RIIT Vol.X. Núm.3. 2009 247-253, ISSN1405-7743 FI-UNAM
(artículo arbitrado)

Mixed Distributions in Low-Flow Frequency Analysis

Distribuciones mezcladas en el análisis de frecuencias de flujos mínimos

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

(Recibido: septiembre de 2007; aceptado: mayo de 2008)

Abstract

Low-flow characteristics are required to solve several water-engineering problems. In this paper, the Mixed Gumbel and the Two Component Extreme Value distributions are presented toward their applications in low-flow frequency analysis. A region in southern Mexico, with 39 gauging stations was selected to analyze the lowest 1 day flows. Additionally, in order to calculate the stream design flow ($_7Q_{10}$) for water quality standards downstream of the Hydroelectric project La Parota, the lowest 7 day average flows were used. Results produced by fitting the mixed distribution were compared with those obtained by the Weibull-3, Gumbel, Lognormal-3 and General Extreme Value distributions. Results suggest that mixed distributions are a suitable option to be considered when analyzing minimum flows.

Keywords: Minimum flow frequency analysis, maximum likelihood, heterogeneous samples, water quality.

Resumen

En muchos de los problemas de ingeniería del agua se requiere conocer las características de los flujos mínimos. En el artículo se presenta la aplicación de las distribuciones Gumbel Mixta y de Valores Extremos de Dos Componentes en el análisis de frecuencias de gastos mínimos. Una región localizada en el sureste de México, con un total de 39 estaciones de aforos, fue seleccionada para analizar los gastos mínimos anuales con duración de un día. Adicionalmente se utilizaron los gastos mínimos anuales promedio de siete días consecutivos con el fin de obtener el gasto de diseño ($_7Q_{10}$) para cumplir con los estándares de calidad hídrica aguas abajo del proyecto hidroeléctrico La Parota. Los eventos estimados por las distribuciones mezcladas fueron comparados con aquellos obtenidos por las distribuciones Weibull-3, Gumbel, Lognormal-3 y General de Valores Extremos. Los resultados sugieren que las distribuciones mezcladas son una opción adecuada a ser considerada en el análisis de flujos mínimos.

***Desciptores**: Análisis de frecuencias de flujos mínimos, máxima verosimilitud, muestras heterogéneas, calidad del agua.*

Introduction

Low-flow is the flow of water in a stream during prolonged dry weather. By contrast, a drought is a natural event that results from an extended period of below average precipitation. While droughts include low-flows, a continuous seasonal low-flow event is not necessarily a drought. A summary about the status of low-flow hydrology can be found in Smakhtin (2001) and Pyrce (2004).

Quantiles of annual low-flow are commonly used as design flows on which to base the design of structures such as wastewater-treatment plants or for describing the capability of a stream to supply requirements for the regulation of fluvial transport, water supply, hydropower, liquid waste disposal, irrigation systems, or assessing the impact of prolonged droughts on aquatic ecosystems.

For instance, when assessing the suitable conditions for aquatic life, a hydrologically-based design flow (EPA, 2006) is computed using the single lowest flow event from each year of record and then examining these flows for a series of years. When a sufficiently long discharge record is available at a river site, low-flow statistics, such as the lowest 7 day average flow that occurs on average once every 10 years (${}_7Q_{10}$), can be obtained through the use of probability distributions. By contrast, the biologically-based design flows (EPA, 2006) use durations and frequencies specified in water quality criteria for individual pollutants and whole effluents; they can be based on the available biological, ecological and toxicological information concerning the stresses that aquatic organisms, ecosystems, and their uses can tolerate.

This method is empirical, not statistical, because it deals with the actual flow record itself, not with a statistical distribution intended to describe the flow record. The biologically-based definition also recognizes that drought imposes severe stress on aquatic organisms, whether pollutants are present or not.

Low-flows typically aggravate the effects of water pollution. During a low flow event, there is less water available to dilute effluent loadings, resulting in higher in-stream concentration of pollutants. Aquatic life criteria are expressed in terms of the intensity of concentration, duration of averaging period, and average frequency of allowed excursions, which are defined as any flow lower than the design flow. Two concentrations, a continuous (CC) and a maximum (MC) are used to protect aquatic life from chronic and acute effects, respectively. From a hydrological point of view, EPA (2006) recommends the use of the ${}_1Q_{10}$ flow as the design flow for the MC and the ${}_7Q_{10}$ as the design flow for the CC.

Extensive literature is available on the application of probability distributions for prediction of flood frequencies, while the number of studies reported on frequency of low flow is rather limited. The modest interest in finding the most appropriate distribution of low flow is due to the relatively short return periods used in low flow design (less than 50 years).

Gumbel (1958) discussed the use of the 3-parameter Weibull distribution (W3) for fitting low flows:

$$f(x)\ [\ /\ ][(x\ \ )/\ ]^{\ 1}\exp\{\ [(x\ \ )/\ ]\ \},\qquad x\ \ 0;\ \ 0;\ \ 0;\ \ ,\qquad (1)$$

where , and are the shape, scale and location parameters.

The Gumbel distribution (EV1) is commonly used for low flow frequency analysis (Al-Mashidani *et al*., 1980):

$$f(x)\ (1/\ )\exp\{\ [(\ \ x)/\ ]\}\ \exp\ \exp\{\ [(\ \ x)/\ ]\},\qquad (2)$$

where and are the location and scale parameters.

This distribution is not bounded in the lower or upper tail. The smallest values of the EV1 distribution have a high probability of negative values.

Chow (1964) provided a theoretical justification for the use of the 3-parameter Lognormal distribution (LN3) in low flow analysis:

$$f(x)\ \frac{1}{(x\ \ x_0)\ {}_y\sqrt{2}}\exp\ \frac{[1n(x\ \ x_0)\ \ {}_y]^2}{2\ {}_y^2},\qquad (3)$$

where x_o, ${}_y$ and ${}_y$ are the location, scale and shape parameters. This distribution is not bounded in the lower or upper tail.

Kroll and Vogel (2002) used L-moments diagrams to examine 1505 gauged river sites in the United States, and recommended the LN3 distribution for describing low streamflow statistics at no intermittent (perennial) sites.

The General Extreme Value distribution (GEV) has been widely used in flood frequency and less in low-flow frequency analysis (Raynal, 1987):

$$f(x)\ (1/\ )\{\ [1\ \ (\ \ x)/\ ]^{1/}\ {}^{\frac{1}{}\ 1}\exp\ \{\ [1\ \ (\ \ x)/\ ]^{1/}\ \},\qquad (4)$$

If <0 then x $/$ and if 0 then $/$ x .

where , and are the scale, shape and location parameters.

Onoz and Bayazit (1999) examined the fit of various probability distributions to low flows at European rivers, and recommended the GEV distribution.

Pearson (1995) analyzed 1 day annual minimum stream flows at over 500 river sites in New Zealand, concluding that no single 2- or 3-parameter distribution provided a superior fit. Same conclusions were reported by Kroll and Vogel (2002). In order to achieve more flexibility in modeling low flows, two mixed distributions with four and five parameters are proposed in this paper.

Mixed distributions

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

$$\Pr(X \quad x) \quad F(x) \quad pF_1(x) \quad (1 \quad p)F_2(x) \tag{5}$$

where p is the proportion of x in the mixture ($0 < p < 1$), and $F(x)$ is said to be a mixture of distributions.

Annual low flows are attributed to a continued depletion of basin water storage until the minimum level of discharge is attained.

The annual low flows of some rivers are related entirely to one process leading to water depletion (e.g. evaporation). In other basins, it may be caused by one process in some years, and another process in others (e.g. fall low flow due to evaporative loss, combined with spring low due to continental drainage without water replenishment from rain) (Waylen and Woo, 1987).

The events from each process form two separate annual minimum subpopulations can be combined to follow a distribution that reflects both sub-samples.

If the EV1 distribution is used in equation 5, the mixed Gumbel distribution (EV1MIX) for the minima is

$$F(x)_{\min} \quad p\exp^{\exp(\ _1\ x)/\ _1} + (1 \quad p)\exp^{\exp(\ _2\ x)/\ _2} \tag{6}$$

where $_1,\ _1$ and $_2,\ _2$ are the location and scale parameters for the first and second population.

The corresponding density function is

$$f(x)_{\min} \quad \frac{p}{_1}\exp^{\exp(\ _7\ x)/\ _7}\exp^{(\ _1\ x)/\ _1} + \frac{(1 \quad p)}{_2}\exp^{\exp(\ _2\ x)/\ _2}\exp^{(\ _2\ x)/\ _2}, \tag{7}$$

Parameters can be computed by the maximum likelihood procedure:

$$\ln L \quad \ln \prod_{i\ 1}^{n} f(x_i;\ _1,\ _1\ _2,\ _2, p), \quad _i \quad 0 \tag{8}$$

where L is called the likelihood function and ln is the natural logarithm.

Given the complexity of the resulting likelihood function and the partial derivatives with respect to the parameters, the constrained multivariable Rosenbrock method (Kuester and Mize, 1973) was applied to obtain the estimators of the five parameters by the direct maximization of equation (8).

The Two Component Extreme Value distribution for the minima (TCEVMIN) is obtained by using the version for the maxima (Rossi *et al*., 1984) and the symmetry principle, Gumbel (1958):

$$F(x)_{\min} \quad \exp(\ _1\exp^{x/\ _1} \quad \exp^{x/\ _2}), \quad x \quad 0 \tag{9}$$

and

$$f(x)_{\min} \quad \exp(\ _1 e^{x/\ _1} \quad _2 e^{x/\ _2}) \quad \frac{_1}{_1}e^{x/\ _1} \quad \frac{_2}{_2}e^{x/\ _2}, \quad x \quad 0. \tag{10}$$

The parameters of the TCEVMIN distribution can be estimated from site-specific data set by the direct maximization of equation (11) by using the Rosenbrock method.

$$\ln L \quad \ln \prod_{i\ 1}^{n} f(x_i;\ _1\ _1,\ _2,\ _2) \tag{11}$$

Case study

A region located in southern Mexico with 39 gauging stations was selected to apply the mixed distributions to lowest 1 day flows. For each station, mixed and standard distributions were fitted and the best one was chosen according to the criterion of minimum standard error of fit (SEF), as defined by Kite (1988):

$$SEF_j \quad \left[\sum_{i\ 1}^{n}(g_1 \quad h_1)^2/(n \quad q)\right]^{1/2}, \tag{12}$$

where g_i; $i = 1,\ldots,n$ are the recorded events, h_i; $i = 1,\ldots,n$ are the events computed from the probability distribution; q is the number of parameters for each distribution j, and n is the length of record.

Table 1 shows the available length of record and catchments area for each station in the region. Equations

(1), (2), (3), (4), (8) and (11) were used in order to fit all samples in the region. The corresponding SEF_j were computed and the best distribution was selected according to its minimum value (table 2).

Table 1. Some characteristics of the stations whose annual minimum data are analyzed

Station	Lenght (years)	area (km²)	Station	Lenght (years)	area (km²)
Achotal	18	2333	la Estrella	18	774
Amapa	17	468	La Junta	18	11878
Amate	14	102	Las Prietas	18	216
A. Cabadas	18	125	Matamba	18	2143
Apoala	18	341	Mazatlan	8	53
Axusco	18	788	Monte Rosa	18	2870
Azueta	18	4656	Otapa	9	64
Bellaco	18	2917	Papaloapan	18	21236
Cabrito	8	48	Pumexcatan	18	821
Camelpo	13	2072	Quiotepec	18	4832
Canton	18	14038	S.J. Evangelista	8	5651
Cuatotolapan	18	7090	Santo Domingo	17	12681
Cuichapa	18	1732	Suchicatlan	13	93
Culebra	13	138	Teopixca	12	27
Dominguillo	18	695	Tepelmeme	18	167
Hamaca	10	30	Tomellin	18	780
Inguirjo	10	21	Tuxtepec	18	15719
Jacatepec	18	1117	Xiquila	18	1078
Lauchapan	18	1478	Zapote	18	633
La Angostura	18	6574			

Table 2. Standard error of fit in m³/s and selected distribution for each gauging station

Station	LN3	EV1	GEV	W3	EV1MIX	TCEVMIN	Desicion
Achotal	3.439	1.478	0.611	1.149	0.821	0.495	TCEVMIN
Amapa	0.258	0.178	0.040	0.018	0.058	0.087	W3
Amate	0.086	0.087	0.025	0.037	0.038	0.039	GEV
A. Cabadas	2.876	0.663	0.310	0.482	0.270	0.323	EV1MIX
Apoala	0.040	0.030	0.010	0.018	0.012	0.008	TCEVMIN
Axusco	0.141	0.008	0.002	0.003	0.003	0.003	GEV
Azueta	6.710	2.514	1.671	2.622	1.835	1.743	GEV

Table 2. Standard error of fit in m³/s and selected distribution for each gauging station (...continuation)

Station	LN3	EV1	GEV	W3	EV1MIX	TCEVMIN	Desicion
Bellaco	4.697	2.373	0.899	1.986	1.143	0.831	TCEVMIN
Cabrito	1.595	0.046	0.011	0.019	0.017	0.015	TCEVMIN
Camelpo	0.160	0.016	0.007	0.009	0.012	0.008	GEV
Canton	2.123	1.076	0.957	1.435	0.585	1.645	EV1MIX
Cuatotolapan	7.522	2.227	1.105	2.166	1.102	1.188	EV1MIX
Cuichapa	4.469	1.142	0.607	1.011	0.422	0.541	EV1MIX
Culebra	0.006	0.005	0.002	0.003	0.004	0.006	GEV
Dominguillo	0.197	0.092	0.019	0.039	0.027	0.016	TCEVMIN
Hamaca	0.001	0.001	0.001	0.001	0.001	0.002	W3
Inguirjo	0.002	0.002	0.001	0.001	0.001	0.002	W3
Jacatepec	4.662	2.525	1.096	1.895	0.781	1.119	EV1MIX
Lauchapan	9.033	1.631	0.929	1.576	0.639	1.094	EV1MIX
La Angostura	0.244	0.216	0.086	0.166	0.099	0.087	GEV
La Estrella	5.200	1.971	0.746	1.308	0.696	0.756	EV1MIX
La Junta	8.098	0.430	0.315	0.454	0.468	0.334	GEV
Las Prietas	0.336	0.126	0.025	0.046	0.049	0.057	GEV
Matamba	5.353	0.350	0.235	0.353	0.122	0.278	EV1MIX
Mazatlan	0.086	0.053	0.020	0.026	0.033	0.028	GEV
Monte Rosa	2.558	1.009	0.580	0.969	0.656	0.664	GEV
Otapa	0.489	0.023	0.012	0.019	0.036	0.030	GEV
Papaloapan	23.31	10.84	9.253	17.16	9.737	14.47	GEV
Pumexcatan	2.164	0.443	0.326	0.489	0.182	0.399	EV1MIX
Quiotepec	3.295	0.271	0.192	0.324	0.222	0.224	GEV
S.J. Evangelista	20.30	10.17	7.120	8.314	9.116	9.324	GEV
Santo Domingo	1.151	0.882	0.412	0.778	0.372	0.498	EV1MIX
Suchicatlan	0.008	0.004	0.001	0.002	0.004	0.005	GEV
Teopixca	0.071	0.124	0.057	0.072	0.075	0.070	GEV
Tepelmeme	0.012	0.016	0.006	0.008	0.009	0.007	GEV
Tomellin	0.012	0.011	0.004	0.006	0.008	0.005	GEV
Tuxtepec	24.26	35.96	16.88	22.54	21.86	22.21	GEV
Xiquila	0.130	0.117	0.049	0.091	0.046	0.062	EV1MIX
Zapote	1.647	1.336	0.695	1.057	0.251	0.758	EV1MIX

In 2003, the *Comisión Federal de Electricidad* proposed the construction of La Parota Dam in the southern State of Guerrero, Mexico. The 180 m and 765-megawatt dam located in the Papagayo River watershed would flood close to 17,000 hectares of land. Communities around the site of the project are concerned because of the expected changes to the river ecosystem downstream of the dam. Major losses in fisheries could occur all the way downstream of the dam until the river's delta at the Pacific Ocean.

In order to do an integral assessment of the environmental impact associated with the hydroelectric project, it is necessary to account with an estimate of the possible ecological flow of the river. According to the one of the aquatic life criteria proposed by the United States Environmental Protection Agency (EPA, 2006), the hydrologically-based design flow ${}_7Q_{10}$ is obtained by using the lowest 7-day average flows computed through the gauged data at station La Parota (table 3). The EV1, GEV , W3, and mixed distributions were used to fit the sample. The corresponding SEF_j and the design events for different return periods are shown in table 4. By considering the criterion of the minimum standard error of fit and from an hydrological point of view, the EV1MIX distribution was selected, and the low flow ${}_7Q_{10}$ = 10.4 m^3/s would be the minimum condition to maintain the water quality and the aquatic life downstream of the dam.

When a short record is used, there is an increased risk that the low-flow estimate will not provide adequate protection of designated uses. One way to reduce the bias or uncertainty in the low-flow estimate is to use a regional data set with observations from several sites. Mixed distributions can be easily used to obtain

Table 3. The lowest 7 day average flows in m^3/s at gauging station La Parota

year	7Q(m3/s)	year	7Q(m3/s)	year	7Q(m3/s)	year	7Q(m3/s)
1963	19.8	1973	15.3	1983	13.1	1993	18.7
1964	15.1	1974	19.3	1984	14.2	1994	15.0
1965	0.3	1975	19.1	1985	17.1	1995	15.2
1966	19.1	1976	13.0	1986	15.8	1996	9.8
1967	19.0	1977	16.4	1987	3.2	1997	21.1
1968	14.4	1978	15.3	1988	13.4	1998	15.7
1969	17.5	1979	22.3	1989	17.7	1999	11.9
1970	15.4	1980	17.4	1990	21.5		
1971	18.9	1981	16.9	1991	9.8		
1972	16.5	1982	23.2	1992	21.1		

Table 4. At-site design events 7Q in m^3/s for different return periods and the SEF_j for the hydroelectric project La Parota

Return period	Distribution				
T (years)	EV1	GEV	W3	EV1MIX	TCEVMIN
2	16.6	15.0	14.9	16.5	15.5
5	12.7	12.1	11.0	12.8	10.4
10	10.0	10.9	9.1	10.4	8.3
20	7.6	10.0	7.6	8.1	6.7
50	4.3	9.2	6.0	5.1	5.0
100	1.9	8.7	5.1	2.9	4.0
SEF_j	1.5	2.4	2.0	1.4	3.8