Articles

Fitting with L-moments of the non–stationary distributions GVE$_1$ and GVE$_2$ to *PMD* series
Ajuste con momentos L de las distribuciones no estacionarias GVE$_1$ y GVE$_2$ a series de *PMD*

Daniel Francisco Campos-Aranda[1]

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

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

Abstract

Design Floods allow the hydrological sizing of hydraulic works. When hydrometric data is not available, design floods are estimated using hydrological methods that are based on *Design Rainfalls*. The most common records used to estimate design rainfalls are the *annual maximum daily precipitations* (*PMD*), this, due to the scarcity of rainfall recorder stations. The impacts of climate change and/or the alteration of the geographic environment of rain–gauge stations cause *PMD* records to show trends and therefore these records become non–stationary. In order to estimate *predictions* of low probability of exceedance a probabilistic analysis of the non–stationary *PMD* records can be performed. A simple approach without computational difficulties is based on the extension of the method of L moments applied to the General of Extreme Values (GVE) distribution with its location parameter (*u*) variable with time (*t*) in years, which is entered as a *covariate*. When the trend in the *PMD* register is linear, the probabilistic model GVE$_1$ is applied in which $u_t = \mu_0 + \mu_1.t$ and when the trend is curve the model GVE$_2$ with $u_t = \mu_0 + \mu_1.t + \mu_2 \cdot t^2$ is used. Thus, the GVE$_1$ distribution has four fit parameters (μ_0, μ_1, a, k) and five for the GVE$_2$ distribution (μ_0, μ_1, μ_2, a, k). Four numerical applications are described and the analysis of their results shows the simplicity of the extension of the L moments

method and its versatility to estimate *predictions* within the historical record and to the future.

Keywords: L moments, GEV distribution, standard error of fit, linear regression, parabolic regression, determinants, multiple linear regression.

Resumen

Las *crecientes de diseño* permiten el dimensionamiento hidrológico de las obras hidráulicas. Cuando no existen datos hidrométricos, las crecientes de diseño se estiman con métodos hidrológicos que se basan en las *lluvias de diseño*. La escasez de estaciones pluviográficas origina que los registros más comunes empleados para estimar las lluvias de diseño sean los de *precipitación máxima diaria* (*PMD*) anual. Debido a los impactos del cambio climático y/o a la alteración del entorno geográfico de las estaciones pluviométricas, los registros de *PMD* están mostrando tendencias y por lo tanto son no estacionarios. El análisis probabilístico de los registros de *PMD* no estacionarios, orientado a estimar *predicciones* de baja probabilidad de excedencia que puede realizar, de manera simple y sin dificultades computacionales, con base en la extensión del método de los momentos L para aplicar la *distribución general de valores extremos* (GVE) con su parámetro de ubicación (*u*) variable con el tiempo (*t*) en años, se introduce como covariable. Cuando la tendencia en el registro de *PMD* es lineal, se aplica el modelo probabilístico GVE$_1$, en el cual $u_t = \mu_0 + \mu_1 \cdot t$ y cuando es curva el modelo GVE$_2$ con $u_t = \mu_0 + \mu_1 \cdot t + \mu_2 \cdot t^2$. Entonces, la distribución GVE$_1$ tiene cuatro parámetros de ajuste (μ_0, μ_1, α, k) y la GVE$_2$ cinco (μ_0, μ_1, μ_2, α, k). Se describen cuatro aplicaciones numéricas, y a través del análisis de sus resultados se demuestra la sencillez de la extensión del método de los momentos L y su versatilidad para estimar predicciones dentro del registro histórico y a futuro.

Palabras clave: momentos L, distribución GVE, error estándar de ajuste, regresión lineal, regresión parabólica, determinantes, regresión lineal múltiple.

Introduction

Generalities

Design rainfalls are maximum precipitations of a certain duration associated to low exceedance probabilities, based on which, by means of hydrological methods, the *Design Floods* are estimated, when hydrometric records are not available. Details of the above can be found in Teegavarapu (2012) and Mujumdar and Nagesh Kumar (2012). Design floods allow the sizing of hydraulic works, during their planning or when they are checked for safety. The most reliable estimate of design rainfalls is based on the probabilistic analysis (AP) of the annual maximum data, which assumes that the random process that generates the observed precipitations is *stationary* and therefore, its statistical properties do not change over time. In the AP, a known probability distribution function (FDP) is adjusted to the precipitation record and based on it, the *predictions* sought or design rainfalls are made. Due to the scarcity of rainfall recorder stations and the relative abundance of rain–gauge stations, the most commonly processed extreme rainfall records are those of the annual maximum daily precipitation (*PMD*).

At present, all rain–gauge stations are subject to the effects of global climate change, or their geographical environment suffers physical changes generated by human activities, among the most striking are urbanization, deforestation, dewatering of lagoons and the construction of reservoirs, which alter local atmospheric processes, giving rise to series or records of annual non–stationary *PMD*, since they present trends (Strupczewski & Kaczmarek, 2001; Khaliq, Ouarda, Ondo, Gachon, & Bobée, 2006; El Adlouni, Ouarda, Zhang, Roy, & Bobée, 2007; El Adlouni & Ouarda, 2008).

To perform the AP of non–stationary records, the statistical theory of extreme values has been extended to adjust the classical model which follows asymptotically the maximum hydrological data series, that is,

the FDP General of Extreme Values (GVE_0) stationary with three fit parameters (u,a,k). The above based on the introduction of so–called covariables, one of the most used is the time (t) in years, through which the *trend* observed in the data series can be taken into account, adopting *variable* the location parameter (u_t). When the trend is linear ($u_t = \mu_0 + \mu_1 \cdot t$) the GVE_1 model of four fit parameters (μ_0, μ_1, a, k) is fitted. If the trend is curved ($u_t = \mu_0 + \mu_1 \cdot t + \mu_2 \cdot t^2$) the GVE_2 model of five fit parameters (μ_0, μ_1, μ_2, a, k) is applied; in this model two covariables ($u_t = \mu_0 + \mu_1 \cdot t + \mu_2 \cdot h$) can be used (Khaliq *et al*., 2006; Prosdocimi, Kjeldsen, & Miller, 2015).

Regarding to other non–stationary distributions, there are several, for example the Log-Normal (Vogel, Yaindl, & Walter, 2011; Aissaoui-Fqayeh, El Adlouni, Ouarda, & St. Hilaire, 2009) and the Generalized Logistics models (Kim, Nam, Ahn, Kim, & Heo, 2015) and Generalized Pareto (Rao & Hamed, 2000), which are susceptible to an *identical* treatment to the one that will be exposed for the GVE distribution. Other covariates can also be used instead of time (t), for example some global or regional climate indexes (Prosdocimi, Kjeldsen, & Svensson, 2014; López-de-la-Cruz & Francés, 2014; Álvarez-Olguín & Escalante-Sandoval, 2016; Campos-Aranda, 2018).

Objective

The *objective* of this study was to present the fitting theory of the non–stationary distributions GVE_1 and GVE_2, by means of the generalization of the method of L moments, initially proposed by El Adlouni and Ouarda (2008) and subsequently applied by Gado and Nguyen (2016). In addition, four series of annual *PMD* are exposed and processed, all with trend and the results of the fitting of the model GVE_1 or GVE_2 are described, highlighting the simplicity and usefulness of the extension of the L–moments method to obtain the *predictions* within the historical record and to the future.

Methods and materials

General of Extreme Values Distribution

The maximum annual values are events that occur in the right tail of the FDP, which defines the behavior of the hydrological random variable. Therefore, the design rainfalls can be *predicted* based on the FDP, as the maximum value corresponding to a certain *average* interval of recurrence or return period (*Tr*), whose probability of exceedance is $p = 1/Tr$. The theory of extreme values established that the extreme data follow asymptotically to the FDP General of Extreme Values (GVE), whose application has been recommended (Stedinger, Vogel, & Foufoula-Georgiou, 1993; Hosking & Wallis, 1997; Coles, 2001; Khaliq *et al*., 2006; Papalexiou & Koutsoyiannis, 2013; Gado & Nguyen, 2016) to model extreme *hydrological data* (*dh*) with a probability of non–exceedance [$F(dh)$] that is:

$$F(dh) = exp\left\{-\left[1 - \frac{k(dh-u)}{\alpha}\right]^{1/k}\right\} \text{ when } k \neq 0 \quad (1)$$

In the previous expression, *u, α* and *k* are the location, scale and shape parameters of the GVE_0. When $k = 0$ the Gumbel distribution is obtained, which is a straight line in the Gumbel–Powell probability paper, whereby the interval of the variable is: $-\infty < dh < \infty$. When $k > 0$ the distribution is Weibull which is a curve with concavity downwards and upper limit, whereby: $-\infty < dh \leq u + \alpha/k$. Finally, if $k < 0$ the distribution is Fréchet which is also a curve but with a concavity upwards and lower border, so: $u + \alpha/k \leq dh < \infty$. The *predictions* sought (DH_{Tr}) are obtained with the inverse solution of equation 1:

$$DH_{Tr} = u + \frac{\alpha}{k}\{1 - [-\ln(1 - p)]^k\} \text{ when } k \neq 0 \quad (2)$$

L-Moments of the data sample

The L–moments method is perhaps the simplest of the reliable procedures for estimating the fitting parameters of the PDFs used in hydrology. This is due to the fact that the L–moments, which are linear combinations (Hosking & Wallis, 1997) of the weighted probability moments (βr), are not affected significantly by the scattered values (*outliers*) of the sample. The first three L–moments of a sample (l_1, l_2, l_3) and the asymmetry L–ratio (t_3), are estimated through the unbiased estimator (b_r) of the β_r, as follows:

$l_1 = b_0$ (3)
$l_2 = 2 \cdot b_1 - b_0$ (4)
$l_3 = 6 \cdot b_2 - 6 \cdot b_1 + b_0$ (5)
$t_3 = l_3/l_2$ (6)

The unbiased estimator of the β_r is (Hosking & Wallis, 1997):

$b_r = \frac{1}{n} \sum_{i=r+1}^{n} \frac{(i-1)(i-2)\cdots(i-r)}{(n-1)(n-2)\cdots(n-r)} dh_i$ (7)

where r = 0, 1, 2,…. and dh_i are the data of the available sample or series of hydrological data of size n, ordered from low to high ($dh_1 \le dh_2 \le \cdots \le dh_n$).

Fit parameters of the stationary distribution GVE

For the GVE_0 model with the L–moments method, its three fit parameters are calculated with the equations (Stedinger *et al*., 1993; Hosking & Wallis, 1997; Rao & Hamed, 2000; Campos-Aranda, 2018):

$k \cong 7.8590 \cdot c + 2.9554 \cdot c^2$ (8)
being:
$c = \frac{2}{3+t_3} - 0.63093$ (9)

$\alpha = \frac{l_2 \cdot k}{(1-2^{-k}) \cdot \Gamma(1+k)}$ (10)

$u = l_1 - \frac{\alpha}{k}[1 - \Gamma(1 + k)]$ (11)

For the estimation of the Gamma $\Gamma(\omega)$ function the Stirling formula (Davis, 1972) was used, shown in the following equation (12):

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

AP non–stationary with GVE$_1$ and GVE$_2$

According to Strupczewski and Kaczmarek (2001) the fit of the non–stationary distributions GVE$_1$ and GVE$_2$ implies two assumptions: (1) it is accepted that the non–stationarity of the annual *PMD* series is caused by gradual changes in the geographical environment and/or global climate change, generating a slight alteration of its statistical parameters and (2) it is accepted that the FDP is independent of time, so the distribution GVE, with fit parameters that are variable over time, is acceptable for modeling extreme non–stationary data. Therefore the mean (μ_t) of the distribution GVE must be considered variable over time whose expression is (Rao & Hamed, 2000):

$$\mu = u + \frac{\alpha}{k}[1 - \Gamma(1 + k)] \quad (13)$$

Fit of the distribution GVE$_1$

The inspection of the *PMD$_i$* time series will allow to define if a trend of the linear mean (μ_t) or curve is adopted. When the trend in the mean is *linear* and is introduced in equation 13, the location parameter (u) can be cleared that will now be variable with respect to time t, which varies of 1 to n, its expression is (El Adlouni & Ouarda, 2008; Gado & Nguyen, 2016):

$$u_t = \mu_0 + \mu_1 \cdot t - \frac{\alpha}{k}[1 - \Gamma(1 + k)] \quad (14)$$

The magnitudes μ_0 and μ_1 of the line representing the linear trend of the sample or series of data PMD_i are obtained based on the equations in Appendix 1. To prove that the slope μ_1 is statistically different from zero, the equations in Appendix 2 are applied. To estimate the values of the scale parameters (a) and shape (k) of the previous expression that correspond to the stationary FDP (GVE$_0$), the trend of the PMD_i series is removed to obtain a stationary series PMD_e^{μ}, based on the following equation (Khaliq *et al*., 2006):

$$PMD_e^{\mu} = PMD_i - \mu_1 \cdot t \quad (15)$$

Then equations 8 to 10 are applied to obtain the sought values of a and k. Finally, equation 2 is applied using expression 14 to take the variable location parameter (u_t) into account over time and make *predictions* within the historical record ($t < n$) and at the end of it ($t = n$), as well as several years later ($t > n$), for example in the future at the end of the *useful life* of the hydraulic work analyzed, which takes into account the trend observed in the series of hydrological data (Mudersbach & Jensen, 2010).

Fit of the distribution GVE$_2$

When the trend of the mean (μ_t) is *curve* and is introduced into the equation 13, the parameter of variable location (u_t) with respect to time, is equal to (El Adlouni & Ouarda, 2008; Gado & Nguyen, 2016):

$$u_t = \mu_0 + \mu_1 \cdot t + \mu_2 \cdot t^2 - \frac{\alpha}{k}[1 - \Gamma(1 + k)] \quad (16)$$

The magnitudes μ_0, μ_1 and μ_2 of the curve equation representing the trend of the sample or series of data PMD_i are obtained based on the equations of Appendix 3. To estimate the values of the scale (a) and shape (k) parameters of the previous expression that correspond to the FDP stationary (GVE$_0$), the trend of the series PMD_i is removed to obtain a stationary series PMD_e^{μ}, based on the following equation:

$$PMD_e^{\mu} = PMD_i - \mu_1 \cdot t - \mu_2 \cdot t^2 \quad (17)$$

Then, equations 8 to 10 are applied to obtain the sought values of *a* and *k*. Finally, equation 2 is applied using expression 16 to take the variable location parameter (u_t) into account over time and make *predictions* within the historical record ($t < n$), at the end of this ($t = n$) and at a future time, as indicated for the GVE$_1$ model.

The non-stationary distribution GVE$_2$ can also be applied with two covariates (*t* and *h*), leaving the variable location parameter (u_t) equal to (Prosdocimi *et al*., 2014; Prosdocimi *et al*., 2015):

$$u_t = \mu_0 + \mu_1 \cdot t + \mu_2 \cdot h - \frac{\alpha}{k}[1 - \Gamma(1 + k)] \ (18)$$

The magnitudes μ_0, μ_1 y μ_2 of the multiple linear regression equation representing the trend of the sample or series of *PMD$_i$* data are obtained based on the equations in Appendix 4. To estimate the values of the scale (*a*) and shape (*k*) parameters of the previous expression that correspond to the FDP stationary (GVE$_0$), the trend of the *PMD$_i$* series is removed to obtain a stationary series PMD_e^{μ} based on the following equation:

$$PMD_e^{\mu} = PMD_i - \mu_1 \cdot t - \mu_2 \cdot h \ (19)$$

Then, equations 8 to 10 are applied to obtain the sought values of *a* and *k*. Then equation 2 is applied using expression 18 to take the variable location parameter (u_t) into account over time and make *predictions* within the historical record ($t < n$), at the end of it ($t = n$), and at a future time ($t > n$) as indicated for the GVE$_1$ model.

Standard error of fit

Since the mid–1970s the *standard error of fit* (*EEA*) was formulated as a quantitative measure that estimates the *descriptive ability* of the fit probabilistic model (Meylan, Favre, & Musy, 2012) and that also allows the objective comparison between the various FDPs that are tested or fitted to a series or sample of data, since it has the data units (*PMD$_i$*). Its expression is the following (Kite, 1977):

$$EEA = \sqrt{\frac{\sum_{i=1}^{n}\left(PMD_i^o - PMD_i^e\right)^2}{n - npa}} \quad (20)$$

where, n is the number of data of the available series, npa is the number of fit parameters of the FDP that is tested, with four for the model GVE_1 and five for the two models GVE_2, PMD_i^o are the data ordered from lowest to highest and PMD_i^e are the values estimated with equation 2 of the FDP (GVE_1 or GVE_2), for non–exceedance probability $P(X < x)$ estimated with the Weibull formula (Benson, 1962):

$$P(X < x) = \frac{m}{n+1} \quad (21)$$

in which, m is the order number of the data, with 1 for the lowest and n for the highest. The calculation of the EEA with equation 20, allows the comparison of other non–stationary probabilistic models in the series that is processed. When the EEA values are similar, a non–stationary model can be adopted in a subjective manner, for example, the one that leads to the most unfavorable predictions.

Approach of the probabilistic analysis

Based on equation 2, *predictions* with return periods (Tr) of 2, 10, 25, 50 and 100 years were estimated through the record period, applying variable the location parameter (u_t). The first prediction corresponds to the *median*, since its probability of non–exceedance ($1-p$) is 50% and the following three are calculated for the following values: 0.90, 0.96, 0.98 and 0.99, respectively. In addition, predictions are made in the processed series for the future, in the years 2025 and 2050. It is considered that extrapolating the observed behavior of the historical trend over 30 years is quite risky. It is also shown in the data graphs and predictions the estimates of the extreme return periods (2 and 100 years) with the stationary model GVE_0, which are horizontal straight lines that are indicated dotted.

Non-stationary AP with LOG y PAG

The FDPs Generalized Logistics (LOG) and Generalized Pareto (PAG) are two probabilistic models used regularly in the analysis of extreme hydrological data frequencies which are applicable in their non–stationary versions with variable location parameter (u_t) by means of generalization of the L–moments method, as has been exposed from equation 13. This key equation of the method has the following expressions in the LOG and PAG distributions (Rao & Hamed, 2000):

$$\mu = u + \frac{\alpha}{k}[1 - \Gamma(1+k) \cdot \Gamma(1-k)] \quad (22)$$

$$\mu = u + \frac{\alpha}{1+k} \quad (23)$$

Regard to equations 1 and 2 corresponding to the LOG and PAG distributions can be consulted in Hosking and Wallis (1997), and Campos-Aranda (2018).

Criterion of severity of the linear trend

Equations in Apendix 2 lead to two statistics associated with the slope of linear regression, the calculated *DS* and its corresponding critical value DS_c. When the *DS* slightly exceeds DS_c, the slope is significant and is *mild*. When the *DS* exceeds by one unit or more to the DS_c, the slope is *severe*. *Moderate* slopes occur in the intermediate case.

Series 1 with severe downward linear trend

This annual *PMD* record belongs to the Abritas rain–gauge station, located in the municipality of Ciudad del Maíz in the northern zone of the Huasteca region of the state of San Luis Potosí, Mexico. It covers 54 years in the period from 1961 to 2015, since it has incomplete the year of 1998. With the exception of series 2, the information of this record

and of the others was provided by the San Luis Potosí Local Office of the National Water Commission. Its values are cited in Table 1.

Table 1. Maximum annual data of *PMD* in millimeters, in three stations of the state of San Luis Potosí, Mexico.

No. da-to	Abri-tas	Las Adjun-tas	Los Fil-tros	No. da-to	Abri-tas	Las Adjun-tas	Los Fil-tros	No. da-to	Los Fil-tros
1	150.0	57.4	15.9	28	140.0	38.0	58.0	55	40.2
2	131.5	75.0	20.6	29	90.0	182.5	42.9	56	111.
3	127.5	42.0	50.9	30	140.0	134.7	26.4	57	43.3
4	183.0	65.5	40.5	31	168.0	69.4	65.5	58	76.9
5	87.0	72.2	63.6	32	194.0	118.2	22.0	59	42.8
6	215.5	63.2	41.9	33	80.0	111.4	51.2	60	46.1
7	203.5	131.5	60.0	34	85.0	92.3	66.5	61	42.5
8	110.0	70.9	35.9	35	138.0	86.0	26.0	62	45.3
9	187.5	100.3	48.6	36	138.0	108.7	31.5	63	44.5
10	254.0	141.5	63.0	37	84.0	67.5	46.5	64	26.0
11	102.0	58.1	35.5	38	56.0	72.0	44.0	65	59.1
12	130.0	74.7	40.0	39	85.0	76.8	41.0	66	44.1
13	164.0	87.0	63.2	40	33.0	58.0	55.0	67	63.0
14	280.0	134.0	39.4	41	23.1	83.0	21.5	68	51.6
15	147.0	78.4	27.2	42	35.0	58.9	29.8	–	–
16	170.0	237.5	59.0	43	70.4	116.8	41.5	–	–
17	193.0	80.0	32.0	44	57.0	141.0	24.9	–	–
18	85.0	59.2	30.0	45	71.0	119.3	59.0	–	–
19	116.5	93.0	40.2	46	120.0	86.3	33.5	–	–
20	87.0	72.2	31.5	47	223.2	146.1	46.5	–	–
21	150.0	85.0	31.5	48	90.9	115.5	51.0	–	–
22	97.0	68.0	52.0	49	133.0	132.7	40.0	–	–
23	120.0	86.4	52.3	50	100.0	303.8	35.5	–	–
24	64.0	66.7	31.3	51	102.0	75.3	45.5	–	–
25	90.0	80.5	35.0	52	98.0	70.0	25.9	–	–
26	175.0	100.0	28.5	53	98.0	151.5	20.7	–	–
27	100.0	133.5	57.2	54	95.0	125.1	37.5	–	–

Series 2 with mild upwards linear trend

Campos-Aranda (2016) presented the annual *PMD* record of the Zacatecas rain–gauge station in the state capital of the same name in Mexico, with 58 data in the period from 1953 to 2010, which showed a

linear upward logarithmic trend; due to this, it was processed in such a reference, based on the FDP Log–normal non–stationary of two fitting parameters, suggested by Vogel *et al*. (2011).

Series 3 with moderate upward linear trend

This record of annual *PMD* belongs to the rain–gauge station Las Adjuntas, located where the rivers Tampaón and Moctezuma join to form the Pánuco river, in the eastern zone of the Huasteca region of the state of San Luis Potosí, Mexico. It covers 54 years in the interval from 1961 to 2015, because the year of 1986 is missing, its values are shown in Table 1.

Series 4 with curved convex trend

This record of annual *PMD* belongs to the Los Filtros rain–gauge station, located within the city of San Luis Potosí, capital of the state of the same name in Mexico. It covers 68 years in the interval from 1949 to 2016. Its values are in the Table 1.

Description of results

Predictions in the Abritas station, S.L.P.

First, equations of Appendix 1 were applied and $\mu_0 = 167.6315$, $\mu_1 = -1.6057$ y $r_{xy} = -0.4619$ were obtained. Then the linear trend was tested with equations of the Appendix 2, a $DS = -3.7558$ and $DS_c = 2.0066$ were obtained, which was highly significant. The stationary series (PMD_e^μ, equation 15) accepts a GVE$_0$ model with scale and shape parameters of: $a = 41.4364$ and $k = 0.0691$; however, equation 2 using the variable

location parameter (equation 14) leads to a 34.8 mm standard error of fit (*EEA*). Table 2 shows part of the predictions within the historical record and an extrapolation to a decade (year 2025), since at future they are smaller due to the downward slope. Predictions with the stationary model GVE_0 of return periods 2 and 100 years are: 114.9 and 290.8 mm, with an *EEA* = 6.0 mm. Figure 1 shows the chronological series of the data and the lines of the predictions, which are plotted, based on the predictions in Table 2, within the historical record with the values of the covariate *t* that vary from 1 to *n*.

Table 2. Predictions in millimeters in the historical and to future period in the station Abritas, S.L.P., based on the non–stationary distribution GVE_1.

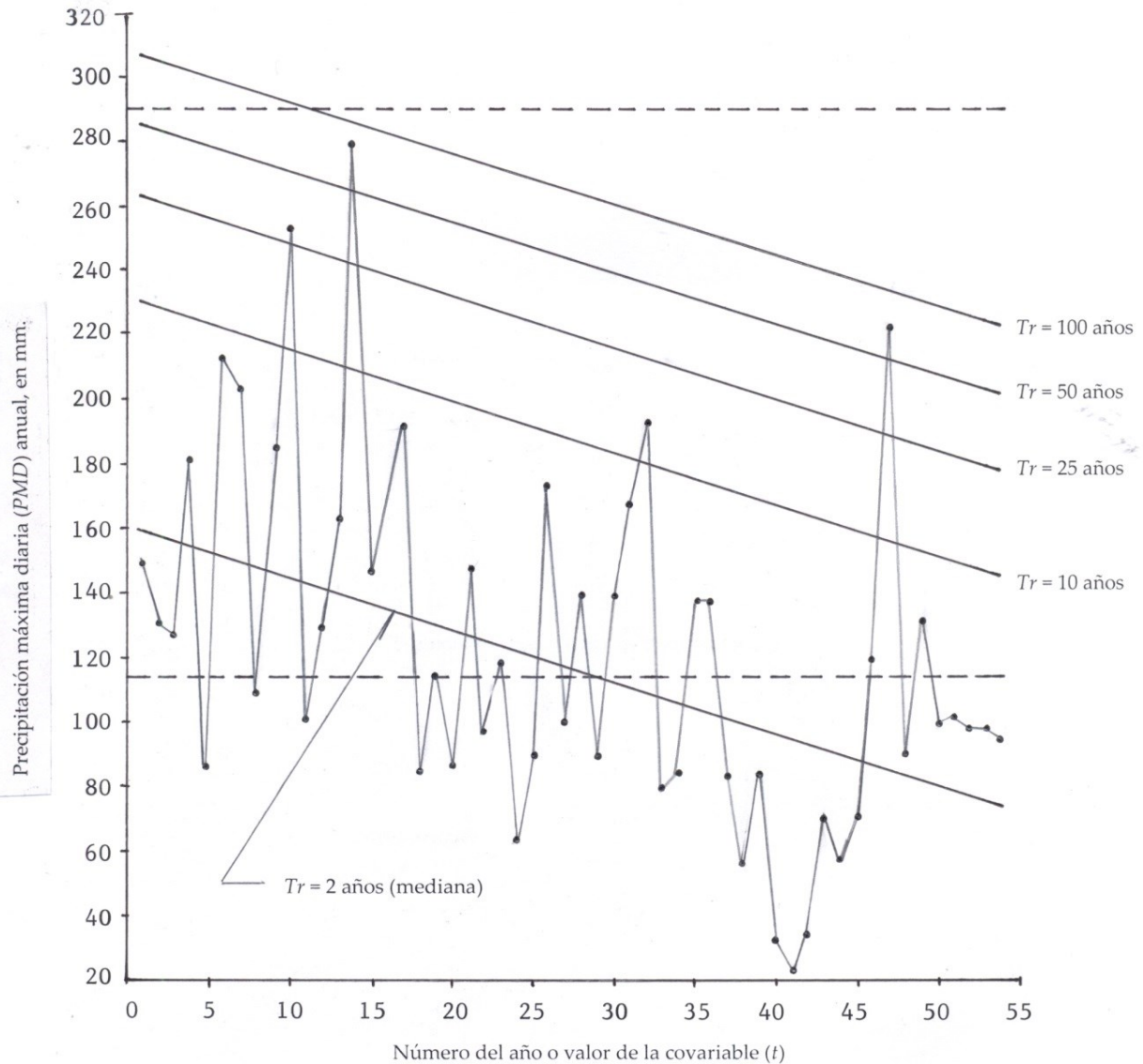| t	Year	Return periods (years)				
		2	10	25	50	100
1	1961	159.6	230.9	263.5	286.3	307.8
10	1970	145.1	216.5	249.0	271.8	293.4
20	1980	129.0	200.4	233.0	255.8	277.3
30	1990	113.0	184.3	216.9	239.7	261.3
40	2001	96.9	168.3	200.8	223.6	245.2
50	2011	80.9	152.2	184.8	207.6	229.2
54	2015	74.5	145.8	178.4	201.2	222.7
64	2025	58.4	129.8	162.3	185.1	206.7

Figure 1. Chronological series of annual *PMD* and estimated prediction lines with the distribution GVE$_1$ in the Abritas station, S.L.P., Mexico.

Predictions in the Zacatecas, Zac. Station

When applying the equations of Appendix 1: μ_0 = 40.1247, μ_1 = 0.2246 and r_{xy} = 0.2638 was obtained. The linear trend was then tested with

equations of Appendix 2, obtaining a DS = 2.0463 and a DS_c = 2.0032, whereby it was scarcely significant. The stationary series (PMD_e^μ equation 15) accepts a GVE_0 model with scale and shape parameters of: a =11.8358 and k = 0.0709: however, equation 2 using the variable location parameter (equation 14) leads to a 3.0 mm standard error of fit. Table 3 shows a part of the predictions within the historical and to future record in the years 2025, 2050 and 2100. As this record of 58 data ends in the year 2010, then, the value of time t in 2025 is 73, in 2050 is 98 and 2100 is 148. The predictions with the stationary model GVE_0 of return periods 2 and 100 years are: 45.2 and 86.7 mm, with an EEA = 1.5 mm. Figure 2 shows the chronological series of data and lines of the predictions, plotted according to the results of Table 3.

Table 3. Predictions (mm) in the historical period and to future in the Zacatecas, Zac., station, based on the non–stationary distribution GVE_1.

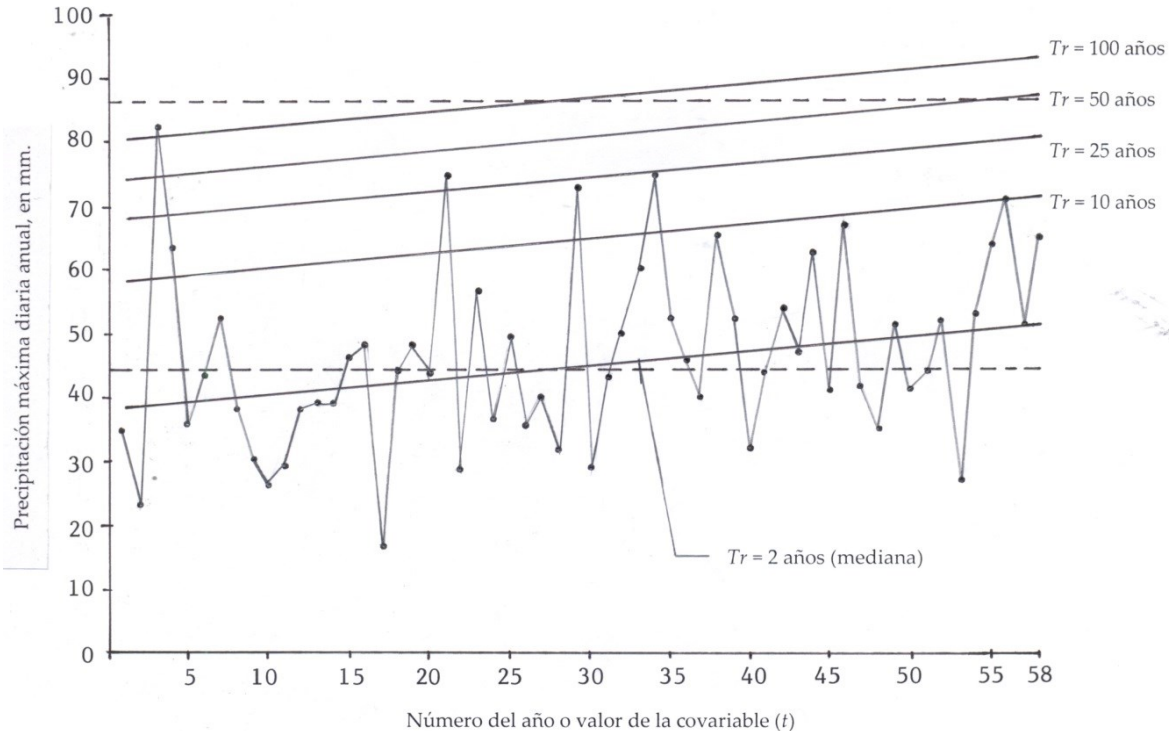| t	Year	Return periods (years)				
		2	10	25	50	100
1	1953	38.5	58.9	68.1	74.6	80.7
10	1962	40.5	60.9	70.1	76.6	82.7
20	1972	42.8	63.1	72.4	78.9	85.0
30	1982	45.0	65.4	74.6	81.1	87.2
40	1992	47.3	67.6	76.9	83.3	89.5
50	2002	49.5	69.9	79.1	85.6	91.7
58	2010	51.3	71.7	80.9	87.4	93.5
73	2025	54.7	75.0	84.3	90.8	96.9
98	2050	60.3	80.6	89.9	96.4	102.5
148	2100	71.5	91.9	101.1	107.6	113.7

Figure 2. Chronological series of annual *PMD* and estimated prediction lines with the distribution GVE$_1$ in Zacatecas station, Zac., Mexico.

Predictions in the Las Adjuntas station, S.L.P.

When applying equations of Appendix 1, μ_0 = 74.9213, μ_1 = 0.8813 and r_{xy} = 0.2981 was obtained. The linear trend was then tested with the equations in Appendix 2, a *DS* = 2.2520 and a *DS$_c$* = 2.0066 were obtained, therefore it was moderately significant. The stationary series (PMD_e^μ equation 15) accepts a GVE$_0$ model with scale and shape parameters of a = 26.5840 and k = −0.1617; however, equation 2 using the variable location parameter (equation 14) leads to a 15.6 mm standard error of fit. Table 4 shows a part of the predictions within the historical record and to future in the years 2025, 2050 and 2100. As this record of 54 data ends in the year 2015, then, the value of time *t* in 2025 is 64, in 2050 is 89 and 2100 is 139. The predictions with the stationary model GVE$_0$ of return periods 2 and 100 years are: 87.4 and

280.9 mm, with an *EEA* = 11.3 mm. Figure 3 shows the chronological series of data and lines of the predictions, plotted according to the results of Table 4.

Table 4. Predictions (mm) in the historical period and to future in Las Adjuntas, S.L.P., station, based on the non-stationary distribution GVE_1.

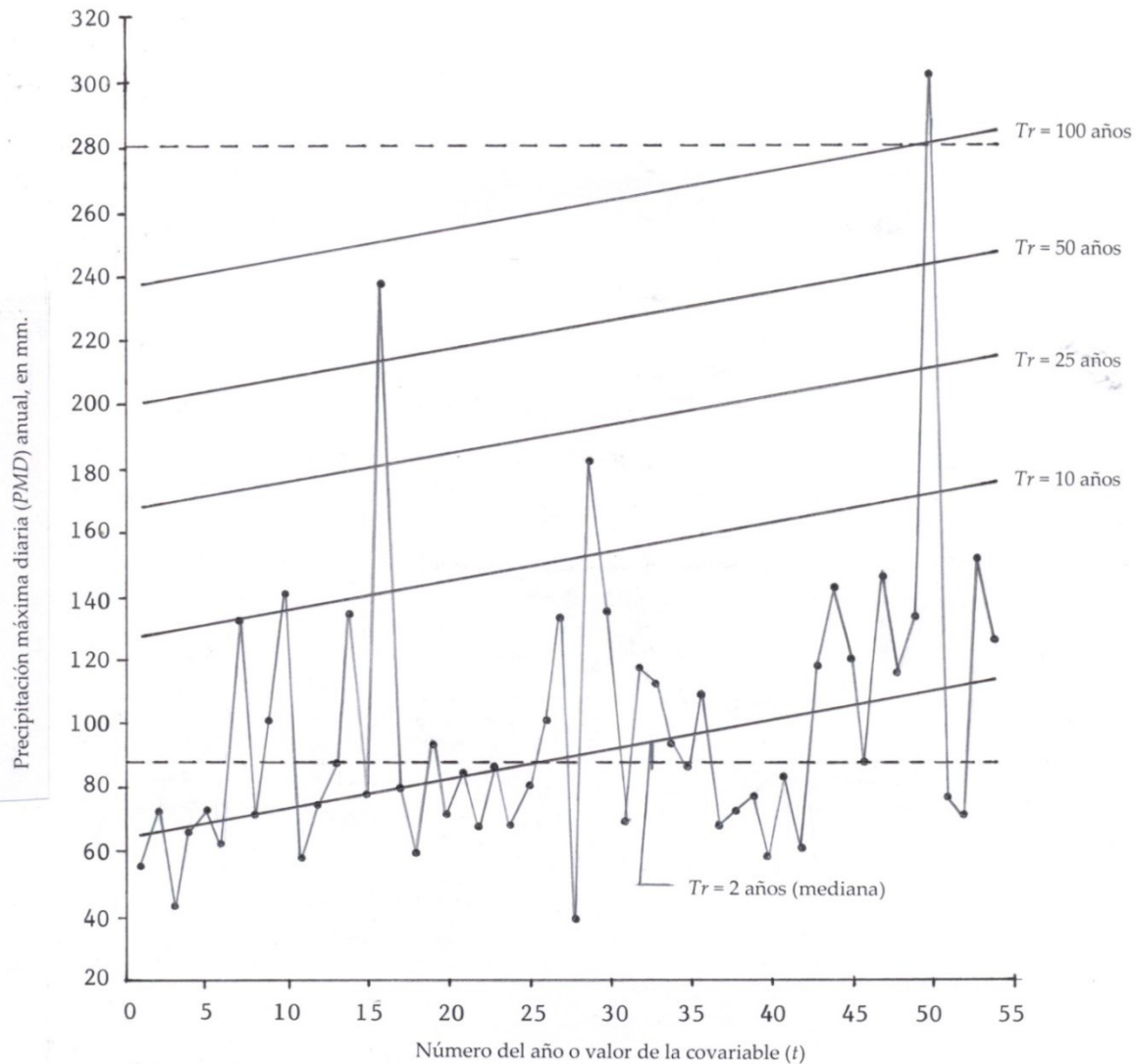| *t*	Year	Return periods (Years)				
		2	**10**	**25**	**50**	**100**
1	1961	65.7	127.8	167.0	200.2	237.2
10	1970	73.6	135.7	174.9	208.2	245.1
20	1980	82.4	144.5	183.7	217.0	253.9
30	1991	91.2	153.4	192.6	225.8	262.7
40	2001	100.0	162.2	201.4	234.6	271.5
50	2011	108.9	171.0	210.2	243.4	280.3
54	2015	112.4	174.5	213.7	246.9	283.9
64	2025	121.2	183.3	222.5	255.8	292.7
89	2050	143.2	205.4	244.6	277.8	314.7
139	2100	187.3	249.4	288.6	321.8	358.8

Figure 3. Chronological series of annual *PMD* and estimated prediction lines with the GVE1 distribution in Las Adjuntas station, S.L.P., Mexico.

Predictions in Los Filtros station, S.L.P.

Equations in Appendix 3 lead to the following values: μ_0 = 43.7097, μ_1 = −0.2339 and μ_2 = 0.0049 as coefficients of the polynomial regression (equation A3.1). The stationary series (PMD_e^{μ} equation 17) accepts a

GVE_0 model with scale and shape parameters of: a =12.6389 and k = 0.05071: however, equation 2 using the variable location parameter (equation 16) leads to a 4.3 mm standard error of fit. Table 5 shows a part of the predictions within the historical record and to future in the years 2025 and 2050. As this record of 66 data ends in the year 2016, then, the value of time t in 2025 is 77 and in 2050 is 102. The predictions with the stationary model GVE_0 of return periods 2 and 100 years are: 41.5 and 87.5 mm, with an *EEA* = 3.7 mm. Figure 4 shows the chronological series of data and curves of predictions, plotted based on the results of Table 5.

Table 5. Predictions (mm) in the historical period and to future in Los Filtros station, S.L.P., based on the non-stationary distribution GVE_2.

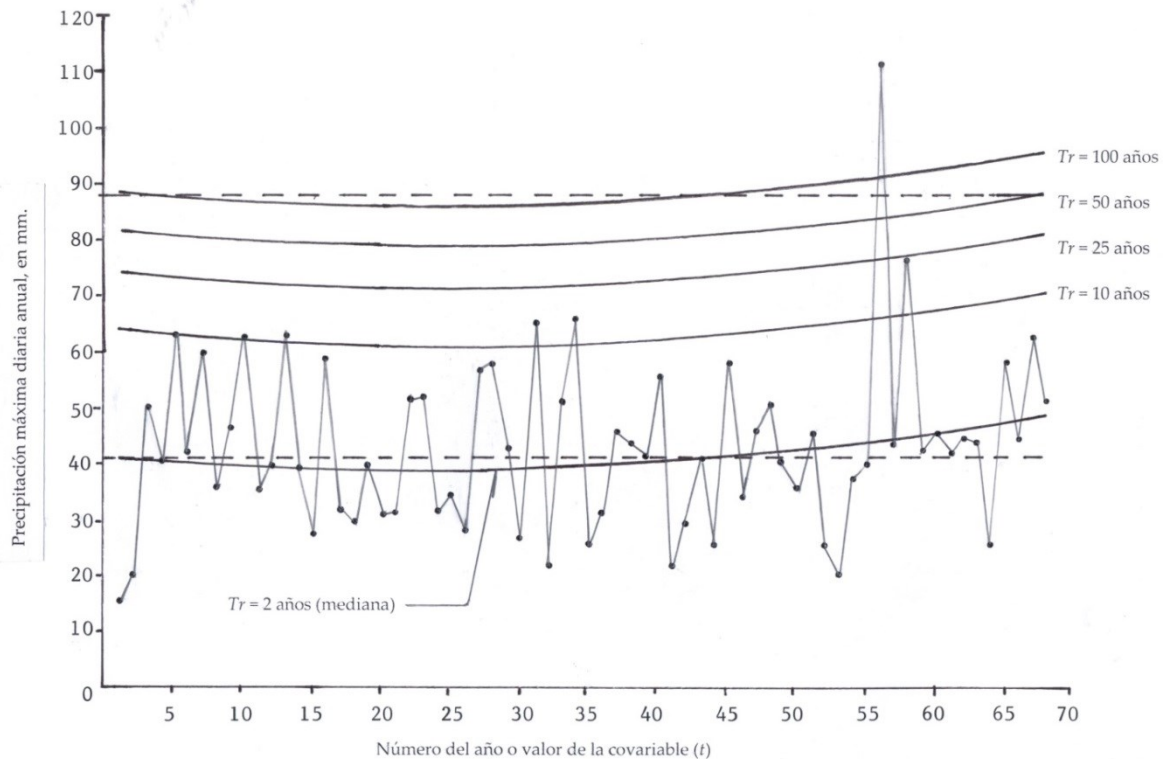| t	Year	Return periods (years)				
		2	10	25	50	100
1	1949	41.3	63.6	74.0	81.4	88.6
10	1958	39.7	62.0	72.4	79.8	86.9
20	1968	38.8	61.1	71.5	79.0	86.1
30	1978	38.9	61.2	71.7	79.1	86.2
40	1988	40.1	62.3	72.8	80.2	87.3
50	1998	42.2	64.5	74.9	82.3	89.4
60	2008	45.3	67.5	78.0	85.4	92.5
68	2016	48.4	70.7	81.2	88.6	95.7
77	2025	52.8	75.1	85.5	92.9	100.0
102	2050	69.0	91.3	101.8	109.2	116.3

Figure 4. Chronological series of annual *PMD* and estimated prediction curves with the distribution GVE$_2$ in Los Filtros station, S.L.P., Mexico.

Predictions in the future

The four numerical applications described cover the most common cases that are known of non–stationary PMD_i records, which correspond to series with downward (severe) and upward (mild and moderate) linear trend and with upward curve trend (mild in an appreciative context) towards the future. Upward trends usually occur in rain–gauge stations located in cities, or where a large nearby reservoir has been built. In the first case there is an increase in temperature due to the effects of the heat island and in the second there may be an increase in relative humidity due to evaporation. Downward trends may be associated to regional climate change.

Naturally, the *predictions* associated with low probabilities of exceedance are less important in records with a downward trend,

because they are lower in the future. The opposite occurs in series with upward trend, in which it is necessary to explain or justify the probable physical origin of such trend, to accept *extrapolations* of the predictions to future and to try to discern about the real scope of them, since it is extremely risky to adopt the behavior of trend in increase, when, for example, it is not known if urban development will continue.

Conclusions

The probabilistic analysis of non–stationary PMD_i records that show trends will be, in the immediate future, increasingly common, due to the impacts of climate change and urban development. A simple approach and without computational difficulties to process such records, is based on the extension of the theory of extreme values through the fit with L–moments, of the non–stationary GVE_1 and GVE_2 distributions with variable location parameter (u) with time (t) or other covariates.

Through the description of the four numerical applications in non–stationary records of PMD_i, the simplicity of the exposed method is observed, as well as the facility to obtain the *predictions* associated with probabilities of non–exceedance. The selection among the exposed models (GVE_1 and GVE_2) depends on the observed trend, so their graphic contrast is basic to validate the descriptive ability of the predictions within the historical record and in the near future (years 2025 and 2050). The numerical results of the standard error of fit will allow the contrast and acceptance of other non–stationary probabilistic models.

Appendix 1: Linear trend line

It is considered that the dependent variable (y) is the maximum annual hydrological data or PMD_i and the times or years t are the abscissa (x), in this case equal to the i–th value i. The slope (μ_1) of the regression line

fitted by least squares of the residuals and the ordinate to the origin (μ_0) are obtained with the following equations (Campos-Aranda, 2003):

$$PMD_i = \mu_0 + \mu_1 \cdot t \text{ (A1.1)}$$

$$\mu_1 = \frac{\text{cov}(PMD,t)}{\text{var}(t)} = \frac{\frac{1}{n}\sum_{i=1}^{n} PMD_i \cdot i - \overline{PMD} \cdot \bar{t}}{\frac{1}{n}\sum_{i=1}^{n} i^2 - \bar{t}^2} \text{ (A1.2)}$$

$$\mu_0 = \overline{PMD} - \mu_1 \cdot \bar{t} \text{ (A1.3)}$$

In the previous equations $\overline{PMD}$ and $\bar{t}$ are the arithmetic means of the series of data and time, which range from 1 to n. The linear correlation coefficient (r_{xy}) measures the degree of dependence or association between the variables PMD_i and t, varies from zero to one, indicating with the unit the perfect regression, its equation is:

$$r_{xy} = \frac{\text{cov}(PMD,t)}{\sqrt{\text{var}(t) \cdot \text{var}(PMD)}} \text{ (A1.4)}$$

being:

$$\text{var}(PMD) = \frac{1}{n}\sum_{i=1}^{n} PMD_i^2 - \overline{PMD}^2 \text{ (A1.5)}$$

Appendix 2: Statistical test of the slope

To test whether the slope (μ_1) of the regression line, obtained with the equation A1.2, is statistically different from zero, a test based on the Student distribution that was proposed by Ostle and Mensing (1975) was used and that applies the statistical DS by means of the following three equations:

$$DS = \frac{\mu_1}{\sqrt{S_\mu^2}} \text{ (A2.1)}$$

$$S_\mu^2 = \frac{S_E^2}{\sum_{i=1}^{n}(i-\bar{t})^2} \text{ (A2.2)}$$

$$S_E^2 = \frac{\sum_{i=1}^{n}\left(PMD_i - PMD_i^e\right)^2}{(n-2)} \text{ (A2.3)}$$

S_E^2 and S_μ^2 are the variances of the errors and of the slope. In the previous equation, PMD_i^e is the value estimated with equation A1.1. If the calculated absolute value DS (equation A2.1) is greater than the critical (DS_c), obtained for the Student distribution with $v = n - 2$ degrees of freedom and $a = 5\%$, in a two–tailed test, the slope μ_1 is significant, that is, there is a linear trend. To estimate the value DS_c the algorithm proposed by Zelen and Severo (1972) is used with $Z = 1.95996$ for a reliability $(1 - a)$ of 95%.

$$DS_c = Z + \frac{G1}{v} + \frac{G2}{v^2} + \frac{G3}{v^3} + \frac{G4}{v^4} \text{ (A2.4)}$$

where:

G1 = $(Z^3 + Z)/4$
G2 = $(5Z^5 + 16Z^3 + 3Z)/96$
G3 = $(3Z^7 + 19Z^5 + 17Z^3 - 15Z)/384$
G4 = $(79Z^9 + 776Z^7 + 1482Z^5 - 1920Z^3 - 945Z)/92160$

Appendix 3: Parabolic Curve for the trend

The dependent variable (y) are the maximum annual hydrological data or PMD_i and the times or years t are the abscissa (x), in this case equal to i–th value i. The coefficients μ_0, μ_1 and μ_2 of the polynomial regression curve are:

$$PMD_i = \mu_0 + \mu_1 \cdot t + \mu_2 \cdot t^2 \text{ (A3.1)}$$

fitted by least squares of the residuals, are obtained based on the following normal equations (Campos-Aranda, 2003):

$$n \cdot \mu_0 + \mu_1 \cdot \sum_{i=1}^{n} i + \mu_2 \cdot \sum_{i=1}^{n} i^2 = \sum_{i=1}^{n} PMD_i \text{ (A3.2)}$$

$$\mu_0 \cdot \sum_{i=1}^{n} i + \mu_1 \cdot \sum_{i=1}^{n} i^2 + \mu_2 \cdot \sum_{i=1}^{n} i^3 = \sum_{i=1}^{n} i \cdot PMD_i \text{ (A3.3)}$$

$$\mu_0 \cdot \sum_{i=1}^{n} i^2 + \mu_1 \cdot \sum_{i=1}^{n} i^3 + \mu_2 \cdot \sum_{i=1}^{n} i^4 = \sum_{i=1}^{n} i^2 \cdot PMD_i \ \text{(A3.4)}$$

In the three previous expressions, *n* is the number of values of the series of extreme hydrological data processed and all the summations lead to numerical magnitudes which are the *coefficients* of μ_0, μ_1 and μ_2 in each row and column (C_{jk}); In addition, there are the magnitudes of the right side of the equations, known as *independent terms*. The determinant Δ of such coefficients is evaluated with the following expression (Chapra & Canale, 1988):

$$\Delta = \begin{vmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{vmatrix} \ \text{(A3.5)}$$

$$\Delta = C_{11} \cdot C_{22} \cdot C_{33} + C_{12} \cdot C_{23} \cdot C_{31} + C_{13} \cdot C_{32} \cdot C_{21} - C_{13} \cdot C_{22} \cdot C_{31} - C_{23} \cdot C_{32} \cdot C_{11} - C_{33} \cdot C_{21} \cdot C_{12} \ \text{(A3.6)}$$

To obtain the sought value of μ_0, a determinant Δ_0 is formed by changing column 1 of coefficients for the three independent terms and evaluating their value with equation A3.6, then:

$$\mu_0 = \frac{\Delta_0}{\Delta} \ \text{(A3.7)}$$

The determinant Δ_1 is formed by changing column 2 of coefficients by the three independent terms, its magnitude is obtained with equation A3.6 and then:

$$\mu_1 = \frac{\Delta_1}{\Delta} \ \text{(A3.8)}$$

Finally, the determinant Δ_2 is formed by changing the third column of coefficients for the three independent terms and their value is quantified with equation A3.6 and now we have that:

$$\mu_2 = \frac{\Delta_2}{\Delta} \ \text{(A3.9)}$$

Appendix 4. Multiple linear regression for the trend

Again, the dependent variable (y) is the maximum annual hydrological data or PMD_i and the times or years t are the abscissa (x), in this case equal to the i–th value i. The coefficients μ_0, μ_1 and μ_2, of the multiple linear regression are:

$$PMD_i = \mu_0 + \mu_1 \cdot t + \mu_2 \cdot h_i \text{ (A4.1)}$$

fitted by least squares of the residuals, are obtained based on the following normal equations (Campos-Aranda, 2003):

$$n \cdot \mu_0 + \mu_1 \cdot \sum_{i=1}^{n} i + \mu_2 \cdot \sum_{i=1}^{n} h_i = \sum_{i=1}^{n} PMD_i \text{ (A4.2)}$$

$$\mu_0 \cdot \sum_{i=1}^{n} i + \mu_1 \cdot \sum_{i=1}^{n} i^2 + \mu_2 \cdot \sum_{i=1}^{n} i \cdot h_i = \sum_{l=1}^{n} i \cdot PMD_i \text{ (A4.3)}$$

$$\mu_0 \cdot \sum_{i=1}^{n} h_i + \mu_1 \cdot \sum_{i=1}^{n} i \cdot h_i + \mu_2 \cdot \sum_{i=1}^{n} h_i^2 = \sum_{i=1}^{n} h_i \cdot PMD_i \text{ (A4.4)}$$

In the three previous expressions, n is the number of values of the series of extreme hydrological data processed and all the summations lead to numerical magnitudes that are the coefficients of μ_0, μ_1 and μ_2 in each row and column (C_{jk}); In addition, there are the magnitudes of the right side of the equations, known as *independent terms*. The determinant Δ of such coefficients is formed with equation A3.5, whose magnitude is obtained with the expression A3.6.
The sought values of μ_0, μ_1 and μ_2 are calculated with the procedures in equations A3.7 to A3.9. Prosdocimi *et al*. (2014, 2015) present the application of two covariables when analyzing maximum flows, using an index related to the extension of the urban area in the basin to model the trend and a climate index based on the *PMD* of return period 100 years of each year, occurred in the basin to reproduce the variability of flows.

Acknowledgments

References

Aissaoui-Fqayeh, I., El Adlouni, S., Ouarda, T. B. M. J., & St. Hilaire, A. (2009). Développement du modèle log-normal non-stationnaire et comparaison avec le modèle GEV non-stationnaire. *Hydrological Science Journal*, 54(6), 141-1156.

Álvarez-Olguín, G., & Escalante-Sandoval, C. A. (2016). Análisis de frecuencias no estacionario de series de lluvia anual. *Tecnología y ciencias de agua*, 7(1), 71-88.

Benson, M. A. (1962). Plotting positions and economics of engineering planning. *Journal of Hydraulics Division*, 88(6), 57-71.

Campos-Aranda, D. F. (2003). Capítulo 5: Ajuste de Curvas. En: *Introducción a los métodos numéricos: software en Basic y aplicaciones en hidrología superficial* (pp. 93-127). San Luis Potosí, México: Editorial Universitaria Potosina.

Campos-Aranda, D. F. (2016). Modelo probabilístico simple para análisis de frecuencias en registros hidrológicos extremos con tendencia. *Tecnología y ciencias del agua*, 7(3), 171-186.

Campos-Aranda, D. F. (2018). Ajuste con momentos L de las distribuciones GVE, LOG y PAG no estacionarias en su parámetro de ubicación, aplicado a datos hidrológicos extremos. *Agrociencia*, 52(2), 169-189*.*

Chapra, S. C., & Canale, R. P. (1988). Tema 7.1: Solución de pocas ecuaciones. En: *Métodos numéricos para ingenieros* (pp. 217-227). México, DF, México: McGraw-Hill/Interamericana de México.

Coles, S. (2001). Chapter 3: Classical Extreme Value Theory and Models. In: *An introduction to statistical modeling of extreme values* (pp. 45-73). London, England: 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.

El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., & Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43(W03410), 1-13.

El Adlouni, S., & Ouarda, T. B. M. J. (2008). Comparaison des méthodes d'estimation des paramètres du modèle GEV non stationnaire. *Revue des Sciences de L'Eau*, 21(1), 35-50.

Gado, T. A., & Nguyen, V. T. V. (2016). An at-site flood estimation in the context of nonstationarity. I: A simulation study. *Journal of Hydrology*, 535, 710-721.

Hosking, J. R., & Wallis, J. R. (1997). Chapter 2: *L*-moments and Appendix: *L*-moments for some specific distributions In: *Regional frequency analysis. An approach based on L-moments* (pp. 14-43, 191-209). Cambridge, England: Cambridge University Press.

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.

Kim, S., Nam, W., Ahn, H., Kim, T., & Heo, J. H. (2015). Comparison of nonstationary generalized logistic models based on Monte Carlo simulation. *Proceedings of the IAHS* (No. 371) (pp. 65-68), 22 June to 2 Jul. Prague, Czech Republic.

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.

López-de-la-Cruz, J., & Francés, F. (2014). La variabilidad climática de baja frecuencia en la modelación no estacionaria de los regímenes de las crecidas en las regiones hidrológicas Sinaloa y Presidio-San Pedro. *Tecnología y ciencias del agua*, 5(4), 79-101.

Meylan, P., Favre, A. C., & Musy, A. (2012). Chapter 6: Validation of the model. In: *Predictive hydrology. A Frequency analysis approach* (pp. 103-117). Boca Raton, USA: CRC Press.

Mudersbach, C., & Jensen, J. (2010). Nonstationary extreme value analysis of annual maximum water levels for designing coastal structures on the German North Sea coastline. *Journal of Flood Risk Management*, 3(1), 52-62.

Mujumdar, P. P., & Nagesh-Kumar, D. (2012). *Floods in a changing climate. Hydrologic modeling*. Cambridge, United Kingdom: International Hydrology Series (UNESCO) and Cambridge University Press.

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.

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research, 49* (1), 187-201.

Prosdocimi, I., Kjeldsen, T. R., & Svensson, C. (2014). Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. *Natural Hazards and Earth System Sciences*, 14(5), 1125-1144.

Prosdocimi, I., Kjeldsen, T. R., & Miller, J. D. (2015). Detection and attribution of urbanization effect on flood extremes using nonstationary flood frequency models. *Water Resources Research*, 51(6), 4244-4262.

Rao, A. R., & Hamed, K. H. (2000). *Flood Frequency Analysis*. Chapter 7: Extreme Value Distributions (pp. 207-257). Boca Raton, USA: CRC Press.

Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. Chapter 18. In: Maidment, D. R. (ed.). *Handbook of hydrology* (pp. 18.1-18.66). New York, USA: McGraw-Hill, Inc.

Strupczewski, W. G., & Kaczmarek, Z. (2001). Non-stationary approach to at-site flood frequency modeling II. Weighted least squares estimation. *Journal of Hydrology*, 248(1-4), 143-151.

Teegavarapu, R. S. V. (2012). *Floods in a changing climate. Extreme Precipitation*. Cambridge, United Kingdom: International Hydrology Series (UNESCO) and Cambridge University Press.

Vogel, R. M., Yaindl, C., & Walter, M. (2011). Nonstationarity: Flood magnification and recurrence reduction factors in the United States. *Journal of the America Water Resources Association*, 47(3), 464-474.

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.