Metodología para el ajuste de modelos de valor extremo tipo I (gumbel) y log pearson tipo III, para series de valores máximos

Sandra Biviana González Fiagá. Licenciada en Matemáticas y Física, Universidad Pedagógica y Tecnológica de Colombia. Especialista en Matemáticas y Estadística Aplicadas, UPTC. Grupo de investigación GIPDCB, Tunja. Docente Departamento de Ciencias Básicas. Universidad Santo Tomás seccional Tunja. sgonzalez@ustatunja.edu.co

Hélver Rincón Márquez. Licenciado en Matemáticas y Física, Universidad Pedagógica y Tecnológica de Colombia. Ingeniero Civil. Universidad Santo Tomás seccional Tunja. Grupo de investigación GIPDCB, Tunja. Docente Tutor VUAD, Universidad Santo Tomás seccional Tunja. helver.rincon@ustatunja.edu.co

Resumen
En el presente trabajo se desarrolla una metodología que permite comparar los modelos de valor extremo tipo I (EVI) mejor conocida como distribución Gumbel y la distribución Log Pearson tipo III, para modelar un conjunto de valores extremos como son una serie de caudales máximos. Para la elección de la función de distribución más adecuada de los datos de la muestra se aplican tres criterios estadísticos tales como: el método gráfico, el cálculo de error cuadrático mínimo y la prueba de bondad de ajuste 2χ. Por último, se presentan conclusiones de tipo general y se muestra un ejemplo concreto vinculado a una base de datos que corresponde a los valores medios mensuales de caudal en m3/s del Río Bogotá, registrados en la estación hidrometeorológica la Balsa.

Palabras clave
Distribución Gumbel, distribución Log Pearson tipo III, intervalo de confianza, periodo de retorno, serie de excedencias, series de máximos.

Abstract
In this paper develops a methodology for comparing models of extreme value type I (EVI) better known as the Gumbel distribution and the Log Pearson Type III, distribution to model a set of extreme values such as a series of maximum flow. For the choice of most appropriate distribution function of the sample data, apply three statistical criteria such as the graphical method, the calculation of minimum square error and the test of goodness of fit. Finally, conclusions are general in nature and shows a specific example related to a database that corresponds to the average monthly flow rate m3/s of the Bogota River, in the Balsa hydrometeorological station.

Keywords
Gumbel distribution, Log Pearson type III distribution, confiance interval, return periods, exceedance series, maximum series.

1. INTRODUCCIÓN

En el estudio de la hidrología es posible observar cómo en repetidas ocasiones los sistemas hidrológicos se ven directamente afectados por eventos extremos, tales como aumento de caudales en ríos, presencia de tormentas severas y sequías, entre otras. La magnitud de un evento extremo está inversamente relacionado con su frecuencia de ocurrencia, es decir que los eventos moderados ocurren con mayor frecuencia, mientras que los eventos extremos se presentan en pocas oportunidades. Para analizar la probabilidad de ocurrencia de estos eventos se utilizan algunas distribuciones de probabilidad. Éstas son funciones matemáticas que relacionan la magnitud de un evento con su probabilidad de ocurrencia.

La probabilidad puede ser expresada en forma de frecuencia a través del periodo de retorno o recurrencia. El periodo de retorno T de un evento con una magnitud dada se define como el intervalo de recurrencia promedio entre eventos que igualan o exceden una magnitud especificada (Chow, Maidment y Mays, 1994).

La probabilidad $P(X \geq x_T)$ de ocurrencia del evento $X \geq x_T$ en cualquier observación se relaciona con el periodo de retorno de la siguiente forma:

$$T = \frac{1}{P(X \geq x_T)} \qquad (1)$$

Es posible calcular el periodo de retorno del n-ésimo evento para el conjunto de datos que pertenecen a una serie de valores extremos, Aparicio (2001). En particular para calcular el periodo de retorno máximo y mínimo para un conjunto de n elementos, se calcula de acuerdo con la expresión:

$$T = \frac{n+1}{m} \qquad (2)$$

Donde m corresponde a la posición de los datos de la serie, ordenados en forma descendente, para el caso del periodo de retorno máximo $m=1$ y para el periodo de retorno mínimo $m=n$.

Para el análisis de frecuencias de eventos se tienen en cuenta los siguientes supuestos, según Beguería (2002):

- Los eventos hidrológicos extremos son variables aleatorias que pueden expresarse por medio de una distribución de probabilidad.
- Si la magnitud de cada suceso no tiene correlación con los sucesos anteriores, significa que la serie de eventos extremos son independientes.
- La distribución de probabilidad que explica el proceso no varía en el tiempo, ni cambia en función de la magnitud de la variable.

Como el análisis de frecuencias de eventos extremos se relaciona directamente con el estudio de las colas de la distribución de frecuencias de la variable, se hace necesario introducir técnicas de muestreo que permitan extraer de las series de los datos originales, los valores de magnitud excepcional. En Hidrología existen dos técnicas de muestreo muy utilizadas que permiten extraer una serie de valores parciales: la serie de máximos y la serie de excedencias.

Para la serie de máximos se selecciona el valor máximo que ocurre en un intervalo de tiempo fijo, generalmente un año, de modo que el tamaño de la muestra será igual al número de años registrados en la serie original. En el caso de la serie de excedencias los datos deben seleccionarse de tal forma que su magnitud sea mayor que un valor predefinido, según Chow et al. (1994), éste valor corresponde al número de años en el registro.

"El periodo de retorno T_E de magnitudes de evento deducido de una serie de excedencia anual se relaciona con el correspondiente periodo de retorno T para magnitudes deducido de una serie máxima anual como" (Chow et al, 1994, p. 395).

$$T_E = \left[\ln\left(\frac{T}{T-1} \right) \right]^{-1} \qquad (3)$$

Cabe aclarar que la serie de excedencia anual es útil para algunos propósitos ya que hace un uso más eficiente de la información contenida en las series originales. Por ejemplo, puede incluir más de un evento por año, si éste cumple con el requisito para ser considerado extremo. No obstante, está limitada por el hecho de asegurar la independencia de las observaciones.

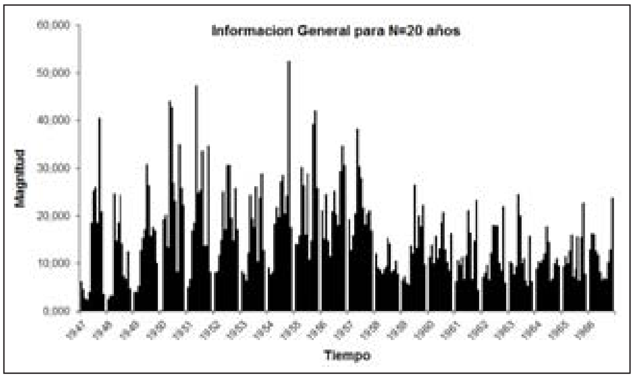
Por el contrario, para la serie de máximos anuales se garantiza la independencia de las observaciones, pues los eventos muestreados corresponderán a eventos igualmente espaciados en el tiempo.

Si por alguna razón se llegara a violar el supuesto de independencia para la serie de excedencia anual, los resultados no se verán afectados significativamente. Es decir, que mientras el periodo de retorno sea mayor, los resultados de los métodos analizados tienden a ser similares, Beguería (2001). Por esta razón, en el desarrollo de este artículo se utilizará la serie de máximos anuales para el análisis de los resultados.

Para calcular la probabilidad de ocurrencia de eventos extremos, a partir de la serie de máximos anuales, se asumirá que éste tipo de datos se ajustan en forma teórica a la función de probabilidad de valor extremo tipo I (EVI) mejor conocida como distribución Gumbel o a la distribución Log Pearson tipo III, Aparicio (2001). Para cada una de las distribuciones propuestas se estimarán sus parámetros, límites de confianza y bondad del ajuste.

La base de datos utilizada para el desarrollo del trabajo corresponde a los valores medios mensuales de caudal en m3/s del Río Bogotá, registrados en la Estación Hidrometeorológica La Balsa, Departamento de Cundinamarca, con registros comprendidos entre 1943 – 1999. Para la aplicación de la metodología se utilizaron únicamente los registros de 20 años, comprendidos entre 1947 a 1966, por estar la totalidad de los datos, ver figura 1.

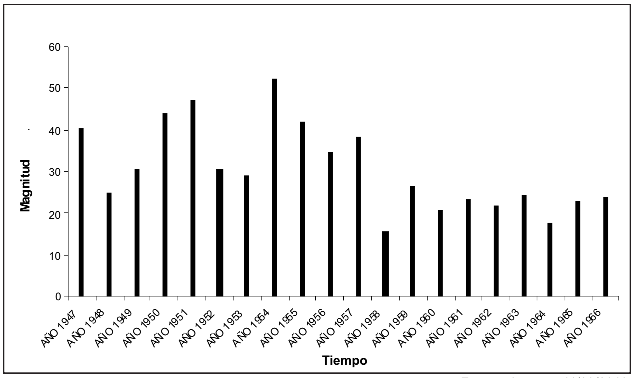
FIGURA 1.- Valores medios mensuales de caudales 1947 - 1966



Fuente: Autores, 11/09/2010

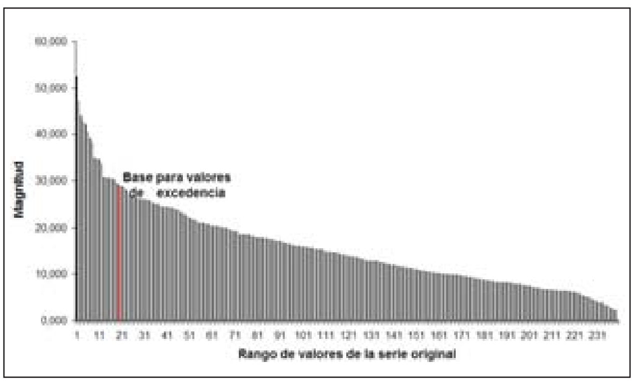
La figura 2, presenta la serie de valores máximos anuales, tomada de la serie original.

FIGURA 2.- Serie de máximos anuales



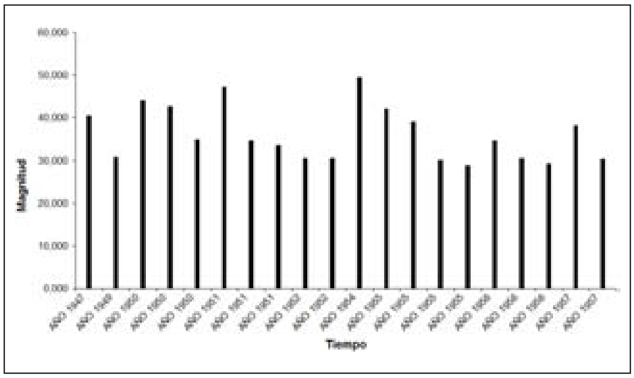
Fuente: Autores, 11/09/2010

FIGURA 3.- Muestreo para la serie de excedencias



Fuente: Autores, 11/09/2010

FIGURA 4.- Serie de excedencias anuales



Fuente: Autores, 11/09/2010

En la figura 3, se presenta el proceso gráfico de selección de la muestra para la serie de excedencias anuales, cuyo valor base corresponde a 28.88 m3/s. Del mismo modo, en la figura 4 se muestran los veinte (20) valores extremos menores a éste punto ordenados por su tiempo de ocurrencia.

2. METODOLOGÍA

2.1 Distribución de probabilidad para serie de máximos anuales: Gumbel (EVI)

Debido a la naturaleza de los eventos hidrológicos extremos la distribución más frecuente utilizada para las series de máximos anuales es la distribución de valores extremos Tipo I, su función de densidad de probabilidad es:

$$f(x) = \frac{1}{\alpha}\exp\left[\frac{-(x-\beta)}{\alpha} - \exp\left(\frac{-(x-\beta)}{\alpha}\right)\right] \quad (4)$$

Donde x puede tomar valores en el rango $-\infty \leq x \leq \infty$, α y β son parámetros de escala y origen respectivamente.

La función de distribución acumulada es:

$$F(x) = \int f(x)dx = \exp\left[-\exp\left(-\frac{(x-\beta)}{\alpha}\right)\right] \quad (5)$$

Donde los parámetros se estiman a través de las siguientes expresiones:

$$\hat{\alpha} = \frac{\sigma_x}{s_x} \quad \text{y} \quad \hat{\beta} = \bar{x} - \frac{\mu_x}{\hat{\alpha}} \quad (6)$$

Donde x y s_x son la media y la desviación estándar estimadas con la muestra. μ_x y σ_x se obtienen de la siguiente tabla.

n	μ_x	σ_x
10	0.4952	0.9496
15	0.5128	1.0206
20	**0.5236**	**1.0628**
25	0.5309	1.0914
30	0.5362	1.1124
⋮	⋮	⋮
100	0.5600	1.2065

Factor de frecuencia: para la distribución de valor extremo Tipo I, Chow et al. (1994, p. 402), dedujo la siguiente expresión:

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\} \quad (7)$$

En esta expresión se observa que los valores $\sqrt{6}/\pi$ y 0.5772 corresponden en forma análoga a σ_x y μ_x para un tamaño de muestra n, luego el factor de frecuencia K_T, se puede expresar como sigue:

$$K_T = -\sigma_x\left\{\mu_x + \ln\left[-\ln\left(\frac{T}{T-1}\right)\right]\right\} \quad (8)$$

Donde T es el periodo de retorno. Para la distribución Gumbel se tiene que el caudal para un período de retorno de 2.33 años es igual a la media de la distribución del valor extremo, es decir cuando $0{=}K_T$.

Para expresar en términos de K_T, la ecuación (7) puede escribirse como:

$$T = \frac{1}{1-\exp\left\{-\exp\left[-\left(\gamma + \frac{\pi K_T}{\sqrt{6}}\right)\right]\right\}} \quad (9)$$

Donde $\gamma = 0.5772$ es una constante.

Los Límites de confianza para el caudal medio de la serie de máximos anuales, corresponden a:

Límite superior de confianza: $L_{T,\alpha} = \bar{x} + t_{1-\alpha} s_e$

Límite inferior de confianza: $\mathrm{U}_{T,\alpha} = \bar{\mathrm{x}} - \mathrm{t}_{1-\alpha}\mathrm{s}_e$

$t_{1-\alpha}$ es la variable normal estandarizada para una probabilidad de no excedencia $\alpha-1$. Donde $_eS$ y δ corresponden a:

$$s_e = \frac{\delta \cdot s_x}{\sqrt{n}} \quad \text{Y} \quad \delta = \sqrt{1 + 1.1396K_T + 1.1{K_T}^2} \quad (9)$$

2.2 Distribución de probabilidad para serie de máximos anuales: Log- Pearson Tipo III

Esta distribución se utiliza para el análisis probabilístico de eventos extremos, en la cual se utiliza como variable $y{=}\log x$ para reducir la simetría. La estimación de los parámetros de esta distribución se calcula de la misma forma que para la distribución Pearson Tipo III, pero con la diferencia que y y s_y corresponden al promedio aritmético y desviación estándar de los logaritmos con base 10 de la variable original x.

La función de densidad de probabilidad para esta distribución corresponde a:

$$f(y) = \frac{1}{\alpha_1\Gamma(\beta_1)}\left(\frac{y-\delta_1}{\alpha_1}\right)^{\beta_1-1}\exp\left(-\frac{y-\delta_1}{\alpha_1}\right) \quad (11)$$

Donde y toma valores en el rango $\delta_1 \le y \le \alpha_1$ para α_1, β_1 y δ_1 son parámetros de la función y $\Gamma(\beta_1)$ es la función Gamma.

La función de distribución de probabilidad acumulada es:

$$F(y) = \frac{1}{\alpha_1\Gamma(\beta_1)}\int_0^y e^{-\left(\frac{y-\delta_1}{\alpha_1}\right)}\left(\frac{y-\delta_1}{\alpha_1}\right)^{\beta_1-1} dy \quad (12)$$

Sustituyendo $$w = \left(\frac{y-\delta_1}{\alpha_1}\right) \quad (13)$$

$$F(w) = \frac{1}{\Gamma(\beta_1)}\int_0^w e^{-w} w^{\beta_1-1} dw \quad (14)$$

La expresión (14) es una función de distribución chi-cuadrada con $2\beta_1$ grados de libertad y $\chi_2{=}2w$.

Los parámetros α_1, β_1 y δ_1 se evalúan a partir de los n datos de la muestra, mediante el sistema de ecuaciones:

$$\bar{y} = \alpha_1 \cdot \beta_1 + \delta_1$$

$$s_y^2 = \alpha_1^2 \cdot \beta_1$$

$$C_S = \frac{2}{\sqrt{\beta_1}}$$

Los parámetros estimados para esta distribución corresponden a las expresiones:

$$\hat{\beta}_1 = \left(\frac{2}{Cs}\right)^2; \quad \hat{\alpha}_1 = \frac{s_y}{\sqrt{\beta_1}}; \quad \hat{\delta}_1 = \bar{y} - \hat{\alpha}_1 \cdot \hat{\beta}_1 \quad (15)$$

El coeficiente de asimetría de la distribución γ se obtiene a partir del tercer momento alrededor de la media, dividiéndolo por el cubo de la desviación estándar para que sea adimensional.

$$\varphi = \frac{E(x-\mu)^3}{\sigma^3} \tag{16}$$

Un estimador muestral del coeficiente de asimetría está dado por:

$$C_s = \frac{n\sum_{i=1}^{n}(y_i-\bar{y})^3}{(n-1)(n-2)s_y^3} \tag{17}$$

Factor de frecuencia: el factor de frecuencia para la distribución Log- Pearson Tipo III depende del periodo de retorno T y del coeficiente de asimetría Cs, Chow et al. (1994, p. 403). Cuando $0 = Cs$, el factor de frecuencia es igual a la variable normal estándar z. Si $Cs \neq o$, K_T esta dada por la expresión:

$$K_T = z+(z^2-1)k+\frac{1}{3}(z^3-6z)k^2-(z^2-1)k^3+zk^4+\frac{1}{3}k^5 \tag{18}$$

Con

$$k = C_S/6$$

El valor de z para un periodo de retorno T es la variable normal estandarizada y k se encuentra tabulado de acuerdo al valor de Cs.

Intervalos de confianza: los estimadores estadísticos por lo general, se presentan con un rango o intervalo de confianza, dentro del cual se espera que incluya el valor correcto. El tamaño del intervalo depende del nivel de confianza β. Los valores extremos del intervalo se conocen como limites de confianza superior e inferior.

A cada nivel de confianza β corresponde un nivel de significancia α dado por

$$\alpha = (1-\beta)/2$$

De acuerdo con Chow et al. (1994, p. 417) los límites de confianza para estimar la magnitud del evento con un periodo de retorno T, están dados por:

$$\mathrm{L}_{\mathrm{T},\alpha} = \bar{y} + s_y k^{\mathrm{L}}_{\mathrm{T},\alpha}$$

Límite superior de confianza:

$$\mathrm{U}_{\mathrm{T},\alpha} = \bar{y} - s_y k^{\mathrm{U}}_{\mathrm{T},\alpha}$$

Límite inferior de confianza:

Donde $k^U_{T,\alpha}$ y $k^L_{T,\alpha}$ son los factores de los límites de confianza los cuales fueron aproximados por las siguientes ecuaciones para la distribución Log- Pearson Tipo III por (Natrella, 1963; U.S. Water Resources Council, 1981):

$$a = 1-\frac{z_\alpha^2}{2(n-1)} \tag{19}$$

$$b = K_T^2-\frac{z_\alpha^2}{n} \tag{20}$$

$$k^U_{T,\alpha} = \frac{K_T+\sqrt{K_T^2-ab}}{a} \tag{21}$$

$$k^L_{T,\alpha} = \frac{K_T-\sqrt{K_T^2-ab}}{a} \tag{22}$$

El valor z_α es la variable normal estándar con una probabilidad de excedencia α.

2.3 Ajuste de distribuciones de probabilidad

Para la modelación de caudales máximos se utilizarán las distribuciones Gumbel y Log-Pearson Tipo III principalmente. A continuación, se presentan algunos de los criterios estadísticos que permiten seleccionar la distribución de probabilidades de la serie histórica que mejor ajusta este tipo de datos:

2.3.1 Análisis gráfico

Para verificar que una función de distribución de probabilidad se ajusta a un conjunto de datos hidrológicos, éstos se grafican en un papel de probabilidad, utilizando una escala de graficación que linealice la función de distribución.

Como el objetivo del artículo es seleccionar la función de probabilidad que mejor ajuste la serie de máximos anuales, se elige la función de probabilidad para la que el conjunto de datos sea semejante a una línea recta.

Este método presenta un alto grado de subjetividad, es por esto que se recomienda usarlo paralelamente con otros métodos.

2.3.2 Método del error cuadrático medio

Este método consiste en calcular para cada función el error cuadrático de la distribución, donde x_{e_i} corresponde al i-ésimo dato estimado y x_{o_i} es el i-ésimo dato calculado con la función de distribución bajo análisis, donde C es el error cuadrático medio, para cada distribución. (Montgomery y Runger, 2008).

$$C = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{e_i}-x_{o_i}\right)^2}{n}} \tag{22}$$

2.3.3 Prueba de bondad de ajuste

La bondad de ajuste de una distribución de probabilidad puede probarse comparando los valores teóricos y muestrales de las funciones de frecuencia relativa. Para éste caso se utiliza la prueba chi–cuadrado χ^2. Antes de aplicar la prueba se construye una tabla de distribución de frecuencias con k intervalos de clase, donde el valor de k es el valor entero de calcular $1+33.3\log(n)$. El estadístico de prueba corresponde a:

$$D = \sum_{i=1}^{k} \frac{(\theta_i - \varepsilon_i)^2}{\varepsilon_i} \tag{24}$$

Donde θ_i es el número de observaciones en el intervalo i, y ε_i es el número esperado de eventos en el mismo intervalo. ε_i se calcula de la siguiente forma:

$$\varepsilon_i = n\left[f(S_i) - f(I_i)\right] \quad para\ i = 1, 2, \ldots, k \tag{25}$$

$f(S_i)$ es la función de probabilidad evaluada en el límite superior del intervalo I, $f(I_i)$ es la misma función pero evaluada en el límite inferior del intervalo i.

Una vez calculado el estadístico de prueba D para cada función de distribución, se determina el valor de la variable aleatoria con distribución χ^2, para $\nu = k - m - 1$ grados de libertad y un nivel de significancia α, donde m es el número de parámetros estimados a partir de los datos.

Para aceptar una función de distribución dada, se debe cumplir:

$$D \leq \chi^2_{1-\alpha,\nu} \tag{25}$$

Este valor se debe buscar en la tabla de valores de cuantiles de la distribución χ^2.

3. RESULTADOS

3.1 Estimación de los parámetros para la Distribución de probabilidad para serie de máximos anuales: Gumbel (EVI)

Para la estimación de los parámetros es necesario tener las estadísticas básicas de la serie de valores de máximos anuales que pueden calcularse a través de cualquier software de estadística. En este caso se muestra la salida del SPSS Statistic.

TABLA 2.- Estadísticas para caudales máximos

Descriptive Statistics

	N	Mean	Std. Deviation	Variance	Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
MAX	20	30,4995	10,42907	108,765	-,613	,992
Valid N (listwise)	20					

Fuente: Autores, 11/09/2010

El coeficiente de asimetría para $n = 20 años$ es $Cs = 0.613$. Los parámetros estimados según la expresión (6) con $\mu_x = 0.5236$ y $\sigma_x = 1.0628$ corresponden a:

$$\hat{\alpha} = \frac{1.0628}{10.4291} = 0.1019$$

$$\hat{\beta} = 30.49 - \frac{0.5236}{0.1019} = 25.36$$

La función de distribución estimada y acumula son respectivamente:

$$f(x) = \frac{1}{0.1019} \exp\left[\frac{-(x-25.36)}{0.1019} - \exp\left(\frac{-(x-25.36)}{0.1019}\right)\right]$$

$$F(x) = \exp\left[-\exp\left(-\frac{(x-25.36)}{0.1019}\right)\right]$$

A partir de la ecuación (2) el periodo de retorno máximo $m=1$, es $T=21$, mientras que para el periodo de retorno mínimo $m=20$, $T=1.05$.

Para un periodo de retorno $T=10 años$ el valor del factor de frecuencia es:

$$K_T = -1.0628\left\{0.5236 + \ln\left[-\ln\left(\frac{10}{10-1}\right)\right]\right\} = 1.84$$

El intervalo de confianza para el caudal medio poblacional de la serie de máximos anuales para un periodo de retorno $T=10$, de aproximadamente el 95% de confianza está dado por:

$$\bar{x} = 30.5$$

$$L_{T,\alpha} = 40.5$$

$$U_{T,\alpha} = 20.5$$

$$20.5 \leq \bar{x} \leq 40.5$$

3.2 Estimación de los parámetros para la Distribución de probabilidad para serie de máximos anuales: Log-Pearson Tipo III

TABLA 3.- Estadísticas para Log_{10} de caudales máximos

Descriptive Statistics

	N	Mean	Std. Deviation	Variance	Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
LOG10MAX	20	1,4606	,14721	,022	-,790	,992
Valid N (listwise)	20					

Fuente: Autores, 11/09/2010

La tabla 3, muestra las estadísticas básicas para la serie de los $\log_{10}$ de la serie de caudales máximos.

El coeficiente de asimetría para la serie de caudales máximos es:

$$C_s = \frac{n\sum_{i=1}^{n}(y_i - \bar{y})^3}{(n-1)(n-2)s_y^3} = 1.202$$

La estimación de los parámetros para la distribución Log Pearson tipo III, se calculan como sigue:

$$\hat{\alpha}_1 = \frac{0.1472}{\sqrt{2.77}} = 0.088$$

$$\hat{\beta}_1 = \left(\frac{2}{1.202}\right)^2 = 2.77$$

$$\hat{\delta}_1 = 1.4609 - (0.088)(2.77) = 1.22$$

La función de densidad para esta serie está dada por:

$$f(y) = \frac{1}{0.088\,\Gamma(2.77)}\left(\frac{y-(1.22)}{0.088}\right)^{2.77-1}\exp\left(\frac{y-(1.22)}{0.088}\right)$$

La función de distribución de probabilidad acumulada es:

$$F(w) = \frac{1}{\Gamma(2.77)}\int_0^w e^{-w}w^{\beta-1}\,dw$$

$$w = \frac{y-1.22}{0.088}$$

El valor de χ^2 y el número de grados de libertad son respectivamente:

$$\chi^2_{2\beta_1} = 2w$$

$$\nu = 2\beta_1 = 5.5 \quad \text{Grados de libertad}$$

El factor de frecuencia para un $C_s = 1.202$ y un periodo de retorno $T=10$ años se puede calcular también por medio de la tabla de valores K_T para la distribución Log Pearson tipo III. En este caso como el valor de C_s no se encontró directamente, se calculó mediante una interpolación.

$$K_{10} = 1.2896$$

El intervalo de confianza para el caudal medio poblacional de la serie de $\log_{10}$ de máximos anuales para un periodo de retorno $T=10$, de aproximadamente el 95% de confianza se determina calculando las expresiones (19), (20), (21) y (22), para un valor $z_{(0.05)} = 1.645$ el cual se toma de la tabla de probabilidad acumulada de la distribución normal estándar.

$$a = 1 - \frac{1.645^2}{2(20-1)} = 0.9288$$

$$b = 1.2896^2 - \frac{1.645^2}{20} = 1.5278$$

$$k^U_{T,\alpha} = \frac{1.2896 + \sqrt{1.2896^2 - 0.9288\cdot 1.5278}}{0.9288} = 1.9204$$

$$k^L_{T,\alpha} = \frac{1.2896 - \sqrt{1.2896^2 - 0.9288\cdot 1.5278}}{0.9288} = 0.8565$$

Intervalos de confianza para el Log10 de los caudales máximos

$$L_{T,\alpha} = 1.4606 + 0.1472\cdot 0.8565 = 1.5867$$

$$U_{T,\alpha} = 1.4606 - 0.1472\cdot 1.9204 = 1.1779$$

El límite de confianza para el promedio de los caudales máximos, con un periodo de retorno $T=10$ corresponde a:

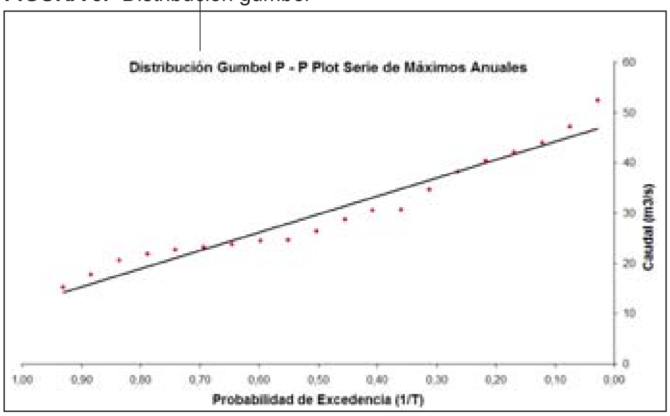
$$15.0628 \le \bar{x} \le 36.6087$$

3.3 Ajuste de distribuciones de probabilidad

3.3.1 Análisis gráfico

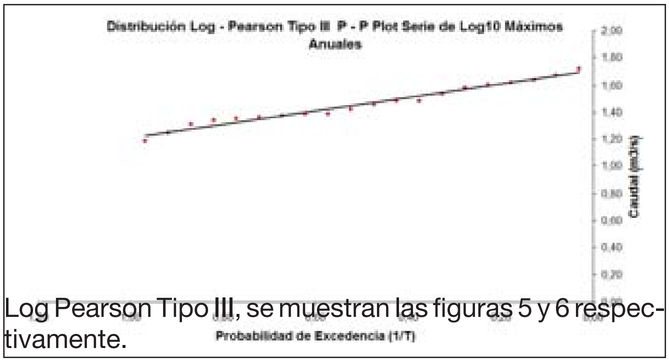
El método gráfico permite comparar el ajuste de una distribución a la serie de máximos anuales, para la distribución de valor extremo Tipo I (Gumbel) y la distribución

FIGURA 5.- Distribución gumbel



Fuente: Autores, 11/09/2010

FIGURA 6.- Distribución gumbel



Log Pearson Tipo III, se muestran las figuras 5 y 6 respectivamente.

En la figura 6, se observa que la línea ajustada es consistente con la mayoría las observaciones. Sin embargo, también se aprecia que en la figura 5, ocurre un comportamiento similar pero las observaciones se alejan más de la línea de tendencia. Teniendo en cuenta el análisis visual de las dos gráficas, se puede decir que la distribución que mejor ajusta el conjunto de datos es la Log Pearson Tipo III, por tener el mayor número de observaciones sobre la línea de tendencia.

3.3.2 Método del error cuadrático medio

El error cuadrático medio para las distribuciones Gumbel y Log Pearson Tipo III, se relaciona en la tabla 4; x_0 corresponde al valor de la serie de máximos anuales.

Para el caso de la distribución Gumbel, x_e se calcula en forma analítica, despejando el valor de la variable de la función de probabilidad acumulada.

El i-ésimo valor calculado para la función de distribución Gumbel, según Aparicio (2001, p. 265), está dado por la expresión:

$$x = \beta - \frac{1}{\alpha}\ln\left(-\ln\left(\frac{T-1}{T}\right)\right)$$

En el caso de la distribución Log Pearson Tipo III, cada uno de los valores x_e se obtiene entrando al gráfico (Figura 6), correspondiente al valor de probabilidad dado por la frecuencia experimental, e interceptando la recta para obtener el valor de la variable.

TABLA 4.- Errores cuadráticos medios para las distribuciones Gumbel y Log Pearson Tipo III

		Gumbel		Log Pearson Tipo III	
T	$x_0\ (m^3/s)$	$x_e\ (m^3/s)$	$(x_e-x_0)^2$	$x_e\ (m^3/s)$	$(x_e-x_0)^2$
21.0	52.4	55	6.75	48.49	15.32
10.5	47.2	47.95	0.56	45.91	1.66
7.0	44.0	43.71	0.08	43.48	0.28
5.3	42.09	40.61	2.18	41.17	0.85
4.2	40.5	38.14	5.57	38.98	2.30
3.5	38.25	36.05	4.84	36.91	1.79
3.0	34.7	34.22	0.23	34.95	0.06
2.6	30.7	32.55	3.51	33.10	5.75
2.3	30.5	31.06	0.31	31.34	0.71

TABLA 4.- Errores cuadráticos medios para las distribuciones Gumbel y Log Pearson Tipo III

		Gumbel		Log Pearson Tipo III	
T	$x_0\ (m^3/s)$	$x_e\ (m^3/s)$	$(x_e-x_0)^2$	$x_e\ (m^3/s)$	$(x_e-x_0)^2$
1.9	26.4	28.29	3.57	28.10	2.90
1.8	24.6	26.99	5.7	26.61	4.05
1.6	24.49	25.71	1.49	25.20	0.5
1.5	23.84	24.44	0.36	23.86	0.0
1.4	23.25	23.15	0.01	22.60	0.43
1.3	22.68	21.82	0.74	21.40	1.65
1.2	21.94	20.40	2.38	20.26	2.82
1.2	20.64	18.83	3.28	19.18	2.12
1.1	17.69	16.97	0.52	18.17	0.23
1.1	15.32	14.44	0.78	17.20	3.54
Error cuadrático medio			1.47		1.54

Fuente: Autores, 11/09/2010

En la tabla 4, se observa que la distribución que mejor ajusta la serie de máximos anuales es la distribución Gumbel, por tener el menor error cuadrático medio (1.47); aunque la diferencia entre los errores cuadráticos medios de las distribuciones no es muy significativa.

4. CONCLUSIONES

Se recomienda ajustar para una muestra de series de máximos anuales, homogénea (varianza mínima) la distribución de valor extremo tipo I Gumbel.

El intervalo de confianza (95%) para el promedio de caudal máximo anual para las distribuciones analizadas corresponden a:

Gumbel $20.5 \le \bar{x} \le 40.5$

Log Pearson Tipo III $15.06 \le \bar{x} \le 36.6$

Se observa que el estimador promedio muestral $\bar{x}=88.28$ está contenido en estos intervalos. Por lo que se recomienda ajustar la distribución Gumbel por tener la menor longitud en el intervalo.

El tamaño de muestra influye directamente sobre la confiabilidad de los resultados, se sugiere usar la serie de máximos anuales pues ésta garantiza independencia en los eventos.

Para el ajuste de los datos a la distribución Log Pearson, se requiere transformar la variable al campo logarítmico para modelarla, con el objetivo de disminuir la varianza muestral

En la prueba de ajuste de forma gráfica se dibujan los valores registrados en la serie de máximos anuales contra la distribución teórica de probabilidad y de manera visual (subjetiva), se determina si el ajuste es adecuado o no.

REFERENCIAS

Aparicio, F.J. (1987). Fundamentos de Hidrología de Superficie. Limusa, 303 pp.

Beguería, S. (2002). Revisión de métodos paramétricos para la estimación de la probabilidad de ocurrencia de eventos extremos en climatología e hidrología: El uso de series de excedencias y su comparación con las series de máximos anuales. 10 pp.

Chow, V.T., Maidment, D.R., Mays (1993). L.W. Hidrología aplicada. McGraw-Hill, 580 pp.

Dudewicz, E.J., Mishra, S.N. (1998). Modern mathematical statistics. John Wiley & Sons, 838 pp.

Montgomery, D.C., Runger, G.C. Probabilidad y estadística aplicadas a la ingeniería. Limusa Wiley, 817 pp.

Monsalve, G. (2002). Hidrología en la ingeniería. Escuela colombiana de ingeniería, 382 pp.