

# DISTRIBUCIÓN DE LA SEQUÍA MÁS SEVERA EN UN INTERVALO DE TIEMPO DADO

Jesús ABAURREA y Ana Carmen CEBRIÁN

*Dpto. Métodos Estadísticos, Universidad de Zaragoza. E-mail: [acebrian@posta.unizar.es](mailto:acebrian@posta.unizar.es)*

## RESUMEN

El objetivo de este trabajo es caracterizar la máxima sequía que cabe esperar en un determinado periodo de tiempo. Para ello es necesario disponer de un modelo estocástico que describa el proceso de sequías (proponemos un proceso Poisson cluster para describir la ocurrencia y tres series de variables aleatorias, Longitud, Déficit e Intensidad Máxima, para describir la severidad) y desarrollar los resultados teóricos necesarios sobre la distribución del máximo en una muestra de tamaño Poisson.

**Palabras clave:** Análisis de sequías, proceso Poisson cluster, máximo en muestras de tamaño Poisson, valores extremos.

## ABSTRACT

*This work aims to characterize the largest drought event to occur in a given period of time. A Poisson cluster process is used to model drought occurrence and three series of random variables (Length, Deficit and Maximum Intensity) to describe their severity. Some theoretical results on the distribution of the maximum in a random Poisson size sample are developed for describing the largest drought events.*

**Key words:** Drought analysis, Poisson cluster process, maximum in random size samples, extreme values.

## 1. INTRODUCCIÓN

La sequía es un fenómeno que afecta de forma recurrente a muchas zonas, en particular a la mayor parte del este y sur de España. La descripción de las características de las sequías mediante distribuciones de probabilidad proporciona medidas que pueden ser de gran ayuda e interés en la gestión de los recursos hídricos. En particular, disponer de información sobre la distribución de la máxima sequía esperada en un determinado periodo de tiempo sería de gran utilidad para el diseño de la capacidad de los pantanos que requiere una región y, en general, para cualquier decisión relativa al máximo riesgo al que está sometido una zona. Conociendo la distribución de las características de la máxima sequía esperada se pueden obtener estimaciones no sólo de los valores medios sino también de percentiles altos de la distribución, que representan los valores críticos extremos.

El objetivo de este trabajo es proporcionar una metodología que permita caracterizar la distribución del máximo suceso de sequía en un periodo de tiempo dado. Para calcular esta distribución no se puede hacer uso de los métodos usuales de estimación debido a la escasez de datos disponibles; en

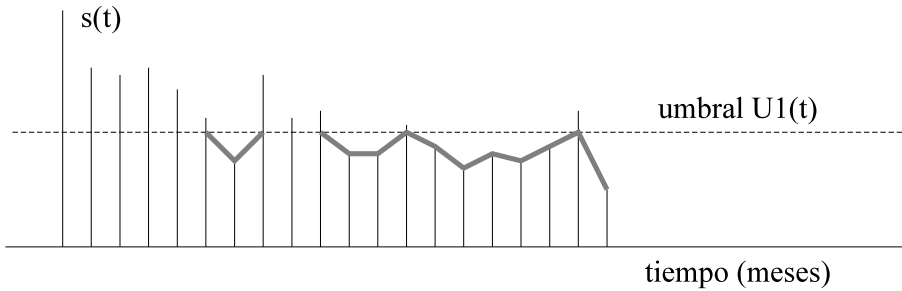


Figura 1: Definición de periodo seco basada en la aproximación EOT.

efecto, la distribución del máximo  $M$  de una variable  $X$  en un periodo de cien años, por ejemplo, no se puede estimar a partir de una muestra de valores de  $M$ , ya que es muy poco frecuente disponer de registros suficientemente largos. Por esta razón es necesario recurrir a resultados de la teoría de valores extremos que permitan estimar esta distribución a partir de una muestra de valores de  $X$ .

Para caracterizar la sequía más severa es necesario disponer de un modelo estadístico que describa tanto la ocurrencia como la severidad del proceso de las sequías. La descripción de este modelo se presenta en las secciones 1 y 2 mientras que el desarrollo de los resultados teóricos se realiza en la sección 3. Finalmente en la sección 4 se presenta la aplicación de la metodología propuesta en el observatorio de Huesca.

## 2. DEFINICIÓN DE SEQUÍA

En general la sequía meteorológica se puede definir como la escasez de precipitación en una región durante un largo periodo de tiempo. En la práctica, sin embargo, la definición de sequía debe reflejar las diferencias causadas por el tipo de clima, las características y las necesidades regionales, por lo que no se puede establecer una única definición válida en todas las circunstancias. Por otra parte, el análisis del proceso de sequías requiere una definición de carácter operativo que permita identificar el inicio, el final y el grado de severidad de la sequía.

Definimos sequía como un cluster o agrupación de periodos secos próximos y dependientes; aplicando la aproximación EOT (*Excess Over Threshold*), diremos que se produce un periodo seco cuando una señal  $s(t)$ , relacionada con la precipitación u otra variable que describa el estado hídrico del sistema, toma valores inferiores a un umbral  $u_1(t)$  que representa un valor crítico para el proceso (véase figura 1).

### 2.1. Señal

Dado que la sequía es un fenómeno que requiere un cierto periodo de tiempo para hacer notar sus efectos, utilizamos como señal  $s(t)$  una serie mensual en la que cada observación es la lluvia acumulada en los  $p$  meses anteriores. El uso de diferentes valores de  $p$  permite caracterizar

distintos tipos de sequías. En este trabajo consideramos  $p = 12$  y, en consecuencia, la serie reflejará deficiencias de agua en procesos basados en la precipitación a largo plazo, como el nivel de los pantanos. El análisis de la sequía en otros campos, por ejemplo la agricultura, requeriría periodos de acumulación más cortos. Sea cual sea el valor de  $p$  utilizado, la actualización mensual de la serie permite una evaluación frecuente del estado de sequía.

## 2.2. Umbral

Hemos considerado como umbral  $u_1(t)$  de sequía el percentil 10 de la serie, denotado  $p_{10}$ , que de acuerdo con el criterio climático de GIBBS y MAHER (1948) define periodos extremadamente secos.

## 2.3. Formación de los clusters

Uno de los problemas que presenta esta definición de sequía es determinar la composición de los clusters, es decir cómo decidir cuando dos periodos secos son dependientes. DAVISON y SMITH (1990) sugieren el uso de modelos paramétricos como los de Neyman-Scott y Bartlett-Lewis pero concluyen que en general, estos métodos no mejoran los resultados obtenidos aplicando reglas de carácter empírico. Después de ensayar varios criterios, aplicamos una regla basada en las de MADSEN y ROSBJERG (1998) y RASMUSSEN *et al.* (1994), que incorpora información sobre la magnitud de los periodos no secos de separación: Consideramos que dos periodos pertenecen al mismo cluster si el tiempo entre su ocurrencia es inferior a 7 meses y la intensidad en ese periodo no supera un umbral  $u_2(t)$  que representa un valor normal de precipitación; en las figuras 2 y 3 se pueden ver ejemplos de esta definición de cluster. Aplicando de nuevo el criterio de Gibbs y Maher utilizamos como umbral  $u_2(t)$  el percentil 30 de la serie.

## 2.4. Severidad

Para caracterizar la severidad de la sequía hemos utilizado tres variables, la duración o longitud  $L$ , la máxima intensidad observada en el periodo  $IM$ , y el déficit al umbral  $u_2(t)$ ,  $D$  (véase figura 2). Esta definición del déficit respecto a un umbral normal de precipitación presenta dos ventajas:

- Al incluir más información sobre la magnitud de los periodos de separación entre los periodos secos que forman la sequía, se representa mejor la severidad global que con otras definiciones, como la de ZELENHASIC y SALVAI (1987).
- El déficit no puede tomar valores negativos como sucede con la definición propuesta por MADSEN y ROSBERG (1998).

## 3. MODELO PARA EL PROCESO DE SEQUÍAS: PROCESO POISSON CLUSTER COMPUUESTO

Una vez que se ha establecido la definición de sequía es necesario encontrar un modelo estocástico que permita describir todo el proceso, tanto la ocurrencia como la severidad, de forma adecuada.

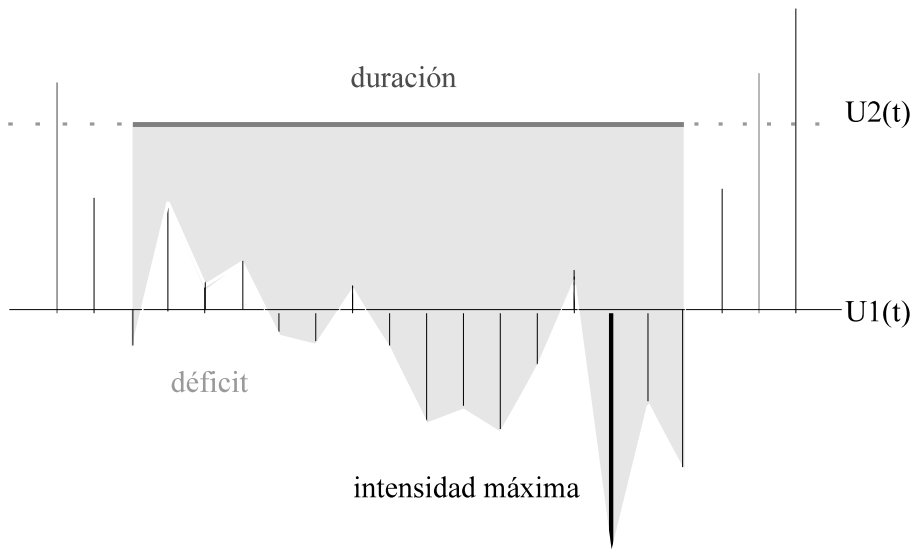


Figura 2: Variables utilizadas para caracterizar la severidad de la sequía.

### 3.1. Ocurrencia

Dado que las sequías se han definido como agrupaciones de periodos secos, un modelo adecuado para describir su ocurrencia es el proceso Poisson cluster. En este modelo los clusters, las sequías, ocurren de acuerdo a un proceso Poisson (es decir los tiempos entre dos ocurrencias consecutivas tienen una distribución exponencial) y cada uno de ellos está formado por un número aleatorio de puntos, los periodos secos. La estructura del modelo está representada en la figura 3.

### 3.2. Severidad

Para caracterizar la severidad, el modelo se completa con tres series de variables aleatorias independientes e idénticamente distribuidas asociadas a los clusters: la longitud  $L$ , el déficit  $D$  y la intensidad máxima  $IM$ . Para cada una de estas tres variables se ajusta una distribución de probabilidad adecuada utilizando un procedimiento de estimación máximo verosímil.

La justificación de la utilización de este modelo para representar el proceso de sequías se basa en la teoría de Valores Extremos. Un desarrollo exhaustivo de los principales resultados generales de misma se puede encontrar en EMBRECHS *et al.* (1997) y la justificación de este modelo en CEBRIÁN (1999). La validez del mismo se comprobó aplicándolo a seis series españolas de precipitación: Burgos, Daroca, Madrid, Murcia y San Fernando, con resultados satisfactorios, CEBRIÁN (1999).

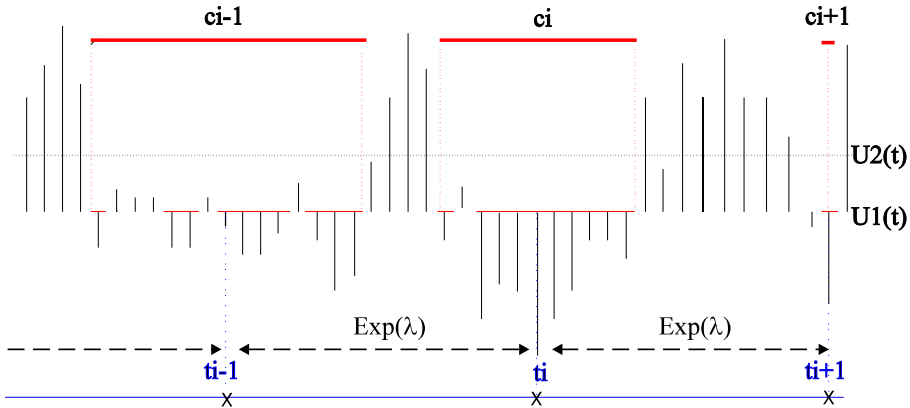


Figura 3: Estructura del modelo de ocurrencia de sequías (proceso Poisson cluster).

De acuerdo con este modelo, tenemos la siguiente situación:

- Dado que el proceso de sequías es un proceso Poisson, el número de sequías en un periodo de tiempo de longitud  $n$  es una variable aleatoria con distribución Poisson de parámetro  $\lambda(n) = \lambda n$ .
- Cada suceso, que corresponde a una sequía, tiene asociadas tres variables,  $L$ ,  $D$  e  $IM$ , con distribuciones  $F_L$ ,  $F_D$ ,  $F_{IM}$ .

En esta situación, la caracterización de la sequía más grave en un intervalo de tiempo dado, más exactamente la de la máxima duración, el máximo déficit o la máxima intensidad, se puede obtener a partir de la distribución del máximo de una variable con distribución  $F$  en una muestra de tamaño aleatorio  $N$ , con  $N$  Poisson( $\lambda$ ).

#### 4. DISTRIBUCIÓN DEL MÁXIMO DE UNA VARIABLE EN UNA MUESTRA DE TAMAÑO POISSON

En este apartado se presentan los resultados necesarios para la caracterización del máximo de las distribuciones más frecuentes en el análisis de datos climáticos y medioambientales: Exponencial, Pareto generalizada, Gamma, Lognormal, Weibull, etc. Los detalles de esos resultados se pueden consultar en CEBRIÁN (1999) y en ABAURREA y CEBRIÁN (2002).

##### 4.1. Distribuciones exactas

Dada una muestra de una variable  $X$  con función de distribución  $F$  y tamaño Poisson ( $\lambda$ ), la función de distribución del máximo,  $F_M$ , se puede expresar en términos de la distribución de la muestra y

el parámetro  $\lambda$ , de la siguiente forma:

$$F_M(x) = e^{-\lambda[1-F(x)]}$$

Esta expresión permite obtener la distribución exacta del máximo para algunas distribuciones. En particular se puede probar que,

- Si la distribución  $F$  de la muestra es Exponencial( $\alpha$ ), el máximo tiene una distribución Gumbel( $1/\alpha, \ln(\lambda)/\alpha$ ).
- Si la distribución  $F$  de la muestra es Pareto Generalizada, PG( $\gamma, \sigma$ ), el máximo tiene una distribución de Valor Extremo, VE( $\gamma, \sigma\lambda^\gamma, \sigma(\lambda^\gamma - 1)/\gamma$ ) con el mismo parámetro de forma  $\gamma$ .

#### 4.2. Distribuciones aproximadas

Desafortunadamente, la aplicación a otras distribuciones de la expresión anterior para  $F_M$  proporciona expresiones demasiado complicadas por lo que es preferible hacer uso de resultados aproximados. En condiciones bastante generales, se puede probar el siguiente resultado:

- El máximo, convenientemente normalizado, en una muestra con distribución Gamma, Log-normal o Weibull y tamaño Poisson tiene una distribución asintótica Gumbel( $1, \ln(\lambda)$ ).

Este resultado es válido para más distribuciones, en particular para cualquier distribución que pertenezca al máximo dominio de atracción de la distribución Gumbel.

A pesar de su generalidad, este resultado no es muy útil debido a su carácter asintótico y al hecho de que es difícil disponer de muestras suficientemente grandes. Por esta razón es preferible utilizar una nueva aproximación, denominada *penúltima aproximación*, que proporciona mejores resultados en la práctica.

#### 4.3. Penúltima aproximación de la distribución del máximo

GOMES (1984) y CASTILLO (1987) estudiaron el comportamiento de la distribución del máximo en muestras de tamaño no aleatorio; con distribuciones verificando condiciones muy generales establecieron que:

- La convergencia del máximo a su distribución límite es muy lenta.
- Una distribución VE( $\gamma_n, \mu_n, \sigma_n$ ) con sus tres parámetros dependiendo de la muestra, distribución que denominaron penúltima aproximación, proporciona una mejor aproximación que la distribución límite VE( $\gamma, \mu_n, \sigma_n$ ) con parámetro de forma constante.

Dado el interés práctico de este resultado, hemos desarrollado resultados similares para el caso de muestras de tamaño aleatorio Poisson.

- Dada una muestra con distribución Gamma, Lognormal o Weibull y tamaño Poisson, la penúltima aproximación del máximo es  $VE(\gamma_n, \mu_n, \sigma_n)$ , cuyos tres parámetros se pueden estimar a partir de la muestra.

La estimación de los tres parámetros se puede realizar utilizando, por ejemplo, el método de los percentiles.

## 5. RESULTADOS

Como ejemplo de la aplicación de los resultados anteriores a la caracterización de la máxima sequía esperada en un periodo de tiempo dado, presentamos los resultados correspondientes a la serie de Huesca, observatorio localizado en el noreste de España.

La validación del modelo en predicciones a largo plazo utilizando datos independientes es difícil ya que no hay disponibles series con un registro suficientemente largo. Por esta razón sólo se han podido realizar algunas comparaciones entre las predicciones del máximo en  $n$  años basadas en el modelo y los valores observados en el registro en periodos de  $n=50$  y  $n=100$  años. En particular comparamos la mediana y la media muestrales con los valores paramétricos correspondientes.

### 5.1. Duración

Respecto a la variable duración, la distribución que proporcionó un mejor ajuste fue la distribución Exponencial. En consecuencia, aplicando el resultado exacto visto en 4.1 se obtiene que la distribución del máximo es Gumbel con parámetros que dependerán de la longitud del periodo de tiempo considerado. Los parámetros estimados correspondientes a periodos de 50, 100 y 500 años se muestran en la tabla 1 junto a la media y la mediana de la distribución, así como los correspondientes valores empíricos, cuando estos están disponibles. De acuerdo con el modelo, la duración esperada de la sequía más larga en un periodo de 100 años es de 26 meses y la mediana casi 20 mientras que los correspondientes valores empíricos son iguales a 26 meses. En un periodo de 50 años, la predicción del modelo respecto a la media es de 21.5 meses y el valor observado en el registro 24.

### 5.2. Déficit

Para ajustar el déficit de una sequía se seleccionó la distribución Lognormal; en consecuencia, aplicando los resultados sobre la penúltima aproximación, la distribución del máximo será VE con los correspondientes parámetros que se presentan en la tabla 2. Utilizando esta distribución se estima que el máximo déficit esperado en 100 años es  $4911 \text{ l/m}^2$  y la mediana  $3876$ , mientras que los valores empíricos son iguales a  $3829 \text{ l/m}^2$ . La media y la mediana del máximo déficit en un periodo de 50 años son  $3590$  y  $2712 \text{ l/m}^2$  respectivamente y la media empírica calculada para ese mismo periodo es  $3683 \text{ l/m}^2$ .

Tabla 1: Duración: parámetros estimados para 50, 100 y 500 años, media y mediana (de la distribución y de los valores empíricos).

Duración	50 años	100 años	500 años
$\lambda$	11.4	22.8	114.0
$\mu$	17.6	22.3	33.2
$\sigma$	6.8	6.8	6.8
$E(M)$ (meses)	21.5	26.2	37.2
Mediana ( $M$ )	15.1	19.8	30.7
Media empírica	23.9	26.0	–
Mediana empírica	26	26	–

Tabla 2: Déficit: parámetros estimados para 50, 100 y 500 años, media y mediana (de la distribución y de los valores empíricos).

Déficit	50 años	100 años	500 años
$\lambda$	11.4	22.8	114.0
$\gamma$	0.33	0.31	0.27
$\mu$	21.978	32.304	66.989
$\sigma$	13.191	16.644	26.989
$E(M)$ ( $l/m^2$ )	3.590	4.911	9.227
Mediana ( $M$ )	2.712	3.876	7.739
Media empírica	3.683	3.829	–
Mediana empírica	3.829	3.829	–



Tabla 3: Intensidad Máxima: parámetros estimados para 50, 100 y 500 años, media y mediana (de la distribución y de los valores empíricos).

Intensidad máxima	50 años	100 años	500 años
$\lambda$	11.4	22.8	114.0
$\gamma$	-0.5	-0.5	-0.5
$\mu$	1.437	1.614	1.850
$\sigma$	302	214	96
$E(M)$ ( $l/m^2$ )	151	166	187
Mediana ( $M$ )	154	169	188
Media empírica	163	178	–
Mediana empírica	155	181	–

### 5.3. Intensidad Máxima

Finalmente la distribución que mejor ajustaba la intensidad máxima fue la distribución Pareto Generalizada; en consecuencia, el máximo tiene una distribución VE con los parámetros que se indican en la tabla 3. La media y la mediana de la intensidad máxima que cabe esperar en 100 años de acuerdo con el modelo estimado son 166 y 169  $l/m^2$  respectivamente y los correspondientes valores calculados a partir del registro 178 y 181  $l/m^2$ . Considerando un periodo de 50 años, las predicciones del modelo son ligeramente mayores que 150  $l/m^2$  y los valores empíricos se encuentran en torno a 160  $l/m^2$ .

## 6. CONCLUSIONES

En resumen, los resultados obtenidos llevan a las siguientes conclusiones:

- Dado un proceso aleatorio Poisson, la metodología propuesta permite caracterizar el máximo suceso, en particular la máxima sequía, que cabe esperar en un periodo de tiempo dado.
- Las distribuciones de probabilidad de las características del máximo suceso calculadas tanto a partir de los resultados teóricos exactos como de la penúltima aproximación proporcionan predicciones acordes con lo observado en el registro.
- La aplicabilidad de esta metodología es bastante general ya que se han presentado resultados para algunas de las distribuciones más frecuentes en las ciencias medioambientales.

## 7. REFERENCIAS

- ABAURREA, J. y CEBRIÁN, A.C. (2002): Drought Analysis based on a cluster Poisson model: distribution of the most severe drought. *Climate Research* (preprint).
- CASTILLO, E. (1987): *Extreme value theory in Engineering*. Academic Press.
- CEBRIÁN, A.C. (1999): *Análisis, modelización y predicción de episodios de sequía*. Tesis doctoral, Universidad de Zaragoza.
- DAVISON, A.C. y SMITH, R. L. (1990): Models for exceedances over high thresholds. *J. R. Statist. Soc. (Serie B, 52)*, pp. 393-442.
- EMBRECHTS, P., KLUPPELBERG, C. y MIKOSCH, T. (1997): *Modelling extremal events*. Springer-Verlag.
- GIBBS, W.J. y MAHER, J.V. (1967): Rainfall deciles as drought indicators. *Bureau of Meteorology Bulletin*, Melbourne. Australia, (48).
- GOMES, M. I. (1984): Penultimate limiting forms in extreme value theory. *Ann. Inst. Statist. Math.* (36), pp. 71-85.
- MADSEN, H. y ROSBJERG, D. (1998): A regional Bayesian method for estimation of extreme streamflow droughts. En PARENT, E., BOBÉE, B., HUBERT, P. y MIQUEL, J. (Eds.): *Studies and reports in Hydrology*. Unesco, pp. 327-340.
- RASMUSSEN, P. F., ASHKAR, F., ROSBJERG, D. y BOBÉE, B. (1994): The POT method for flood estimation: a review. En HIPEL, K.W., (Eds.): *Extreme values: floods and droughts*. Kluwer Academic Publishers, pp. 15-26.
- ZELENHASIC, E. y SALVAI, A. (1987): A method of streamflow drought analysis. *Water Resour. Res.* (23, 1), pp. 156-168.