

DOI: 10.24850/j-tyca-2021-03-05

Articles

Fitting of non-stationary distribution GEV_{11} through L moments

Ajuste de la Distribución No Estacionaria GVE_{11} a través de momentos L

Daniel Francisco Campos Aranda¹

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

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

Abstract

Hydrological dimensioning or revision of the hydraulic works and the elaboration of flood risk maps is based on the so-called *Design Floods*, which are maximum flows of the river associated with low probabilities of exceedance. The most reliable way of estimating such *predictions* is through Flood Frequency Analysis (FFA). Its fundamental assumption is

that the stochastic process under study is *stationary*, i.e., it does not change with time. Construction of reservoirs, urbanization and changes in land use in the basin, as well as global or regional climate change, alter the hydrological processes and generate records of annual flows that are *non-stationary*, showing trends and changes in variability. For the FFA of such registries, the extreme value theory has been extended to apply its classical distribution, the Generalized Extreme Values (GEV), with parameters of location (u) and scale (a) varying with time (t), which is introduced as a *covariate*. In this work the L-moment method is presented for the fit of the probabilistic model GVE_{11} whose parameters u and a vary linearly over time. Three numerical applications are described. Conclusions highlight the simplicity of the exposed method and its importance in the estimation of the expected predictions in non-stationary annual maximum data series.

Keywords: Covariate, L-moments, GEV distribution, standard error of fit, trend, linear regression, residuals.

Resumen

El dimensionamiento o la revisión hidrológica de las obras hidráulicas y la elaboración de mapas de riesgo por inundación se realizan con base en las llamadas *crecientes de diseño*, que son gastos máximos del río asociados con bajas probabilidades de excedencia. La manera más confiable de estimar tales *predicciones* es a través del Análisis de Frecuencias de Crecientes (AFC), cuya suposición fundamental es que el proceso estocástico bajo estudio es *estacionario*, es decir, que no cambia

con el tiempo. La construcción de embalses pequeños, la urbanización y los cambios de uso del suelo en la cuenca, así como el cambio climático global o regional alteran los procesos hidrológicos y generan registros de crecientes anuales que son *no estacionarios*, al mostrar tendencias y cambio en su variabilidad. Para el AFC de tales registros se ha extendido la teoría de valores extremos, a fin de aplicar su distribución clásica, la General de Valores Extremos (GVE), con parámetros de ubicación (u) y escala (a) variables con el tiempo (t), que se introduce como *covariable*. En este trabajo se expone el método de momentos L para el ajuste del modelo probabilístico GVE₁₁, cuyos parámetros u y a varían linealmente con el tiempo. Se describen tres aplicaciones numéricas. Las conclusiones destacan la sencillez del método expuesto y su importancia en la estimación de las predicciones buscadas en series de datos máximos anuales no estacionarios.

Palabras clave: covariable, momentos L, distribución GVE, error estándar de ajuste, tendencia, regresión lineal, residuales

Received: 28/08/2019

Accepted: 11/08/2020

Introduction

Flood risk estimation of a certain area and hydrological dimensioning of water works: reservoirs, bridges, walls and protective embankments and channelings, is carried out based on so-called *Design Floods*, which are extreme flows of the river associated with low Exceedance Probability. The most reliable estimation of design floods is carried out through Flood Frequency Analysis (FFA) of the available record or series of maximum annual flows. FFA assumes that the random process that generates such floods is *stationary* and therefore its statistical properties do not change over time (Khaliq, Ouarda, Ondo, Gachon, & Bobée, 2006; El-Adlouni, Ouarda, Zhang, Roy, & Bobée, 2007; El-Adlouni & Ouarda, 2008).

All physical changes that occur in drainage basins, such as the construction of reservoirs, urbanization, channel straightening, deforestation, and land use change alter the hydrological processes, giving rise to series or records of *non-stationary* annual maximum flows, because they present trends and changes in variability. In addition, the impacts of global climate change can exacerbate hydrological process alterations, favoring the formation of non-stationary extreme hydrological data sets. Since early 1990s, non-stationarity due to urbanization and/or climate change was taken into account, correcting the available records, applying incremental factors to predictions, or making predictions in the rural basin and then adjusting the results for their level of urbanization (Jakob, 2013; Prosdocimi, Kjeldsen, & Miller, 2015).

Currently, the statistical Extreme Value Theory (EVT) has been extended to non-stationary conditions, to fit the classical distribution that

follow asymptotically the maximum hydrological data series, i.e. the stationary probabilistic model *Generalized Extreme Value* (GEV_0) with three fitting parameters (u, a, k) . This is based on the introduction of the so-called *covariates*, and one of the most used is time (t) in years, through which the trend observed in data series can be taken into account, adopting the location parameter variable (u_t). When the trend is linear, $u_t = \mu_0 + \mu_1 \cdot t$ and then the four fitting parameter GVE_1 model is fitted (μ_0, μ_1, a, k) . If the trend is curve, $u_t = \mu_0 + \mu_1 \cdot t + \mu_2 \cdot t^2$ and the five fitting parameter GVE_2 model is applied $(\mu_0, \mu_1, \mu_2, a, k)$. In the GVE_{11} model, both the location parameter and the scale parameter ($a_t = \sigma_0 + \sigma_1 \cdot t$) are linear functions of t time and then, in addition to the trend, the change over time of series variability is considered (Khaliq *et al.*, 2006; El-Adlouni & Ouarda, 2008; Aissaoui-Fqayeh, El-Adlouni, Ouarda, & St. Hilaire, 2009; Jakob, 2013; Gado & Nguyen 2016).

Khaliq *et al.* (2006), and Meylan, Favre and Musy (2012) have presented a review of various methods used in FFA, in the non-stationary context. Such techniques include the incorporation of trends in distribution parameters or at statistical moments, the quantile regression method, local likelihood processes, and the Bayesian approach. In the probabilistic models of the first approach, the use of *covariates* aims to integrate changes that have occurred in the past directly into the techniques of FFA, with the aim of extrapolating them to the future. In this context, models that use time as a covariate, case of the one presented here, are simple and useful to reproduce the non-stationary behavior of a flood record, but the acceptance of their predictions is not entirely correct, given that, as indicated by López and Francés (2013),

trends may change in the short or long term, due to climate variability or the intensification of human activities.

Due to the above, the possibility of incorporating climate indices as external forcing into the FFA models has been explored over the past two decades, assuming linear or non-linear dependence. This approach has shown that the use of such *covariates* leads to models that better describe temporal changes occurred in flood records. In fact, a remote connection has been found between the observed changes in flood regimes and the anomalies in the indices (ENSO: El Niño–Southern Oscillation; PDO: Pacific Decadal Oscillation and NAO: North Atlantic Oscillation) describing the temporal evolution of low-frequency atmospheric circulation patterns (Khaliq *et al.*, 2006; López-de-la-Cruz & Francés, 2014).

The *objective* of this work was to expose the theory of fitting Generalized Extreme Values (GEV) distribution in non-stationary conditions, by generalizing the L-moment method, proposed by Gado and Nguyen (2016) for the GVE₁₁ model. Three series of extreme hydrological data taken from the specialized literature are exposed and processed, and their results discussed, highlighting the simplicity and usefulness of the L-moment method to obtain *predictions* within the historical and future record, when they are important, due to the series upward trend and/or its increasing variability.

Operational theory

Generalized Extreme Values Distribution

Annual extreme values are rare events occurring in the right extremity or tail of the probability distribution function (PDF), which is the probabilistic model that defines the behavior of the hydrological random variable under study, such as: floods and maximum values of rainfalls, temperatures, winds and sea levels. The extreme values can be *predicted* based on the PDF, such as the maximum of the random variable that corresponds to a certain *average* interval of recurrence or period of return (Tr), whose exceedance probability is $p = 1/Tr$.

Extreme value theory justifies and states that extreme data follows in an *asymptotic* manner any of the three types of distributions called: Gumbel, Fréchet or Weibull (Clarke, 1973; Stedinger, Vogel, & Foufoula-Georgiou, 1993; Coles, 2001; Khaliq *et al.*, 2006). These three probabilistic models can be represented in one, called *Generalized Extreme Value distribution* (GEV), whose application has been widely recommended for modeling annual maximum flows (q) and other extreme data (Hosking & Wallis, 1997; Papalexiou & Koutsoyiannis, 2013; Gado & Nguyen, 2016; Stedinger, 2017). The PDF of the GEV with a non-exceedance probability $[F(q)]$ is:

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

In the above expression, u , a , and k are the location, scale, and shape parameters of the GEV distribution. When $k = 0$ the Gumbel distribution is obtained, which is a straight line on the Gumbel–Powell probability paper (Chow, 1964), so the interval of the variable is: $-\infty < q < \infty$. When $k > 0$ the distribution is Weibull, which is a curve with downward concavity and upper limit, so: $-\infty < q \leq u + a / k$. Finally, if $k < 0$ the distribution is Fréchet, which is also a curve, but with an upward concavity and lower border, so: $u + a / k \leq q < \infty$. The expected *predictions* (Q_{Tr}) are obtained with the inverse solution of Equation (1):

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

Data sample L-moments

L-moment method is perhaps the simplest one, and it has become one of the reliable procedures for estimating the PDF fitting parameters used in hydrology. This is because L-moments, which are linear combinations

(Hoking & Wallis, 1997) of weighted probability moments (β_r), are not significantly affected by sample *outliers*. The first three L-moments of a sample (l_1, l_2, l_3) and the asymmetry L-ratio (t_3) are estimated by the unbiased estimator (b_r) of the β_r , as follows:

$$l_1 = b_0 \quad (3)$$

$$l_2 = 2 \cdot b_1 - b_0 \quad (4)$$

$$l_3 = 6 \cdot b_2 - 6 \cdot b_1 + b_0 \quad (5)$$

$$t_3 = l_3/l_2 \quad (6)$$

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

$$b_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\cdots(j-r)}{(n-1)(n-2)\cdots(n-r)} q_j \quad (7)$$

where $r = 0, 1, 2, \dots$ and q_j are the data of the available sample or series of size n , ordered from minor to major ($q_1 \leq q_2 \leq \dots \leq q_n$).

Stationary GEV distribution fitting parameters

With the L-moment method for the GVE₀ model, the following equations are obtained for its three fitting parameters (Stedinger *et al.*, 1993; Hosking & Wallis, 1997; Rao & Hamed, 2000; Stedinger, 2017):

$$k \cong 7.8590 \cdot c + 2.9554 \cdot c^2 \quad (8)$$

being:

$$c = \frac{2}{3+t_3} - 0.63093 \quad (9)$$

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

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

The Stirling formula (Davis, 1972) was used to estimate the value of Gamma $\Gamma(\omega)$ function:

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

Non-stationary analysis approach

It has two inherent assumptions, the first accepts that non-*stationarity* of the hydrological series of extreme annual values is caused by gradual changes in land use or by global or regional climate change; generating a slight alteration of its statistical parameters. The second assumption accepts that the PDF is time-independent, so the GEV distribution, with variable fitting parameters over time, is acceptable for modeling extreme non-stationary data.

Due to non-stationarity, there is great statistical variability associated with the sample or series of available extreme annual data and it is therefore appropriate to adopt that mean (μ) and standard deviation (σ) are functions of time. The simplest way, is carried out by means of the probabilistic model GEV₁₁ with linear variations, so: $\mu_t = \mu_0 + \mu_1 \cdot t$ and $\ln(\sigma_t) = \sigma_0 + \sigma_1 \cdot t$. In the above expressions, t is the time-in-years *covariate* that encompasses the n -size data series. The use of natural logarithms of σ ensures positive values of the scale parameter (a) in the GEV distribution (Gado & Nguyen, 2016). The general expressions of the mean

and standard deviation of the GEV_0 model, the second one obtained through the second central moment, are (Rao & Hamed, 2000):

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

$$\sigma^2 = \frac{\alpha^2}{k^2} [\Gamma(1 + 2 \cdot k) - \Gamma^2(1 + k)] \quad (14)$$

GEV₁₁ distribution fitting

Since the mean and the natural logarithm of the standard deviation are linear functions of time, the expression: $\mu_t = \mu_0 + \mu_1 \cdot t$ is first entered in Equation (13) to clear the variable location parameter expression over time (u_t), in which the scale parameter (α_t) is also variable, and where FK_1 is the first function factor of the form parameter (k), such an equation is (Gado & Nguyen, 2016):

$$u_t = \mu_0 + \mu_1 \cdot t - \frac{\alpha_t}{k} [1 - \Gamma(1 + k)] = \mu_0 + \mu_1 \cdot t - FK_1 \cdot \alpha_t \quad (15)$$

where:

$$FK_1 = \frac{1 - \Gamma(1+k)}{k} \quad (16)$$

From Equation (14), you get the variable scale parameter expression over time:

$$\alpha_t^2 = \frac{k^2 \cdot \sigma_t^2}{\Gamma(1+2 \cdot k) - \Gamma^2(1+k)} \quad (17)$$

So:

$$\alpha_t = FK_2 \cdot \sigma_t = FK_2 \cdot \exp(\sigma_0 + \sigma_1 \cdot t) \quad (18)$$

being:

$$FK_2 = \sqrt{\frac{k^2}{\Gamma(1+2 \cdot k) - \Gamma^2(1+k)}} \quad (19)$$

Substituting Equation (18) in the Equation (15), the variable scale parameter expression over time is obtained:

$$u_t = \mu_0 + \mu_1 \cdot t - FK_1 \cdot FK_2 \cdot \exp(\sigma_0 + \sigma_1 \cdot t) \quad (20)$$

The magnitudes μ_0 and μ_1 of the straight line representing the linear trend of the sample or data series q_t are obtained based on the equations of the linear regression straight line, as follows. The dependent variable (y) is considered to be the maximum annual hydrological data q_t and the times or years t are the abscissa (x), in this case equal to the i -th value of i . The slope (μ_1) of the regression straight line adjusted by least squares of the residuals and the y -intercept (μ_0) are obtained with the following equations (Campos-Aranda, 2003):

$$q_t = \mu_0 + \mu_1 \cdot t \quad (21)$$

$$\mu_1 = \frac{\text{cov}(q,t)}{\text{var}(t)} = \frac{\frac{1}{n} \sum_{i=1}^n q_i i - \bar{q} \cdot \bar{t}}{\frac{1}{n} \sum_{i=1}^n i^2 - \bar{t}^2} \quad (22)$$

$$\mu_0 = \bar{q} - \mu_1 \cdot \bar{t} \quad (23)$$

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

$$r_{xy} = \frac{\text{cov}(q,t)}{\sqrt{\text{var}(t) \cdot \text{var}(q)}} \quad (24)$$

being:

$$\text{var}(q) = \frac{1}{n} \sum_{i=1}^n q_i^2 - \bar{q}^2 \quad (25)$$

To estimate the linear trend in the standard deviation (σ_t), first the residuals (ε_t) are obtained with the following equation (Cunderlik & Burn, 2003; Gado & Nguyen, 2016):

$$\varepsilon_t = q_t - \mu_1 \cdot t \quad (26)$$

Mean value is obtained ($\bar{\varepsilon}$) and absolute deviations (da) of the residuals are calculated, with the following expression:

$$\varepsilon_t^{da} = |\varepsilon_t - \bar{\varepsilon}| \quad (27)$$

Natural logarithms are applied to ε_t^{da} and represented by a straight line whose equation is:

$$\ln(\varepsilon_t^{da}) = \sigma_0 + \sigma_1 \cdot t \quad (28)$$

The values of σ_1 and σ_0 are calculated based on equations (22) and (23). The degree of association between $\ln(\varepsilon_t^{da})$ and t is estimated with Equation (24) and is designated r_y , since the y series is the left side of Equation (28). The following equations apply to obtain the stationary series without trends in mean and standard deviation ($q_e^{\mu,\sigma}$) (Cunderlik & Burn, 2003; Khaliq *et al.*, 2006; Gado & Nguyen, 2016):

$$q_e^{\mu,\sigma} = \varepsilon_t - \sigma_t \text{ if } \varepsilon_t \geq \bar{\varepsilon} \quad \text{when } \sigma_t > 0 \quad (29)$$

$$q_e^{\mu,\sigma} = \varepsilon_t + \sigma_t \text{ if } \varepsilon_t < \bar{\varepsilon} \quad \text{when } \sigma_t > 0 \quad (30)$$

$$q_e^{\mu,\sigma} = \varepsilon_t + \sigma_t \text{ if } \varepsilon_t \geq \bar{\varepsilon} \quad \text{when } \sigma_t < 0 \quad (31)$$

$$q_e^{\mu,\sigma} = \varepsilon_t - \sigma_t \text{ if } \varepsilon_t < \bar{\varepsilon} \quad \text{when } \sigma_t < 0 \quad (32)$$

In the above expressions, $\sigma_t = \exp(\sigma_0 + \sigma_1 \cdot t)$. Equations (9) and (8) are applied to the series $q_e^{\mu,\sigma}$ to estimate the form parameter (k) of the stationary GEV_0 distribution. Finally, Equation (2) is applied to carry out the desired predictions, using the expressions of the variable location and scale parameters over time, equations (20) and (18), respectively.

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 fitting probabilistic model (Meylan *et al.*, 2012). The *EEA* allows objective comparison between the various PDFs that are tested or adjusted to a series or sample of data, since it has the units of the data (q_i). Its expression is as follows (Kite, 1977):

$$EEA = \sqrt{\frac{\sum_{i=1}^n (q_i - \hat{q}_i)^2}{n - npa}} \quad (33)$$

where n is the number of data in the available series, npa is the number of PDF fitting parameters being tested (five for the GVE₁₁), q_i are the data sorted from lowest to highest and $\hat{q}_i$ are the values estimated with the inverse solution of the PDF (Equation (2)), for the non-exceedance probability $P(X \leq x)$ estimated with the Weibull formula (Benson, 1962):

$$P(X \leq x) = \frac{m}{n+1} \quad (34)$$

in which m is the data order number, with 1 for the lowest and n for the highest.

Approach to probabilistic analysis

For the GEV₁₁ distribution to be applicable to non-stationary record, its graph of data values (q_i) against time (t_i) must show linear trend and variability or standard deviation that increases or decreases over time. The calculation of the standard error of fit (EEA) with Equation (33), allows the comparison or matching of the various non-stationary models tested in the series being processed. When the EEA values are similar, a non-stationary model can be adopted in a subjective manner, for example, which leads to the most unfavorable predictions.

For other non-stationary distributions, there are several such as Log-Normal (Vogel, Yaoundi, & Walter, 2011; Aissaoui-Fqayeh *et al.*, 2009) and the Generalized Logistics (Kim, Nam, Ahn, Kim, & Heo, 2015) and Generalized Pareto models, which are susceptible to treatment identical to that described in this paper for GEV distribution, changing equations (13) and (14) to those corresponding to these models (Rao & Hamed, 2000). Other covariates can also be used instead of time (t), for example some global or regional climate indices (López-de-la-Cruz & Francés,

2014; Álvarez-Olguín & Escalante-Sandoval, 2016; Campos-Aranda, 2018a).

Based on Equation (2), *predictions* with return periods (Tr) of 2, 10, 25, 50, and 100 years are estimated through the registration period, applying equations (20) and (18). The first prediction corresponds to the *median*, since its non-exceedance probability ($1-p$) is 50 % and the following four are calculated for the following values: 0.90, 0.96, 0.98 and 0.99 respectively. In addition, future predictions are made in one of the processed series, in years 2025 and 2050. Extrapolating over 30 years of the observed behavior in the historical trend and its variability is considered to be extremely risky.

In this regard, *mobile L-moment* method suggested by Mudersbach and Jensen (2010) and applied by Campos-Aranda (2018b), whose main limitation is to require fairly extensive records, is considered much more reliable. Also shown in data and prediction graphs the estimates of extreme Tr (2 and 100 years) with stationary GEV_0 model, which are horizontal dotted straight lines.

Processed hydrological series

Series 1: With a downward trend and a decrease in variability

This record was exposed by Katz (2013) as an impressive example of the reduction in the winter (May to October) maximum daily precipitation (*MDP*) measured at the Manjimup station in the extreme southwest of Australia. The record covers 75 annual values (1930 to 2004), with a declining trend and greater variability at the beginning of the record. Their approximate magnitudes are shown in Table 1, as they were read from a bar graph and are shown in Figure 1.

Table 1. Maximum annual data to be processed for maximum daily precipitation (*MDP*, in millimeters) and flow (*q*, m³/s) at the indicated stations.

Data No.	(<i>MDP</i>) Manjimup station	(<i>q</i>) Aberjona River	(<i>q</i>) Mercer Creek	Data No.	(<i>MDP</i>) Manjimup station	(<i>q</i>) Aberjona River	(<i>q</i>) Mercer Creek
1	35.7	9.1	6.8	39	34.7	18.7	5.2
2	49.0	5.9	5.1	40	24.8	39.6	8.9
3	37.8	7.4	6.7	41	33.6	7.1	15.5
4	47.1	6.8	6.2	42	24.8	9.3	17.6

5	84.0	5.0	5.9	43	29.9	24.4	9.1
6	44.9	6.2	5.4	44	28.3	12.7	12.3
7	54.1	9.1	4.8	45	34.8	22.9	11.0
8	45.9	9.1	4.3	16	36.1	8.5	8.6
9	64.8	10.2	6.4	47	34.1	14.7	15.5
10	35.1	3.4	5.4	48	34.9	24.6	18.1
11	52.3	7.1	5.2	49	35.0	8.8	21.3
12	63.9	6.8	7.2	50	30.0	8.8	10.9
13	42.5	6.2	5.0	51	35.0	19.8	13.1
14	31.4	7.1	7.1	52	28.1	13.9	–
15	54.6	13.9	5.3	53	28.0	6.8	–
16	97.5	19.5	5.7	54	51.5	13.9	–
17	49.6	10.8	11.5	55	32.1	15.0	–
18	95.9	4.8	6.9	56	37.0	5.9	–
19	50.0	11.0	9.1	57	24.1	9.1	–
20	42.9	4.8	9.6	58	32.0	34.0	–
21	51.8	5.4	7.8	59	63.7	31.1	–
22	35.6	6.8	7.5	60	33.3	13.3	–
23	37.2	10.8	11.4	61	46.3	11.0	–
24	57.0	22.4	13.3	62	51.8	45.3	–

25	53.9	6.5	14.7	63	37.9	6.8	–
26	54.8	8.8	11.9	64	41.2	8.8	–
27	43.5	2.8	19.0	65	34.1	28.3	–
28	47.0	10.8	17.4	66	35.0	6.5	–
29	36.7	18.4	11.5	67	35.0	36.8	–
30	29.2	13.0	10.0	68	43.4	15.3	–
31	34.3	21.2	23.6	69	36.0	8.2	–
32	48.2	4.0	14.2	70	51.4	–	–
33	50.7	15.6	9.5	71	30.0	–	–
34	37.6	8.5	6.5	72	34.1	–	–
35	42.7	7.4	18.9	73	37.0	–	–
36	31.9	6.5	15.1	74	36.0	–	–
37	38.8	13.0	7.8	75	27.3	–	–
38	29.4	10.2	7.9	–	–	–	–

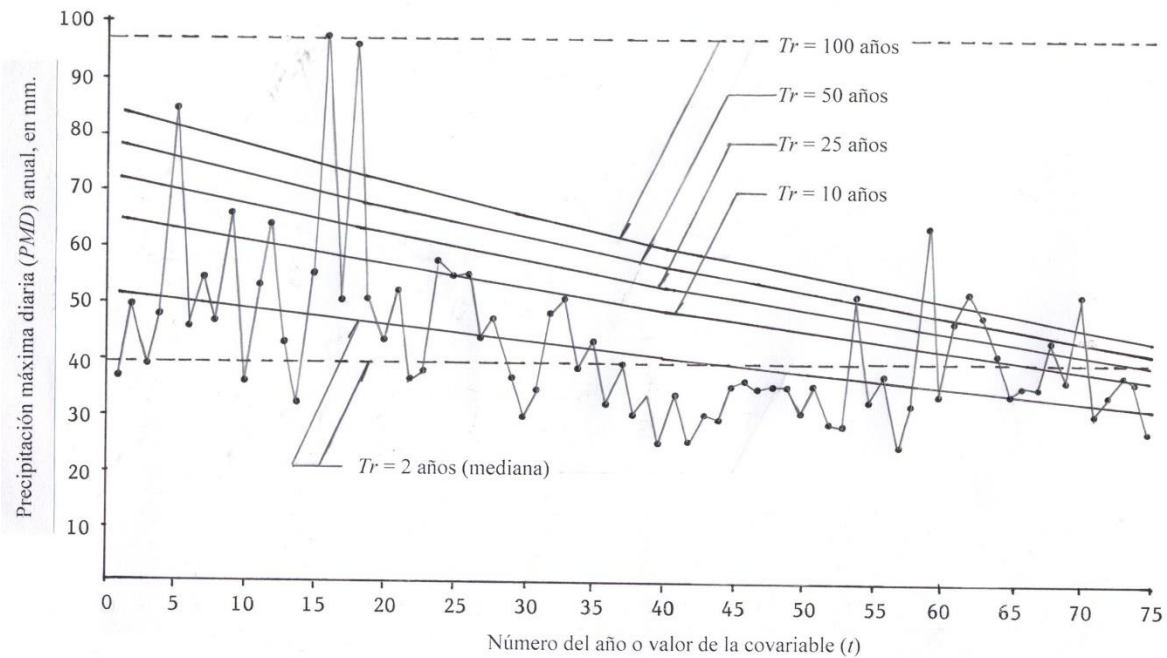


Figure 1. Time series of data (*MDP*) and estimated prediction curves with GVE_{11} distribution at Manjimup pluviometric station, Australia.

Series 2: With an upward trend and an increase in variability

This annual maximum flows record was processed by Vogel *et al.* (2011), it has 69 values in the period from 1940 to 2008, and comes from a gauging station located on the Aberjona River just after the city of Boston,

Massachusetts, USA, so its basin has always been under the increasing impact of urban development. Their approximate values are shown in Table 1, as they were read in a dispersion diagram and are shown in Figure 2.

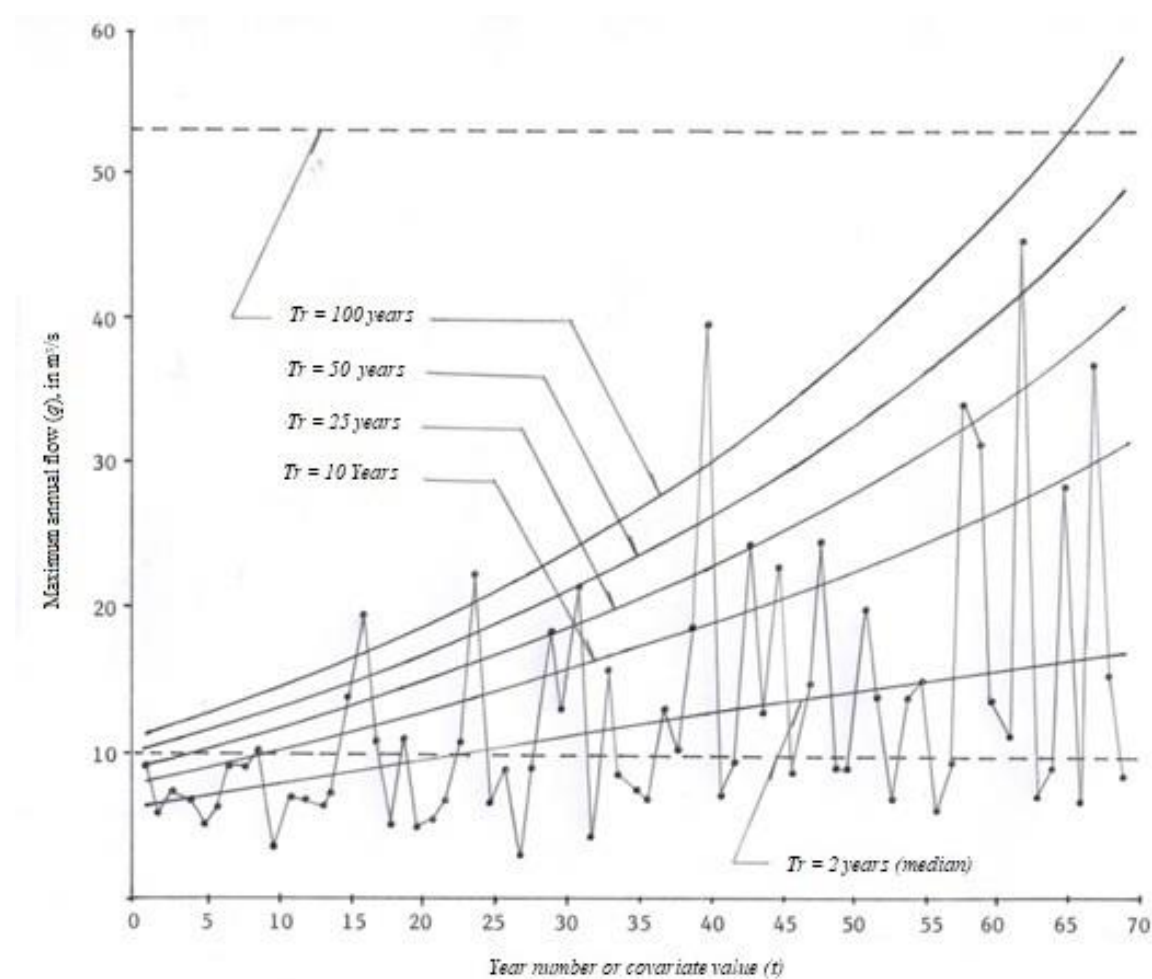


Figure 2. Time series of data (q) and estimated prediction curves with the GEV₁₁ distribution at a gauging station on the Aberjona River, USA

Series 3: Segment with an upward trend and an increase in variability

This annual maximum flow record was presented by Gilleland and Katz (2011), and Katz (2013), from a basin in Washington, USA whose Mercer Creek gauging station has 51 values in the period from 1956 to 2006. In the 15-year interval between 1971 and 1985, this record was influenced by an increasing and intense urbanization. Because of this, only such a lapse should be considered non-stationary. The approximate data of that series are given in Table 1, as they were read from a bar graph (Katz, 2013) and are shown in Figure 3.

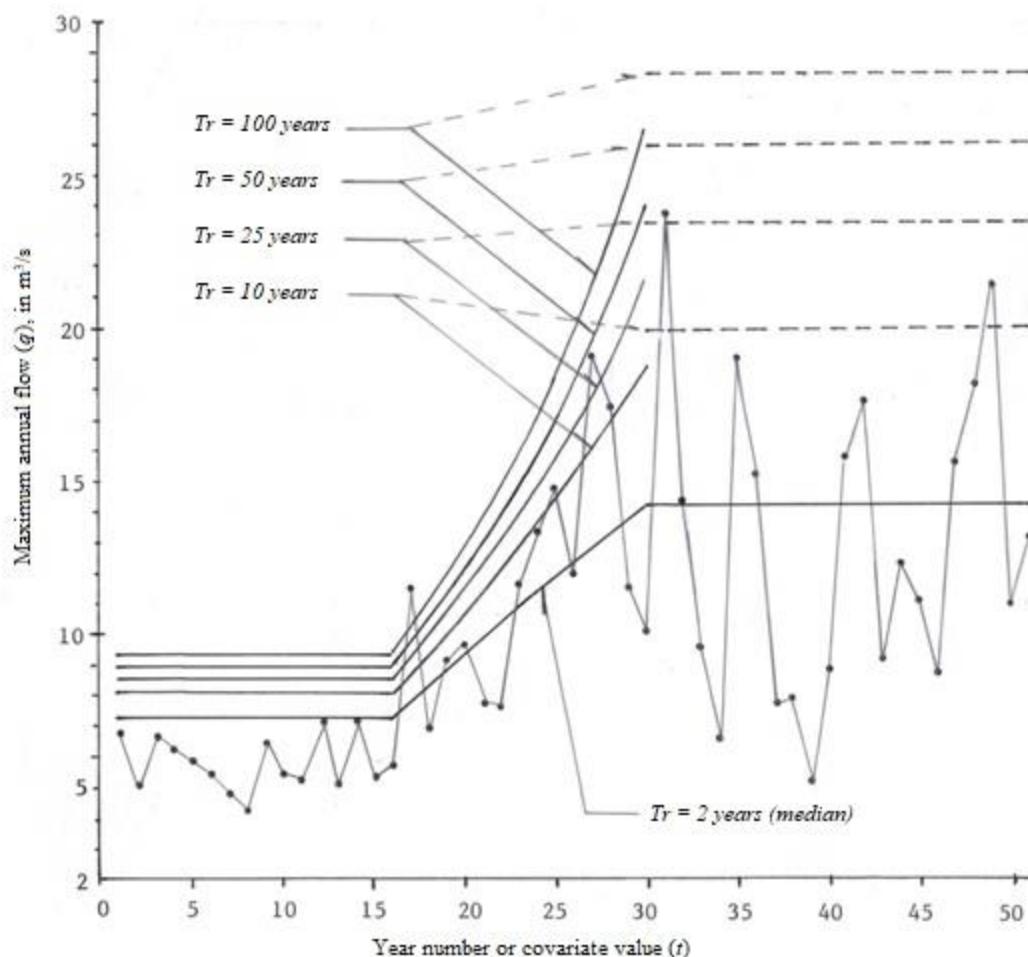


Figure 3. Time series of data (q) and estimated prediction curves with the GEV_{11} distribution at Mercer Creek gauging station, Washington, USA.

Description of results

Predictions at Manjimup station

The application of equations (22) to (25) to the 75 *MDP* data in Table 1 led to the following results: $\mu_0 = \bar{\varepsilon} 52.8382$, $\mu_1 = -0.2811$, $r_{xy} = -0.4313$, $\sigma_0 = 2.2767$, $\sigma_1 = -0.0132$, and $r_y = -0.2701$. The stationary series $q_e^{\mu,\sigma}$ according to equations (29) to (32) accepts a GEV_0 model with shape parameter: $k = -0.0580$.

Equation (2) using the variable location (Equation (20)) and scale (Equation 18) parameters leads to a standard error of fit (*EEA*) of 15.5 mm. Table 2 shows some of the predictions within the historical record, and in the future its predictions are not important due to the downward linear trend and the decrease in variability. The predictions with GVE_0 stationary model of return periods 2 and 100 years are: 38.7 and 96.6 mm. Figure 1 shows the data and the downward prediction curves.

Table 2. Predictions (mm) in the historical period at the Manjimup station, based on non-stationary GEV_{11} distribution.

No. (<i>t</i>)	Year	Data (mm)	Return periods in years				
			2	10	25	50	100
1	1930	35.7	50.9	64.4	71.8	77.6	83.5

10	1939	35.1	48.5	60.6	67.1	72.2	77.5
20	1949	42.9	45.9	56.4	62.2	66.7	71.3
30	1959	29.2	43.2	52.5	57.5	61.5	65.5
40	1969	24.8	40.6	48.7	53.1	56.5	60.1
50	1979	30.0	37.9	45.0	48.9	51.9	55.0
60	1989	33.3	35.2	41.4	44.8	47.5	50.2
70	1999	51.4	32.5	37.9	40.9	43.2	45.6
75	2004	27.3	31.1	36.2	39.0	41.2	43.4

Predictions on the Aberjona River

The application of equations (22) to (25) to the 69 annual maximum flows (q) in Table 1 led to the following results: $\mu_0 = \bar{\varepsilon} = 6.1849$, $\mu_1 = 0.1902$, $r_{xy} = 0.4256$, $\sigma_0 = 0.2563$, $\sigma_1 = 0.0304$, and $r_y = 0.5496$. The stationary series $q_e^{\mu, \sigma}$ according to equations (29) to (32) accepts a GEV_0 model with a shape parameter: $k = -0.1857$; instead, Equation (2) using the variable location (Equation (20)) and scale (Equation (18)) parameters leads to a standard error of fit (EEA) of 2.39 m³/s. Table 3 shows some of the predictions within the historical record and in the future their predictions are important due to the upward linear trend and the increase in

variability. Since this record of 69 data ends in 2008, then the value of time t in 2025 is 86 and in 2050 is 111. The predictions with GVE₀ stationary model of return periods 2 and 100 years are: 10.1 and 53.3 m³/s. Figure 2 shows the data and the upward prediction curves.

Table 3. Predictions (m³/s) in the historical period and in the future at Aberjona River, based on the non-stationary GEV₁₁ distribution.

No. (t)	Year	Data (m ³ /s)	Return periods in years				
			2	10	25	50	100
1	1940	9.1	6.1	7.9	9.1	10.1	11.3
10	1949	3.4	7.7	10.1	11.7	13.0	14.6
20	1959	4.8	9.4	12.7	14.8	16.7	18.8
30	1969	13.0	11.1	15.6	18.5	21.0	23.8
40	1979	39.6	12.8	18.8	22.7	26.1	29.9
50	1989	8.8	14.3	22.5	27.8	32.4	37.6
60	1999	13.3	15.7	26.8	34.0	40.2	47.2
69	2008	8.2	16.8	31.4	40.8	49.0	58.3
86	2025	–	18.4	42.7	58.6	72.4	87.9
111	2050	–	18.4	70.5	104.6	134.0	167.2

Predictions at Mercer Creek Station

The application of equations (22) to (25) to the 15 data from 1971 to 1985, i.e., data with numbers 16 to 30 in Table 1, led to the following results: $\mu_0 = \bar{\varepsilon} = 6.9048$, $\mu_1 = 0.5311$, $r_{xy} = 0.6299$, $\sigma_0 = -0.7769$, $\sigma_1 = 0.1292$ y $r_y = 0.3835$. The stationary series $q_e^{\mu, \sigma}$ according to equations (29) to (32) accepts a GEV_0 model with a shape parameter: $k = -0.1451$; instead, Equation (2) using the variable location (Equation (20)) and scale (Equation (18)) parameters leads to a standard error of fit (EEA) of $0.85 \text{ m}^3/\text{s}$.

Table 4 shows a portion of the predictions within the historical record and Figure 3 illustrates the data, the upward curves of predictions, and their respective extrapolations.

Table 4. Predictions (m^3/s) in the historical period at the Mercer Creek station, based on the non-stationary GEV_{11} distribution.

No. (t)	Year	Data (m^3/s)	Return periods in years				
			2	10	25	50	100
1	1971	5.7	7.3	8.1	8.5	8.9	9.3
4	1974	9.1	8.9	9.9	10.6	11.2	11.8
7	1977	7.5	10.4	12.0	13.0	13.8	14.7

10	1980	14.7	11.8	14.2	15.7	16.9	18.2
13	1983	17.4	13.3	16.7	18.9	20.7	22.7
15	1985	10.0	14.2	18.7	21.5	23.8	26.3

For this data series, the predictions with the stationary GEV_0 model are of capital importance to *approve* the extrapolation, toward the beginning and end of the record, of the predictions of GEV_{11} model that will be the most critical one (Gilleland & Katz, 2011). Therefore, two fittings were made: (1) with the initial stationary record of 15 data (1956–1970) and (2) with the lapse of 21 values (1986–2006) occurring after the trend and increasing variability segment (1971–1985). The predictions of the first fitting were: 5.7, 7.0, 7.6, 7.9, and 8.2 m³/s for the return periods (Tr) of 2, 10, 25, 50, and 100 years, with an $EEA = 0.193$ m³/s; the predictions for the second fitting were: 12.2, 19.9, 23.4, 25.9 and 28.2 m³/s, with an $EEA = 0.691$ m³/s.

Then, according to Table 4 results, all predictions of GEV_{11} are extrapolable toward the start of the record, being greater than those of GEV_0 . Towards the end of the record, only the median is extrapolated ($Tr = 2$ years), which is 14.2 m³/s, due to the predictions of the other Tr ; those obtained with GEV_0 are more extreme than those in the last line of Table 4. The above is shown in Figure 3.

Discussion of results

From a practical point of view, the three numerical applications described cover the common cases in which is *applicable* the GVE₁₁ model to non-stationary extreme hydrological data records; these are:

1. Series with downward trend and variability, present in records of floods when several small reservoirs have been built within the drainage basin; such behavior may also be associated with global climate change, which was the case with the Manjimup pluviometric station.
2. Records of flows with upward trend, usually associated with basins that have intense urban development or deforestation, which is the case with Mercer Creek station.

If both processes occur in the basin, or are enhanced by the impacts of climate change, variability also increases toward the end of the sample, which was the case with the Aberjona River.

Predictions, beyond the historical record ($t > n$), associated with low exceedance probabilities (*Design Floods*) lack importance in records with downward trend and variability, as they are smaller in the future. The opposite is true for series with upward trend and increasing variability, but in these series it is necessary to explain or account for the probable physical origin of such behavior in order to accept future predictions and try to discern their actual extent. This, because it is extremely risky to

accept the behavior of increasing variability, when, for example, it is not known whether urban development will continue in the drainage basin.

To further explain the above, it is indicated that Meylan *et al.* (2012) have noted that the various approaches or new methods of the FFA, which accept nonstationarity, have generated techniques that allow predictions to be made, associated with low exceedance probability to face the design of hydraulic works. However, there is an urgent need to replace the concept of return periods (Tr , in years), to adapt it to the context of nonstationarity. This is because the Tr is defined with the average recurrence interval, measured over a large number of occurrences. As it happens now, in nonstationarity, mean changes over time, the concept of Tr is meaningless. Details and solutions have been proposed in the references of Sivapalan and Samuel (2009); Salas and Obeysekera (2014); Serinaldi (2015), and Salas, Obeysekera and Vogel (2018).

Conclusions

Flood Frequency Analysis (FFA) in non-stationary records showing non-constant trend and/or variability will be increasingly common in the immediate future, due to increased demand for drinking water and food, as well as to the impacts of climate change and urban development. A

simple and uncomplicated computational approach that allows the FFA to be performed on such records is based on the extension of the extreme values theory through the fitting with L-moments, of non-stationary distribution GEV_{11} with location parameter and scale variables linearly with the time that is entered as *covariate*. Therefore, GEV_{11} model is suitable when the series of extreme hydrological data shows trend and increase or decrease in variability that can be accepted as linear.

Based on the results of the three numerical applications in non-stationary records, the simplicity of the above method is observed, as well as the ease of obtaining the *predictions* associated with non-exceedance probabilities. Graphical contrast is basic for validating the descriptive ability of predictions within the historical record and in the near future. The numerical results of the standard error of fit will allow the contrast and acceptance of other probabilistic non-stationary models. The application of the GEV_{11} probabilistic model to other extreme hydrological data, such as winds, temperatures, droughts and sea levels, is fully feasible.

Acknowledgements

The corrections suggested by the three anonymous arbitrators A, C and E were appreciated, which allowed the concepts of non-stationary and covariates to be expanded, and the type of method exposed and its advantages to be placed among those available for non-stationary FFA.

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. (2018a). 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.
- Campos-Aranda, D. F. (2018b). Ajuste con momentos L móviles de la distribución GVE con parámetros variables de ubicación y escala. *Agrociencia*, 52(7), 933-950.
- Chow, V. T. (1964). Statistical and probability analysis of hydrologic data. Frequency analysis. In: Chow, V. T. (ed.). *Handbook of applied hydrology* (pp. 8.1-8.42). New York, USA: McGraw-Hill Book Co.

- Clarke, R. T. (1973). Chapter 5: The estimation of floods with given return period. In: *Mathematical models in hydrology* (pp. 130-146) (Irrigation and Drainage Paper 19). Rome, Italy: FAO.
- Coles, S. (2001). Chapter 3: Classical Extreme Value Theory and Models. In: *An introduction to statistical modeling of extreme values* (pp. 45-73). London, UK: Springer-Verlag.
- Cunderlik, J. M., & Burn, D. H. (2003). Non-stationary pooled flood frequency analysis. *Journal of Hydrology*, 276(1-4), 210-223.
- 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.
- 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(3), 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 non-stationarity. I: A simulation study. *Journal of Hydrology*, 535, 710-721.
- Gilleland, E., & Katz, R. W. (2011). New software to analyze how extremes change over time. *Eos*, 92(2), 13-14.

- Hosking, J. R., & Wallis, J. R. (1997). Chapter 2: L-moments. Regional frequency analysis and Appendix: *L-moments for some specific distributions*. In: *An approach based on L-moments* (pp. 14-43, 191-209). Cambridge, UK: Cambridge University Press.
- Jakob, D. (2013). Non-stationarity in extremes and engineering design. In: AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (eds.). *Extremes in a changing climate* (pp. 363-417). Dordrecht, The Netherlands: Springer.
- Katz, R. W. (2013). Statistical methods for non-stationary extremes. In: AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (eds.). *Extremes in a changing climate* (pp. 15-37). Dordrecht, The Netherlands: Springer.
- Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J. C., Gachon, P., & Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, 329(3-4), 534-552.
- 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* (pp. 65-68) (No. 371). 22 June to 2 July, 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, J., & Francés, F. (2013). Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. *Hydrology and Earth System Sciences*, 17(8), 3189-3203.
- 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. Chapter 9: Perspectives. In: *Predictive hydrology. A frequency analysis approach*. (pp. 103-117, pp. 161-187). 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.
- 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., & 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). Chapter 7: Extreme Value Distributions. In: *Flood frequency analysis* (pp. 207-257). Boca Raton, USA: CRC Press.
- Salas, J. D., & Obeysekera, J. (2014). Revisiting the concepts of return period and risk for nonstationary hydrology extreme events. *Journal of Hydrologic Engineering*, 19(3), 554-568.
- Salas, J. D., Obeysekera, J., & Vogel, R. M. (2018). Techniques for assessing water infrastructure for nonstationary extreme events: A review. *Hydrological Sciences Journal*, 63(3), 325-352.
- Serinaldi, F. (2015). Dismissing return periods! *Stochastic Environmental Research and Risk Assessment*, 29(4), 1179-1189.
- Sivapalan, M., & Samuel, J. M. (2009). Transcending limitations of stationary and the return period: Process-based approach to flood estimation and risk. *Hydrological Processes*, 23(11), 1671-1675.
- Stedinger, J. R. (2017). Chapter 76. 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.
- Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In: Maidment, D. R. (ed.). *Handbook of Hydrology* (pp. 18.1-18.66) New York, USA: McGraw-Hill, Inc.
- Vogel, R. M., Yaindl, C., & Walter, M. (2011). Nonstationarity: Flood magnification and recurrence reduction factors in the United States. *Journal of the American Water Resources Association*, 47(3), 464-474.