INGENIERÍA Investigación y Tecnología VIII. 3. 123-133, 2007
(artículo arbitrado)

A Mixed distribution with EV1 and GEV components for analyzing heterogeneous samples

C. Escalante-Sandoval
División de Ingeniería Civil y Geomática
Facultad de Ingeniería, UNAM
E-mail: caes@servidor.unam.mx

(Recibido: agosto de 2006; aceptado: abril de 2007)

Abstract

Flood characteristics are required to solve several water-engineering problems. Traditional flood frequency analysis involves the assumption of homogeneity of the flood distribution. However, floods are often generated by distributions composed of a mixture of two or more populations. Differences between the populations may be the result, for instance, of the ENSO phenomenon. If these physical processes are not considered in conventional flood frequency analysis, the T-year flood estimate can be inefficient for design purposes. In order to model heterogeneous samples, a mixed distribution with Extreme Value Type I (EV1 or Gumbel) and General Extreme Value (GEV) components is proposed. A region in Northwestern Mexico with 35 gauging stations has been selected to apply the model and at-site quantiles were estimated based on the maximum likelihood procedure. Results produced by fitting the EV1-GEV distribution were compared through the use of a goodness-of-fit test with those obtained by the mixed Gumbel and mixed GEV distributions. The EV1-GEV distribution was the best option for the 40% of analyzed samples and thus it is suggested its application when modeling heterogeneous series in flood frequency analysis.

Keywords: *Heterogeneous samples, flood frequency analysis, mixed distributions, maximum likelihood parameter estimation.*

Resumen

Muchos problemas en ingeniería hidráulica requieren conocer las características de una creciente. El análisis tradicional de frecuencias implica la consideración de homogeneidad de la serie. Sin embargo, en ocasiones los gastos máximos anuales son generados por distribuciones formadas por dos o más poblaciones. La diferencia entre poblaciones puede ser el resultado, entre otros, de la presencia del fenómeno ENSO. Si estos procesos físicos no se consideran en el análisis convencional, el evento estimado de cierto período de retorno puede ser ineficiente para propósitos de diseño. Con el fin de modelar muestras heterogéneas se propone la aplicación de una distribución mezclada, cuyas componentes son la distribución de Valores Extremos Tipo 1 (VE1 o Gumbel) y la General de Valores Extremos (GVE). Para aplicar el modelo se eligió una región del Noroeste de México que cuenta con 35 estaciones de aforos y se empleó la técnica de máxima verosimilitud para la estimación de los eventos de diseño. Los resultados de la distribución VE1-GVE, se compararon con aquellos obtenidos con las distribuciones Gumbel mixta y GVE mixta, a través de un criterio de bondad de ajuste. La distribución EV1-GVE fue la de mejor ajuste en el 40% de las muestras analizadas, por lo que se sugiere su aplicación en el caso de requerir estimar eventos de diseño a partir de series no homogéneas.

Descriptores: Muestras heterogéneas, análisis de frecuencias de crecientes, distribuciones mezcladas, estimación de parámetros por máxima verosimilitud.

Introduction

The objective of flood frequency analysis is to estimate the flood magnitude corresponding to any return period of occurrence through the use of probability distributions, which are needed in many studies and projects such as flood plain delineation, flood protection works, river crossings, and channel improvements.

Most flood studies have been analyzed through the use univariate distributions. Several efforts have been made to provide physical and statistical basics for selecting the type of probability distribution function that best fits the frequency distribution of the actual data. One common assumption in statistical analysis of flood frequency is the homogeneity of flood distributions. However, floods are often generated by distributions composed of a mixture of two or more populations. Differences between the populations may be the result of El Niño or La Niña oscillations. The occurrences of this phenomenon modify the normal precipitation patterns in Mexico (Cavazos and Hastenrath, 1990; Magaña *et al.*, 2003; Magaña and Ambrizzi, 2005). Its signal reflects in more intense winter precipitation in the Northern states, particularly in Northwestern Mexico. As mentioned by Alila and Mtiraoui (2002) if these physical processes are not considered in conventional flood frequency analysis, the T-year flood estimate can be inefficient for design purposes.

The Mexican government has recognized that climate variability affects many of the its socio-economical activities and has begun to implement actions to diminish the negative effects of extreme climate conditions (floods and droughts). However, poverty has forced people to live almost on the water of rivers, situation that becomes an additional problem for the local governments. In order to protect their lives and goods is very important to account with an additional mathematical tool that might reduce the uncertainties in computing the design events for different return periods, which are needed in many studies and projects such as flood plain delineation.

In order to estimate more efficient quantiles of short or heterogeneous samples, a mixed distribution with Extreme Value Type I (EV1 or Gumbel) and General Extreme Value (GEV) components for the maxima is proposed and it will be called EV1-GEV distribution.

Mixed distributions

The use of a mixture of probability distributions functions for modeling samples of data coming from two populations have been proposed long time ago (Mood *et al*, 1974):

$$Pr(X \leq x) = F(x) = pF_1(x) + (1-p)F_2(x) \qquad (1)$$

Where p is a factor used to weigh the relative contribution of each population ($0<p<1$), and $F(x)$ is the composite exceedance probability. $F_1(x)$ and $F_2(x)$ are the components in the mixture.

Mixed Gumbel Distribution

If $F_1(x)$ and $F_2(x)$ of equation (1) are Gumbel distributions (NERC, 1975) then the five-parameter mixture model of annual floods is (Raynal and Guevara, 1997):

$$F(x) = p\exp^{-\exp^{-\left(\frac{x-\upsilon_1}{\alpha_1}\right)}} + (1-p)\exp^{-\exp^{-\left(\frac{x-\upsilon_2}{\alpha_2}\right)}} \qquad (2)$$

where υ_1, α_1 and υ_2, α_2 are the location and scale parameters for the first and second population, respectively

The corresponding probability density function is

$$f(x) = \frac{p}{\alpha_1}\exp^{-\left(\frac{x-\upsilon_1}{\alpha_1}\right)}\exp^{-\exp^{-\left(\frac{x-\upsilon_1}{\alpha_1}\right)}} + \frac{(1-p)}{\alpha_2}\exp^{-\left(\frac{x-\upsilon_2}{\alpha_2}\right)}\exp^{-\exp^{-\left(\frac{x-\upsilon_2}{\alpha_2}\right)}} \qquad (3)$$

Mixed General Extreme Value Distribution

If $F_1(x)$ and $F_2(x)$ of equation (1) are GEV distributions (NERC, 1975) then the seven- parameter mixture model of annual floods is (Raynal and Santillan, 1986):

$$F(x) = p\exp\left\{-\left[1-\left(\frac{x-\omega_1}{\lambda_1}\right)\beta_1\right]^{1/\beta_1}\right\} + (1-p)\exp\left\{-\left[1-\left(\frac{x-\omega_2}{\lambda_2}\right)\beta_2\right]^{1/\beta_2}\right\} \qquad (4)$$

Where $\omega_1, \lambda_1, \beta_1$ and $\omega_2, \lambda_2, \beta_2$ are the location, scale and shape parameters for the first and second population, respectively.

The corresponding probability density function is

$$f(x) = \frac{p}{\lambda_1}\exp\left\{-\left[1-\left(\frac{x-\omega_1}{\lambda_1}\right)\beta_1\right]^{1/\beta_1}\right\}\left[1-\left(\frac{x-\omega_1}{\lambda_1}\right)\beta_1\right]^{1/\beta_1-1} +$$

$$\frac{(1-p)}{\lambda_2}\exp\left\{-\left[1-\left(\frac{x-\omega_2}{\lambda_2}\right)\beta_2\right]^{1/\beta_2}\right\}\left[1-\left(\frac{x-\omega_2}{\lambda_2}\right)\beta_2\right]^{1/\beta_2-1} \qquad (5)$$

EV1-GEV Distribution

Assuming that first and second populations behave as EV1 and GEV distributions, respectively, equation (1) yields to the six-parameter mixture model of annual floods:

$$F(x) = p\exp^{-\exp^{-\left(\frac{x-\upsilon}{\alpha}\right)}} + (1-p)\exp\left\{-\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta}\right\} \qquad (6)$$

Where υ, α and ω, λ are the location and scale parameters for the first and second population, respectively; β is the shape parameter for the second population.

The corresponding probability density function is

$$f(x)=\frac{p}{\alpha}\exp^{-\left(\frac{x-\upsilon}{\alpha}\right)}\exp^{-\exp^{-\left(\frac{x-\upsilon}{\alpha}\right)}}+\frac{(1-p)}{\lambda}\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta-1}\exp\left\{-\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta}\right\} \quad (7)$$

Estimation of parameters by maximum likelihood

The likelihood function of *n* random variables is defined to be the joint density of *n* random variables and it is a function of the parameters. If $X_1, X_2, ..., X_n$ is a random sample of a univariate density function, the corresponding likelihood function is (Mood *et al.*, 1974):

$$L(x,\underline{\theta})=\prod_{i=1}^{n} f(x_i,\underline{\theta}) \quad (8)$$

The logarithmic function will be used instead of the likelihood function because it is easier to handle. So, equation (8) is transformed:

$$InL(x,\underline{\theta})=In\prod_{i=1}^{n} f(x_i,\underline{\theta}) \quad (9)$$

Where L is called the likelihood function, ln is the natural logarithm, $\underline{\theta}$ is the set of parameters to be estimated and $f(x,\underline{\theta})$ is the EV1-GEV density function, thus

$$In\,L(x,\underline{\theta})=\sum_{i=1}^{n} In\left\{\begin{array}{c}\frac{p}{\alpha}\exp^{-\left(\frac{x-\upsilon}{\alpha}\right)}\exp^{-\exp^{-\left(\frac{x-\upsilon}{\alpha}\right)}}\\ +\frac{(1-p)}{\lambda}\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta-1}\exp\left\{-\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta}\right\}\end{array}\right\} \quad (10)$$

And the corresponding first order partial derivatives of such function with respect to each of the parameters are

$$\frac{\partial InL}{\partial\upsilon}=\frac{p}{\alpha^2}\sum_{i=1}^{n}\frac{1}{f(x)}\left\{\exp^{-\exp^{-\left(\frac{x-\upsilon}{\alpha}\right)}}\left[\exp^{-2\left(\frac{x-\upsilon}{\alpha}\right)}-\left(\frac{x-\upsilon}{\alpha}\right)\right]\right\} \quad (11)$$

$$\frac{\partial InL}{\partial\alpha}=-\frac{p}{\alpha^2}\sum_{i=1}^{n}\frac{1}{f(x)}\left\{\exp^{-\exp^{-\left(\frac{x-\upsilon}{\alpha}\right)}}\left[\exp^{-\left(\frac{x-\upsilon}{\alpha}\right)}+\exp^{-2\left(\frac{x-\upsilon}{\alpha}\right)}-(x-\upsilon)\right]\right\} \quad (12)$$

$$\frac{\partial InL}{\partial\omega}=-\frac{(1-p)}{\lambda^2}\sum_{i=1}^{n}\frac{1}{f(x)}\left\{\begin{array}{c}\exp\left\{-\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta}\right\}\\ \left\{(1-\beta)\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta-2}-\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{2/\beta-2}\right\}\end{array}\right\} \quad (13)$$

$$\frac{\partial InL}{\partial \lambda} = \frac{(1-p)}{\lambda^2} \sum_{i=1}^{n} \frac{1}{f(x)} \left\{ \begin{array}{c} -\exp\left\{ -\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta} \right\}\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta-1} + \\ (x-\omega) f(x) \left\{ (1-\beta) \Big/ \left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right] - \left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta-1} \right\} \end{array} \right\} \quad (14)$$

$$\frac{\partial InL}{\partial \beta} = \frac{(1-p)}{\lambda^2} \sum_{i=1}^{n} \frac{1}{f(x)} \left\{ \exp\left\{ -\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta} \right\} \left\{ \begin{array}{c} \frac{1}{\beta}\left(\frac{x-\omega}{\lambda}\right)\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta-2} \\ +\frac{1}{\beta^2} In\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta-1} \\ -\left(\frac{x-\omega}{\lambda}\right)\left(\frac{1}{\beta}-1\right) \Big/ \left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right] \\ -1\frac{1}{\beta^2} In\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right] \end{array} \right\} \right\} \quad (15)$$

$$\frac{\partial InL}{\partial p} = \frac{1}{\alpha\lambda} \sum_{i=1}^{n} \left\{ \frac{1}{f(x)} \left\{ \exp^{-\left(\frac{x-\upsilon}{\alpha}\right)} \exp^{-\exp^{-\left(\frac{x-\upsilon}{\alpha}\right)}} - \left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta-1} \exp\left\{ -\left[1-\left(\frac{x-\omega}{\lambda}\right)\beta\right]^{1/\beta} \right\} \right\} \right\} \quad (16)$$

The exact solution provided by the system of equations (11)-(16) is not known, so the maximum likelihood estimators of the parameters were obtained by the direct maximization of the log-likelihood function (eq. 10), which is constrained to $\alpha > 0, \lambda > 0, 0 < p < 1$, and $x > 0$. The suggested procedure is the constrained multivariable Rosenbrock method (Kuester and Mize, 1973).

As it is known, in any of the multivariable constrained non-linear optimization techniques, global optimality is never assured. Therefore, care must be taken in order to avoid a local optimum. It is suggested to start always with values of the location, scale and shape parameters computed by considering the sample divided into two equal parts. If sample is sorted in decreasing order of magnitude, the first set of data is fitted to the univariate GEV distribution (Prescott and Walden, 1980), and the second one to the univariate Gumbel distribution (NERC, 1975). The initial value of the association parameter p will be equal to 0.5.

For the mixed Gumbel and the mixed GEV distributions parameters are estimated following the same optimization procedure.

Case study

A region located in Northwestern Mexico, with a total of 35 gauging stations was selected to apply the EV1-GEV distribution to flood frequency analysis. Table 1 shows statistical characteristics of data for each station in the region.

In the area considered in this study, flood outliers correspond to observed rainfall values much higher than the other annual maxima. Such extremely heavy rainfall is due to special meteorological conditions in connection with ENSO events in the Pacific Ocean. In the analyzed area, 62% of the highest annual maximum discharges gauged were generated in an El Niño year and 38% for its counterpart, La Niña.

Results provided by the EV1-GEV distribution were compared with those produced by the mixed Gumbel and mixed GEV distributions. For each station the best one was chosen according to the criterion of minimum standard error of fit (*SE*), as defined by Kite (1988):

$$SE = \left[\sum_{i=1}^{n} (g_i - h_i)^2 / (n - q) \right]^{1/2} \qquad (17)$$

Where $g_{i,} i = 1, \dots n$ are the $h_{i,} i = 1, \dots n$ recorded events; are the event magnitudes computed from the probability distribution at probabilities obtained from the sorted ranks of, $g_i, i = 1, \dots, n$, n is the length of record, and q is the number of parameters estimated for the mixed distribution. For the mixed distributions, Gumbel, GEV and EV1-GEV q will be equal to 5, 7 and 6, respectively.

In table 2 is depicted the *SE* for all mixed distributions along with the best model for the sample of data considered.

The final at-site design events Q (m^3/s) for different return periods T(years) in each station are presented in Table 3.

In some sites a comparison is made among different at-site design events (Table 4). For instance, in station Chinipas the computed SE are very similar, however, as return period increases, differences among flood estimates are more significant. A bad selection of the best distribution in the analyzed site can substantially modify the design event and that the hydraulic project might become economically unfeasible or unsafe.

An additional problem is when a short record is used (less than 30 years), because there is an increased risk that the flood estimate will not provide adequate protection of designated uses. One way to reduce the bias or uncertainty in the flood estimate is to use a regional data set with observations from several sites.

Mixed Gumbel, GEV and EV1-GEV distributions can be easily used to obtain regional at-site estimates of floods by using the station-year method in regions with heterogeneous sample data. The general procedure of this regional technique can be found in paper written by Cunnane (1988).

This regional technique was not applied in the paper and it just was mentioned to be considered for users in their hydrological analyses.

Table 1. Statistical characteristics of flood data for each analyzed gauging station

m	Gauging Station	Years of record	Period of record	Mean	Standard Deviation	Coefficient of Skewness	Coefficient of Kurtosis	Coefficient of Variation
1	Acatitan	31	1955-1985	1031.6	864.5	2.63	12.79	0.84
2	Alamos	22	1948-1969	247.6	178.8	0.83	3.60	0.72
3	Badiraguato	27	1959-1985	957.9	1853.6	3.92	20.65	1.94
4	Bamicori	32	1951-1982	194.5	176.7	1.39	4.46	0.91
5	Cazanate	19	1967-1985	555.0	727.9	3.06	15.04	1.31
6	Chinipas	21	1965-1985	1061.0	524.4	0.31	2.83	0.49
7	Choix	29	1955-1983	392.9	336.5	2.59	11.84	0.86
8	El Bledal	48	1938-1985	286.0	273.4	2.83	13.92	0.96
9	El Mahone	20	1966-1985	198.4	26.3	0.69	5.96	0.13
10	El Naranjo	47	1939-1985	621.9	655.5	1.87	6.98	1.05
11	El Quelite	26	1960-1985	468.5	445.2	1.72	6.27	0.95
12	Guamuchil	36	1938-1973	605.9	630.4	3.09	16.22	1.04
13	Guatenipa	21	1965-1985	1888.8	1393.2	0.84	3.27	0.74
14	Huites	53	1941-1993	2942.0	3124.3	2.63	10.39	1.06
15	Ixpalino	31	1953-1983	1317.8	1218.2	2.79	12.68	0.92
16	Jaina	46	1941-1986	1197.4	1189.9	3.20	16.30	0.99
17	La Huerta	17	1969-1985	934.2	574.3	0.29	2.43	0.61
18	La Tina	24	1960-1983	106.5	152.3	4.00	22.39	1.43
19	Las Cañas	24	1948-1971	2497.0	3194.2	1.60	4.84	1.28
20	Palo Dulce	29	1957-1985	975.9	1195.7	4.40	25.89	1.23
21	Palos Blancos	47	1939-1985	1481.8	1726.4	1.92	7.85	1.17
22	Pericos	26	1960-1985	201.0	95.1	0.14	2.56	0.47
23	Piaxtla	16	1958-1973	1419.8	1587.8	2.48	10.79	1.12
24	Pte Sud Pacífico	35	1924-1958	2961.0	2204.9	1.41	7.20	0.74
25	Puente Cañedo	22	1932-1953	1116.0	932.7	0.76	3.35	0.84
26	San Francisco	33	1941-1973	1724.6	1450.1	1.89	7.03	0.84
27	San Ignacio	19	1967-1985	1622.4	813.4	1.70	7.38	0.50
28	Sanalona	42	1944-1985	447.3	505.4	2.99	13.49	1.13
29	Santa Cruz	43	1943-1985	1269.7	1216.5	2.94	14.86	0.96
30	Tamazula	23	1962-1984	583.6	278.0	1.38	5.46	0.48
31	Tecusiapa	17	1958-1974	975.7	792.7	1.66	6.49	0.81
32	Toahayana	29	1957-1985	1048.9	629.7	0.67	3.19	0.60
33	Urique	19	1967-1985	302.6	148.0	1.16	6.81	0.49
34	Zapotitlán	22	1960-1981	624.6	645.4	2.03	9.09	1.03
35	Zopilote	47	1939-1985	363.2	275.9	0.69	2.81	0.76

Table 2. The computed SE (in m^3/s) for each analyzed gauging station

Gauging Station	EVI-GVE	Mixed Gumbel	Mixed GVE	Best distribution
Acatitan	308.9	337.0	386.3	EV1-GVE
Alamos	25.9	28.4	25.8	Mixed GVE
Badiraguato	769.5	921.3	*	EV1-GVE
Bamicori	30.1	46.5	30.8	EV1-GVE
Cazanate	377.9	443.8	372.1	Mixed GVE
Chinipas	89.2	81.6	91.2	Mixed Gumbel
Choix	152.7	118.5	146.2	Mixed Gumbel
El Bledal	74.7	69.4	86.3	Mixed Gumbel
El Mahone	24.0	8.5	8.6	Mixed Gumbel
El Naranjo	112.0	174.4	125.9	EV1-GVE
El Quelite	126.1	143.1	116.3	Mixed GVE
Guamuchil	241.9	263.2	352.1	EV1-GVE
Guatenipa	225.1	355.9	337.0	EV1-GVE
Huites	614.1	987.0	805.6	EV1-GVE
Ixpalino	390.3	370.8	*	Mixed Gumbel
Jaina	402.8	411.2	*	EV1-GVE
La Huerta	364.0	99.0	*	Mixed Gumbel
La Tina	122.4	105.4	127.3	Mixed Gumbel
Las Cañas	*	2139.1	790.5	Mixed GVE
Palo Dulce	884.9	963.3	900.5	EV1-GVE
Palos Blancos	340.1	550.9	294.4	Mixed GVE
Pericos	19.2	18.1	26.3	Mixed Gumbel
Piaxtla	641.0	502.5	828.6	Mixed Gumbel
Pte Sud Pacífico	614.8	658.0	624.5	EV1-GVE
Puente Cañedo	149.8	135.6	174.8	Mixed Gumbel
San Francisco	323.4	333.8	302.8	Mixed GVE
San Ignacio	296.6	344.4	273.8	Mixed GVE
Sanalona	115.8	214.6	214.5	EV1-GVE
Santa Cruz	390.9	341.4	*	Mixed Gumbel
Tamazula	93.5	77.2	*	Mixed Gumbel
Tecusiapa	278.5	211.9	308.8	Mixed Gumbel
Toahayana	104.6	106.6	101.8	Mixed GVE
Urique	92.2	44.1	*	Mixed Gumbel
Zapotitlán	219.0	301.1	238.6	EV1-GVE
Zopilote	37.4	37.6	38.6	EV1-GVE

* No convergence was attained in the estimation of parameters

Table 3. Design events $Q(m^3/s)$ for the best fitted distribution in each analyzed gauging station

	Return period Tr(years)							
Gauging Station	**2**	**5**	**10**	**20**	**50**	**100**	**500**	**1000**
Acatitan	693.2	1610.3	2175.5	2720.9	3502.6	4168.2	6049.8	7039.0
Alamos	198.1	418.5	522.0	589.4	651.0	685.1	742.8	770.0
Alamos	203.9	396.2	496.0	585.3	697.6	780.7	971.7	1053.6
Badiraguato	415.0	686.5	2369.1	4770.6	7736.6	10461.1	19586.4	25268.6
Bamicori	116.4	353.7	499.8	574.5	635.0	665.2	706.8	717.3
Cazanate	304.9	728.5	1295.8	1944.6	2934.4	3824.0	6550.4	8097.6
Chinipas	1050.5	1484.1	1751.1	2003.3	2327.4	2569.5	3128.1	3368.2
Choix	299.7	500.4	713.5	1013.9	1419.1	1709.0	2359.6	2636.2
El Bledal	209.0	375.9	572.8	830.2	1147.9	1373.2	1879.4	2094.8
El Mahone	197.7	214.6	226.9	240.4	260.5	276.8	315.8	332.8
El Naranjo	377.2	937.0	1647.7	2001.4	2450.2	2808.3	3753.9	4223.0
El Quelite	307.1	749.0	1052.7	1384.2	1896.1	2357.1	3757.6	4543.1
Guamuchil	431.4	880.3	1236.2	1728.5	2472.7	3114.0	4953.5	5935.7
Guatenipa	1593.2	3211.7	3903.2	4451.4	5041.7	5414.1	6096.2	6327.7
Huites	1865.7	3428.2	6929.2	9949.8	13232.7	15504.8	20439.9	22464.1
Ixpalino	963.7	1580.2	2334.1	3720.8	5550.2	6835.9	9716.7	10941.7
Jaina	807.3	1576.7	2480.2	3401.5	4707.4	5809.4	8894.1	10502.4
La Huerta	801.9	1474.2	1627.1	1764.4	1937.2	2065.3	2359.5	2485.7
La Tina	71.0	127.2	173.6	254.1	563.3	726.9	1051.5	1184.8
Las Cañas	929.2	4382.3	7643.8	8899.5	9703.0	10022.1	10355.7	10416.1
Palo Dulce	676.4	1097.1	1518.3	2137.7	3160.1	4012.9	6300.9	7450.2
Palos Blancos	871.2	2682.6	3850.1	5076.2	6882.0	8432.5	12820.8	15123.2
Pericos	199.4	274.6	319.7	362.0	416.2	456.7	550.0	590.1
Piaxtla	781.0	1490.9	3902.2	5156.8	6582.3	7602.4	9910.7	10895.7
Pte Sud Pacifico	2833.8	4573.0	5709.2	6866.1	8493.5	9824.7	13327.1	15039.5
Puente Cañedo	894.5	1914.3	2424.0	2889.1	3478.1	3915.5	4921.8	5353.7
San Francisco	1149.8	2590.9	3717.4	4733.3	6042.8	7045.4	9475.7	10577.5
San Ignacio	1509.3	2246.5	2754.5	3308.9	4156.9	4911.2	7156.5	8391.5
Sanalona	301.4	517.5	764.3	1651.6	2296.9	2512.4	2728.8	2768.1
Santa Cruz	927.0	1667.6	2625.1	3833.9	5249.3	6245.8	8480.8	9431.4
Tamazula	510.4	747.8	939.4	1136.9	1398.4	1595.2	2049.7	2245.0
Tecusiapa	698.5	1400.1	2112.2	2732.2	3480.1	4023.9	5262.5	5792.2
Toahayana	926.5	1666.1	1977.3	2214.4	2462.4	2615.7	2895.6	2995.6
Urique	294.5	354.2	480.4	602.6	757.5	872.6	1137.5	1251.2
Zapotitlán	533.7	1099.7	1517.6	1969.9	2646.9	3233.8	4911.4	5798.8
Zopilote	316.2	610.6	748.6	858.5	977.0	1052.0	1189.8	1236.8

Table 4. Comparison of design events Q(m^3/s) and SE (in m^3/s) for some selected stations of case study

		Return period Tr (years)								
Gauging Station	**Distribution**	**2**	**5**	**10**	**20**	**50**	**100**	**500**	**1000**	***SE***
Chinipas	EV1-GVE	1040	1542	1752	1912	2079	2181	2359	2417	89.2
	Mixed GVE	1040	1551	1763	1925	2096	2203	2427	2573	91.2
	Mixed Gumbel*	1051	1484	1751	2003	2327	2569	3128	3368	81.6
Palo Dulce	EV1-GVE*	676	1097	1518	2138	3160	4013	6301	7450	88.49
	Mixed GVE	672	1114	1539	2137	3262	4321	7402	9039	90.05
	Mixed Gumbel	702	1094	1395	1776	2806	4097	6984	8191	96.33
San Francisco	EV1-GVE	1155	2524	3543	4626	6256	7687	11879	14151	32.34
	Mixed GVE*	1150	2591	3717	4733	6043	7045	9476	10578	30.28
	Mixed Gumbel	1141	2613	3676	4605	5760	6613	8566	9404	33.38

* Best distribution according to the minimum value of *SE*.

Conclusions

Floods are often generated by heterogeneous distributions composed of a mixture of two populations. Differences between the populations may be the result of a number of factors such as the El Niño/La Niña oscillations. In the analyzed area 62% of the highest annual maximum discharges (outliers) were generated in an El Niño year. The magnitude of these events is very important and floods can seriously affect people. For this reason, it is necessary to account with an additional mathematical tool that be able to reduce the uncertainty in estimating of de- sign events, which are needed in many water-engineering studies and projects.

In this paper a mixed distribution has been derived by considering different components in an opposite way as usually do. $F_1(x)$ and $F_2(x)$ of equation (1) were the EV1 and the GEV distributions, respectively.

Results shown that there exists a reduction in the standard error of fit when using the EV1-GEV distribution in comparison with the mixed Gumbel or mixed GEV distributions, and just in one out of the 35 analyzed cases, the proposed distribution could not reach convergence in the estimation of parameters process. By contrast, the Mixed GEV distribution had seven failures with the same estimation process.

In 13 sample data the EV1-GEV distribution produced the least standard error of fit (40% of analyzed cases) and in other different cases it was very close to the mixed Gumbel and mixed GEV distributions, However, as it was shown, differences between at-site design events can be significant as return period increases. A bad selection of the best distribution in the analyzed site can substantially modify the design event and also that the hydraulic project might become economically unfeasible or unsafe. Thus, by taking into consideration the mixed flood distributions a more accurate, physically based flood frequency analysis can be obtained and sensible savings in costs of construction of flood protection structures can be expected. This can also improve the setting of flood plain limits and the safety of control structures.

References

Alila Y. and Mtiraoui A. (2002). Implications of Heterogeneous Flood-Frequency Distributions on Traditional Stream-Discharge Prediction Techniques. *Hydrological Processes*, 16:1065-1084.

Cavazos T. and Hastenrath S. (1990). Convection and Rainfall Over Mexico and their Modulation by the Southern Oscillation. *International Journal of Climatology*, 10: 377-386.

Kite G.W. (1988). *Frequency and Risk Analyses in Hydrology*. Water Resources Publications, Littleton, Colorado, USA.

Kuester J.L. and Mize J.H. (1973). *Optimization Techniques with FORTRAN*. McGraw-Hill.

Magaña V. and Ambrizzi T. (2005). Dynamics of Subtropical Vertical Motions Over the Americas During El Niño Boreal Winters. *Atmósfera*, 18(4): 211-233.

Magaña V., Vázquez J., Pérez J. and Pérez J.B. (2003). Impact of El Niño on Precipitation in Mexico. *Geofísica Internacional*, 42(3): 313-330.

Mood A., Graybill F. and Boes D. (1974). *Introduction to the Theory of Statistics*. Third Ed., McGraw-Hill.

NERC (1975). *Natural Environment Research Council. Flood Studies Report I, Hydrologic Studies*. Whitefriars Press Ltd., London, United Kingdom.

Prescott P. and Walden A. (1980). Maximum Likelihood Estimation of the Parameters of the Generalized Extreme Value Distribution. *Biometrika*, 67(3): 723-724.

Raynal J. and Guevara J. (1997). Maximum Likelihood Estimators for the two Populations Gumbel distribution. *Hydrological Science and Technology Journal*, 13(1-4):47-56.

Raynal J. and Santillan O. (1986). Maximum Likelihood Estimators of the Parameters of the Mixed GEV Distribution. IX Congreso Nacional de Hidráulica. AMH. Querétaro, Qro., Mex. pp. 79-90. (In Spanish)

Semblanza del autor

Dr. Carlos Agustín Escalante-Sandoval. Es doctor en ingeniería hidráulica por la Facultad de Ingeniería de la UNAM. Actualmente es profesor titular "C" de tiempo completo definitivo. Ha impartido 85 cursos en el Posgrado de la UNAM; dirigido 38 tesis de maestría y cinco de doctorado. Dentro de su producción académica se encuentran: 30 publicaciones en revistas con arbitraje, 45 en congresos nacionales e internacionales, 3 capítulos en libro, 2 libros como autor y otro como co-editor. Recibió la medalla Gabino Barreda por sus estudios de doctorado, el premio Distinción Universidad Nacional para Jóvenes Académicos en Docencia en Ciencias Exactas 1999 que otorga la UNAM y el Premio Nacional Enzo Levi "Investigación y Docencia en Hidráulica 2002", por parte de la Asociación Mexicana de Hidráulica. Es miembro del Sistema Nacional de Investigadores, Academia Mexicana de Ciencias, Academia de Ingeniería, Colegio de Ingenieros Civiles de México y la Asociación Mexicana de Hidráulica.