

# UTILIZACIÓN DE LA FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD TCEV CON INFORMACIÓN NO SISTEMÁTICA DENTRO DE UN ANÁLISIS REGIONAL. APLICACIÓN A LOS RÍOS JÚCAR Y TURIA

Félix Francés García<sup>1</sup>

**RESUMEN:** Debido a sus implicaciones sociales y económicas, en el análisis de la frecuencia de las avenidas es exigible la mayor precisión posible. Para ello debe seleccionarse el modelo estadístico más adecuado y utilizarse la máxima información disponible. Las avenidas de los ríos mediterráneos se originan mediante dos mecanismos distintos, lo cual exige el empleo de una función de distribución no tradicional como es la TCEV. El incremento de información puede obtenerse mediante la utilización de información no sistemática adicional, o un análisis regional, o ambos (como en el caso del ejemplo presentado). A través del concepto de ganancia estadística se demuestra que el empleo de información no sistemática adicional puede suponer fácilmente disminuciones de más del 50% en el error de estimación de un cuantil. En un análisis regional, el beneficio de la información no sistemática en una estación de aforos, se transmite al resto de estaciones con sólo una pequeña disminución respecto de la ganancia estadística de un análisis local equivalente.

## INTRODUCCIÓN

Uno de los objetivos más habituales de los estudios hidrológicos es la realización del análisis de la frecuencia de las avenidas en un punto de un río. Este análisis es utilizado en el diseño de obras hidráulicas (aliviaderos de presas, túneles de desvío, encauzamientos, etc.), sistemas de drenaje transversal de infraestructuras lineales (puentes y alcantarillas), zonificación de riesgos de inundación, etc. El análisis de la frecuencia de las avenidas consiste básicamente en obtener la relación entre los caudales de avenida (o cuantiles) y su probabilidad de no excedencia (o período de retorno, o riesgo), aunque en la mayor parte de los casos basta con obtener los caudales correspondientes a unos pocos períodos de retorno. Por ejemplo, si se fija el riesgo admitido en el funcionamiento del aliviadero de una presa, el valor del caudal de proyecto debe determinarse con la mayor precisión posible, ya que un error por defecto de este cuantil da lugar a un aumento del riesgo realmente asumido por parte de la seguridad de la presa y aguas abajo de la misma. En el otro extremo, si el error es por exceso, lo que se está provocando es un incremento innecesario del coste del proyecto.

Los ríos mediterráneos presentan avenidas muy peligrosas, pudiendo ser provocadas por lluvias tanto de tipo ciclónico como de tipo convectivo. Las lluvias más torrenciales de éste último tipo se presentan principalmente

durante el otoño, dando lugar a caudales varios órdenes de magnitud superiores a los caudales medios del río. Estas avenidas poco frecuentes, pero de gran magnitud, provocan daños incalculables en las poblaciones ribereñas. Los 3000 millones de dólares en daños producidos por las avenidas en España de los años 1982 y 1983 y las más de 1100 vidas humanas segadas en la cuenca mediterránea española en los últimos 30 años (Berga, 1991), no sólo justifican la inversión en medidas estructurales y no estructurales de defensa contra las avenidas, sino que dado el alto coste de dichas inversiones hacen además exigible la realización de un análisis cuidadoso del comportamiento de las avenidas del río.

El método más directo de estimar el riesgo de una avenida en un punto donde exista una estación de aforos, es el de ajustar a la serie de caudales instantáneos máximos anuales una determinada función de distribución de probabilidad. Desafortunadamente, esta metodología puede dar lugar a estimadores de los cuantiles altamente variables debido a:

1. La incertidumbre del modelo estadístico
2. Los errores en las series de datos (especialmente los correspondientes a las mayores avenidas)
3. La utilización de series mucho más cortas que los períodos de retorno que se desean obtener
4. El hecho de que las poblaciones de máximas cauda-

---

<sup>1</sup> Departamento de Ingeniería Hidráulica y Medio Ambiente. Universidad Politécnica de Valencia Apdo. 22012, 46071 Valencia. Tel.: 96 3877612 Fax: 96 3877618 e-mail: ffrances@hma.upv.es

Artículo publicado en *Ingeniería del Agua*. Vol.5 Num. 1 (marzo 1998), páginas 47-58, recibido el 5 de septiembre de 1997 y aceptado para su publicación el 18 de febrero de 1998. Pueden ser remitidas discusiones sobre el artículo hasta seis meses después de la publicación del mismo. En el caso de ser aceptadas, las discusiones serán publicadas conjuntamente con la respuesta de los autores en el primer número de la revista que aparezca una vez transcurrido el plazo indicado.

les anuales presentan coeficientes de variación y sesgo altos (lo que da lugar a una alta variabilidad de las propiedades estadísticas de las muestras observadas).

Para incrementar la fiabilidad de los estimadores de los cuantiles, en los últimos años se han realizado esfuerzos importantes en la búsqueda de la función de distribución más adecuada, el método de estimación más preciso y robusto, y en el incremento de la información utilizada. El incremento de la información puede realizarse mediante un análisis regional, la incorporación de información local no sistemática de tipo histórica o sobre paleoavenidas, o empleando ambas técnicas.

Es precisamente en este último aspecto donde se centra este artículo, es decir, en presentar un modelo estadístico que permita incorporar en un análisis regional de la frecuencia de las avenidas de ríos mediterráneos la información histórica y de paleoavenidas que pueda existir a nivel local. Además, se van a comprobar las mejoras que se producen en la estimación de los cuantiles frente a la utilización exclusiva de la información sistemática local, incluyendo por último un ejemplo de aplicación práctica.

### EL MODELO ESTADÍSTICO

#### La función de distribución Two Component Extreme Value

Las series de caudales máximos caudales anuales de muchos ríos en el mundo se caracterizan por lo que se ha dado en llamar el “efecto pata de perro” (Potter, 1958) y la “condición de separación” (Matalas *et al.*, 1975). Los ríos mediterráneos también presentan estas características debido fundamentalmente a la existencia de dos poblaciones de avenidas. Una parte de las avenidas, que denominaremos avenidas ordinarias, son originadas por lluvias frontales, habitualmente en invierno y primavera, son más frecuentes y de menor magnitud. Por otra parte, las avenidas que denominaremos extraordinarias, son las avenidas de mayor magnitud, menos frecuentes, y generadas por fenómenos de precipitación de tipo convectivo durante el verano y otoño.

Las funciones de distribución clásicas, como pueden ser la Gumbel o la LogPearson Tipo III, no son capaces de reproducir estos fenómenos, por lo que son inadecuadas, y como se verá en la aplicación práctica incluso son contraproducentes en el análisis de la frecuencia de las avenidas de ríos mediterráneos. Por contra, funciones de nueva generación como la distribución “Two Component Extreme Value” (en adelante TCEV) han sido desarrolladas teniendo en cuenta estas características estadísticas (Rossi *et al.*, 1984).

La forma más inmediata de obtener la expresión de una TCEV es asumir que las poblaciones de las avenidas máximas anuales ordinarias (X1) y extraordinarias (X2) derivan de poblaciones Gumbel independientes. Por tanto, la avenida máxima anual será la máxima de estas dos, siendo su función de distribución el producto de las funciones de distribución originales:

$$F_X(x) = \exp(-\lambda_1 e^{-\theta_1 x} - \lambda_2 e^{-\theta_2 x}) \quad (1)$$

donde  $\lambda_1$  y  $\theta_1$  son los parámetros de forma y escala correspondientes a las avenidas ordinarias, y  $\lambda_2$  y  $\theta_2$  los de las avenidas extraordinarias. La función de distribución Gumbel puede ser considerada como un caso particular, en el que los dos parámetros de escala son iguales. La función de densidad de probabilidad es la derivada de la ecuación (1):

$$f_X(x) = F_X(x) \psi(x) \quad (2)$$

donde:

$$\psi(x) = \lambda_1 \theta_1 e^{-\theta_1 x} + \lambda_2 \theta_2 e^{-\theta_2 x} \quad (3)$$

Se pueden obtener unos parámetros estandarizados (Beran *et al.*, 1986), definidos por las siguientes ecuaciones:

$$\theta = \frac{\theta_2}{\theta_1} \quad (4)$$

$$\lambda = \lambda_2 \lambda_1^{-\theta} \quad (5)$$

El parámetro estandarizado  $\theta$  representa la relación en orden de magnitud entre las avenidas ordinarias y extraordinarias. Por otra parte, la probabilidad de que la avenida máxima anual sea del tipo extraordinario resulta ser:

$$P_2 = P[X_1 < X_2] = \int_0^{\infty} \exp(-\lambda \xi^\theta - \xi) d\xi \quad (6)$$

que en el rango habitual de  $\theta$  se puede aproximar por  $\lambda$  (Francés, 1995). Es decir,  $\lambda$  representa aproximadamente la probabilidad de presentación de una avenida extraordinaria.

En cuanto a los momentos, mientras que el coeficiente de variación es función de  $\lambda, \lambda_2$  y  $\theta$  el coeficiente de sesgo sólo es función de los parámetros estandarizados (Francés, 1995), siendo la TCEV capaz de reproducir el efecto de separación (Beran *et al.*, 1986).

#### Análisis regional

El análisis regional de la frecuencia de las avenidas consiste en estandarizar previamente las series de caudales provenientes de diferentes estaciones de aforo, de tal forma que se puedan analizar todos los datos conjuntamente. Al incrementar el tamaño de la muestra cabe esperar que disminuya el error de estimación de los parámetros en la medida en que la región sea homogénea estadísticamente y la correlación espacial sea baja, aunque este segundo condicionante tiene una menor influencia (Hosking y Wallis, 1988). Los distintos modelos de análisis regional se diferencian en el tipo de estandarización, en la función de distribución asumida o en el método de estimación. Cunnane (1988) realiza una interesante recopilación de una buena parte de ellos.

Para la TCEV la estandarización habitualmente propuesta consiste en (Rossi *et al.*, 1984):

$$Z_{i,j} = \theta_{i,1} X_{i,j} - L\lambda_{i,1} \quad (7)$$

donde:  $X_{i,j}$  = serie original en la estación de aforos  $i$ , y  $\lambda_{i,1}$  y  $\theta_{i,1}$  = parámetros correspondientes a las avenidas ordinarias locales. Se puede demostrar que dicha estandarización equivale a que los parámetros estandarizados  $\lambda$  y  $\theta$  definidos por las ecuaciones (4) y (5) permanecen constantes dentro de una misma región.

Si se sustituye la ecuación (7) en la ecuación (1) resulta que los parámetros regionales  $\lambda'_1$  y  $\theta'_1$  tienen que valer la unidad. Arnell y Gabrielle (1988) relajan esta condición y obtienen los parámetros por un método iterativo, mientras que Ferrari (1994) y Ferrer (1996) emplean un método mixto en el que incluyen el concepto de la variable índice para la mejora de la estimación de los parámetros locales  $\lambda'_1$  y  $\theta'_1$ . Para no complicar en exceso la obtención de las expresiones de las varianzas asintóticas de los estimadores, la metodología propuesta en este trabajo para un análisis regional es la siguiente:

1. Para la estandarización de los datos locales según la ecuación (7), se ajusta una distribución Gumbel a las avenidas ordinarias en cada estación de aforos, obteniendo una estimación de los parámetros  $\lambda_{i,1}$  y  $\theta_{i,1}$ . Previamente es necesaria la selección de las avenidas ordinarias, proponiéndose el test del sesgo definido por Kottegoda (1984) de eliminación de "outliers".
2. Ajuste de una TCEV a los datos estandarizados obteniendo los 4 parámetros regionales  $\lambda'_1$ ,  $\theta'_1$ ,  $\lambda'_2$  y  $\theta'_2$ , sin forzar a que los correspondientes a las avenidas ordinarias sean unidad.

**Información no sistemática**

La información no sistemática es aquella información censurada correspondiente a un periodo anterior al registro sistemático de caudales en una estación de aforos. Según el modo de obtenerla se puede distinguir entre información histórica e información sobre paleoavenidas. Información histórica se puede encontrar en marcas en los edificios, fotografías, noticias recogidas en los periódicos y libros, comunicaciones verbales, etc., siendo empleada por primera vez por Benson (1950). Datos referentes a paleoavenidas nos llegan a través de evidencias en la vegetación del llano de inundación (como pueden ser anomalías en los anillos de los árboles) y de paleoniveles (como son los denominados depósitos de remanso o "slackwater deposits"). Las técnicas paleohidrológicas han sido descritas en detalle por Baker (1987) y Baker *et al.* (1988).

La información no sistemática se clasifica desde el punto de vista de su tratamiento estadístico en función del tipo de censura que la origina (Francés *et al.*, 1994). Los tres casos habituales se esquematizan en la Figura 1. Cuando existe un límite censor  $X_H$  conocido se trata de información censurada Tipo 1. Si la fuente de información es histórica éste suele ser el caso, ya que existe un umbral de percepción por debajo del cual las avenidas no tienen la suficiente magnitud como para ser recorda-

das. El valor de las avenidas no censuradas puede ser conocido o no. Como simplificación se seguirá la clasificación de Stedinger y Cohn (1986), denominándose información censurada propiamente dicha si las  $K$  avenidas que durante el periodo no sistemático de longitud  $M$  superaron el umbral de percepción histórico son conocidas (Figura 1b). Si se desconocen sus valores, la denominaremos binomial censurada (Figura 1a). Si no existe un límite censor, la información es censurada Tipo 2. En este caso se conocen las  $K$  mayores avenidas durante el periodo no sistemático, siendo  $K$  una variable determinística. Dado que la mayor paleoavenida suele borrar la información correspondiente al resto de grandes paleoavenidas,  $K$  suele adoptar el valor de 1, denominándose en este caso información de máxima avenida (Figura 1c).

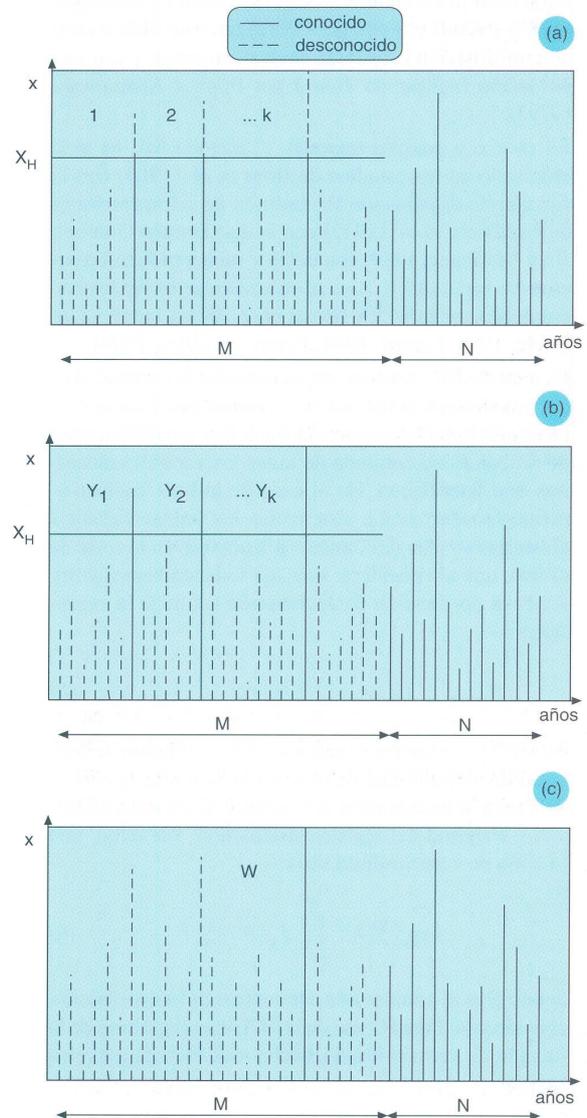


Figura 1. Tipos de información no sistemática: (a) binomial censurada, (b) censurada, y (c) máxima avenida

**Estimación de parámetros por Máxima Verosimilitud**

La elección del método de estimación de parámetros por Máxima Verosimilitud (en adelante ML) no sólo se debe a sus propiedades (como la existencia de una varianza asintótica y la facilidad de incorporación de cualquier tipo de información adicional cuantitativa), sino también a su bondad general frente a otros métodos de estimación, tanto con información adicional no sistemática como en un análisis regional.

El método ML junto con información adicional no sistemática ha sido empleado por numerosos investigadores: con una población Gumbel por Leese (1973), Hosking y Wallis (1986a), y Guo y Cunnane (1991); con la familia EV de dos parámetros por Francés *et al.* (1994); con la GEV por Phien y Fang (1989); con la lognormal por Condie y Lee (1982), Cohn y Stedinger (1987) y Kroll y Stedinger (1996) (en este último caso con información censurada exclusivamente); y con una población logPearson Tipo 3 por Pilon y Adamowski (1993).

En cuanto a análisis regional, el método ML ha sido empleado en los estudios de Boes *et al.* (1989) con la función de distribución Weibull; en las comparaciones de Landwehr *et al.* (1979) con la distribución Gumbel; Jin y Stedinger (1989) con la GEV incluyendo información no sistemática; y es el método mayoritariamente empleado con la TCEV (Rossi *et al.*, 1984; Arnell y Gabriele, 1988; Ferrari, 1994; Ferrer y Ardiles, 1994). El método ML consiste en seleccionar los parámetros que maximicen la función de verosimilitud  $L(\cdot)$ , que es cualquier función proporcional a la función de densidad de probabilidad conjunta de todas las variables aleatorias que intervienen. En el caso de utilizar exclusivamente la información sistemática las únicas variables aleatorias son las diferentes realizaciones en la serie de aforos, que al considerar que son independientes entre sí tienen por función de distribución conjunta la expresión:

$$f_{\underline{x}}(\underline{x}) = \prod_{i=1}^N f_X(x_i) \tag{8}$$

donde:  $\underline{x}$ ,  $x_i$  = registro sistemático,  $f_X(\cdot)$  = función de densidad de probabilidad de la variable aleatoria X, común para toda la muestra con la hipótesis de estacionaridad, y  $N$  = longitud del registro sistemático. Por tanto, una función de verosimilitud sería:

$$L_{SY}(\Theta) = \prod_{i=1}^N f_X(x_i; \Theta) \tag{9}$$

donde  $\Theta$  es el conjunto de parámetros de la distribución considerada. Para el caso de la TCEV, el logaritmo de la función de verosimilitud (mucho más fácil de maximizar) es:

$$LL_{SY}(\Theta) = -\lambda_1 \sum_{i=1}^N e^{-\theta_1 x_i} - \lambda_2 \sum_{i=1}^N e^{-\theta_2 x_i} + \sum_{i=1}^N L\Psi(x_i) \tag{10}$$

Si además se está utilizando información binomial censurada, el número  $K$  de avenidas por encima del umbral de percepción  $X_H$  sigue una distribución binomial, cuya función de densidad de probabilidad es:

$$f_K(k) = \binom{M}{k} (1-p)^{M-k} p^k \quad k = 0, 1, \dots, M \tag{11}$$

con lo que la función de distribución conjunta quedaría como:

$$f_{\underline{x},K}(\underline{x},k) = \binom{M}{k} (1-p)^{M-k} p^k \prod_{i=1}^N f_X(x_i) \tag{12}$$

donde:  $p$  = probabilidad de excedencia de  $X_H$  función de los parámetros del modelo,  $M$  = longitud del período no sistemático. Por tanto, el logaritmo de la función de verosimilitud particularizada para la TCEV tiene por expresión:

$$LL_{BC}(\Theta) = (M-k) (-\lambda_1 e^{-\theta_1 Q_H} - \lambda_2 e^{-\theta_2 Q_H}) + kLp + LL_{SY}(\Theta) \tag{13}$$

Para el caso de utilizar información adicional censurada, aparece una nueva variable aleatoria  $Y$ , que se corresponde con las avenidas que superan el umbral de percepción. Se puede demostrar que:

$$f_Y(y) = \frac{f_X(y)}{p} \tag{14}$$

y la función de distribución conjunta es ahora:

$$f_{\underline{x},\underline{y},K}(\underline{x},\underline{y},k) = \binom{M}{k} (1-p)^{M-k} p^k \prod_{i=1}^N f_X(x_i) \prod_{j=1}^k f_Y(y_j) \tag{15}$$

donde:  $y_j$  = avenidas históricas,  $f_Y(\cdot)$  = función de densidad de probabilidad de  $Y$ , de tal forma que la función  $LL$  a emplear, después de sustituir la ecuación (14) en la ecuación (15), es:

$$LL_{CE}(\Theta) = (M-k) (-\lambda_1 e^{-\theta_1 Q_H} - \lambda_2 e^{-\theta_2 Q_H}) - \lambda_1 \sum_{j=1}^k e^{-\theta_1 y_j} - \lambda_2 \sum_{j=1}^k e^{-\theta_2 y_j} + \sum_{j=1}^k L\Psi(y_j) + LL_{SY}(\Theta) \tag{16}$$

Si la única información disponible durante el período no sistemático es la máxima avenida  $W$ , su función de densidad de probabilidad es:

$$f_W(w) = Mf_X(w)F_X^{M-1}(w) \tag{17}$$

Por lo tanto, la distribución conjunta resulta ser para este nivel de información:

$$f_{\underline{x},W}(\underline{x},w) = \prod_{i=1}^N f_X(x_i; \Theta) f_W(w; \Theta) \tag{18}$$

Que operando da lugar a que la función  $LL$  para la TCEV con esta información sea:

$$LL_{MF}(\Theta) = -M\lambda_1 e^{-\theta_1 w} - M\lambda_2 e^{-\theta_2 w} + L\Psi(w) + LL_{SY}(\Theta) \tag{19}$$

En el caso de tener más de un tipo de información no sistemática (por ejemplo, al modificarse el umbral de percepción histórico o al incluir simultáneamente información histórica y sobre paleoavenidas en el mismo punto), sería inmediato obtener la función de verosimilitud por combinación de las anteriores.

Por otra parte, la única variación si se está realizando un análisis regional es que toda la información sistemática y no sistemática de cada estación de aforos tiene que estandarizarse de acuerdo con la ecuación (7).

La búsqueda del máximo de una función en un espacio de 4 dimensiones, puede ser una tarea compleja a poco complicada que sea la función y, desafortunadamente, éste es el caso. Como el algoritmo de búsqueda del máximo iba a ser utilizado decenas de miles de veces en la fase de simulación de esta investigación, era conveniente el que fuera rápido y a la vez efectivo sobre una superficie compleja, para una casuística casi infinita. Dados los problemas encontrados con el método iterativo propuesto por Rossi *et al.* (1984) y después de diversas pruebas, el algoritmo de maximización seleccionado fue el de Powell de dirección es conjugadas (Press *et al.*, 1989).

**EL ERROR DE ESTIMACIÓN DE LOS CUANTILES**

Una medida de la fiabilidad de la estimación de un cuantil  $X_T$  es el error cuadrático medio (MSE) del estimador, que se define como:

$$MSE(\hat{X}_T) = E[(\hat{X}_T - X_T)^2] \tag{20}$$

donde  $E[.]$  es el operador valor esperado. Este error es muy complejo de obtener analíticamente, por lo que se recurre o bien a simulaciones de tipo Monte Carlo para obtener el error cuadrático medio muestral, o bien a la obtención de su valor asintótico.

**Error cuadrático medio muestral**

Se define el error cuadrático medio muestral del estimador del cuantil, SMSE, como:

$$SMSE_i(\hat{X}_T) = \sum_{j=1}^S (\hat{X}_{Tij} - X_T)^2 \tag{21}$$

donde:  $X_T$  = verdadero valor del cuantil de período de retorno  $T$ ,  $\hat{X}_{Tij}$  = estimador del cuantil en la simulación  $j$  utilizando la información  $i$ ,  $S$  = número total de simulaciones.

Sin embargo, el error cuadrático medio definido por la ecuación (21) es a su vez una variable aleatoria. Se trata por tanto de una aproximación que tenderá al verdadero MSE cuando las simulaciones tiendan a infinito. Para obtenerlo es necesario la generación de series sintéticas en número muy elevado. En este trabajo se han simulado múltiples escenarios cada uno de ellos con 5000 simulaciones. El algoritmo empleado para al obtención de números aleatorios con distribución uniforme en el intervalo [0, 1] ha sido un generador congruencial lineal (Bartley *et al.*, 1987).

**Varianza asintótica**

Otra forma de resolver el problema es mediante el uso de la varianza asintótica. La varianza asintótica o límite inferior de Cramer-Rao (CRLB) es un límite inferior de la varianza de todos los estimadores insesgados. El método ML es asintóticamente insesgado, por lo que el CRLB es el límite inferior de su MSE cuando la información utilizada tiende a infinito.

La matriz de varianzas y covarianzas asintóticas de los estimadores de los parámetros de un modelo estadístico viene dada como la inversa de la matriz de información de Fisher (Kendall y Stuart, 1967). Los elementos de esta matriz se obtienen como:

$$I_i(j,k) = E \left[ - \frac{\partial^2 LL_i(\Theta)}{\partial \Theta_j \partial \Theta_k} \right] \tag{22}$$

donde:  $LL_i(.)$  = logaritmo de la función de verosimilitud,  $\Theta_j, \Theta_k$  = parámetros del modelo,  $i$  = tipo de información utilizada: sistemática, binomial censurada, censurada, de máxima avenida, o una combinación de las anteriores. Desafortunadamente para la función de distribución TCEV esta matriz es excesivamente compleja y no expresable en una forma compacta. En cualquier caso, las expresiones completas se encuentran en Francés (1995).

Por otra parte, la varianza asintótica de los cuantiles o límite inferior de Cramer-Rao (CRLB) viene dada por la ecuación:

$$CRLB_i(\hat{X}_T) = \underline{V} I_i^{-1} \underline{V}' \quad i = 0,1,2,3 \tag{23}$$

donde  $\underline{V}$  es el vector de las derivadas del cuantil respecto de los parámetros. Aunque  $X_T$  no puede ser obtenido explícitamente, si que se pueden obtener sus derivadas, que son:

$$\underline{V} = \left( \frac{\partial X_T}{\partial \lambda_1}, \frac{\partial X_T}{\partial \theta_1}, \frac{\partial X_T}{\partial \lambda_2}, \frac{\partial X_T}{\partial \theta_2} \right) = \left( \frac{e^{-\theta_1 X_T}}{\psi(X_T)}, \frac{\lambda_1 X_T e^{-\theta_1 X_T}}{\psi(X_T)}, \frac{e^{-\theta_2 X_T}}{\psi(X_T)}, \frac{\lambda_2 X_T e^{-\theta_2 X_T}}{\psi(X_T)} \right) \tag{24}$$

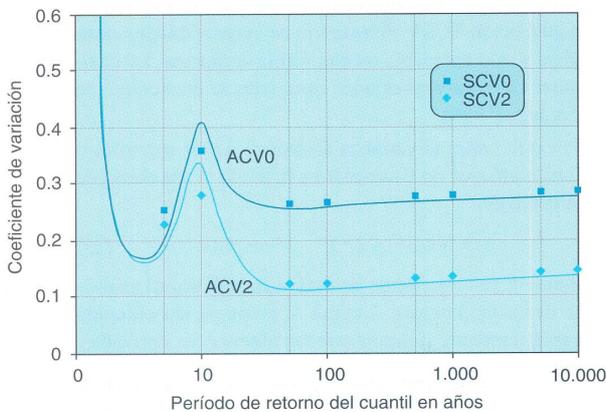
**Comparación de las varianzas asintóticas y muestrales de los cuantiles**

Debido a que el error cuadrático medio muestral SMSE requiere la realización de simulaciones por Monte Carlo, es preferible emplear la varianza asintótica CRLB como una aproximación del error cuadrático medio MSE. Si los resultados asintóticos y muestrales son similares, los primeros también lo serán de los verdaderos y el CRLB podría ser utilizado en lugar del MSE. Para realizar la comparación entre el CRLB y el SME se han representado gráficamente los coeficientes de variación asintótico y muestral del estimador del cuantil, cuyas expresiones respectivamente son:

$$ACVi(\hat{X}_T) = \frac{\sqrt{CRLBi(\hat{X}_T)}}{X_T} \quad (25)$$

$$SCVi(\hat{X}_T) = \frac{\sqrt{SMSEi(\hat{X}_T)}}{X_T} \quad (26)$$

En la *Figura 2* se comparan los coeficientes de variación muestral y asintótico para los casos de utilizar exclusivamente información sistemática y con información adicional de tipo censurada. El escenario representado, que se ha denominado escenario inicial, tiene un período sistemático  $N$  de 100 años de longitud, el período no sistemático es de 500 años, el período de retorno del umbral de percepción histórico  $H$  es de 50 años, y los parámetros de la función TCEV son  $\lambda_1=1$ ,  $\theta_1 = 1$ ,  $\lambda_2 = 0.2$  (que equivale a una  $p_2 = 0.18$ ) y  $\theta_2 = 0.1$ . A la vista de la figura, las diferencias entre el ACV y el SCV son muy pequeñas para cuantiles medios y altos, mientras que para cuantiles bajos la diferencia se incrementa.



**Figura 2.** Comparación en el escenario inicial del coeficiente de variación asintótico (ACV) y muestral (SCV) con información sistemática (0) y con información censurada adicional (2).

Resultados similares se obtuvieron con otros tipos de información adicional y un amplio rango de escenarios. Por tanto, se puede deducir que la varianza asintótica es una buena aproximación del error cuadrático medio. Sin embargo, este resultado es contradictorio con el obtenido por Phien y Fang (1989), que concluían que para la distribución GEV el CRLB se quedaba muy por debajo de la varianza muestral, recomendando el empleo de la matriz de información observada en vez de la de Fisher.

**GANANCIA ESTADÍSTICA ASINTÓTICA EMPLEANDO INFORMACIÓN ADICIONAL NO SISTEMÁTICA**

El concepto de ganancia estadística fue introducido por Francés et al. (1991), siendo una forma de medir el in-

cremento de fiabilidad en la estimación de un cuantil al utilizar algún tipo de información adicional al registro sistemático.

La ganancia estadística asintótica se define como:

$$ASGi = 1 - \frac{CRLB0}{CRLB1} \quad i=1,2,3 \quad (27)$$

donde el sufijo 0 indica que solo se utiliza la información sistemática y el sufijo i corresponde a los diferentes usos de información no sistemática adicional.

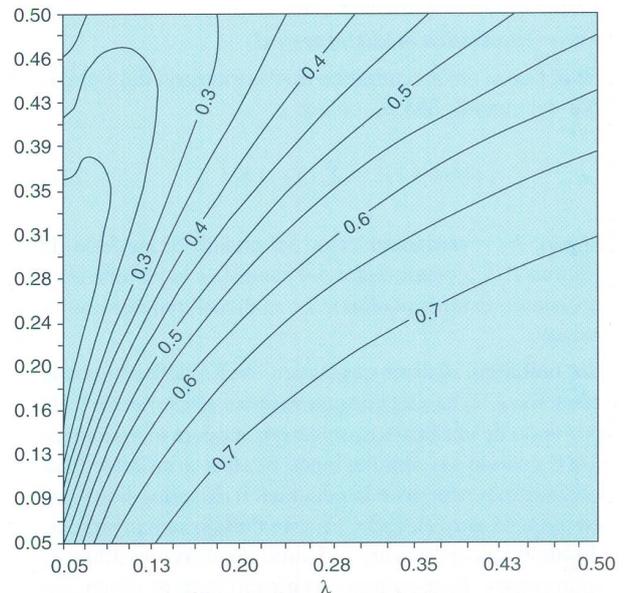
Para la TCEV no han podido obtenerse las expresiones analíticas, pero numéricamente se ha comprobado que las ganancias estadísticas asintóticas con información censurada *Tipo 1* son función exclusiva de:

1. la relación  $r$  entre las longitudes de los períodos no sistemático y sistemático;
2. el período de retorno  $H$  del umbral de percepción histórico;
3. el período de retomo  $T$  del cuantil considerado;
4. los parámetros estandarizados  $\lambda$  y  $\theta$ , definidos en las ecuaciones (4) y (5).

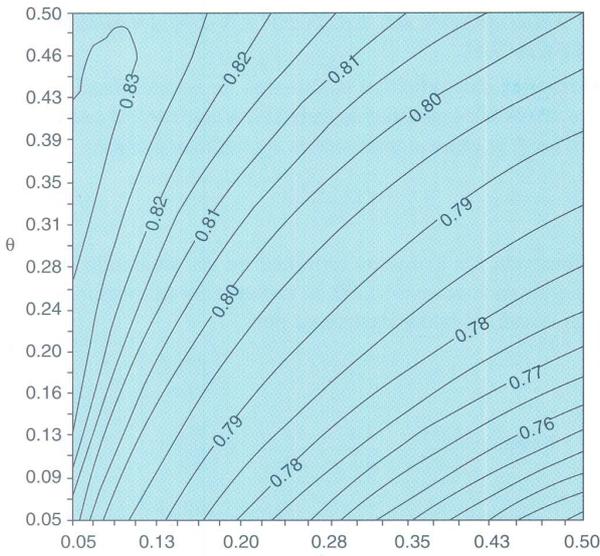
Mientras que la ganancia estadística asintótica utilizando información censurada *Tipo 2* es función exclusiva de:

1. la longitud  $N$  del registro sistemático;
2. la longitud  $M$  del período histórico;
3. el período de retomo  $T$  del cuantil considerado;
4. los parámetros estandarizados  $\lambda$  y  $\theta$ .

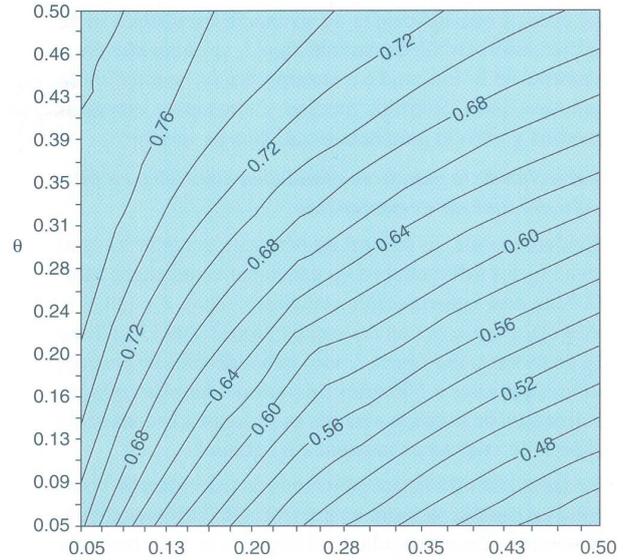
La influencia de estos factores en la ganancia estadística se refleja a continuación desde un punto cualitativo.



**Figura 3.** Ganancia estadística asintótica con información adicional binomial censurada ( $ASG1$ ) en función de los parámetros estandarizados, para el cuantil de 500 años y escenario inicial.



**Figura 4.** Ganancia estadística asintótica con información adicional censurada (ASG2) en función de los parámetros estandarizados, para el cuantil de 500 años y escenario inicial.



**Figura 5.** Ganancia estadística asintótica con información adicional de máxima avenida (ASG3) en función de los parámetros estandarizados, para el cuantil de 500 años y escenario inicial.

**Influencia de los parámetros estandarizados en la ganancia estadística asintótica**

Como se puede observar en la Figura 3, en el rango habitual de los parámetros estandarizados la ganancia estadística utilizando información adicional binomial censurada (ASG1) aumenta a medida que aumenta la proporción de avenidas extraordinarias, es decir, a medida que aumenta  $\lambda$ . Por contra, disminuye a medida que se parecen en magnitud las dos poblaciones de avenidas (incremento de  $\theta$ ).

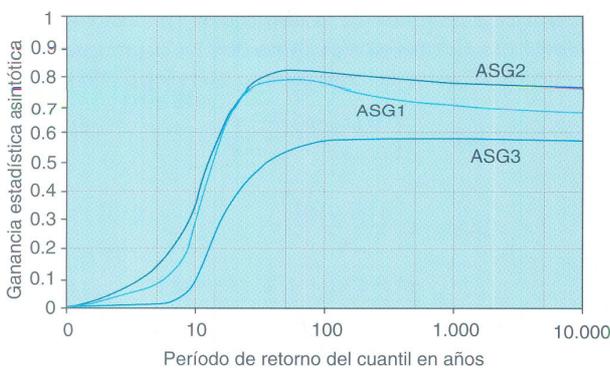
En lo que respecta a la ganancia estadística utilizando información adicional censurada (Figura 4), dado que las diferencias son mínimas se puede considerar a efectos prácticos que es independiente de los 4 parámetros. La Figura 5 representa la ganancia estadística asintótica utilizando la información adicional de máxima avenida. Dicha ganancia disminuye a medida que aumenta

la proporción de avenidas extraordinarias, siendo menor la influencia del parámetro estandarizado  $\theta$ .

**Variación de las ganancias estadísticas asintóticas en función del periodo de retorno del cuantil**

Esta variación queda reflejada para el escenario inicial en la Figura 6 en la que se puede observar que la ganancia estadística es en todos los casos mínima para los cuantiles de bajo periodo de retorno, siendo sólo apreciable para cuantiles medios y altos.

Con información censurada Tipo 1, la máxima ganancia asintótica se alcanza con ASG1 en el cuantil igual al umbral de percepción  $H$ , o ligeramente por encima en el caso de ASG2, decreciendo muy lentamente a partir de este valor máximo. Obsérvese la escasa diferencia entre ASG1 y ASG2 en el escenario inicial para todo el rango de cuantiles analizado.



**Figura 6.** Variación de la ganancia estadística asintótica en función del cuantil, para diferentes niveles de información adicional y escenario inicial



**Figura 7.** Variación de la ganancia estadística asintótica en función del umbral de percepción, para diferentes niveles de información adicional y escenario inicial

El cuantil que maximiza su ganancia estadística si se utiliza información censurada *Tipo 2*, se encuentra en el entorno de la longitud no sistemática *M*, aunque en la práctica, para cuantiles medios y altos para cuantiles medios y altos la ganancia estadística es similar.

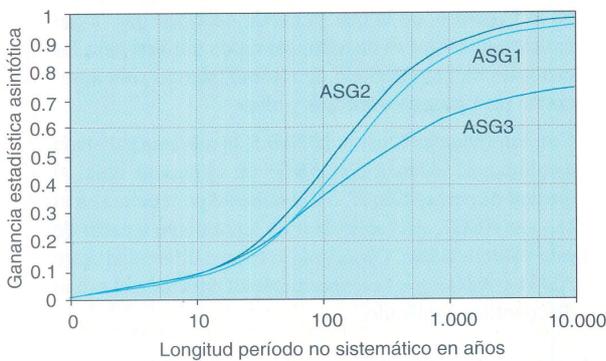
**Variación de la ganancia estadística asintótica en función del umbral de percepción**

Evidentemente, el umbral de percepción histórico solo es un factor influyente en las ganancias estadísticas asintóticas con información censurada *Tipo 1* (ASG1 y ASG2), y no lo es con censurada *Tipo 2* (ASG3). Como se observa en la *Figura 7*, sólo con umbrales de percepción superiores a 500 años de período de retorno la información de máxima avenida es superior a la censurada *Tipo 1*, lo cual no es la situación habitual.

La ganancia estadística utilizando la información binomial censurada siempre es inferior a la correspondiente a información censurada. La diferencia es muy pequeña para umbrales de percepción medios y altos en todos los casos, mientras que para umbrales de percepción bajos ASG1 es muy pequeña.

**Variación de la ganancia estadística asintótica en función de la relación de longitudes entre el período no sistemático y sistemático**

Las tres ganancias estadísticas aumentan con el ratio *r*, como queda reflejado en la *Figura 8*, aunque si *r* es inferior a 1 (en la figura *M* < 100 años), la información adicional comienza a tener escaso valor.



**Figura 8.** Variación de la ganancia estadística asintótica en función de la longitud del período no sistemático, para diferentes niveles de información adicional y escenario inicial.

Con información censurada *Tipo 1*, las ganancias estadísticas tienen un comportamiento asintótico, de tal forma que con información censurada ASG2 tiende a 1 a medida que crece *r*; es decir, si la longitud histórica es infinita la varianza de los estimadores de los cuantiles utilizando la información censurada es nula. Este comportamiento asintótico de ASG1 y ASG2 hace que estadísticamente no resulte rentable incrementar en exceso el período no sistemático de búsqueda de la información. Por contra, ASG3 crece con mayor lentitud, presentando incluso un valor máximo para *M* muy alto.

**UTILIZACIÓN DE LA TCEV EN UN ANÁLISIS REGIONAL**

Una vez obtenidos los parámetros de estandarización locales y los 4 parámetros de la TCEV regional, el cuantil local se obtiene mediante la siguiente expresión:

$$X_{i,T} = \frac{Y_{i,T} + L\lambda_{i,l}}{\theta_{i,l}} \tag{28}$$

Tomando los primeros términos de un desarrollo de Taylor en la ecuación (28), la varianza del estimador de un cuantil se puede aproximar por:

$$Var(X_{i,T}) \approx \left( \frac{\partial X_{i,T}}{\partial Y_{i,T}}, \frac{\partial X_{i,T}}{\partial \lambda_{i,l}}, \frac{\partial X_{i,T}}{\partial \theta_{i,l}} \right) S \begin{pmatrix} \frac{\partial X_{i,T}}{\partial Y_{i,T}} \\ \frac{\partial X_{i,T}}{\partial \lambda_{i,l}} \\ \frac{\partial X_{i,T}}{\partial \theta_{i,l}} \end{pmatrix} \tag{29}$$

Siendo *S* la matriz de varianzas y covarianzas de las variables estadísticas *Y<sub>i,T</sub>*, *λ<sub>i,l</sub>* y *θ<sub>i,l</sub>*. Si se aproximan estas varianzas y covarianzas por los CRLB correspondientes el resultado es:

$$\begin{aligned} Var(X_{i,T}) \approx & \frac{1}{\theta^2} CRLB(Y_{i,T}) + \\ & + \frac{1}{\lambda^2_{i,l} \theta^2_{i,l}} CRLB(\lambda_{i,l}) - \\ & - 2 \frac{Y_{i,T} + L\lambda_{i,l}}{\lambda_{i,l} \theta^3_{i,l}} CRLB(\lambda_{i,l}, \theta_{i,l}) + \\ & + \frac{(Y_{i,T} + L\lambda_{i,l})^2}{\theta^4_{i,l}} CRLB(\theta_{i,l}) \end{aligned} \tag{30}$$

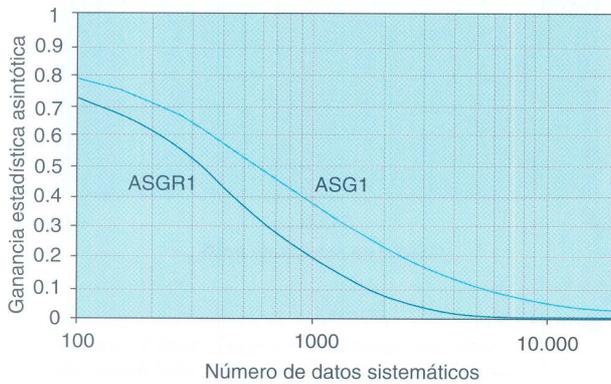
donde *CRLB(Y<sub>i,T</sub>)* es la varianza asintótica del cuantil de una distribución TCEV por aplicación de la ecuación (23). Por otra parte, *CRLB(λ<sub>i,l</sub>)*, *CRLB(θ<sub>i,l</sub>)* y *CRLB(λ<sub>i,l</sub>, θ<sub>i,l</sub>)* son las varianzas y covarianzas asintóticas de los parámetros de una distribución Gumbel, es decir, los elementos de la matriz inversa de la matriz de información de Fisher de la distribución Gumbel con información sistemática exclusivamente, que en este caso si pueden obtenerse de forma compacta (Francés, 1995):

$$I_0(1,1) = -E \left[ \frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda^2} \right] = \frac{N}{\lambda^2} \tag{31}$$

$$\begin{aligned} I_0(2,2) = -E \left[ \frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta^2} \right] = \\ = \frac{N [I + \Gamma''(2) + L^2 \lambda - 2\Gamma'(2) L \lambda]}{\theta^2} \end{aligned} \tag{32}$$

$$\begin{aligned} I_0(1,2) = I_0(2,1) = -E \left[ \frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta \partial \lambda} \right] \\ = \frac{N [\Gamma'(2) - L \lambda]}{\lambda \theta} \end{aligned} \tag{33}$$

A la vista de la ecuación (30) parece bastante evidente que cualquier mejora en la estimación de los cuantiles regionales repercute automáticamente en los cuantiles locales, con las mismas propiedades que si se realiza un análisis local con la misma longitud de años. La única diferencia estriba en que la ganancia estadística regional tiene que ser menor que su equivalente local (Figura 9) al existir en el primer caso dos parámetros más a estimar en cada estación de aforos.



**Figura 9.** Comparación de las ASG con análisis local (1) y regional (R1) en función del número total de datos sistemáticos, con información adicional binomial censurada y escenario inicial.

Como era de esperar, el efecto de la información adicional local disminuye con el aumento del número de estaciones de aforo consideradas en la región, análogamente al caso de análisis local en el que al incrementar la longitud sistemática disminuye la ganancia estadística. En la Figura 9 se puede observar que la ganancia estadística producida por una información adicional local de tipo binomial censurada se reduce por debajo del 10% sólo si se regionalizan más de 20 estaciones de 100 años de longitud cada una.

Estos resultados son similares a los obtenidos por Jin y Stedinger (1989), pero resultan contradictorios con el escepticismo que mantienen Hosking y Wallis (1986a y 1986b) sobre las ventajas del uso de información no sistemática dentro de un análisis regional. Sin embargo hay que tener en cuenta que Hosking y Wallis consideran como información adicional solamente la correspondiente a la máxima avenida (que es la que habitualmente da lugar a menores ganancias estadísticas, según se ha analizado en apartados anteriores), no considerando otras fuentes de información. Además de esto, añaden errores de medición exclusivamente a esa única avenida histórica o paleoavenida referida, cuando en la práctica los errores cometidos en la determinación de una gran avenida histórica y una gran avenida sistemática son del mismo orden de magnitud, y en cualquier caso, si los valores de la información no sistemática presentan errores apreciables pueden utilizarse en la forma binomial censurada.

### APLICACIÓN A LOS RÍOS JÚCAR Y TURIA

La razón de una aplicación a un caso práctico radica fundamentalmente en la comprobación de la utilidad real de la metodología propuesta, tanto desde el punto de vista de su aplicabilidad como de los beneficios resultantes. El ejemplo que se va a presentar es el análisis conjunto de las avenidas de los ríos Júcar y Turia, haciendo un uso parcial de la información histórica disponible.

La cuenca vertiente del río Júcar tiene una superficie de 22.000 km<sup>2</sup>, aunque en realidad la mitad superior de la cuenca no contribuye a las avenidas de la cuenca baja. La estación de aforos utilizada (E89 en Huerto Mulet) se localiza cercana a la desembocadura, teniendo la serie sistemática empleada una longitud de 42 años. En cuanto a su información histórica, se tiene noticia que desde el año 1388 se han producido 70 avenidas que han causado daños en las poblaciones ribereñas (Calvo, 1989). El Centro de Estudios Hidrográficos (1983) cuantifica las 6 avenidas más importantes acaecidas desde el siglo XVII, de tal forma que si se sitúa el límite censor en 6.200 m<sup>3</sup>/s, se conoce el valor de las 5 avenidas que lo superaron con una aproximación similar a la de las grandes avenidas registradas durante el período sistemático. La información histórica utilizada del río Júcar es por tanto de tipo censurada con  $M=154$  años.

El río Turia, que se encuentra al norte del anterior, tiene una cuenca de 6.300 km<sup>2</sup> y desemboca en la ciudad de Valencia. La estación de aforos utilizada es la de Manises (E25), siendo el total de datos sistemáticos empleados de 41 años. Según Carmona (1990) desde 1321 hasta 1977 se han producido 22 desbordamientos. A finales del siglo XVI se comienza la construcción del encauzamiento del río a su paso por la ciudad, sin sufrir variaciones importantes desde principios del siglo XVIII hasta el comienzo de la serie sistemática. En este período, las 4 inundaciones que afectan a la ciudad son aquellas que superan la capacidad del encauzamiento, que se puede evaluar en 2.300 m<sup>3</sup>/s. Sin embargo, el valor exacto de estas avenidas no es posible conocerlo con una precisión suficiente. Por ello, se ha empleado la información histórica como de tipo binomial censurada, siendo  $M=235$  años.

El coeficiente de correlación espacial entre estos dos ríos es de -0,0384, por lo que pueden considerarse como series independientes. Por otra parte, el análisis conjunto de la información sistemática e histórica de estos dos ríos fue posible debido no sólo a su proximidad geográfica, sino también a sus similares características hidrogeomorfológicas. Dado el pequeño tamaño de la región (2 estaciones) la comprobación de su homogeneidad estadística se realizó una vez finalizado el proceso de estimación del modelo, comparando la función de distribución empírica con la ajustada tanto de la serie estandarizada como de las series locales. De hecho los parámetros regionales resultan ser  $\lambda'_1=0,9633$ ,  $\theta'_1=1,0434$ ,  $\lambda'_2=0,0906$  y  $\theta'_2=0,0329$ , casi cumpliéndose en una sola iteración la hipótesis de regionalización.

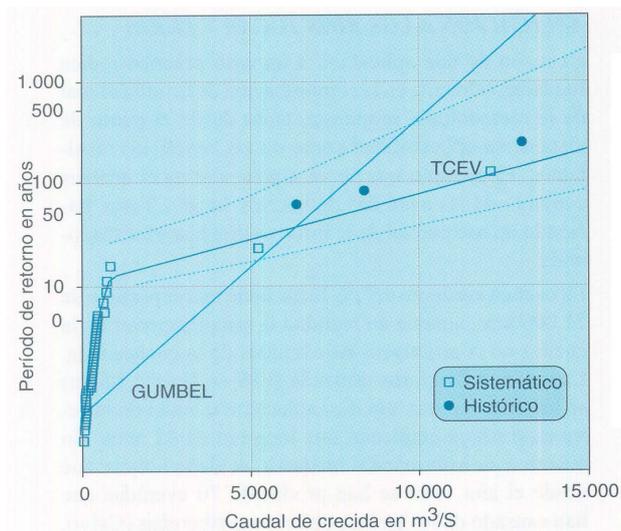


Figura 10. Resultados en el río Júcar, comparando la función de distribución empírica local con las ajustadas Gumbel local y TCEV regional (líneas de puntos límites del 95%)

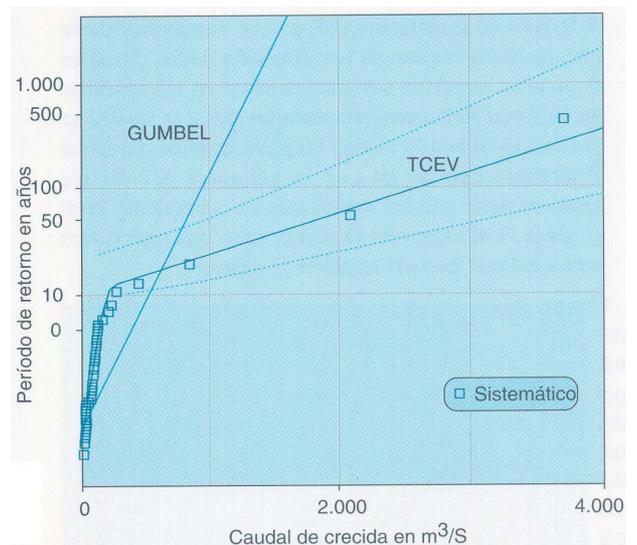


Figura 11. Resultados en el río Turia, comparando la función de distribución empírica local con las ajustadas Gumbel local y TCEV regional (líneas de puntos límites del 95%)

	$X_{1000}$ (m <sup>3</sup> /s)	Info, local sistemática	Info. local sistemática + histórica	Análisis regional con info. sisteniática + histórica
Júcar	22.711	± 24.700	± 12.700	± 10.700
Turia	2.222	±6.130	± 3.250	±2.650

Tabla I. Evolución del intervalo de confianza del 95% del cuantil de 1000 años de período de retorno con diferentes niveles de información utilizada

En las Figuras 10 y 11 se representan el ajuste TCEV con información histórica y análisis regional, sus límites de confianza del 95% asumiendo una distribución normal del error de estimación con una varianza igual al CRLB correspondiente, la función de distribución de probabilidad empírica obtenida sólo con la información local mediante la expresión de Hazen (Hirsh, 1987) y el ajuste de una función de Gumbel utilizando la información sistemática local exclusivamente. Es claro que el modelo TCEV es mucho más apropiado que el Gumbel, fundamentalmente debido a que es capaz de reproducir el “efecto pata de perro” presente en estos ríos.

El empleo de información histórica adicional y el análisis conjunto ha disminuido apreciablemente el error de estimación del modelo estadístico. En la tabla adjunta se refleja la disminución del error conforme se incrementa la información utilizada en la estimación del cuantil de 1000 años. Se puede observar que el empleo de la información histórica local disminuye el error en algo menos del 50%, y si además se realiza un análisis regional la disminución se incrementa en un 10% adicional.

**CONCLUSIONES**

En este trabajo se ha presentado una metodología que permite incrementar la información utilizada en la estimación de las avenidas de un río mediante la incorporación de información no sistemática dentro de un análisis regional. Como función de distribución se ha empleado la TCEV, que como se ha podido comprobar en el ejemplo estudiado, se adapta muy bien a las propiedades estadísticas de los ríos mediterráneos. por contra, funciones tradicionales como la Gumbel pueden dar lugar a resultados desastrosos, infravalorando los cuantiles estimados para períodos de retorno medios y altos, y sobreestimando los de períodos de retorno bajos.

Las mejoras introducidas con la información adicional y con el análisis regional respecto al empleo de la información sistemática local exclusivamente se han evaluado a través del concepto de ganancia estadística. Para ello en primer lugar se ha comprobado que la varianza asintótica (CRLB) obtenida analíticamente es una buena aproximación del error cuadrático medio de estimación de los cuantiles.

Para un determinado cuantil, el valor de la información no sistemática censurada *Tipo 1* es función de su período de retorno, de la relación entre el período no sistemático y sistemático, del período de retorno del umbral de percepción y de los parámetros estandarizados. En el caso de información no sistemática censurada *Tipo 2*, su valor es función de las longitudes de los períodos sistemático y no sistemático y de los parámetros estandarizados. Estos parámetros estandarizados representan la proporción de avenidas extraordinarias y la relación de magnitudes entre las avenidas ordinarias y extraordinarias.

Si el límite censor tiene un período de retorno medio o alto, las diferencias entre ASG1 y ASG2 son pequeñas, por lo que lo razonable en estos casos es emplear la información adicional en la forma binomial censurada. En el ejemplo analizado, el límite censor del Turia tiene un período de retorno de 70 años, mientras que el del Júcar es de 35 años. Por ello, en el primer caso se ha empleado la información histórica como binomial censurada sin evaluar la magnitud de las avenidas, y en el segundo como censurada.

En la práctica no tiene sentido comparar las ganancias estadística *Tipo 1* y *2* por tener fuentes distintas. Sin embargo, si se tiene información histórica suficiente puede no ser rentable estadísticamente el ampliar el período de investigación histórico o añadir costosa información sobre paleoavenidas.

Por último, con el modelo estadístico presentado el empleo de información adicional de cualquier estación de aforos se traduce en una mejora de la fiabilidad en la estimación de cuantiles en toda la región analizada, en magnitud similar al obtenido en un análisis local equivalente.

#### A GRADECIMIENTOS

El presente trabajo ha estado auspiciado parcialmente por el proyecto de la CICYT con referencia HID96-1318.

#### REFERENCIAS

- Arnell, N.W., S. Gabriele, (1988) *The Performance of the Two-Component Extreme Value Distribution in Regional Flood Frequency Analysis*. Water Res. Res. 24, 879-887.
- Baker, V.R., (1987) *Paleofloodhydrology and extraordinary flood events*. Jour. of Hydrol. 96, 79-99.
- Baker, V.R., R. C. Kochel and P. C. Patton (eds.), (1988) Flood geomorphology. Wiley, New York.
- Bartley, P., B.L. Fox, L.E. Schräge, (1987) A Guide to Simulation. Springer-Verlag.
- Benson, M.A., (1950) *Use of Historical Data in Flood Frequency Analysis*. Eos. Trans. AGU. 31 (3), 419-424.
- Beran, M., J. R. M. Hosking, and N. Arnell, (1986) *Comment on "Two-Component Extreme Value Distribution for Flood Frequency Analysis"*. Water Res. Res. 22, 263-266.

- Berga, L., (1991) *Flood Forecasting in Spain*. Proc. of XXIV IAHR International Congress, Madrid, A 79-88.
- Boes, D.C., J. Heo, J.D. Salas, (1989) *Regional Flood Quantité Estimation for a Weibull Model*. Water Res. Res. 25, 979-990.
- Calvo, F., (1989) Grandes Avenidas e Inundaciones Históricas. Reunión Científica Internacional sobre Avenidas Fluviales e Inundaciones en la Cuenca del Mediterráneo. Universidad de Alicante.
- Carmona González, P., (1990) La Formado de la Plana Al.luvial de València. Institució Valenciana d'Estudis i Investigació, Valencia.
- Centro de Estudios Hidrográficos, (1983) Actualización de los Estudios de Desagüe de la Autopista Tarragona-Alicante. Tomo IV: Cruce con el río Júcar. MOPU, Madrid.
- Cohn, T.A., J.R. Stedinger, (1987) *Use of Historical Information in a Maximum Likelihood Framework*. Jour. of Hydrol. 96, 215-233.
- Condie, R., y K.H. Lee, (1982) *Flood Frequency Analysis with Historic Information*. Jour. of Hydrol. 58,47-61.
- Cunnane, C., (1988) *Methods and merits of regional flood frequency analysis*. Jour. of Hydrol. 100, 269-290.
- Ferrari, E, (1994) *Regional Rainfall and Flood Frequency Analysis in Italy*. Preprocc. UNESCO IPH-IV "Developments in Hydrology of Mountainous Areas: Regionalization of Rare Extreme Floods and Precipitation". Stará Lesná, Eslovaquia.
- Ferrer, J, (1996) El modelo de función de distribución SQRT-ET max en el análisis regional de máximos hidrológicos: aplicación a lluvias diarias. Tesis doctoral.
- Ferrer, J y L. Ardiles, (1994) *Análisis estadístico de las series anuales de máximas lluvias diarias en España*. Ingeniería Civil. 95, 87-100.
- Francés, F., (1995) Utilización de la Información Histórica en el Análisis Regional de las Avenidas. Centro Internacional de Métodos Numéricos en Ingeniería, monografía n1 27, 242 pp, Barcelona.
- Francés, F., J.D. Salas y D.C. Boes, (1991) *Flood Frequency Analysis by Using Historical Information*. Proc. of XXIV IAHR International Congress, Madrid. A 11-21.
- Francés, F., J.D. Salas y D.C. Boes, (1994) *Flood frequency analysis with systematic and historical or paleoflood data based on the two-parameter general extreme value models*. Water Res. Res. 30(6), 1653-1664.
- Hirsch, R. M., (1987) *Probability Plotting Position Formulas for Flood Records with Historical Information*. Jour. of Hydrol. 96, 185-199

- Hosking, J.R.M., J.R. Wallis, (1986a) *Paleoflood Hydrology and Flood Frequency Analysis*. Water Res. Res., 22 (4), 543-550.
- Hosking, J.R.M., J.R. Wallis, (1986b) *The Value of Historical Data in Flood Frequency Analysis*. Water Res. Res., 22 (11), 1606-1612.
- Hosking, J.R.M. y J.R. Wallis, (1988) *The effect of intersite dependence on regional flood frequency analysis*. Water Res. Res., 24 (4), 588-600.
- Guo, S. L. y C. Cunnane, (1991) *Evaluation of the usefulness of historical and paleological floods in quantile estimation*. Jour. of Hydrol., 129, 245-262.
- Jin, M. y J. R. Stedinger, (1989) *Flood Frequency Analysis with Regional and Historical Information*. Water Res. Res., 25 (5), 925-936.
- Kendall, M. G. y A. Stuart, (1967) The Advanced Theory of Statistics. Vol. II, 2nd Edition, Hafner Publ. Co., New York.
- Kottegoda, N.T., 1984. *Investigation of Outliers in Annual Maximum Flow Series*. Journal of Hydrology, 72, 105-137.
- Kroll, C.N. y J.R. Stedinger, 1996. *Estimation of moments using censored data*. Water Res. Res., 32 (4), 1005-1012.
- Landwehr, J.M., N.C. Matalas y J.R. Wallis, (1979) *Probability Weighted Moments Compared with some Traditional Techniques in Estimating Gumbel Parameters and Quantiles*. Water Res. Res., 15, 1055-1064.
- Leese, M. N., (1973) *Use of Censored Data in the Estimation of Gumbel Distribution Parameters for Annual Maximum Flood Series*. Water Res. Res., 9, 1534-1542.
- Matalas, N. C., J. R. Slack, and J. R. Wallis, (1975) *Regional Skew in Search of a Parent*. Water Res. Res., 11 (6), 815-826.
- Phien, H. N. y T. E. Fang, (1989) *Maximum likelihood estimation of the parameters and quantiles of the general extreme-value distribution from censored samples*. Jour. of Hydrol., 105, 139-155.
- Pilon, P. J. y K. Adamowski, (1993) *Asymptotic variance of flood quantile in log Pearson type III distribution with historical information*. Jour. of Hydrol., 143, 481-503.
- Potter, W.D., (1958) *Upper and Lower Frequency Curves for Peak Rates of Runoff*. EOS. Trans. AGU. 39, 100-105.
- Press, W.H., B.P. Flannery, S.A. Tevkolsky y W.T. Vetterlig, (1989) Numerical Recipes. Cambridge University Press, 702 p.
- Rossi, F., M. Fiorentino y P. Versace, (1984) *Two-Component Extreme Value Distribution for Flood Frequency Analysis*. Water Res. Res., 20, p. 847-856.
- Stedinger, J.R. y T.A. Cohn, 1986. *Flood Frequency Analysis with Historical and Paleoflood Information*. Water Res. Res., 22 (5), 785-793.