

AJUSTE DE LA DISTRIBUCIÓN DE PROBABILIDAD DE MÁXIMA PRECIPITACIÓN DIARIA EN SANTA CRUZ DE TENERIFE A LA DISTRIBUCIÓN GENERALIZADA DE PARETO

Ana MORATA GASCA y Carlos ALMARZA MATA

*Instituto Nacional de Meteorología. Camino de la Moreras s/n. Madrid 28040. Spain
e-mail: anamg@fis.ucm.es, almarza@inm.es*

RESUMEN

Las colas de las distribuciones de probabilidad permiten modelizar sucesos con poca probabilidad de ocurrencia y por tanto largos períodos de retorno. La cola derecha de las distribuciones de lluvia es de gran importancia en la evaluación de riesgos relacionados con avenidas. Entre las diferentes funciones de distribución de probabilidad que permiten modelizar sucesos extremos en climatología se ha escogido la distribución de Pareto generalizada por permitir la modelización de los sucesos a partir de un cierto umbral. La serie de precipitación de Santa Cruz de Tenerife ha sido modelizada usando la distribución de Pareto.

Palabras clave: Distribución de Pareto, umbral, precipitación, Santa Cruz de Tenerife.

ABSTRACT

The tail distribution is of special interest because extreme events characterised by a low probability of occurrence and long return period, have a capital importance in dealing with various problems including design. The right tail of rainfall data distributions is of great importance in assessing risks related to floods. From among the different models of probability distribution functions available to model extreme hydrologic events, the generalized Pareto distribution has been chosen for its adequacy in modelling an excess beyond a threshold. The Santa Cruz de Tenerife precipitation serie has been modelised by means the Pareto distribution.

Key words: Pareto distribution, threshold, precipitation, Santa Cruz de Tenerife.

1. INTRODUCCIÓN

Las colas de las distribuciones de probabilidad de las variables climatológicas son de especial interés ya que permiten modelizar sucesos extremos con poca probabilidad de ocurrencia. En el caso de la precipitación diaria es la cola derecha de la distribución de probabilidad correspondiente la que permite este tipo de modelización. En este estudio se han escogido diferentes umbrales de precipitación con el fin de realizar un estudio comparativo de los periodos de retorno de eventos por encima del umbral. Se ha estudiado la serie de precipitación de Santa Cruz de Tenerife que en dato diario cuenta con registro desde 1938 a abril de 2002 y en dato máximo mensual desde 1925 hasta abril de 2002. En el proceso seguido para el ajuste del modelo se han tenido que tratar problemas como la elección del valor umbral, la inferencia estadística de los parámetros de la distribución y la estimación de la probabilidad de sucesos con un largo período de retorno.

2. DATOS Y METODOLOGÍA

Se tienen series de datos observacionales de precipitación diaria de Santa Cruz de Tenerife (1865-4/2002) y La Laguna (1906-4/2002). Los datos se han obtenido del archivo documental del Instituto Nacional de Meteorología así como de los contenidos en la base de datos climatológica. La serie diaria de Santa Cruz solo se considerará desde 1938, año a partir del cual aparece completa, ya que para períodos anteriores el registro presenta innumerables lagunas. La serie de La Laguna presenta los siguientes años incompletos: 1935, 1936, 1953 (de abril a diciembre), 1957 (de abril a diciembre), 1964-74, 1977-85, de modo que esta serie ha sido desechada en el estudio con datos diarios. Se intentó utilizar, sin embargo, en la reconstrucción desde 1925 hacia el pasado de la serie de máximos mensuales de precipitación diaria de Santa Cruz, pero tras el estudio comparativo de ambas series en períodos de solape se descartó por completo esta idea ya que la correlación entre las dos series es de $r = 0.21$, valor que estadísticamente no es significativo. Los datos comprendidos entre 1925 y 1937, ambos inclusive, se han extraído del fondo antiguo de datos que el Instituto Nacional de Meteorología tiene de la serie de Santa Cruz. Con todo, el registro del que se dispone para la serie de máximas del observatorio de Santa Cruz de Tenerife es desde el año 1925 hasta abril de 2002.

En trabajos dirigidos al estudio de sucesos extremos es esencial disponer de series de datos homogéneas. Una serie temporal es homogénea si las medidas han sido tomadas de forma consistente mediante el mismo método, con la misma instrumentación, en el mismo lugar y hora y en el mismo ambiente. Para detectar inhomogeneidades en series temporales se utilizan diferentes tests estadísticos, tests univariantes como el de las rachas y el de Mann-Kendall o tests que requieran series de referencia *a priori* homogéneas con respecto a las cuales se analizan las series a estudiar, como el método de Alexansersson. La homogeneidad de la serie de precipitación anual acumulada ha sido previamente estudiada por Almarza et. al. 1996, trabajo que concluyó en que la serie tiene un comportamiento homogéneo y sin tendencia desde el año 1925. Por otro lado se ha estudiado el comportamiento de la serie de máximos anuales (fig. 1) obteniendo resultados análogos.

3. DESARROLLO DEL MODELO

Se extrae de la serie un histograma de frecuencias para cada uno de los umbrales escogidos y su estadística correspondiente, y con ello se construye el modelo para cada umbral. El modelo escogido ha sido la distribución Generalizada de Pareto (DGP) ya que permite fijar un umbral y modelizar sucesos por encima del mismo con poca probabilidad de ocurrencia y por tanto largos períodos de retorno. Este modelo ya ha sido utilizado en hidrología para el estudio de avenidas (ASHKAR y OUARDA, 1996). La DGP tiene como función de distribución:

$$F(x) = \begin{cases} 1 - \left(1 - \frac{kx}{\alpha}\right)^{\frac{1}{k}}, & k \neq 0 \\ 1 - e^{-x/\alpha}, & k = 0 \end{cases}$$

y como función de densidad:

$$f(x) = \begin{cases} \alpha^{-1} \left(1 - \frac{kx}{\alpha}\right)^{\frac{1-k}{k}}, & k \neq 0 \\ \alpha^{-1} e^{-x/\alpha}, & k = 0 \end{cases}$$

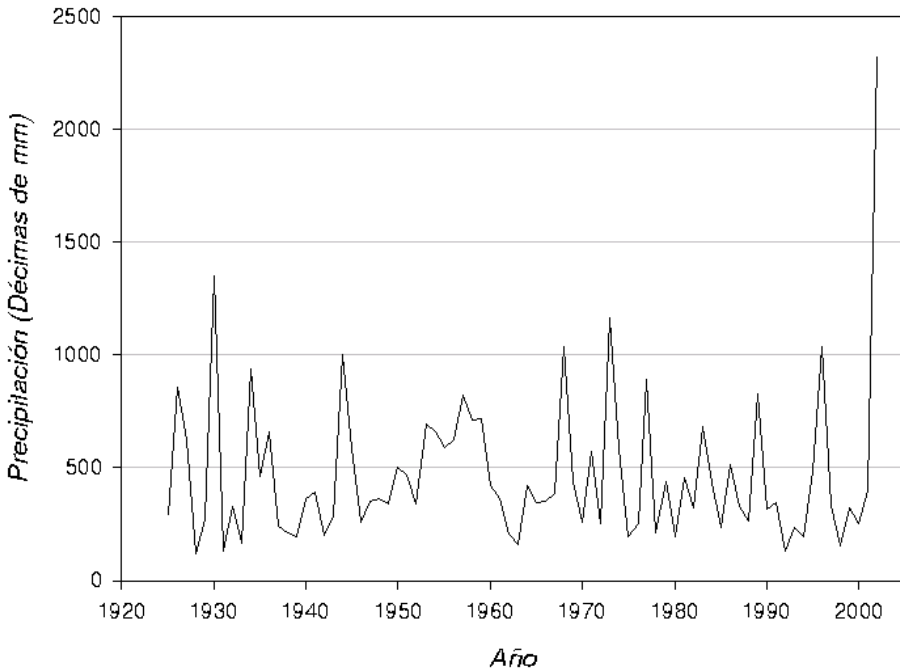


Figura 1: Máximos anuales de precipitación diaria

donde

$$\begin{cases} 0 \leq x < \inf & \text{para } k \leq 0 \\ 0 \leq x \leq \frac{\alpha}{k} & \text{para } k > 0 \end{cases}$$

La DGP se define para dos parámetros: α es el parámetro de escala y k es el parámetro de forma. Cuando $k = 0$ se convierte en la distribución exponencial. El valor del parámetro k no varía con la elección del umbral (BRADSON y PALUTIKOV, 2000). La estimación de los parámetros de escala y forma se realiza mediante el método de los momentos (HOSKING y WALLIS, 1987):

$$\hat{\alpha} = \frac{1}{2} \bar{x} \left(\frac{\bar{x}^2}{s^2} + 1 \right) \quad \hat{k} = \frac{1}{2} \left(\frac{\bar{x}}{s^2} - 1 \right)$$

Para el cálculo de los períodos de retorno se define un valor x para el que la probabilidad de ocurrencia, de un suceso $X > x$, en un año sea igual a $1/T$. La probabilidad de un exceso $Y(Y = X - u)$ es mayor que un valor especificado y viene dada por:

$$P(Y > y) = 1 - P(Y \leq y) = 1 - F(y)$$

La probabilidad de que este suceso ocurra en un día es:

$$P(Y > y) = p_1(1 - F(y))$$

donde p_1 es la probabilidad de que la lluvia diaria supere cierto umbral. La probabilidad de que se de un suceso en un año es:

$$P(Y > y) = [p_1(1 - F(y))]^{p_2}$$

donde p_2 es el número de sucesos medio por año. El período de retorno, en años, para el suceso $Y > y$ con un máximo de precipitación diaria de $y + u$ es:

$$T = \frac{1}{P(Y > y)}$$

4. RESULTADOS

Las figuras 2 y 3 muestran las predicciones sobre los datos de los modelos generados con las distribuciones de Pareto, con el parámetro k estimado por el método de los momentos, y exponencial, con un parámetro $k = 0$ respectivamente, obtenidos a partir de la serie diaria de Santa Cruz de Tenerife, para distintos umbrales de precipitación extrema. El eje de ordenadas representa en ambos casos, precipitación esperada en décimas de milímetro.

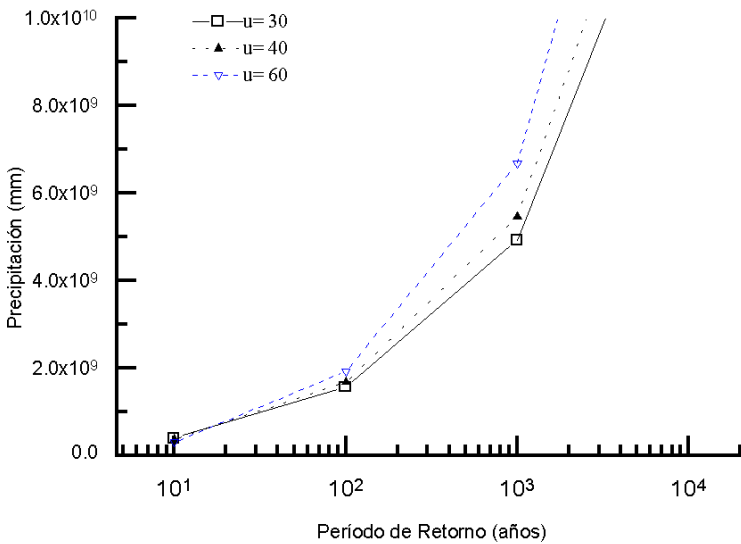


Figura 2: Modelos generados con la distribución de Pareto para distintos valores umbral.

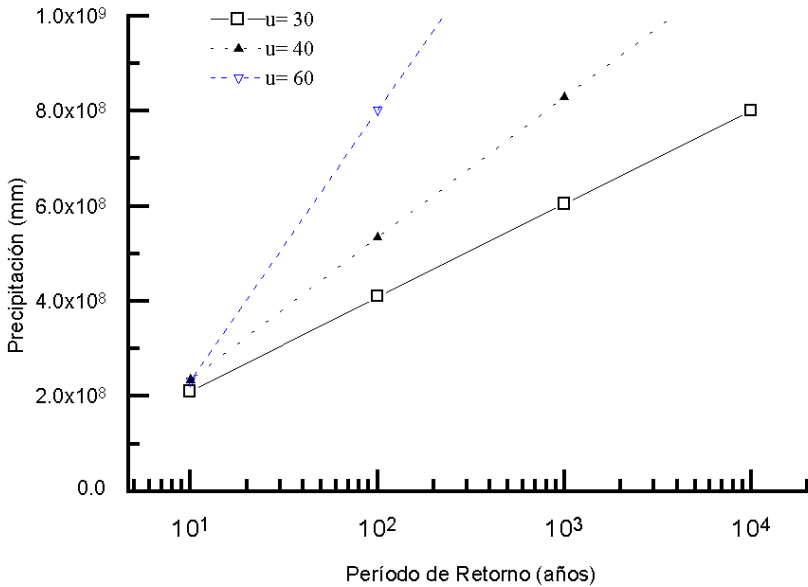


Figura 3: Modelos generados con la distribución exponencial para distintos valores umbral.

Se puede observar que cuanto mayor es el umbral seleccionado, más alto estima el modelo el valor de la precipitación para un período de retorno fijado. Al comparar estos resultados con los obtenidos por MORATA y ALMARZA (2002) para la serie de precipitación diaria de Badajoz, que cuenta con un registro de 122 años (1880-2001), se tiene también una sobrestimación mayor de la precipitación esperada por los modelos en el caso de Santa Cruz que en el caso de Badajoz (no mostrado). Estas sobrestimaciones aumentan notablemente al disminuir el número de datos con los que se ha parametrizado el modelo, de manera que se tendrá más certidumbre al bajar el umbral de precipitación para crear el modelo

Por otro lado se comparan los modelos calculados con las distribuciones exponencial y de Pareto, para la serie de precipitación diaria, con el obtenido de la ley General de Valores Extremos (L.G.V.E.) para la serie de máximos anuales de precipitación caída en un solo día (figura 4). Los modelos exponencial y de Pareto son los generados para el umbral de 40 mm fijado el período de retorno. Se puede observar que ambos modelos prácticamente coinciden en sus predicciones, mientras que con la L.G.V.E estas quedan infraestimadas. Por otro lado se han fijado los valores de precipitación esperados, que se han hecho coincidir con los umbrales seleccionados para el cálculo de los modelos, y se han calculado los períodos de retorno estimados para cada caso, los resultados se muestran en la figura 4. Se puede ver como, de nuevo, los modelos exponen-

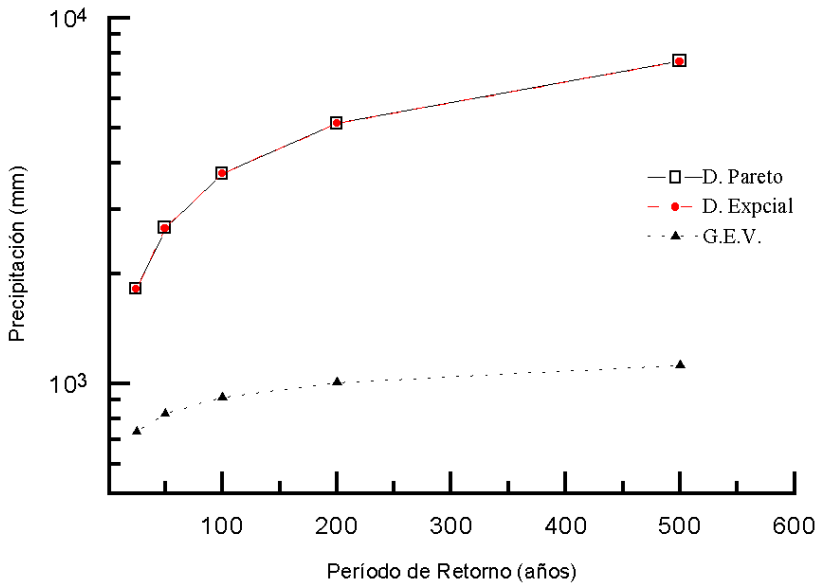


Figura 4: Comparación de los modelos generados con las distribuciones de Pareto, exponencial y la L.G.V.E.

cial y de Pareto prácticamente coinciden y crecen de manera paralela, manteniendo la tendencia, a los períodos de retorno obtenidos empíricamente para cada uno de los umbrales escogidos. En la figura 4 se puede observar que la L.G.V.E. estima mejor los valores esperados para períodos de retorno inferiores a dos años, mientras que para escalas temporales superiores sobreestima los valores alejándose de los valores empíricos y de las estimaciones de los modelos exponencial y de Pareto (la representación de esta figura se ha realizado fijando los umbrales y ajustando el período de retorno).

Los valores cuantitativos de la figura 5 se presentan en la tabla 1, en la que para cada valor umbral de precipitación fijado se tienen los períodos de retorno obtenidos. Los modelos estiman bastante bien los tiempos de recurrencia al comparar los resultados.

5. CONCLUSIONES

La serie de precipitación de Santa Cruz de Tenerife no presenta oscilaciones ni tendencias de tipo determinista, por lo que la serie de valores extremos anuales se puede considerar como un conjunto de sucesos independientes que se distribuye con arreglo a las leyes propuestas. La serie de

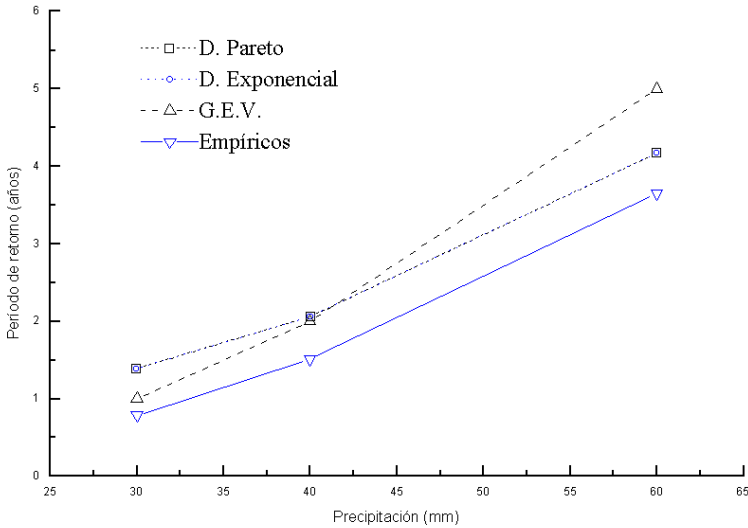


Figura 5: Comparación de los modelos con los resultados empíricos.

Santa Cruz de Tenerife analizada prácticamente garantiza la independencia de los sucesos debido a la longitud del registro estudiado, de 78 años. La ventaja que presenta la distribución de Pareto frente a la L.G.V.E. es que para valores de precipitación inferiores a 50 mm el modelo de Pareto resulta más realista que el modelo clásico. Este resultado cobra gran importancia en el caso de cuencas pequeñas y de gran pendiente ya que la respuesta de los sistemas hidrológicos puede tener graves consecuencias, como es el caso de avenidas. La elección de un umbral puede ser un factor determinante en la obtención de resultados para la estimación de posibles sucesos extremos.

Tabla 1: Períodos de retorno obtenidos con los modelos para valores de precipitación fijados.

Precip. (mm)	T (años)			
	D. Pareto	D. Expcial.	L.G.V.E	D. Empírica
30	1.390	1.390	1	0.788
40	2.061	2.062	2	1.508
60	4.173	4.173	5	3.651

6. AGRADECIMIENTOS

Trabajo realizado gracias a una beca de formación de postgrados del Instituto Nacional de Meteorología

7. REFERENCIAS

ALMARZA, C., LOPEZ, J. A. y FLORES C. (1996): *Homogeneidad y variabilidad de los registros históricos de precipitación de España*. Series Monográficas del Instituto Nacional de Meteorología (A-143). Ministerio de Medio de medio Ambiente.

ASHKAR, F. y OUARDA, T. B. M. J. (1996): On some methos of fitting the generalized Pareto distribution. *Journal of Hydrology*, 78, pp. 305-315.

BRADSON, B. B. y PALUTIKOV, J. P., (2000): Test of the generalized Pareto distribution for predicting extreme wind speeds. *Journal of Applied Meteorology*, 39(9), pp. 1627- 1640.

HOSKING, J. R. M. y WALLIS, J. R. (1987): Parameters and quantiles estimation for the generalized Pareto Distribution. *Technometrics*, 29, pp. 339-349.

MORATA, A., ALMARZA, C., (2002): Modelización de la probabilidad máxima de precipitación diaria en Badajoz mediante la distribución generalizada de Pareto. Comparación con los métodos clásicos. *Comunicaciones de la III Asamblea Hispano-Portuguesa de Geodesia y Geofísica*. (In Press).