

02-020

NEW STATISTICAL DISTRIBUTION APPLIED HYDROLOGY IN THE SOUTHWEST AREA OF BADAJOZ

García Conde, Antonio Jesús ¹; Cortés Pérez, Juan Pedro ²; Lopez Rodriguez,
Fernando ²; Moral García, Francisco Jesús ²

¹ Universidad de Extremadura/Diputación de Badajoz, ² Universidad de Extremadura

In the study of flood plains of rivers, channeling sizing, and also for the design of drainage works in general is common to use the statistical distribution of Gumbel, log Pearson 3 or QRTSmáx. It is a statistical tool of proven effectiveness in the calculation of the distribution starting from the highs of a series. But there are other statistical DISTRIBUTIONS used in other scientific fields that fits perfectly, even better, as the fit test employees to extreme values given by the rain and that will develop in the article presented. Thus it has more tools to justify hydrological calculations in sensitive projects such as urban or other civil engineering works.

Keywords: *Statistics; Statistical distributions; hydrology*

NUEVA DISTRIBUCIÓN ESTADÍSTICA APLICADA EN HIDROLOGÍA EN LA PROVINCIA DE BADAJOZ

En el estudio de las llanuras de inundación de los ríos, el dimensionamiento de encauzamientos, y también para el dimensionamiento de obras de drenaje en general, es común el uso de la distribución estadística de Gumbel, log Pearson 3, VGE o QRTSmáx . Se tratan de unas herramienta estadísticas de contrastada eficacia en el cálculo de la distribución partiendo de los máximos históricos de una serie. Pero existen otra distribución estadística usada en otros campos científicos que se adapta perfectamente, incluso mejor, según los test de ajuste empleados, a los valores extremos dados por las lluvia y que se desarrollarán en el artículo presentado. De esta forma se dispone de más herramientas para justificar los cálculos hidrológicos en proyectos tan sensibles como los urbanos u otras obras de ingeniería civil.

Palabras clave: *Estadística; Distribuciones estadísticas; Hidrología*

Correspondencia: Antonio Jesús García Conde coligarcia@gmail.com

1. Introducción

Los estudios de inundabilidad de los ríos, así como, los estudios hidrológicos se realizan para asegurar que las aguas quedarán a una determinada cota durante una lluvia estadística con un determinado periodo de retorno. La lluvia de cálculo es utilizada para obtener la posible llanura de inundación, y es calculada a través de diversas distribuciones estadísticas (clásicas) a partir de una serie de datos de lluvia máximos anuales registrados. Las distribuciones más utilizadas en la actualidad son, la distribución desarrollada por el matemático alemán (Gumbel, 1960) y posteriormente aplicada a la hidrología (Chow, 1994), la distribución Log-Pearson 3 (Ponce, 1989) y (Pearson, 1916) propuesta por varios autores para utilizar en hidrología (Ferrer, 1991) y (Témez, 1991), la distribución de Valores Extremos generalizados (Kotz & Nadarajah, 2000), y la distribución SQRT-ET máx. (Zorraquino, 2004) utilizada por la publicación del Ministerio de Fomento "Máximas lluvias diarias en la España Peninsular" (Ministerio de Fomento, 1999) como la que mejor se ajusta a las características de las lluvias españolas; esta publicación se apoya en esta distribución para establecer un análisis regional basado en la geoestadística.

Actualmente, en España, el método más extendido para el cálculo del caudal de avenida (m^3/s) para un periodo de retorno concreto se encuentra regulada por la Instrucción de Carreteras, Norma 5.2-IC de Drenaje Superficial (Ministerios de Obras Públicas, 1990), en adelante Instrucción 5.2-IC. Esta instrucción propone el Método Racional (Catalá, 1997) para el cálculo del caudal, método que se basa en que el tiempo de duración del aguacero es mayor o igual que el tiempo de concentración de la cuenca y, por lo tanto, determina caudales maximizados. El Método Racional es utilizado en España para cuencas reducidas. Para cuencas grandes con periodos de concentración de más de 6 horas o más de 1 Km² y hasta 3.000 km² se utiliza el método modificado (Témez, 1991) citado por numerosos hidrólogos (Ferrer, 1993).

El presente artículo estudia el comportamiento de una nueva distribución estadística, la distribución de Burr (Singh & Maddala, 1976), que es actualmente utilizada en otras ramas de la ciencia (Kleiber & Kotz, 2003) y nunca ha sido empleada en hidrología. Se calcula el ajuste de las mencionadas distribuciones mediante los test de bondad de Anderson-darling (Anderson & Darling, 1952) y el test de Kolmogorov-Smirnov (Kolmogorov, 1933).

2. Objetivos

Como se ha mencionado, el presente artículo estudia el comportamiento de una nueva distribución estadística, la distribución de Burr, que es actualmente utilizada en otras ramas de la ciencia y nunca ha sido empleada en hidrología, y de este modo, determinar si se adapta estadísticamente mejor que las utilizadas tradicionalmente hasta ahora a las lluvias de la Provincia de Badajoz.

Se analiza de este modo, si la Distribución de Burr se ajusta mejor según los test de bondad a las distribuciones de las máximas lluvias anuales que el ajuste que nos proporcionan las distribuciones típicas utilizadas frecuentemente, sobre la base de los datos de lluvias máximas anuales de las estaciones meteorológicas de la Provincia de Badajoz.

3. Estado actual del conocimiento

LA DISTRIBUCIÓN DE GUMBEL.

Según su propio autor (Gumbel, 1935) su función de densidad es :

$$F(x) = e^{-e^{-\alpha(x-\mu)}} \quad (1)$$

Donde x es el valor de la variable aleatoria, α y μ son parámetros del ajuste y $F(x)$ representa la probabilidad de que se produzca una precipitación con un valor menor o igual que x .

La distribución de Log-Pearson tiene como formulación:

$$f(x) = \frac{\lambda^\beta (y - \varepsilon)^{\beta-1} e^{-\lambda(y-\varepsilon)}}{x\Gamma(\beta)} \quad (2)$$

Donde β , λ y ε son los parámetros de forma, escala y posición respectivamente y que responden a las siguientes fórmulas:

$$\beta = \left(\frac{2}{C_s} \right)^2 \quad (3)$$

$$C_s = \frac{n}{(n-1) * (n-2) * (S_{\log(x)})^3} * \sum (\log(x) - \overline{\log(x)})^3 \quad (4)$$

$$\lambda = \frac{S_y}{\sqrt{\beta}} \quad (5)$$

LA DISTRIBUCIÓN DE SQRT-ET MÁX

Tiene como formulación (Zorraquino, 2004) :

$$F(x) = \exp\left[-k \cdot (1 + \sqrt{\alpha \cdot x}) \cdot \exp(\sqrt{-\alpha \cdot x})\right] \quad (6)$$

$F(x)$ es la probabilidad de que se presente un valor inferior a x ,

k y α son parámetros a estimar que dependen de la media y la desviación típica de la serie de datos.

LA DISTRIBUCIÓN LOG-PEARSON 3

La distribución Log-Pearson Tipo 3 (Pearson, 1901) describe la distribución de la probabilidad de ocurrencia de un evento determinado en un proceso de Poisson. Cuando la población de los acontecimientos presenta un sesgo positivo, los datos suelen ser transformados a logaritmos y la distribución se llama distribución de Log Pearson Tipo 3.

Su función de densidad es la siguiente:

$$f(x) = \frac{\lambda^\beta (y - \varepsilon)^{\beta-1} e^{-\lambda(y-\varepsilon)}}{x\Gamma(\beta)} \quad (7)$$

Donde:

$y = \log(x)$

$\Gamma(\beta)$ = Función

β, λ y ε son los parámetros de forma, escala y posición respectivamente y que responden a las siguientes fórmulas:

LA DISTRIBUCIÓN DE VGE

Tiene como función (Kotz & Nadarajah, 2000)

$$f(x) = \left\{ \frac{1}{\sigma} e^{-(1+kz)^{-1/k}} (1 + k \cdot z) \right\}^{-1-k \cdot z} k \neq 0 \quad (8)$$

Para la obtención de los caudales de cálculo para un periodo de retorno de 500 años se ha de utilizar el método Racional, descrito en la Instrucción de Carreteras 5.2-IC de Drenaje Superficial o el racional modificado o el Témex dependiendo del tamaño de la cuenca, pero es insoslayable mencionar que el presente artículo solo proporciona los datos de la precipitación en mm arrojada por las distribuciones de probabilidad.

LA DISTRIBUCIÓN DE BURR

Es interesante destacar esta distribución estadística de Burr (Maddala, 1996) se conoce desde hace tiempo y se usa en distintos campos como la econometría o ciencias sociales; pero que nunca se han utilizado en hidrología. Es compatible su uso en hidrología ya que es utilizada igualmente para datos extremos como hacen las distribuciones tradicionales, es decir, para datos de lluvia máximos anuales.

En la teoría de la probabilidad, estadística y econometría, la distribución a Burr Tipo XII, o simplemente la Distribución de Burr, es una distribución de probabilidad continua para un valor no negativo de la variable aleatoria. También es conocida como la distribución de Singh-Maddala y es una de la serie de diferentes distribuciones llamadas de distribución logística.

Su formulación corresponde con la siguiente.

Su función de densidad queda definida por:

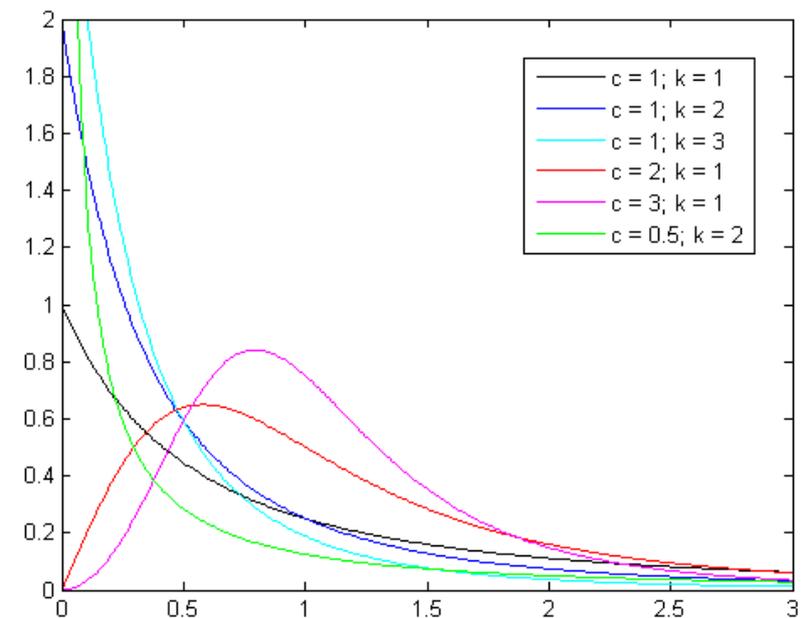
$$f(x) = c \cdot k \frac{\left(\frac{x}{\beta}\right)^{c-1}}{\beta \left[1 + \left(\frac{x}{\beta}\right)^\alpha\right]^{k+1}} \quad (9)$$

Integrando la función de densidad su función de distribución queda de la forma:

$$F(x) = 1 - \left(1 - \left(\frac{x}{\beta}\right)^\alpha\right)^{-k} \quad (10)$$

Donde c y k son parámetros de forma, y β lo es de escala.

Figura 1: Función de densidad de Burr.



Nota: Función de densidad de la distribución de Burr en función de sus parámetros.

EL TEST DE KOLMOGOROV- SMIRNOV

En el test de Kolmogorov-Smirnov (Kolmogorov, 1950) considera dos hipótesis:

$$H_0: F(X) = F_s(X) \quad (11)$$

$$H_1: F(X) \neq F_s(X)$$

Donde $F(X)$ es la función de distribución que se pretende estudiar, y $F_s(X)$ es la probabilidad o proporción teórica de valores que deben ser iguales o menores que x suponiendo cierta la hipótesis planteada.

Muestra: n observaciones independientes.

$$\text{Estadístico de contraste: } |D_n| = \text{máxima } |F_s(X) - S(X)| \quad (12)$$

EL TEST DE ANDERSON- DARLING

En el test de Anderson- Darling (Anderson & Darling, 1954) , como estadístico de prueba se usa la siguiente formulación:

$$A_2 = -N \cdot S \quad (13)$$

$$S = \sum_{i=1}^N \frac{(2i-1)}{N} \cdot \ln(F(x)_i + \ln(1 - F(x)_{N+1-i})) \quad (14)$$

Donde:

A_2 : estadístico de Prueba

N : tamaño de la muestra

$F(x)$: frecuencia

4. Materiales y métodos

Se realiza el presente estudio con los datos de máximas lluvias anuales de 53 estaciones meteorológicas de la provincia de Badajoz aportados por el Instituto meteorológico.

Se ha procedido a desarrollar un exhaustivo estudio utilizando los test estadísticos de bondad de Anderson-Darling (Anderson & Darling, 1954) y de Kolmogorov-Smirnov (Kolmogorov, 1950) para establecer el grado ajuste de esta nueva distribución de Burr (a la serie de lluvias máximas anuales, y así poder comparar los ajustes de esta nueva distribución con los ajustes estadísticos dados por las distribuciones estadísticas clásicas, Gumbel (Gumbel, 1960), SQRT-ET máx (Zorraquino, 2004) , Log Pearson tipo 3 (Ponce, 1989) y la distribución VEG (Kotz & Nadarajah, 2000).

Es decir, el objetivo es comparar mediante los dos ajustes de bondad mencionados, si la función de densidad de la nueva distribución estadística de Burr (Tadikamalla, 1980) se

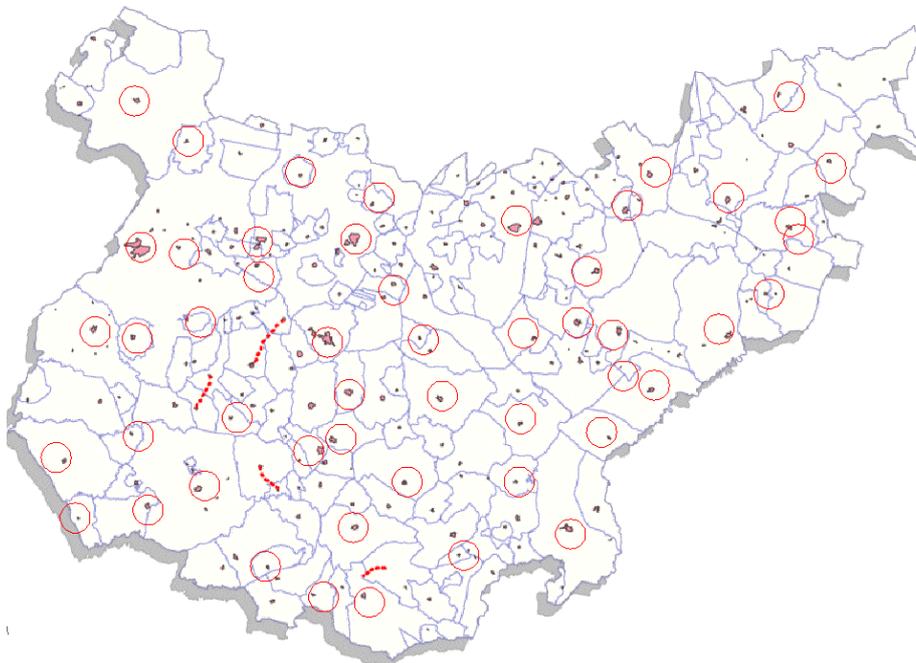
ajusta mejor que las distribuciones tradicionales a las lluvias extremas caídas en la provincia de Badajoz durante una serie de registros anual.

Para el ajuste de las distribuciones se ha empleado el software easyfit, el programa al introducirle los datos de lluvias máximas anuales proporciona el grado de ajuste de cada distribución estadística según los dos test de bondad utilizados.

Como se decía, se han procesado los datos de las siguientes 52 estaciones meteorológicas:

San Vicente Alcántara, Alburquerque, La Roca de la Sierra, Badajoz (Sagrajas), Montijo (Instituto), Olivenza, Cheles, Alconchel, Villanueva del fresno, Valencia Mombuey, Zahínos, Higuera de Vargas, Jerez de los Caballeros, Barcarrota, La Albuera, Talavera de la Reina, Fregenal de la Sierra, Segura de León, Cabeza la Vaca, Fuente de Cantos, Puebla del Maestre, Casas de Reina, Villagarcía de las Torres, Berlanga, Valverde de Llerena, Azuaga, Granja de Torrehermosa, Maguilla, Peraleda del Zaucejo, Puebla del Prior, Puebla de la Reina, Monterrubio de la Serena, Puerto Hurraco, Castuera, Quintana de la Serena, Valle de la Serena, Alange, Manchita, Guareña, Aceuchal, Mérida (centro escolar), Santa Amalia, Coronada, Campanario, Acedera, Orellana de la Sierra, Casas de Don Pedro, Capilla, o Baterno, Herrera del Duque, Villarta de los Montes, Helechosa y Siruela.

Figura 1: Situación de las estaciones meteorológicas.

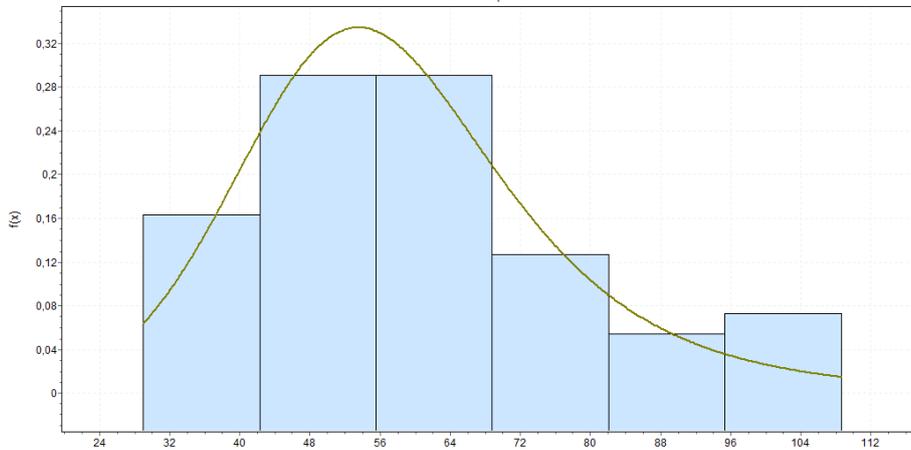


Nota: Distribución espacial de las estaciones meteorológicas.

Estos datos de lluvia de las 52 estaciones meteorológicas se han introducido en el programa estadístico Easyfit y se ha estudiado el grado de ajuste de la función de densidad de cada distribución estadística a los histogramas de lluvias.

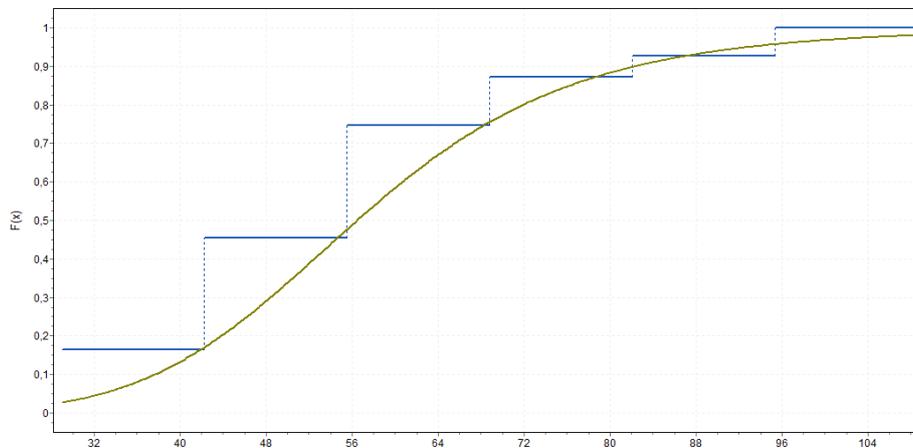
Como ejemplo, en la Figura 2 se puede observar la distribución de probabilidad de la distribución de Burr en la localidad de San Vicente de Alcántara.

Figura 2 : Función de densidad de Burr.



Nota: Distribución Función de Probabilidad de Densidad de Burr en San Vicente de Alcántara.

Figura 3: Función de distribución acumulada de Burr.



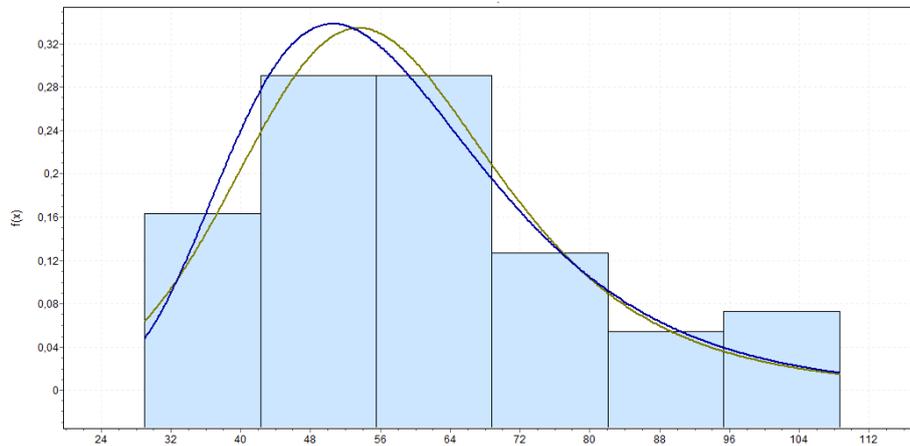
Nota: Función de Distribución Acumulada para la distribución de Burr aplicable a los datos de la localidad de San Vicente de Alcántara.

En la Figura 3 se puede comprobar cómo se ajusta una función de densidad al histograma acumulado de una serie de lluvias de valores máximos anuales. Se evidencia que la función de densidad de la distribución estadística nunca va a tomar exactamente los valores del histograma como sería deseable. Es por esto, que mediante los test de bondad se comprueba cual es la distribución estadística que mejor se ajusta a la serie de datos de lluvia.

Es decir, lo ideal sería que la distribución estadística elegida se ajuste lo máximo posible al histograma de lluvias.

Así mismo, en Figura 4 se puede observar gráficamente el mejor ajuste de la distribución de Burr que el ajuste de la distribución de Gumbel en la localidad concreta.

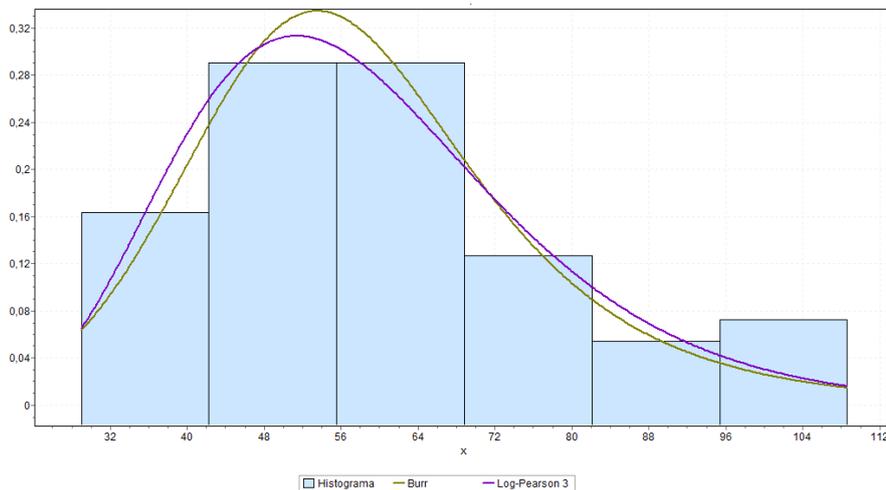
Figura 4: Histograma, Distribución de Gumbel y de Burr



Nota: Diferencia del ajuste entre las Distribuciones de Gumbel y de Burr aplicable al histograma de los datos de la localidad de San Vicente de Alcántara. Azul Gumbel.

En la Figura 5 se puede observar gráficamente que el ajuste es más preciso en la distribución de Burr que el ajuste de la distribución de Log-Pearson 3.

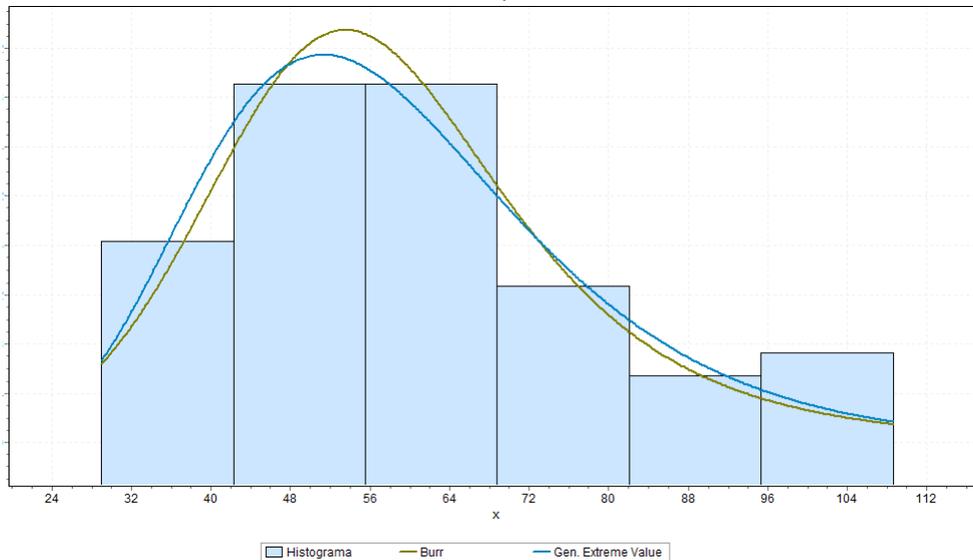
Figura 5: Histograma, Distribución de Log-Pearson 3 y de Burr



Nota: Diferencia del ajuste entre las Distribuciones de Log-Pearson 3 y de Burr aplicable al histograma de los datos de la localidad de San Vicente de Alcántara.

De idéntica manera, en la Figura 6 se puede observar gráficamente que el ajuste es más preciso en la distribución de Burr que el ajuste de la distribución de Log-Pearson 3.

Figura 6: Histograma, Distribución VEG y de Burr



Nota: Diferencia del ajuste entre las Distribuciones de VEG y de Burr aplicable al histograma de los datos de la localidad de San Vicente de Alcántara.

Una vez realizados los cálculos aritméticos del ajuste de las distribuciones se obtienen los siguientes datos:

La distribución de Burr arroja un estadístico de 0,049 según el test de Kolmogorov, y un 0,1746 según el test de Anderson-Darling.

La distribución de Gumbel tiene un estadístico de 0,06037 según el test de Kolmogorov, y un 0,2553 según el test de Anderson-Darling.

La distribución de Log-Perason 3 arroja un estadístico de 0.0584 según el test de Kolmogorov, y un 0,1947 según el test de Anderson-Darling.

La distribución de VGE tiene un estadístico de 0,05637según el test de Kolmogorov, y un 0,1875 según el test de Anderson-Darling.

Esta misma tendencia de presentar la distribución estadística Burr menos estadísticos de bondad de ajuste aparece al estudiar los datos de lluvia de las restantes 51 poblaciones .

Además, y solo para fortalecer las conclusiones anteriores, se ha estudiado la distribución de Burr se han comparado los ajustes de estas distribuciones solo en las diez localidades con mayor serie de datos, llegando a las mismas conclusiones.

5. Resultados y conclusiones

Al analizar los datos de lluvias máximas anuales de 52 estaciones pluviométricas repartidas estratégicamente a lo largo de la provincia y ajustarlos mediante los Test de bondad de Anderson-Darling (Anderson & Darling, 1954) y Kolmogorov-Smirnov (Kolmogorov, 1950), se puede afirmar sin ningún atisbo de duda que además de las distribuciones tradicionalmente usadas en hidrología como la distribución de Gumbel (Gumbel, 1960), Log-Pearson III (Pearson, 1916), y la distribución VEG (Kotz & Nadarajah, 2000), existe la distribución estadística de Burr (Tadikamalla, 1980) usada en otras ramas de la ciencia actual y que es susceptible de ser utilizada en hidrología, que cumple con la formulación de los valores extremos, y que se adapta mejor a los histogramas de lluvia.

Analizando los datos estadísticos arrojados en el estudio, se llega a la conclusión de que la distribución de Burr arroja unos estadísticos menores en los dos test de bondad de ajuste mencionados y por lo tanto, se adapta significativamente mejor a los histogramas de los datos de lluvias máximas anuales que las distribuciones típicas tradicionales. Esto es, esta nueva distribución estadística se adapta mucho mejor al régimen de lluvias que las tradicionales en la provincia de Badajoz.

Por lo tanto, es insoslayable proponer la distribución estadística de Burr como una distribución que mejora los estudios hidrológicos en la provincia de Badajoz, ya que la precipitación arrojada por su función de densidad es mucho más exacta que los datos de lluvia calculados con las distribuciones estadísticas clásicas que se vienen utilizando constantemente y de manera sistemática en los estudios hidrológicos.

Bibliografía

Anderson, T. W & Darling, D. A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*. Institute of Mathematical Statistics.

Anderson, T.W. and Darling, D.A. (1954). A Test of Goodness-of-Fit. *Journal of the American Statistical Association*.

Catalá Moreno, F. (1997). Cálculo de caudales en las redes de saneamiento. Madrid. Colegio de Ingenieros de Caminos, Canales y Puertos.

Chow, V.T. & D.R. Maidment, L.W. Mays. (1994). *Hidrología Aplicada*. Santa Fé de Bogotá. McGraw-Hill.

Ferrer, F.J. (1991). Obtención de la lluvia neta según la metodología del Soil Conservation Service. *Apuntes del Curso de Hidrología General y Aplicada*. Madrid. Cedex.

Ferrer, F.J. (1993). Recomendaciones para el cálculo hidrometeorológico de avenidas. Madrid. Cedex.

Gumbel, E.J. (1935). Les valeurs extrêmes des distributions statistiques. *Ann. Inst. Henri Poincaré*.

Gumbel, E. J. (1960). *Multivariate Extremal Distributions*. París. Bull. Inst. Internat. de Statistique.

Kleiber & Kotz. (2003). *The Burr Distributions*. New York. Wiley.

Kolmogorov, A. (1933). *Els fonaments de la Teoria de la Probabilitat* (2^a ed.). Chelsea. Nova York.

A. N. Kolmogorov. (1950). *Foundations of the Theory of Probability*. New York. Chelsea Publishing Company.

Kotz, S & Nadarajah, S. (2000). *Extreme Value Distributions: Theory and applications*. London. Imperial College Press.

Maddala, G. (1996). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge. Cambridge University Press.

Ministerio de Fomento.(1999). *Máximas Lluvias diarias en la España Peninsular*. Madrid. Ministerio de Fomento.

Instrucción de Carreteras, Norma 5.2-IC de Drenaje Superficial. (1990). Madrid. Ministerio de Obras Públicas y Urbanismo.

Pearson, Karl (1901). *Mathematical contributions to the theory of evolution, X: Supplement to a memoir on skew variation*. London. *Philosophical Transactions of the Royal Society*.

Pearson, Karl. (1916). Mathematical contributions to the theory of evolution, XIX: Second supplement to a memoir on skew variation. London. Philosophical Transactions of the Royal Society.

Ponce, V. M. (1989). Engineering Hydrology, Principles and Practices. Prentice Hall.

Singh, S & Maddala, G. (1976). A Function for the Size Distribution of Incomes. *Econometrica* .

Tadikamalla, P. (1980). A look at the Burr and related distributions. *International Statistical Review*.

Témez, J.R. (1991). Extended and improved rational method. Version of the highways administration of Spain". Madrid. XXIV Congreso Internacional de la IAHR, Madrid.

Zorraquino, C. (2004) La función SQRT-ET max. Madrid. *Revista de Obras Públicas*.

