1997; Campos-Aranda, 2010) (Table 1).

Table 1. Critical values of discordance (D_j) according to the number of stations of the group under analysis (Hosking & Wallis, 1997).

NS	***D_c***	***NS***	***D_c***	***NS***	***D_c***
5	1.333	9	2.329	13	2.869
6	1.648	10	2.491	14	2.971
7	1.917	11	2.632	$\geq$ 15	3.000
8	2.140	12	2.757	–	–

Fortunately, the erroneous values, the scattered events (*outliers*), the trend and the leaps or changes in the data mean, are reflected in the L moments of the sample (equations (2) to (5)). Therefore, a convenient mixture of the L ratios in a single statistic (D_i) that measures the *discordance* between the L ratios of the site and the group averages has been suggested as a basic test to detect sites that are *discordant* with the group as an everything (Hosking & Wallis, 1997).

Assuming there are *NS* sites in the group being analyzed. $\mathbf{u}_j = [t_2^i t_3^i t_4^i]^{\mathrm{T}}$ is defined as a vector containing the sample L ratios: t_2, t_3, and t_4 of each site $_j$, defined by equations (6) to (8). The superscript T means transposed since $\mathbf{u}_j$ is a row vector. The average (unweighted) vector of the group will be (Hosking & Wallis, 1997):

$$\bar{\mathbf{u}} = \frac{1}{N}\sum_{j=1}^{NS} \mathbf{u}_j \tag{10}$$

The matrix **A** of the sum of squares and cross products will be defined as:

$$\mathbf{A} = \sum_{j=1}^{NS} (\mathbf{u}_j - \bar{\mathbf{u}}) \cdot (\mathbf{u}_j - \bar{\mathbf{u}})^{\mathrm{T}} \quad (11)$$

Finally, the measure of the discordancy of each site will be:

$$D_j = \frac{NS}{3} (\mathbf{u}_j - \bar{\mathbf{u}})^{\mathrm{T}} \cdot \mathbf{A}^{-1} \cdot (\mathbf{u}_j - \bar{\mathbf{u}}) \quad (12)$$

So, when D_j is greater than the critical values (D_c) the site will be discordant with the group.

Seasonality indices

Zrinji and Burn (1996), Burn (1997), Cunderlik and Burn (2002), and De-Michele and Rosso (2002) have suggested that the mean date of flood occurrences and the seasonal regularity can be used as measures of *similarity* of the hydrological response of a basin. Therefore, such basins can be integrated into a region for regional frequency analysis. This regionalization approach has the advantage of reserving the use of hydrometric information to verify regional homogeneity. In this *VEA* study, the average occurrence date of the maximum monthly runoff volume of each year and the seasonal regularity were evaluated.

Directional statistics date from the early 1970s and are a simple tool to define measures of similarity from the dates of occurrence of the events analyzed. In Mexico they were introduced by Ramírez–Orozco, Gutiérrez–López and Ruiz-Silva (2009).

There are various conventions or ways of using the circle to define directional statistics; From now on, the one that is similar to the clock hours will be used, since the 12 months of the year are analyzed, so January is one, June six, September nine, and December twelve.

To establish the seasonality indices, the first step is transforming the occurrence month of each maximum monthly runoff volume in each year, to an advanced angle in degrees (θ_i), with a 30º increase in each month; with 15º for January, with 75º for March, 135º for May, 225º for August, 285º for October, and so on.

Taking *n* data from the month in which the maximum monthly runoff volume of each year occurred, the average coordinates of the mean angle ($\bar{\theta}$) of the monthly maximums will be:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} sen(\theta_i) \quad (13)$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} cos(\theta_i) \tag{14}$$

Therefore, the *mean direction* $\bar{\theta}$ of the average occurrence of the months of the maximum monthly runoff in polar coordinates will be:

$$\bar{\theta} = tan^{-1}\left(\frac{\bar{x}}{\bar{y}}\right) \tag{15}$$

The application of the previous equation is carried out by first obtaining the angle alpha (α) tangent of $\bar{x}$ between and $\bar{y}$, both with a positive sign; then if $\bar{x}$ and $\bar{y}$ are positive $\bar{\theta} = \alpha$, if $\bar{x}$>0 and $\bar{y}$<0 $\bar{\theta}$=180°−α, if both are negative $\bar{\theta}$=180°+α and finally, if $\bar{x}$<0 y $\bar{y}$>0 $\bar{\theta}$=360°−α.

Basins with similar values of $\bar{\theta}$, can be expected to show similarities in other important hydrological characteristics. The angle $\bar{\theta}$ may be related to the size of the basin and its geographical location within the analyzed region.

A measure of the *variability* of the *n* occurrence dates of the maximum monthly runoff of each year, to the $\bar{\theta}$, can be estimated by calculating the resulting mean, which is expressed as:

$$\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2} \tag{16}$$

The *seasonality index* is a dimensionless measure of the data dispersion, it takes values between zero and one. A unit value indicates that all the maximum monthly runoffs occur in the same month, while a value close to zero implies great variability of occurrences throughout the year.

Ordinary moments linearity test

Gupta, Mesa, and Dawdy (1994) found that if a region is homogeneous concerning the hydrological variable X_i, then its ordinary moments (MO_s) show logarithmic linearity with the basin areas (A). Therefore, the following expression (Ferro & Porto, 2006) allows for verifying such linearity:

$$MO_s = \frac{\sum_{i=1}^{n}[X_i^s(A)]}{n} \text{ con } s = 1,2,.3,4 \tag{17}$$

being, s the order of the moment, n the number of elements in the record and $X_i^s(A)$ the data associated with the value of the basin area. Taking to a semilogarithmic graph, the basin sizes in the abscissa in the logarithmic scale and the ordinate, in arithmetic scale, the decimal logarithms of MO_s, the slopes of each grouping or cloud of moments of order s can be calculated, using the following type of linear regression lines:

$$MO_s = b + m \cdot (\log A) \tag{18}$$

Test of the linear regression of the β_r

Kumar, Guttarp, and Foufoula-Georgiu (1994) proposed a test that verifies regional hydrological homogeneity, based on an equation similar to the previous one, but that uses weighted probability moments (β_r). This test was applied by De-Michele and Rosso (2002), and Ferro and Porto (2006). A version associated with the concept of Kumar *et al*. (1994) was used and verified by Varas (2000) in the U.S.A., France, England, and Chile, for the linear relations of β_r of a certain order and the immediate lower order. This means that the following linear regression is fulfilled in a hydrologically homogeneous region:

$$\beta_r = b + m \cdot \beta_{r-1} \text{ con } r = 1,2,\ldots,4 \tag{19}$$

in which, b and m are the ordinates to the origin and the slope of the linear regression line (equations (22) and (23)), whose quality of fit is defined by the linear correlation coefficient (r_{xy}):

$$r_{xy} = \frac{\sum_{i=1}^{n} (\beta_{r-1}^{i} - \bar{\beta}_{r-1}) \cdot (\beta_{r}^{i} - \bar{\beta}_{r})}{\left[\sum_{i=1}^{n} (\beta_{r-1}^{i} - \bar{\beta}_{r-1})^2 \cdot (\beta_{r}^{i} - \bar{\beta}_{r})^2\right]^{1/2}} \quad (20)$$

The regression line for a linear trend

The dependent variable (y) is considered to be the annual runoff volumes VEA_i in Mm^3 and the times or years T_i is the abscissa (x), in this case, equal to the i–th value i. To test whether the slope (m) of the regression line adjusted by least-squares of the residuals is statistically different from zero, a test based on the Student's t distribution defined by the following equations is used (Ostle & Mensing, 1975):

$$VEA_i = b + m \cdot T_i \quad (21)$$

$$b = \overline{VEA} - m \cdot \bar{T} \tag{22}$$

$$m = \frac{\frac{1}{n}\sum_{i=1}^{n} VEA_i \cdot i - \overline{VEA} \cdot \bar{T}}{\frac{1}{n}\sum_{i=1}^{n} i^2 - \bar{T}^2} \tag{23}$$

$$t = \frac{m}{\sqrt{S_m^2}} \tag{24}$$

$$S_m^2 = \frac{S_E^2}{\sum_{i=1}^{n}(T_i - \bar{T})^2} \tag{25}$$

$$S_E^2 = \frac{\sum_{i=1}^{n}(VEA_i - \widehat{VEA}_i)^2}{(n-2)} \tag{26}$$

$\overline{VEA}$ and $\bar{T}$ are the means; in the previous equation, $\widehat{VEA}_i$ is the value estimated with the regression line (Equation (21)). S_E^2 and S_m^2 are the variances of the errors and the slope. If the calculated value t is greater than the critical one (t_c), obtained for the student's t distribution with n – 2 degrees of freedom and α = 5 %, in a two-tailed test, the slope m is significant, that is, there is a linear trend. The test has the limitation of not distinguishing between persistence and trend (Adeloye & Montaseri, 2002). To estimate the value of t_c, the algorithm proposed by Zelen and Severo (1972) is used, with Z = 1.95996 for reliability (1 − α) of 95 %:

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

$$t_c = Z + G1/v + G2/v^2 + G3/v^3 + G4/v^4 \tag{27}$$

where:

$G1 = (Z^3 + Z)/4$

$G2 = (5Z^5 + 16Z^3 + 3Z/96)$

$G3 = (3Z^7 + 19Z^5 + 17Z^3 - 15Z)/384$

$G4 = (79Z^9 + 776z^7 + 1482Z^5 - 1920Z^3 - 945Z)/92160$

Available hydrometric information

In Hydrological Region No. 10 (Sinaloa), Mexico there are 42 hydrometric stations with reliable records and only 22 have their runoff regime unaffected by reservoirs and records older than 20 years, therefore, they were used. In the BANDAS system (IMTA, 2002) all the information on annual runoff volumes of such gauging stations was obtained, the general characteristics of which are presented in Table 2, shown in decreasing order of basin size. This information comes from CD–1, it is displayed on a monthly basis and is processed, as already indicated, in millions of cubic meters (Mm^3). As expected, the records of the Guamúchil and Huites hydrometric stations cover even before the year their respective reservoir construction (Eustaquio Buelna and Luis Donaldo Colosio) affected their regime.

Table 2. General characteristics of the 22 hydrometric stations processed in Hydrological Region No. 10 (Sinaloa), Mexico.

No.	Name	River or creek	RP y (NA)	*A* (km^2)	*VEMA* (Mm^3)
1	Huites	R. Fuerte	*1942–1992 (51)	26 057	4 314.818
2	San Francisco	R. Fuerte	*1942–1973 (32)	17 531	2 469.040
3	Santa Cruz	R. San Lorenzo	1944–1987 (44)	8 919	1 655.043
4	Guatenipa II	R. Humaya	*1969–2001 (33)	8 252	1 563.583
5	Jaina	R. Sinaloa	*1942–1998 (57)	8 179	1 372.602
6	Palo Dulce	R. Chinipas	1968–1982 (25)	6 439	932.573
7	Ixpalino	R. Piaxtla	1953–1988 (36)	6 166	1 469.099
8	La Huerta	R. Humaya	*1970–1999 (30)	6 149	1 108.251
9	Toahayana	R. Petatlán	1958–1985 (28)	5 281	1 101.878
10	Urique II	R. Urique	*1968–1994 (27)	4 000	434.820
11	Tamazula	R. Tamzula	1963–1992 (30)	2 241	620.315
12	Naranjo	A. Ocoroni	*1939–1984 (46)	2 064	202.032
13	Acatitán	R. Elota	1955–1994 (40)	1 884	417.784
14	Guamúchil	R. Mocorito	*1939–1971 (33)	1 645	128.734
15	Choix	R. Choix	1955–1985 (31)	1 403	294.790

16	Badiraguato	R. Badiraguato	*1960–1999 (40)	1 018	252.900
17	El Quelite	R. Quelite	1961–1992 (32)	835	148.317
18	Zopilote	A. Cabrera	1939–1986 (48)	666	83.262
19	El Bledal	A. El Bledal	*1938–1994 (57)	371	45.637
20	Pericos	A. Pericos	*1961–1992 (32)	270	47.586
21	La Tina	A. Sivajahui	*1960–1983 (24)	254	7.545
22	Bamícori	A. Baroten	*1951–1983 (33)	223	12.672

Description of acronyms:

RP= years of registration processed.

NA= number of years.

A= basin area.

VEMA= mean annual runoff volume.

*Continuous records.

In the fourth column of Table 2, the 13 records that are continuous have been indicated with an asterisk, that is, they do not present doubts in defining their span of full years. In the remaining nine, which are: Santa Cruz, Palo Dulce, Ixpalino, Toahayana, Tamazula, Acatitán, Choix, El Quelite and Zopilote, there are two or more missing years, generally after the year 1980 and therefore, there were difficulties in defining the period of years to process.

Figure 1 shows the geographic location of the 22 basins that define the hydrometric stations of Hydrological Region No. 10

(Sinaloa), Mexico, cited in Table 2, which will be processed in your *VEA* registry.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

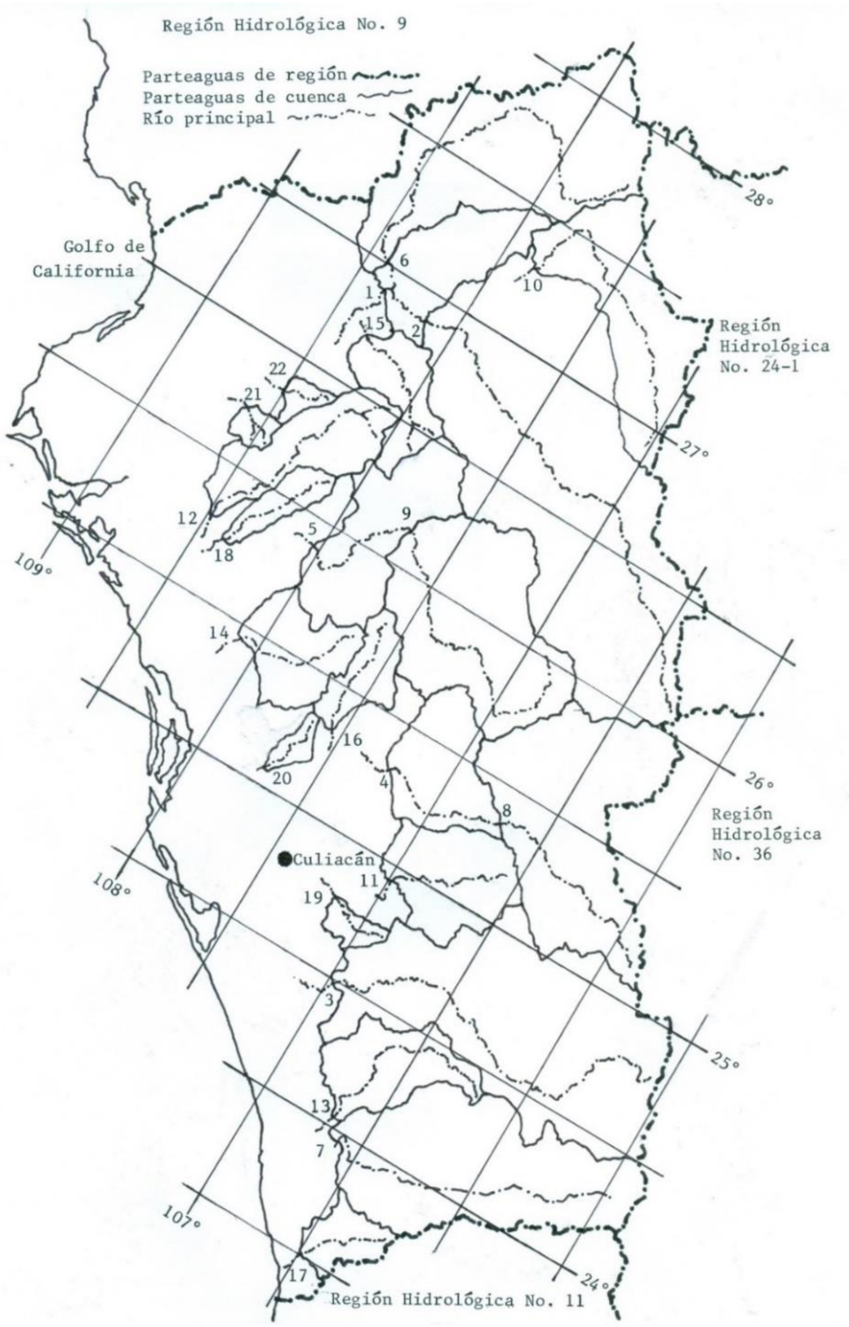

Figure 1. Location in Hydrological Region No. 10 (Sinaloa), Mexico, of the basins of the 22 hydrometric stations processed.

Lapses of records to process

In the definition of the recording periods of annual runoff volume to be processed, two work approaches can be defined: (1) as independent and identically distributed data records and (2) as chronological series. The first format complies with the requirement of frequency analysis and allows the selection of records interrupted by one or more years. The second format involves preserving the sequences of wet and dry years, and this is how runoff should be studied; therefore, it was adopted. The fourth column of Table 2 indicates the *continuous recording* lapses that will be processed.

Once such continuous periods were defined, the missing monthly data for the incomplete years, as well as the missing years, were calculated, as detailed in the following section, to define the *annual mean runoff volume* (*VEMA*), which is cited in the last column of Table 2.

Deduction of missing data

To estimate the missing monthly runoff volumes in incomplete years, a simple method was followed, which consisted of averaging the same month of the previous and subsequent year, to assign such value to the missing month. This is considered acceptable due to the regularity of the runoff season year after year. With this procedure, the records were completed at the stations: San Francisco, Santa Cruz, Guatenipa II, Jaina, Palo Dulce, La Huerta, Toahayana, Naranjo, Choix, and Pericos; in the one or two years.

For the missing years, the weighted transport was applied with the inverse of the distance between hydrometric stations (Campos-Aranda, 2008), using two or more records. Such was the case in 1987 at the El Quelite station, based in Ixpalino and Acatitán; as well as from the year 1985 of Tamazula, using Santa Cruz and Guatenipa II.

When the weighted transport could not be applied, the missing year was deducted based on the mode of the annual register, estimated with the Mixed Gamma distribution; this was the case in 1983 in Urique II, using 26 annual values.

Regional method to be applied

The fourth column of Table 2 highlights the great variety of start and end dates of the annual runoff volume (*VEA*) records. The later highlights the impossibility of defining a common period, of adequate amplitude (> 30 years), of *VEA* records to be processed in the region analyzed with the IE^j method.

Given this situation, it was decided to apply the regional method of stations–years, to process the records that will form a homogeneous region from a hydrological approach.

Search for regional PDF 1: Pearson type III (PT3)

This probabilistic model has three parameters: μ (location), σ (scale), and γ (shape). If $\gamma > 0$, the variation of X is: $\xi \leq X < \infty$ and we have the most common or widespread PT3 distribution; if $\gamma = 0$ the Normal distribution occurs whose range of X is: $-\infty < X < \infty$ and when $\gamma < 0$ the reverse PT3 occurs, with an interval of X of $-\infty < X \leq \xi$ (Hosking & Wallis, 1997). This PDF does not have an explicit solution, so the value X associated with a probability of non–exceedance (F) is estimated through the frequency factor (K_f), with the General equation of Hydrological Frequency Analysis (Chow, 1964):

$$X(F) = \mu + K_f \cdot \sigma \tag{28}$$

the expression for K_f is (Kite, 1977):

$$F_f = Z + (Z^2 - 1)\gamma_c + \frac{1}{3}(Z^3 - 6Z)\gamma_c^2 - (Z^2 - 1)\gamma_c^3 + Z \cdot \gamma_c^4 + \frac{1}{3}\gamma_c^5 \tag{29}$$

in which, $\gamma_c = \gamma/6$ and Z is the standard normal deviation function of F, it is estimated with the following approximation (Zelen & Severo, 1972):

$$Z = w - \frac{2.515517 + 0.802853 \cdot w + 0.010328 \cdot w^2}{1 + 1.432788 \cdot w + 0.189269 \cdot w^2 + 0.001308 \cdot w^3} \tag{30}$$

being:

$$w = \sqrt{\ln(1/F^2)} \tag{31}$$

when $F > 0.50$, $F = 1 - F$ is used in Equation (31) and the result of Equation (30) the sign is changed.

The sign of τ_3 defines that of the asymmetry coefficient (γ) in Equation (36). If $0 < |\tau_3| < 1/3$ we make $z = 3\pi(\tau_3)^2$ to lead to the following expression (Hosking & Wallis, 1997):

$$\alpha \cong \frac{1+0.2906\cdot z}{z+0.1882\cdot z^2+0.0442\cdot z^3} \tag{32}$$

If $1/3 \leq |\tau_3| < 1$ we make $z = 1 - |\tau_3|$ and applies:

$$\alpha \cong \frac{0.36067\cdot z-0.59567\cdot z^2+0.25361\cdot z^3}{1-2.78861\cdot z+2.56096\cdot z^2-0.77045\cdot z^3} \tag{33}$$

Finally:

$$\mu = \lambda_1 \tag{34}$$

$$\sigma = \lambda_2 \cdot \sqrt{\pi \cdot \alpha} \cdot \Gamma(\alpha)/\Gamma(\alpha + 1/2) \tag{35}$$

$$\gamma = 2/\sqrt{\alpha} \tag{36}$$

To evaluate the Gamma function, the Stirling approximation (Davis, 1972) was used:

$$\Gamma(\omega) \cong e^{-\omega} \cdot \omega^{\omega-\frac{1}{2}} \cdot \sqrt{2\pi} \cdot f_1 \tag{37}$$

being:

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

$$f_1 = \left(1 + \frac{1}{12\omega} + \frac{1}{288\omega^2} - \frac{139}{51840\omega^3} - \frac{571}{2488320\omega^4} + \cdots\right)$$

Search for regional PDF 2: Log-Normal (LN3)

In this probabilistic model, the parameters are ξ (location), α (scale), and k (shape). In the Hosking and Wallis (1997) array of variables, the interval of the variable X is: $-\infty < X \leq \xi + \alpha/k$ when $k > 0$, $-\infty < X < \infty$ if $k=0$ (Normal distribution) and $\xi + \alpha/k \leq X < \infty$ when $k<0$; furthermore, X is related to the random variable Z that has a standard Normal distribution (Equation (30)), defined as:

$$X(F) = \xi + \frac{\alpha}{k}\left(1 - e^{-kZ}\right) \text{ when } k \neq 0 \tag{38}$$

The shape parameter k is only a function of τ_3, with the following valid relation for $|\tau_3| \leq 0.94$, corresponding to $|k| \leq 3$:

$$k \approx -\tau_3 \frac{2.0466534 - 3.6544371 \cdot \tau_3^2 + 1.8396733 \cdot \tau_3^4 - 0.20360244 \cdot \tau_3^6}{1 - 2.01582173 \cdot \tau_3^2 + 1.2420401 \cdot \tau_3^4 - 0.21741801 \cdot \tau_3^6} \tag{39}$$

The other parameters are:

$$\xi = \lambda_1 - \frac{\alpha}{k}\left(1 - e^{k^2/2}\right) \tag{40}$$

$$\alpha = \frac{\lambda_2 \cdot k \cdot e^{-k^2/2}}{1 - 2\Phi\left(-k/\sqrt{2}\right)} \tag{41}$$

In the above equation, $\Phi(x)$ is the probability of non–exceedance up to the value *x* of the Normal distribution, equal to:

$$\Phi(x) = \int_{-\infty}^{x} \phi(x)dt \tag{42}$$

where, $\phi(x)$ is the ordinate of the Normal curve at *x*, equal to:

$$\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \tag{43}$$

To evaluate Equation (42), the following approximation was used (Zelen & Severo, 1972):

$$\Phi(x) = 1 - \phi(x) \cdot (c_1 \cdot d + c_2 \cdot d^2 + \cdots + c_5 \cdot d^5) \tag{44}$$

in which:

$$d = \frac{1}{1+0.2316419 \cdot x} \qquad (45)$$

$c_1 = 0.31938153$, $c_2 = -0.356563782$, $c_3 = 1.781477937$, $c_4 = -1.821255978$ y $c_5 = 1.330274429$.

Search for regional PDF 3: Potential Transformation (TP2)

Chander, Spolia, and Kumar (1978) suggest using the Box–Cox transformation to normalize the data X_i to z_i, and then, applying Equation (28) with a Normal distribution and finally, obtaining the desired estimate $X(F)$, with the inverse transformation. The expressions cited are:

$$z_i = \left(X_i^{\varphi} - 1\right)/\varphi \text{ when } \varphi \neq 0, X_i > 0 \qquad (46)$$

$$z_i = \ln(X_i) \text{ when } \varphi = 0, X_i > 0 \qquad (47)$$

Equation (46) is a general potential transformation, in which the logarithmic, reciprocal, and square root conversions are special cases since φ is a constant ranging from −1.0 to 1.0. Obtaining φ is not explicit, but an increase or decrease causes the same effect in the coefficient of asymmetry (*Cs*) of the z_i, then its optimal value will be the one that takes the *Cs* to zero.

Escalante–Sandoval, and Reyes–Chávez (2002) indicate that Equation (46) normalizes the $Cs \cong 0$, but not the kurtosis coefficient ($Ck \neq 3$). The correction of the kurtosis is achieved with a second transformation, exposed by these authors. According to Chander *et al.* (1978) in frequency analysis, the second transformation is not necessary. Equation (28) is now:

$$z(F) = zm + Z \cdot zs \qquad (48)$$

in which, *zm* and *zs* are the mean and standard deviation of the transformed data, and *Z* is estimated with Equation (30). The estimate sought is obtained with the expression:

$$X(F) = [\varphi \cdot z(F) + 1]^{1/\varphi} \qquad (49)$$

Standard error of fit

According to Chai and Draxler (2014), this is the most common indicator for contrasting probability distributions to real data, it was established in the mid-seventies (Kite, 1977). The expression for the standard error of fit (*EEA*, by its acronym in Spanish) is:

$$EEA = \left[\frac{\sum_{i=1}^{n}(VEA_i - \widehat{VEA}_i)^2}{(n-np)}\right]^{1/2} \quad (50)$$

VEA_i is the annual runoff volumes ordered from lowest to highest, whose number is *n* and $\widehat{VEA}_i$ the estimated annual runoff volumes, for the probability of non–exceedance (*F*) estimated with Equation (51) or Cunnane's formula, which according to Stedinger (2017) leads to unbiased estimates in most of the PDFs used in hydrology or probabilistic model that is contrasted. Finally, *np* is the number of fit parameters for the PDF, with three for the PT3 and LN3 and two for the TP2.

$$F = \frac{i-0.40}{n+0.20} \quad (51)$$

Tecnología y Ciencias del Agua

Results and their analysis

Homogeneity tests

To make the results of a regional frequencies analysis reliable, the records of annual runoff volume (*VEA*) to process must have been generated by a *stationary random process*, which implies that it has not changed over time. Therefore, *VEA* records must consist of independent data, free from deterministic components.

To verify the above, seven statistical tests were applied, one general, the Von Neumann Test, and six specifics: two for persistence (Anderson and Sneyers), two for trend (Kendall and Spearman), one for variability (Bartlett), and last, related to the change in the mean (Cramer). These tests can be consulted in WMO (1971), and Machiwal and Jha (2012). All the aforementioned tests were applied with a significance level (α) of 5 % and their results are shown in Table 3, indicating NH loss of homogeneity detected.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Table 3. Two statistical parameters and results of the seven homogeneity tests, in the 22 hydrometric stations of the Hydrological Region No. 10 (Sinaloa), Mexico.

No.	Station	*Cv*	r_1	Homogeneity tests						
				CVN	TA	TS	PK	PS	TB	PC
1	Huites	0.397	0.266	NH	NH	NH	NH	NH	H	NH
2	San Francisco	0.429	−0.041	H	H	H	H	H	H	H
3	Santa Cruz	0.417	0.107	H	H	H	NH	NH	H	NH
4	Guatenipa II	0.419	0.336	NH	NH	NH	H	H	H	H
5	Jaina	0.422	0.216	NH	NH	NH	H	H	H	H
6	Palo Dulce	0.305	−0.021	H	H	H	H	H	H	H
7	Ixpalino	0.378	−0.170	H	H	H	H	H	H	H
8	La Huerta	0.492	0.486	NH	NH	NH	H	H	H	H
9	Toahayana	0.377	−0.141	H	H	H	H	H	H	H
10	Urique II	0.544	0.381	NH	NH	NH	H	H	H	H
11	Tamazula	0.321	−0.092	H	H	H	H	H	H	H
12	Naranjo	0.964	0.345	NH	NH	NH	NH	NH	NH	H
13	Acatitán	0.562	−0.212	H	H	H	H	H	H	H
14	Guamúchil	0.734	−0.018	H	H	H	H	H	H	H
15	Choix	0.395	0.297	NH	NH	H	H	H	H	H
16	Badiraguato	0.611	0.222	H	H	NH	NH	NH	NH	NH
17	El Quelite	1.650	0.032	H	H	H	H	H	NH	H

18	Zopilote	1.009	0.031	H	H	H	NH	NH	H	H
19	El Bledal	0.695	0.039	H	H	H	H	H	H	H
20	Pericos	1.776	0.322	NH	NH	NH	NH	NH	NH	NH
21	La Tina	0.833	0.017	H	H	H	H	H	H	H
22	Bamícori	0.590	−0.204	H	H	H	H	H	H	H

Description of acronyms:

Cv= coefficient of variation $(S/\bar{X})$, dimensionless.

r_1 = coef. of serial correlation of order one, dimensionless.

CVN= Von Neumann quotient.

TA= Anderson test.

PS= Spearman test.

TS= Sneyers test.

TB= Bartlett test.

PK= Kendall test.

PC= Cramer test.

It was found that the following six hydrometric stations showed an upward trend: Huites, Santa Cruz, Naranjo, Badiraguato, Zopilote, and Pericos. Huites, Naranjo and Pericos also show persistence, along with the following five: Guatenipa II, Jaina, La Huerta, Urique II, and Choix. It is observed that persistence is associated with positive values of the serial correlation coefficient of order one (r_1) higher than 0.20, except in Badiraguato.

Discordancy test

After defining the 22 records to be processed (Table 2), equations (2) to (9) were applied to obtain the ratios of L moments, shown in Table 4.

Table 4. L-moment ratios in the 22 hydrometric stations of the Hydrological Region No. 10 (Sinaloa), Mexico.

No.	Name	L ratios		
		t_2	t_3	t_4
1	Huites	0.22402	0.13399	0.10546
2	San Francisco	0.23810	0.19358	0.12958
3	Santa Cruz	0.22581	0.10956	0.17862
4	Guatenipa II	0.23597	0.10004	0.11245
5	Jaina	0.23033	0.17797	0.18571
6	Palo Dulce	0.17121	0.24359	0.04599
7	Ixpalino	0.21747	0.02728	0.10958
8	La Huerta	0.26480	0.27554	0.15224
9	Toahayana	0.21656	0.06209	0.09516

10	Urique II	0.29308	0.25481	0.15198
11	Tamazula	0.18244	0.11958	0.14194
12	Naranjo	0.47655	0.32689	0.19441
13	Acatitán	0.29624	0.20962	0.21919
14	Guamúchil	0.36576	0.37388	0.25532
15	Choix	0.21305	0.18620	0.22018
16	Badiraguato	0.32196	0.27868	0.20039
17	El Quelite	0.48463	0.59616	0.56783
18	Zopilote	0.49516	0.33737	0.21933
19	El Bledal	0.36091	0.22302	0.08793
20	Pericos	0.65688	0.64361	0.45957
21	La Tina	0.45387	0.28892	0.16375
22	Bamícori	0.33300	0.16676	0.11363

Then, based on the values in Table 4 and equations (10) to (12), the Discordance test was developed, which allowed to obtain the results shown in Table 5. The test revealed that the Palo Dulce records (D_6 = 3.62) and El Quelite (D_{17} = 4.12) exceeded the critical value of the Discordance of 3.00 (Table 1) and were therefore discordant. The Pericos record showed a high value (D_{20} = 2.71) and therefore, it was also discarded.

Table 5. Results of the Discordancy Test (D_j) in the 22 hydrometric stations of the Hydrological Region No. 10 (Sinaloa), Mexico.

No.	Name	D_j	No.	Name	D_j
1	Huites	0.21	12	Naranjo	1.04
2	San Francisco	0.25	13	Acatitán	0.27
3	Santa Cruz	0.82	14	Guamúchil	0.44
4	Guatenipa II	0.37	15	Choix	0.74
5	Jaina	0.32	16	Badiraguato	0.06
6	Palo Dulce	3.62*	17	El Quelite	4.12*
7	Ixpalino	1.32	18	Zopilote	1.14
8	La Huerta	0.78	19	El Bledal	0.85
9	Toahayana	0.62	20	Pericos	2.71
10	Urique II	0.26	21	La Tina	1.06
11	Tamazula	0.50	22	Bamícori	0.49

*exceeds critical value.

About the quotient τ_3, it is observed in Table 4, that the value of Palo Dulce was very low and those of El Quelite and Pericos very high, in contrast with their surrounding values by basin size, in the order of 0.22 and 0.33, respectively. On the other hand, in Table 3 and relation to the value of *Cv*, the magnitude of Palo Dulce was very low and those of El Quelite and Pericos quite high; in other words, discordant.

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

Regional homogeneity 1: Seasonality indices

For each year of the 19 records processed (Table 2), the month with the maximum runoff was detected. Such data were processed based on equations (13) to (16) and the results were obtained, which are summarized in Table 6 and drawn in Figure 2. A great seasonality is observed, since in most of the records its average direction ($\bar{\theta}$) occurs in the second fortnight of August and the first of September, with angles that fluctuate between 230.9º in Bamícori (asterisk) and 259º in Urique II (black star).

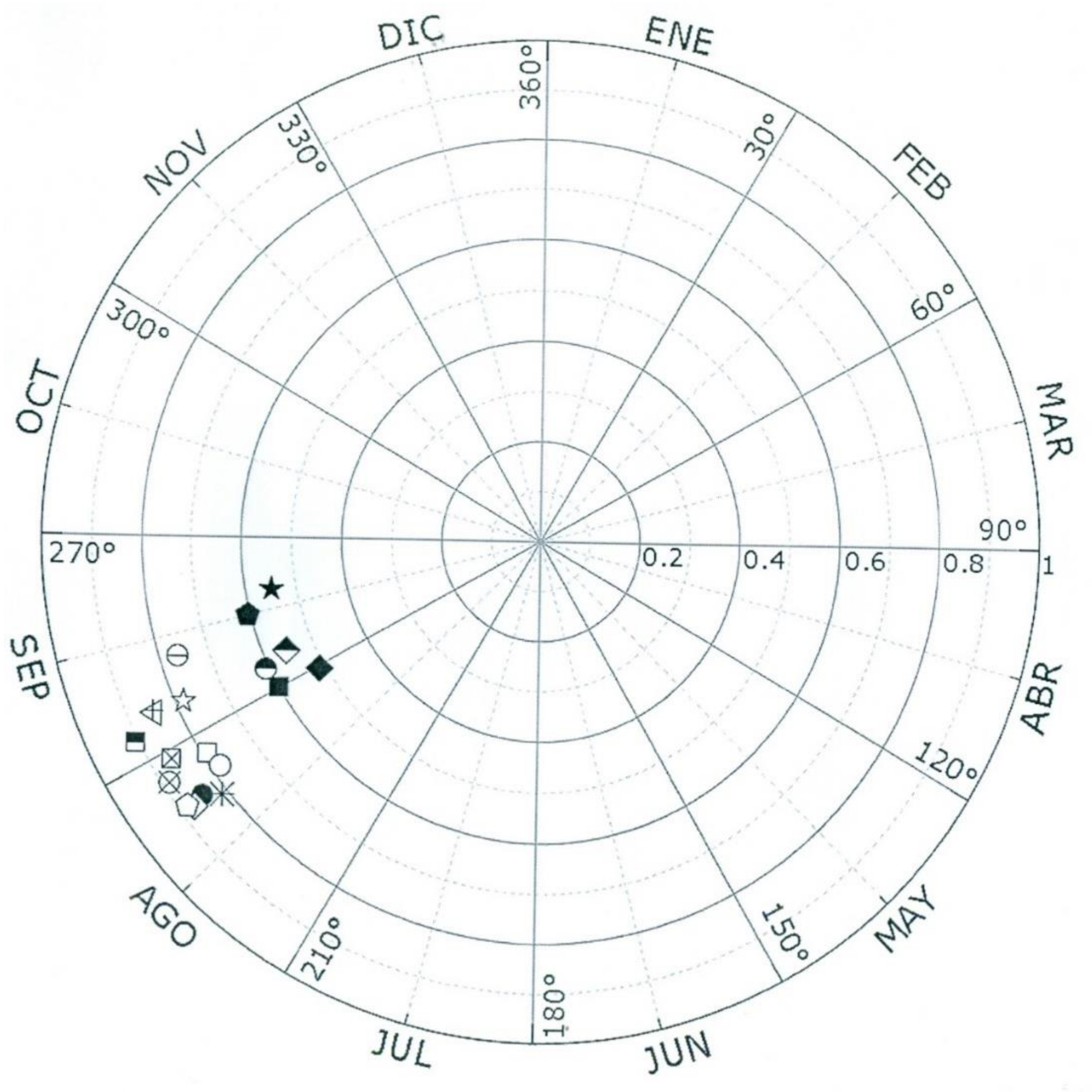

Figure 2. Polar representation of the seasonality indices, in the 19 hydrometric stations of the Hydrological Region No. 10 (Sinaloa), Mexico.

Table 6. Seasonality indices in the 19 hydrometric stations of Hydrological Region No. 10 (Sinaloa), Mexico.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

No.	Name	$(\bar{\theta})$	$(\bar{r})$	No.	Name	$(\bar{\theta})$	$(\bar{r})$
1	Huites	240.1	0.5996	11	Naranjo	234.3	0.7819
2	San Francisco	232.4	0.8444	12	Acatitán	245.7	0.8476
3	Santa Cruz	242.9	0.9055	13	Guamúchil	236.3	0.8857
4	Guatenipa II	244.3	0.6064	14	Choix	231.9	0.8706
5	Jaina	239.3	0.5105	15	Badiraguato	251.7	0.7626
6	Ixpalino	245.3	0.8479	16	Zopilote	238.8	0.8572
7	La Huerta	246.0	0.5544	17	El Bledal	232.5	0.8817
8	Toahayana	255.0	0.6046	18	La Tina	245.1	0.7836
9	Urique II	259.0	0.5474	19	Bamícori	230.9	0.8121
10	Tamazula	236.9	0.7892	–	–	–	–

Regarding variability, the least dispersion is found in Santa Cruz (black-white box) and the greatest in Jaina (black diamond). It is also observed that six stations have a seasonality index of the order of 0.60 or less, they are Huites (black box), Guatenipa II (black-white circle), Jaina, La Huerta (black-white rhombus), and Toahayana (black pentagon) and Urique II. These six stations have large basins greater than 4000 km2.

Based on the results of Table 6, shown in Figure 2, it is concluded that the 19 *VEA* records were processed to form a homogeneous region from a hydrological point of view.

Regional homogeneity 2: Linearity of ordinary moments

For each of the 19 records to be processed (Table 2), Equation (17) was applied to obtain their ordinary moments of order 1 to 4; the natural logarithm was applied to these values to obtain magnitudes easier to chart. Such results are shown in Table 7.

Table 7. Natural logarithms of ordinary moments (MO_s) in the 19 hydrometric stations of Hydrological Region No. 10 (Sinaloa), Mexico.

Station No.	Name	MO_1	MO_2	MO_3	MO_4
1	Huites	8.370	16.883	25.517	34.250
2	San Francisco	7.812	15.788	23.908	32.149
3	Santa Cruz	7.412	14.980	22.696	30.559
4	Guatenipa II	7.355	14.867	22.506	30.251
5	Jaina	7.224	14.610	22.138	29.787
6	Ixpalino	7.292	14.715	22.232	29.820
7	La Huerta	7.011	14.231	21.646	29.219

8	Toahayana	7.005	14.138	21.370	28.678
9	Urique II	6.075	12.400	18.936	25.633
10	Tamazula	6.430	12.955	19.566	26.249
11	Naranjo	5.308	11.264	17.656	24.332
12	Acatitán	6.035	12.339	18.884	25.642
13	Guamúchil	4.858	10.126	15.761	21.605
14	Choix	5.686	11.513	17.475	23.567
15	Badiraguato	5.533	11.376	17.478	23.766
16	Zopilote	4.422	9.536	15.074	20.852
17	El Bledal	3.821	8.029	12.562	17.378
18	La Tina	2.021	4.551	7.346	10.284
19	Bamícori	2.539	5.370	8.387	11.529
Ordered to the origin (b)		−3.723	−6.554	−8.837	−10.787
Slope (m)		2.823	5.466	8.010	10.502
Coeff. correlation (r_{xy})		0.964	0.965	0.964	0.963

In Figure 3, there is a semi-logarithmic graph, where the ordinary moments of Table 7 have been drawn in the ordinate axis, and a logarithmic scale for the basin areas is charted in the abscissa. The values of b and m of Equation (18) are shown in the respective column of s in Table 7. The internal graph of Figure 3 shows the linearity of the slopes (m) and the values of r_{xy} confirm that the 19 hydrometric stations analyzed to form a hydrologically homogeneous region.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

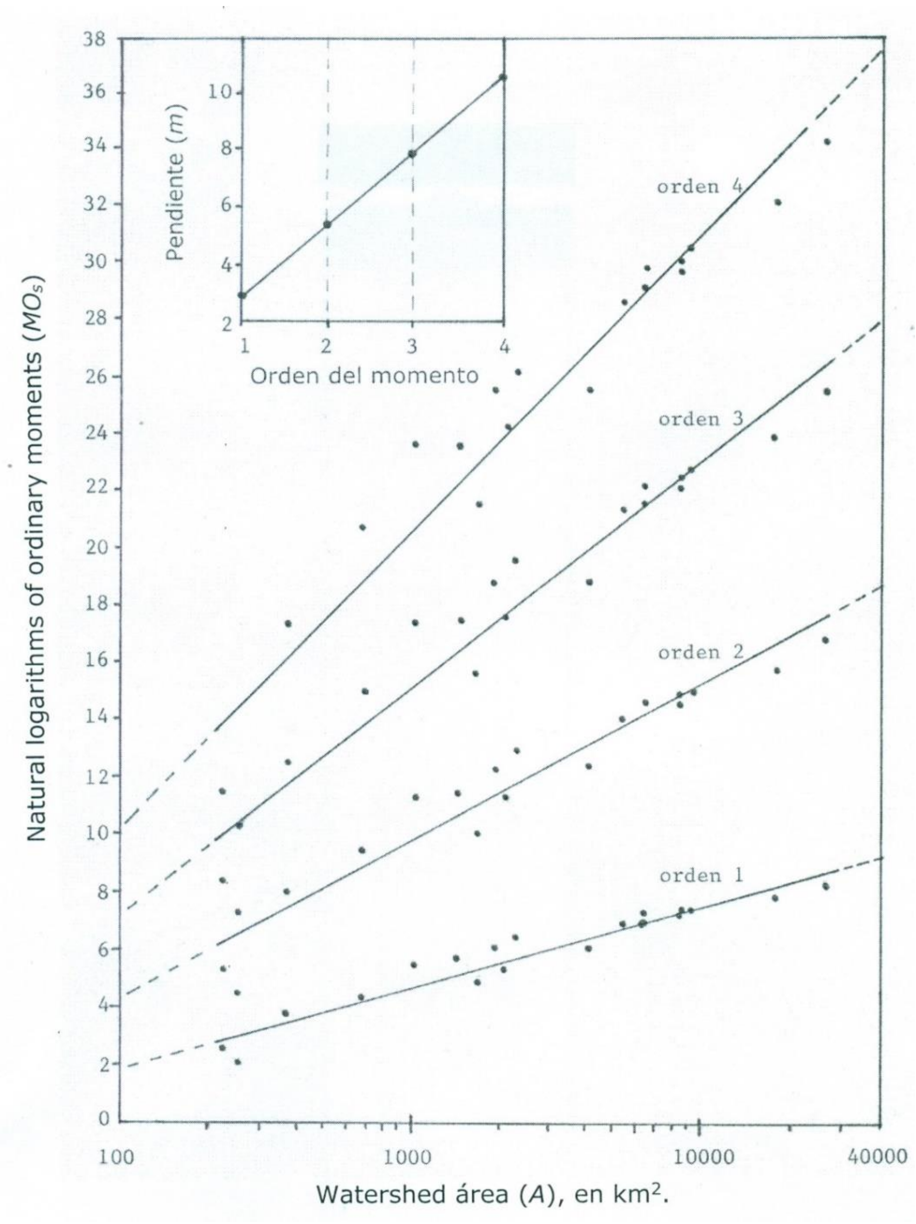

Figure 3. Relationship between ordinary moments and basin area in the 19 hydrometric stations of Hydrological Region No. 10 (Sinaloa), Mexico.

Regional homogeneity 3: Linear regression of the β_r moments

Now, for each of the 19 records processed (Table 2), Equation (9) was applied with *r* from 1 to 4, since *r* = 0 is mean and it is shown in Table 2. Such results are shown in Table 8.

Table 8. Weighted Probability Moments (β_r) in the 19 hydrometric stations of Hydrological Region No. 10 (Sinaloa), Mexico.

Station No.	Name:	β_1	β_2	β_3	β_4
1	Huites	2640.703	1943.151	1551.143	1297.210
2	San Francisco	1528.455	1135.915	914.060	769.370
3	Santa Cruz	1014.385	745.368	595.512	499.392
4	Guatenipa II	966.274	711.829	568.232	475.519
5	Jaina	844.379	624.990	502.423	423.236
6	Ixpalino	894.292	650.894	514.972	427.765
7	La Huerta	700.858	529.626	431.571	366.896

8	Toahayana	670.252	489.076	387.691	322.538
9	Urique II	281.128	214.070	175.137	149.325
10	Tamazula	366.742	265.611	210.191	174.859
11	Naranjo	149.155	120.729	102.637	90.002
12	Acatitán	270.775	205.468	167.983	143.351
13	Guamúchil	87.910	69.388	58.374	50.916
14	Choix	178.797	131.614	105.574	88.851
15	Badiraguato	167.162	128.794	106.354	91.385
16	Zopilote	62.245	50.686	43.298	38.132
17	El Bledal	31.054	24.060	19.812	16.934
18	La Tina	5.485	4.392	3.703	3.220
19	Bamícori	8.446	6.451	5.267	4.473
Ordered to the origin (*b*)		7.4878	3.2501	2.0339	1.4745
Slope (*m*)		0.6110	0.7352	0.7980	0.8362
Coeff. Correlation (r_{xy})		0.9999	0.9999	0.9999	0.9999

The results of applying Equation (19) are shown in the respective column of *r* in Table 8, so, in the first one, there is the linear relationship between β_1 and β_0, and in the last column, the one between β_4 and β_3.

Based on the value of r_{xy} (Equation (20)), it is observed that such relationships are almost perfect. Therefore, this method also guides us to conclude that the 19 hydrometric stations in Table 8 form a homogeneous region from a hydrological perspective.

Elimination of non–homogeneous records

In Table 3, the results of the statistical tests application of local or individual homogeneity showed that six stations or processed records have an upward trend, these are Huites, Santa Cruz, Naranjo, Badiraguato, Zopilote, and Pericos. The last one had already been eliminated in the Discordancy test. Persistence was also detected in the first two and the following five: Guatenipa II, Jaina, La Huerta, Urique II, and Choix. These seven stations with *persistence* are analyzed in the next section.

Based on equations (21) to (27), it was tested, in the six records with a trend, if it was significant; only at the Zopilote station was not. Because the trend detected was *upward*, it was decided to discard such records, since their data would tend to increase the annual runoff and this would increase the uncertainty associated with its estimation, through the regional frequency analysis. Due to the above and finally, 15 records of annual runoff volume were then processed to define their regional probability distribution.

Regarding persistence

In the annual runoff, *persistence* is a statistical property that originates in the occurrence of wet and dry years; Therefore, the processing registration periods of the stations with the highest value of the serial correlation coefficient of order one (r_1) in Table 3 were compared: La Huerta (0.486), Urique II (0.381) and Guatenipa II (0.336), to look for similarity in their lapses and in effect it occurs in the years from 1970 to 2000. The other two stations that showed persistence were Jaina (0.216) and Choix (0.297), with different records both in length and in occurrence.

Khaliq, Ouarda, Ondo, Gachon, and Bobée (2006) indicate that the validity of the flood predictions obtained through probabilistic analysis will be questionable if the observations of the record to be processed are *dependent*, that is, it has *persistence*. The foregoing is valid when looking for floods or predictions associated with very low exceedance probabilities since they are obtained at the extreme right of the probability distribution; This is not the case when generating synthetic annual runoff sequences, which cover the entire domain extension of the probabilistic model used (Salas, 1993; Cannarozzo *et* *al*., 2009). Due to the above reasoning, it was not considered necessary to remove the stations with persistence.

Selection of the regional PDF

All the calculations described below were developed with *dimensionless* data, since the annual runoff volumes of each of the 15 records had been divided by their mean value or the *Runoff Index* (IE^j). The three PDFs recommended by Cannarozzo *et al*. (2009) were applied to each dimensionless record, which is: the Pearson type III (equations (28) to (37)), the Log-Normal of three fit parameters (equations (38) to (45)) and the potential transformation (equations (46) to (49)). The results of the PDF that led to the lowest standard error of fit (*EEA*, Equation (50)) have been condensed in Table 9. It is observed that the Pearson type III PDF did not achieve a better fit in any of the 15 records processed.

Table 9. Best fits and estimated values for the indicated deciles in the 15 hydrometric stations of Hydrological Region No. 10 (Sinaloa), Mexico.

Station	PDF	*EEA*	Probability (*F*) of the indicated deciles:				
			0.100	**0.300**	**0.500**	**0.700**	**0.900**
San Francisco	TP2	0.060	0.528	0.733	0.919	1.151	1.591
Guatenipa II	TP2	0.061	0.504	0.745	0.946	1.180	1.576
Jaina	LN3	0.053	0.532	0.743	0.927	1.150	1.558
Ixpalino	TP2	0.035	0.516	0.789	0.988	1.194	1.502
La Huerta	TP2	0.057	0.514	0.695	0.873	1.117	1.663
Toahayana	TP2	0.059	0.534	0.777	0.969	1.179	1.514
Urique II	LN3	0.080	0.439	0.657	0.869	1.148	1.716
Tamazula	TP2	0.034	0.614	0.808	0.965	1.143	1.439
Acatitán	LN3	0.102	0.412	0.662	0.890	1.175	1.722
Guamúchil	TP2	0.074	0.371	0.567	0.781	1.101	1.903
Choix	LN3	0.062	0.570	0.761	0.929	1.135	1.517
Zopilote	LN3	0.166	0.115	0.402	0.713	1.164	2.199
El Bledal	TP2	0.176	0.315	0.566	0.825	1.181	1.922
La Tina	TP2	0.085	0.147	0.453	0.795	1.261	2.180
Bamícori	TP2	0.070	0.329	0.624	0.897	1.230	1.826
Average value	–	–	0.429	0.665	0.886	1.167	1.722

Table 10 and Table 11 show in detail the results with the three PDFs, in the record with the highest *EEA* (El Bledal, worst fit) and with the lowest one (Tamazula, best fit).

Table 10. Fit parameters and values of the deciles in the El Bledal station (the one with the highest *EEA*) with the three mentioned PDFs.

PDF	**PT3**	**LN3**	**TP2**
Parámetro 1	$\mu = 1.000$	$\xi = 0.857$	$\varphi = 0.145$
Parámetro 2	$\sigma = 0.677$	$\alpha = 0.585$	$Cs = 0.002$
Parámetro 3	$\gamma = 1.346$	$k = -0.462$	$Ck = 2.622$
EEA (adim.)	0.185	0.180	0.176
P(*X*<*x*)	V_i^j/IE^j	V_i^j/IE^j	V_i^j/IE^j
0.10	0.287	0.291	0.315
0.20	0.438	0.449	0.445
0.30	0.574	0.585	0.566
0.40	0.711	0.718	0.690
0.50	0.856	0.857	0.825
0.60	1.020	1.014	0.983
0.70	1.217	1.204	1.181
0.80	1.478	1.459	1.455
0.90	1.899	1.881	1.922

Table 11. Fit parameters and decile values in the Tamazula station (the one with the lowest *EEA*) with the three mentioned PDFs.

PDF	PT3	LN3	TP2
Parámetro 1	$\mu = 1.000$	$\xi = 0.961$	$\varphi = 0.295$
Parámetro 2	$\sigma = 0.329$	$\alpha = 0.315$	$Cs = 0.001$
Parámetro 3	$\gamma = 0.730$	$k = -0.246$	$Ck = 2.910$
EEA (adim.)	0.035	0.035	0.034
P(*X*<*x*)	V_i^j/IE^j	V_i^j/IE^j	V_i^j/IE^j
0.10	0.613	0.614	0.614
0.20	0.719	0.721	0.722
0.30	0.804	0.806	0.808
0.40	0.882	0.883	0.887
0.50	0.961	0.961	0.965
0.60	1.044	1.043	1.048
0.70	1.139	1.137	1.143
0.80	1.258	1.255	1.261
0.90	1.438	1.436	1.439

Finally, Table 12 shows the results of the stations–years method with the three PDFs used, that is, combining the 15 dimensionless records, which add up to 539 data. It is concluded that the FDP Log-Normal leads to the lowest *EEA*.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Table 12. Fit parameters and values of the deciles in Hydrological Region No. 10 (Sinaloa), Mexico, with the three mentioned PDFs, 15 hydrometric stations, and 539 data.

PDF	PT3	LN3	TP2
Parámetro 1	$\mu = 1.000$	$\xi = 0.896$	$\varphi = 0.370$
Parámetro 2	$\sigma = 0.555$	$\alpha = 0.497$	$Cs = -0.006$
Parámetro 3	$\gamma = 1.176$	$k = -0.401$	$Ck = 4.455$
EEA (adim.)	0.105	0.090	0.347
P(*X*<*x*)	V_i^j/IE^j	V_i^j/IE^j	V_i^j/IE^j
0.10	0.395	0.398	0.370
0.20	0.534	0541	0.520
0.30	0.654	0.661	0.649
0.40	0.772	0.777	0.773
0.50	0.896	0.896	0.901
0.60	1.032	1.029	1.043
0.70	1.194	1.186	1.209
0.80	1.405	1.394	1.423
0.90	1.740	1.729	1.759

Based on the results of Table 12, the *regional growth curve* (Equation (1)) was defined, which allowed to generate of the synthetic sequences of dimensionless annual runoff volume, in any site or locality of Hydrological Region No. 10 (Sinaloa), Mexico; this is:

$$q(F) = 0.896 - \frac{0.497}{0.401}[1 - \exp(0.401 \cdot Z)] \quad (52)$$

in which, Z is the standard normal deviation as a function of the probability of non–exceedance F. In Figure 4, the first 500 data of the dimensionless annual runoff volume of 25 in 25 values have been drawn on log–normal probability paper, then the next 30 data from 5 to 5, and finally, the final four values; in addition to the FDP Log-Normal of three fit parameters.

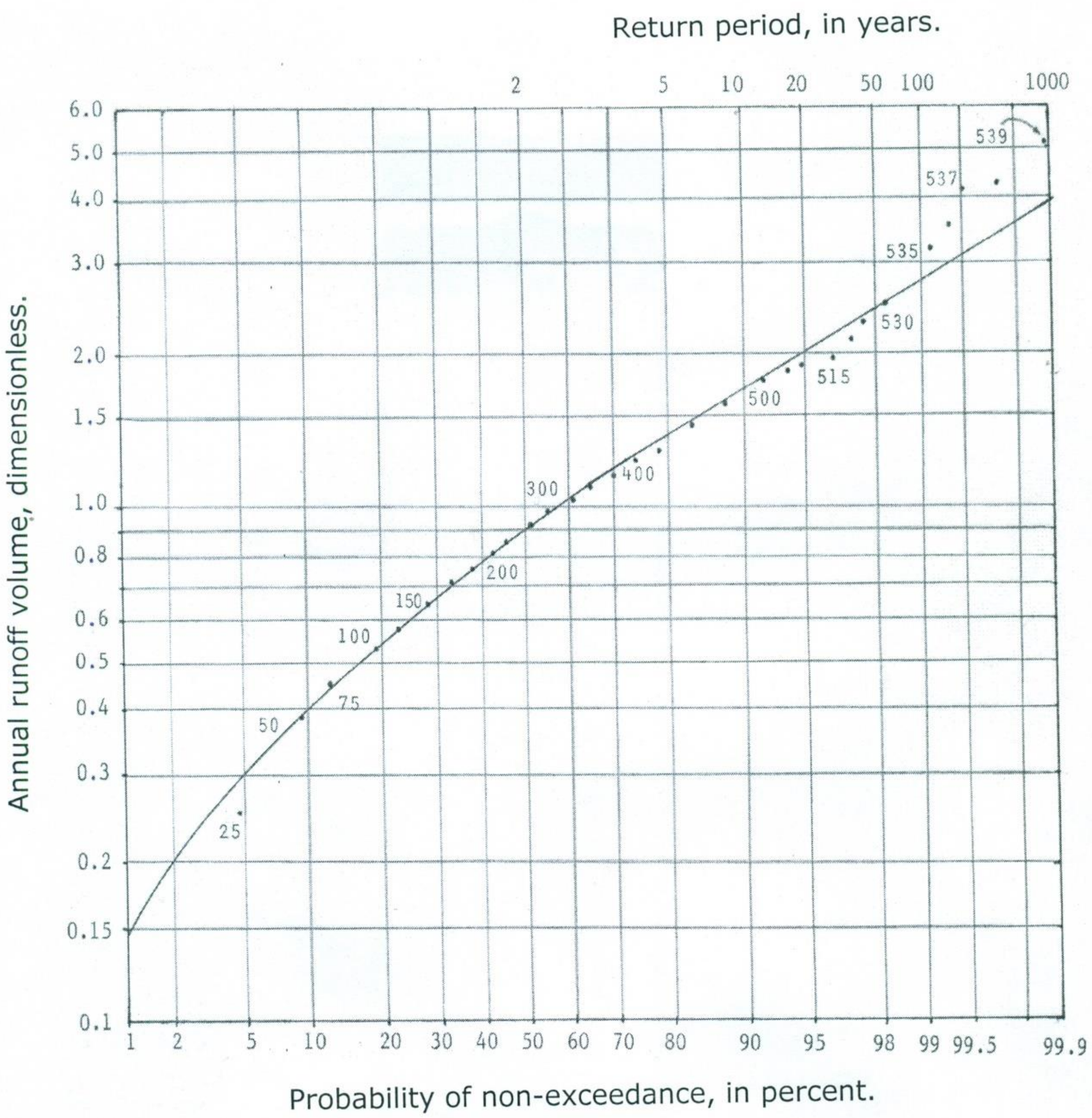

Figure 4. The fit of the Log-Normal distribution to the 539 dimensionless annual runoff volumes of Hydrological Region No. 10 (Sinaloa), Mexico.

Synthetic sequences of dimensionless *VEA*

For the application of Equation (52), first, the size (*ns*) of the generated synthetic sequence of dimensionless annual runoff volume must be defined. Then, making use of an efficient random number generator with uniform distribution (Metcalfe, 1997), the required *ns* numbers u_i are obtained; Such values are made equal to the probability of non–exceedance (*F*), and Equation (30) is applied to obtain the necessary *Z* values, which are taken to Equation (52) to obtain the magnitudes *q*(*F*) sought. With the above process, you can calculate any number of synthetic sequences of V_i^j/IE^j (Equation (1)) equally likely to occur.

Scaling of synthetic sequences

By drawing on a logarithmic paper the two final columns of Table 2, with the basin area (*A*) on the abscissa and the mean annual runoff volume (*VEMA*) on the ordinate, a straight line is defined as shown in Figure 5 and whose equation is:

$$IE^{j} = VEMA = 0.015324 \cdot A^{1.284914} \tag{53}$$

In the previous expression, the correlation coefficient was 0.966, with the 15 pairs of data used.

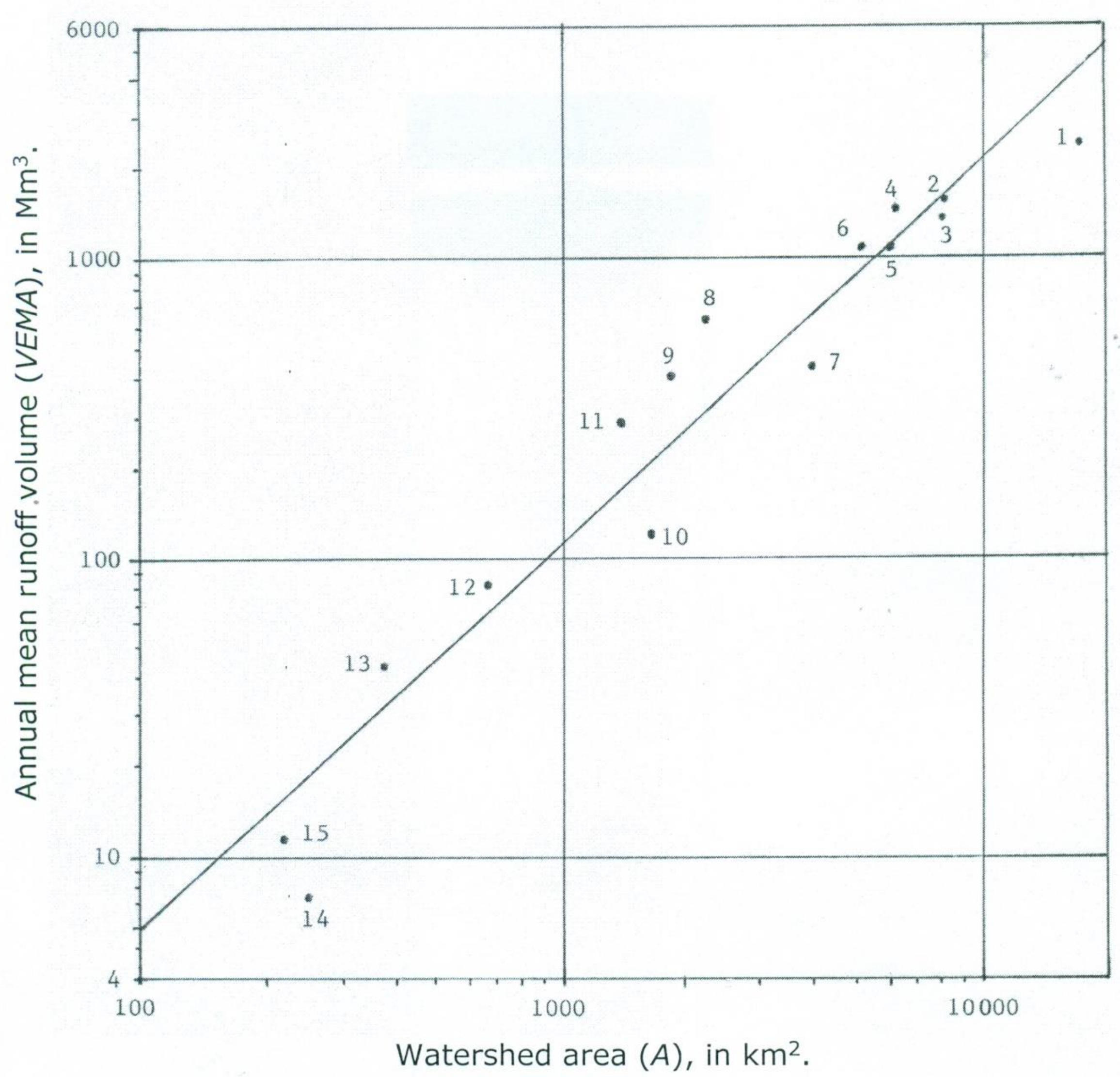

Figure 5. Relationship between the annual mean runoff volume (*VEMA*) and the basin area (*A*), in the 15 homogeneous stations of Hydrological Region No. 10 (Sinaloa), Mexico.

Equation (53) is valid in the entire Hydrological Region No. 10 (Sinaloa), Mexico, and allows estimating the value of the runoff index (IE^j), with which dimension in Mm^3 is assigned to the synthetic sequences of *VEA* generated with Equation (52).

Gathering of main results

From Hydrological Region No. 10 (Sinaloa), Mexico, 22 records of annual runoff volume (*VEA*) were processed. The Discordancy test detected two anomalous records, El Quelite and Palo Dulce; moreover, Pericos station was also eliminated as it showed high discordance.

In this regional study of the *VEA*, the average occurrence date ($\bar{\theta}$) of the maximum monthly runoff volume of each year and its regularity ($\bar{r}$) were evaluated. In Figure 2 two groups of stations were defined, those of low regularity with 0.60 or less have large basins greater than 4000 km^2. The rest of the stations have high regularity between 0.78 and 0.90. However, due to their average date of occurrence, they form a homogeneous region.

The other two regional homogeneity tests, the results of which are shown in Figure 3 and Table 8, confirm that the 19 records

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

processed belong to a homogeneous region from a statistical point of view.

Before starting the search for the regional PDF, four *VEA* records were eliminated, for showing a significant upward trend: Huites, Santa Cruz, Naranjo, and Badiraguato. Table 2 shows that such stations have different basin areas, the first two large and the other small. In addition, from Figure 1 it can be inferred that due to their location they do not define a specific subregion, where it could be accepted that such a deterministic component exists. Therefore, they were eliminated as they would be influenced, by excess, the estimates of the predictions.

The processing of the remaining 15 *VEA* records was carried out by combining such records to apply the station–year method, since a common data period cannot be defined with such records. The annual data (*VEA*) were divided by the mean annual runoff volume (*VEMA*) of each record, to be processed as dimensionless values.

Three probabilistic models have tested: the Pearson type III distributions (PT3), the Log-Normal of three parameters (LN3), and the statistical technique of the potential transformation (TP2), which is equivalent to the Normal PDF. The LN3 distribution was adopted, as it led to a lower standard error of fit.

Finally, a logarithmic equation was defined between the *VEMA* as the dependent variable and the basin area of the hydrometric station as the independent variable. Such an equation allows the "scaling" of the dimensionless predictions obtained with the regional PDF.

Comparison with other studies

Regional frequencies analysis has been carried out predominantly for floods, both in Mexico and in the world. On the other hand, the runoff index method (IE^j) described in this work has not been applied in our country and worldwide, the need to estimate annual runoff volumes for the hydrological design of reservoirs has been resolved by methods that simulate the rainfall-runoff process (Beven, 2001; Wagener, Wheater, & Gupta, 2004).

The application of regional frequencies analysis to estimate the annual runoff volume, in Sicily, Italy, by Cannarozzo *et al*. (2009) supported the objective of this work. However, both studies are completely different. In Sicily, three hydrological regions were defined, based on the altitude and the mean annual rainfall in each basin; its verification was carried out through the *H* test for heterogeneity (Hosking & Wallis, 1997, Campos-Aranda, 2010).

On the other hand, from the analysis of flood frequencies carried out in Sicily by Ferro and Porto (2006), the tests of Gupta *et al*. (1994) and Kumar *et al*. (1994) were taken, to verify the homogeneity of the Hydrological Region No. 10 (Sinaloa), Mexico.

Conclusions

A *regional frequencies analysis* was developed by processing 22 records of annual runoff volume (*VEA*), from Hydrological Region No. 10 (Sinaloa), Mexico. To select the records that form a homogeneous region, according to a hydrological approach, the Discordancy Test was applied and three records were eliminated.

The verification of the *regional hydrological homogeneity* of the 19 remaining records was carried out based on three statistical tests: (1) seasonality indices; (2) linearity of the ordinary moments and (3) linear regression of the probability-weighted moments.

Based on seven tests of *local homogeneity*, another four records that showed an upward linear trend were eliminated. Therefore, 15 *VEA* records were finally processed, with record amplitudes that ranged from 24 to 57 years and amounted to 539 data.

To define the *regional growth curve* [$q(F)$], three probabilistic models were applied: (a) Pearson type III, (b) Log-Normal, and (c) Normal through the Potential Transformation. The first two were adjusted with the L moments method. It was found, through the smallest standard error of fit that the Log-Normal distribution better represented the 539 dimensionless annual runoff values and defined the following equation, a function of Z which is the standard normal deviation: $q(F) = 0.896 - 0.497 \cdot (1 - \exp(0.401 \cdot Z)/0.401$.

Finally, for the *scaling* of the synthetic sequences generated with the previous equation, the following empirical relationship was established: $VEMA = 0.015324 \cdot A^{1.284914}$.

For basins without hydrometry, it is essential to count with a method that allows estimating *VEA* records, of any size and equally likely to occur as a historical record. The additional advantage of such synthetic sequences, having, due to their random nature, absence of persistence and thus, leading to a more reliable reservoir sizing.

Therefore, the systematic application of the *runoff index* method, according to the procedure outlined, is recommended in other hydrological regions of the country to find the corresponding regional equations of $q(F)$ and $VEMA = f(\cdot)$.

Acknowledgments

We are grateful for the observations and corrections suggested by the anonymous referees F and G, which allowed us to improve the writing of the article and avoid important omissions, by appending Figures 1 and 4 and the final two paragraphs in the section of Results and their analysis.

References

Adeloye, A. J., & Montaseri, M. (2002). Preliminary streamflow data analyses prior to water resources study. *Hydrological Sciences Journal*, 47(5), 679-692. DOI: 10.1080/02626660209492973

Beven, K. J. (2001). *Rainfall-runoff modelling. The Primer.* Chichester, England: John Willey & Sons Ltd.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (2013). Chapter 5: Prediction of annual runoff in ungauged basins. In: *Runoff Prediction in Ungauged Basins* (pp. 70-101). Cambridge, UK: Cambridge University Press.

Burn, D. H. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, 202(1-4), 212-230. DOI: 101016/50022-1694(97)00068-1

Campos-Aranda, D. F. (2011). Inciso 8.5: Diseño por simulación estocástica de la operación. En: *Estimación y aprovechamiento del escurrimiento* (pp. 248-250). San Luis Potosí, México: Edición del autor.

Campos-Aranda, D. F. (2010). Verificación de la homogeneidad regional mediante tres pruebas estadísticas. *Tecnología y Ciencias del Agua*, 1(4), 157-165.

Campos-Aranda, D. F. (2008). Contraste de un método simple de transferencia de información para estimación de volúmenes escurridos mensuales. *Ingeniería Hidráulica en México*, 23(2), 153-165.

Cannarozzo, M., Noto, L. V., Viola, F., & La-Loggia, G. (2009). Annual runoff regional frequency analysis in Sicily. *Physics and Chemistry of the Earth*, 34(10-12), 679-687. DOI: 10.1016/j.pce.2009.05.001

Carr, D. P., & Underhill, H. W. (1974). *Los métodos de simulación en el aprovechamiento de los recursos hídricos*. *Estudio sobre Riego y Avenamiento No. 23*. Roma, Italia: Organización de las Naciones Unidas para la Agricultura y la Alimentación.

Cunderlik, J. M., & Burn, D. H. (2002). The use of flood regime information in regional flood frequency analysis. *Hydrological Sciences Journal*, 47(1), 77-92. DOI: 10.1080/02626660209492909

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, 1247-1250. DOI:10.5194/gmd-7-1247-2014

Chander, S., Spolia, S. K., & Kumar, A. (1978). Flood frequency analysis by power transformation. *Journal of Hydraulics Division*, 104(HY11), 1495-1504.

Chow, V. T. (1964). Section 8-I: Frequency Analysis. In: Chow, V. T. (ed.). *Handbook of Applied Hydrology* (pp. 8.1-8.42). New York, USA: McGraw-Hill Book Co.

Dalrymple, T. (1960). *Flood frequency analysis. Water supply paper* 1543-A. Reston, USA: U. S. Geological Survey.

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

De-Michele, C., & Rosso, R. (2002). A multi-level approach to flood frequency regionalization. *Hydrology and Earth System Sciences*, 6(2), 185-194. DOI: 10.5194/hess-6-185-2002

Escalante-Sandoval, C. A., & Reyes-Chávez, L. (2002). Capítulo 6: Análisis de correlación y regresión lineal. En: *Técnicas estadísticas en hidrología* (pp. 105-128). México, DF, México: Facultad de Ingeniería de la Universidad Nacional Autónoma de México.

Ferro, V., & Porto, P. (2006). Flood frequency analysis for Sicily, Italy. *Journal of Hydrologic Engineering*, 11(2), 110-122. DOI: 10.1061/(ASCE)1084-0699(2006)11:2(110)

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.

Gupta, V. K., Mesa, O. J., & Dawdy, D. R. (1994). Multiscaling theory of flood peaks: Regional quantile analysis. *Water Resources Research*, 30(12), 3405-3421. DOI: 10.1029/94WR01791

Hosking, J. R., & Wallis, J. R. (1997). Chapters 1-4 (pp. 1-72) and 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.

IMTA, Instituto Mexicano de Tecnología del Agua. (2002). *Banco Nacional de Datos de Aguas Superficiales* (*BANDAS*). 8 CD's. Jiutepec, México: Comisión Nacional del Agua-Secretaría de Medio Ambiente y Recursos Naturales-Instituto Mexicano de Tecnología del Agua.

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 9: Pearson type III distribution and Chapter 12: Comparison of frequency distributions. In: *Frequency and risk analyses in hydrology* (pp. 105-122, 156-168). Fort Collins, USA: Water Resources Publications.

Kumar, P., Guttarp, P., & Foufoula-Georgiu, E. (1994). A probability weighted moment test to assess simple scaling. *Stochastic Hydrology and Hydraulics*, 8(3), 173-183.

Machiwal, D., & Jha, M. K. (2012). Chapter 4: Methods for time series analysis. In: *Hydrologic time series analysis: Theory and practice* (pp. 51-84). Dordrecht, The Netherlands: Springer.

McMahon, T. A., & Mein, R. G. (1986). Theme 4-17: Simulation using generated data. In: *River and reservoir yield* (pp. 141-142). Littleton, USA: Water Resources Publications.

Metcalfe, A. V. (1997). Chapter 2: Probability distributions and Monte Carlo simulation and Appendix 2: Random number generation. In. *Statistics in Civil Engineering* (pp. 7-38, 319-320). London, UK: Arnold Publishers.

Ostle, B., & Mensing, R. W. (1975). Chapter 7: Regression analysis. In: *Statistics in research* (3rd ed.). (pp. 165-236). Ames, USA: Iowa State University Press.

Ouarda, T. B. M. J. (2017). Regional flood frequency modeling. In: Singh, V. P. (ed.). *Handbook of applied hydrology* (2nd ed.) (pp. 77.1-77.8). New York, USA: McGraw-Hill Education.

Rahman, A., Haddad, K., & Eslamian, S. (2014). Regional flood frequency analysis. In: Eslamian, S. (ed). *Handbook of engineering hydrology: Modeling, climate change and variability* (pp. 451-469). Boca Raton, USA: CRC Press.

Ramírez-Orozco, A. I., Gutiérrez-López, A., & Ruiz-Silva, H. L. (2009). Análisis de la ocurrencia en el tiempo de los gastos máximos en México. *Ingeniería Hidráulica en México*, 24(1), 115-124.

Rao, A. R., & Hamed, K. H. (2000). Chapter 3: Probability weighted moments and L-moments. In: *Flood frequency analysis* (pp. 53-72). Boca Raton, USA: CRC Press.

Salas, J. D. (1993). Analysis and modeling of hydrologic time series. Chapter 19. In: Maindment, D. R. (ed.). *Handbook of Hydrology* (pp. 19.1-19.72). New York, USA: McGraw-Hill, Inc.

Skertchly-Molina, L. (1989). Métodos para calcular capacidades de embalse. *Ingeniería Hidráulica en México,* enero-abril de 1989, 46-52.

Stedinger, J. R. (2017). Flood frequency analysis. In: Singh, V. P. (Ed.). *Handbook of applied hydrology* (2nd ed.) (pp. 76.1-76.8). New York, USA: McGraw-Hill Education.

Varas, E. (2000). Estimación de momentos ponderados regionales de caudales máximos diarios. *Ingeniería Hidráulica en México*, 15(1), 51-61.

Wagener, T., Wheater, H. S., & Gupta, H. V. (2004). *Rainfall-runoff modelling in gauged and ungauged catchments.* London, UK: Imperial College Press.

WMO, World Meteorological Organization. (1971). Annexed III: Standard tests of significance to be recommended in routine analysis of climatic fluctuations. In: *Climatic Change* (pp. 58-71) (Technical Note No. 79). Geneva, Switzerland: World Meteorological Organization.

Zelen, M., & Severo, N. C. (1972). Probability functions. Chapter 26 (pp. 925-995). In: Abramowitz, M., & Stegun, I. A. (eds.). *Handbook of mathematical functions*. New York, USA: Dover Publications.

Zrinji, Z., & Burn, D. H. (1996). Regional flood frequency with hierarchical region of influence. *Journal of Water Resources Planning and Management*, 122(4), 245-252. DOI: 10.1061/(ASCE)0733-9496(1996)122:4(245)

Zsuffa, I., & Gálai, A. (1987). *Reservoir sizing by transition probabilities. Theory, methodology and application*. Littleton, USA: Water Resources Publications.