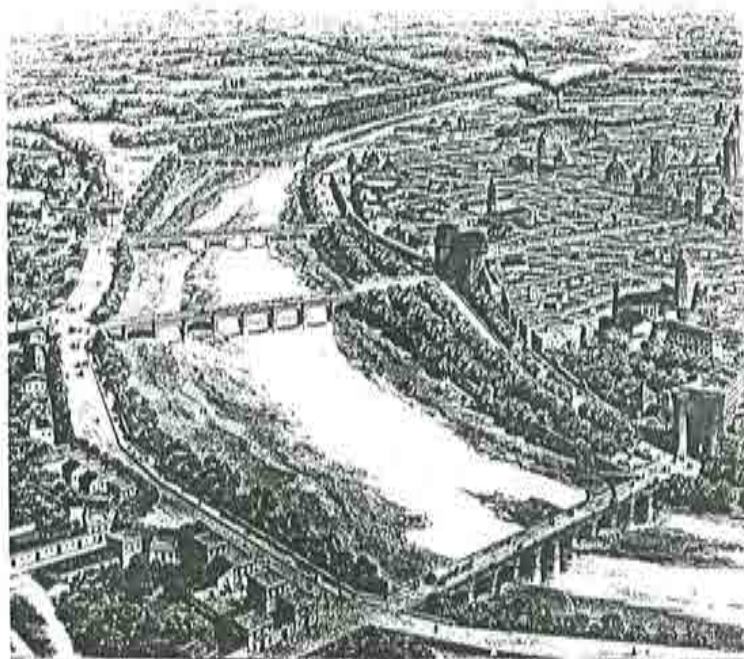


Utilización de la información histórica en el análisis regional de las avenidas

Félix Francés García



Utilización de la información histórica en el análisis regional de las avenidas
Utilización de la información histórica en el análisis regional de las avenidas
Utilización de la información histórica en el análisis regional de las avenidas

Félix Francés García
Félix Francés García
Félix Francés García

27

27

27

Utilización de la información histórica en el análisis regional de las avenidas

Félix Francés García

Monografía Nº 27, Mayo 1995

Diseño de la cubierta: Jordi Pallí

Primera Edición, Mayo 1995

@ El autor

Edita:

Centro Internacional de Métodos Numéricos en Ingeniería

Edificio C1, Campus Norte UPC

Gran Capitán, s/n

08034 Barcelona, España

ISBN: 84-87867-53-7

Deposito Legal: B-24287-95

INDICE

1.- INTRODUCCION Y OBJETIVOS	1
2.- ANTECEDENTES	5
2.1.- Introducción	6
2.2.- Empleo de los datos sistemáticos locales exclusivamente	7
2.2.1.- Modelos de una sola distribución	
2.2.2.- Modelos mixtos	
2.3.- Empleo de datos locales y regionales	13
2.3.1.- Métodos de análisis regional	
2.3.2.- Datos regionales únicamente	
2.3.3.- Obtención de regiones homogéneas	
2.4.- Empleo de datos históricos	22
2.5.- Criterios de selección del modelo	28
2.6.- Conclusiones	30
3.- INCORPORACION DE LA INFORMACION HISTORICA	31
3.1.- Justificación de la metodología adoptada	32
3.1.1.- Introducción	
3.1.2.- Tipos de información histórica considerados	
3.1.3.- Medida de la bondad en la estimación de los cuantiles	
3.2.- Hipótesis consideradas	37
3.3.- Procedimiento estadístico	38
3.3.1.- Función de verosimilitud	
3.3.2.- Obtención de las ganancias estadísticas asintóticas de los cuantiles	
3.3.3.- Obtención de las ganancias estadísticas muestrales	
4.- ANALISIS DE LA INCORPORACION DE LA INFORMACION HISTORICA A LAS DISTRIBUCIONES "EXTREME VALUE"	43
4.1.- Introducción	44
4.2.- Análisis de la distribución LogGumbel	45
4.2.1.- Características estadísticas	
4.2.2.- Estimadores de los parámetros por Máxima Verosimilitud	
4.2.3.- Varianzas asintóticas de los estimadores de los cuantiles	
4.2.4.- Ganancias estadísticas asintóticas de los estimadores de los cuantiles	
4.2.5.- Propiedades de las ganancias estadísticas asintóticas	
4.2.6.- Errores cuadráticos medios y ganancias estadísticas muestrales de los cuantiles	

4.3.- Conclusiones	69
5.- ANALISIS DE LA INCORPORACION DE LA INFORMACION HISTORICA A LA DISTRIBUCION "TWO COMPONENT EXTREME VALUE"	71
5.1.- Introducción	72
5.2.- Génesis de la distribución "Two Component Extreme Value"	73
5.3.- Características estadísticas de la TCEV	76
5.3.1.- Probabilidad de presentación de una avenida extraordinaria	
5.3.2.- Media poblacional	
5.3.3.- Varianza poblacional	
5.3.4.- Sesgo poblacional	
5.4.- Estimadores de los parámetros por Máxima Verosimilitud	86
5.4.1.- Funciones de verosimilitud	
5.4.2.- Obtención de los máximos de las funciones de verosimilitud	
5.5.- Matrices de información de Fisher para los distintos tipos de información	91
5.5.1.- Matriz de información de Fisher con información sistemática exclusivamente	
5.5.2.- Matriz de información de Fisher con información histórica binomial censurada	
5.5.3.- Matriz de información de Fisher con información histórica censurada	
5.5.4.- Matriz de información de Fisher con información histórica de máxima avenida	
5.6.- Varianzas y ganancias estadísticas asintóticas de los cuantiles	106
5.6.1.- Obtención de la varianza y ganancia estadística asintótica	
5.6.2.- Influencia de los parámetros estandarizados en la ganancia estadística asintótica	
5.6.3.- Variación de las ganancias estadísticas asintóticas en función del periodo de retorno del cuantil	
5.6.4.- Variación de las ganancias estadísticas asintóticas en función del límite de percepción histórico	
5.6.5.- Variación de las ganancias estadísticas asintóticas en función de la relación de longitudes entre el periodo histórico y sistemático	
5.7.- Varianzas y ganancias estadísticas muestrales de los cuantiles	147
5.8.- Utilización de la TCEV en un análisis regional	156
5.8.1.- Fases del análisis regional	
5.8.2.- Test SKEW de detección de "high outliers"	
5.8.3.- Varianza asintótica de los cuantiles	
5.9.- Conclusiones	162
6.- APLICACION A LOS RIOS JUCAR, TURIA Y SEGURA	165
6.1.- Introducción	166
6.2.- Información disponible	168
6.2.1.- Río Júcar	

6.2.2.- Río Turia	
6.2.3.- Río Segura	
6.3.- Análisis local del río Júcar	186
6.4.- Análisis local del río Turia	191
6.5.- Análisis conjunto de los ríos Júcar y Turia	197
6.6.- Análisis local del río Segura	202
6.7.- Solución adoptada	208
7.- SUMARIO	215
8.- REFERENCIAS	221
ANEJOS:	
A.- Obtención de la función de densidad de probabilidad acumulada empírica . .	229
B.- Resultados para la distribución Weibull	233
C.- Resultados para la distribución Gumbel	239

1.- INTRODUCCION Y OBJETIVOS

Uno de los objetivos de la Hidrología es el de realizar el análisis de la frecuencia de las avenidas de un río. Este análisis es utilizado en el planeamiento y decisión en sistemas de recursos hidráulicos, diseño de obras hidráulicas, sistemas específicos contra avenidas, zonificación de riesgos, etc... conectándose con el proceso de decisión económico, donde todas estas utilidades están inmersas. Pero esta conexión es un problema aún no resuelto, dada su complejidad y gran variedad de elementos diferentes que intervienen (por ejemplo, ¿cómo incluir las vidas humanas en el proceso de decisión?).

El análisis de la frecuencia de las avenidas consiste básicamente en obtener la relación entre caudales de crecida y su probabilidad de no excedencia, aunque en la mayor parte de los casos, basta con obtener los caudales correspondientes a uno pocos niveles de riesgo. Fijada la frecuencia o riesgo de la avenida de proyecto, ésta debe de determinarse con la mayor precisión posible. Un error por defecto del cuantil da lugar a un aumento del riesgo realmente asumido. En el otro extremo, si el error es por exceso, lo que se está provocando es un aumento no necesario del coste del proyecto.

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

i) Presencia de errores debidos a

- La elección del modelo
- Estimación a partir de series de longitud finita
- Errores en los propios datos

ii) Uso de series excesivamente cortas para los períodos de retorno que se estiman.

iii) Poblaciones con coeficientes de variación y sesgo altos, es decir, que presentan fuertes variaciones en los valores muestrales.

iv) En ciertas regiones, las crecidas pueden ser debidas a dos o más causas diferentes.

Los ríos mediterráneos presentan avenidas muy peligrosas. Son ríos con coeficientes de sesgo muy altos y sus crecidas pueden 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 250.000 millones del año 1982 (Las Provincias, 1988) en daños producidos por una crecida del Júcar no solo justifican la inversión en medidas contra las avenidas de este o de cualquier otro río, sino que también exigen que estas medidas estén derivadas de un análisis cuidadoso de la frecuencia y comportamiento de las crecidas del río.

Históricamente, se puede resumir la investigación en el análisis de crecidas en la existencia de tres etapas en el tiempo:

- i) Búsqueda y evaluación de la "mejor función de distribución de máximos anuales" (años 50 y 60).
- ii) Mejora de los métodos de estimación de los parámetros (años 70).
- iii) Aumentar la fiabilidad de los estimadores incrementando la longitud estadística de las series de las estaciones de aforos (años 80).

Hasta la actualidad, el aumento de la longitud estadística de las series se ha conseguido bien sea a través de un análisis regional, o bien mediante la incorporación de la información que se tenga sobre avenidas históricas. Este trabajo se centra precisamente en este último aspecto, siendo sus objetivos fundamentales:

- i) Realizar una revisión exhaustiva del estado del arte en el análisis de la frecuencia de las crecidas.
- ii) Desarrollar un modelo estadístico que sea capaz de incorporar la información histórica dentro de un análisis regional, teniendo en cuenta las particularidades de los ríos mediterráneos.
- iii) Obtener y analizar las mejoras estadísticas que se producen con dicha incorporación. Con ello no solo se podrá calcular la bondad de los estimadores de las avenidas, sino que también se deducirán recomendaciones de utilización de la información histórica.
- iv) Aplicar las nuevas metodologías a un caso práctico.

La revisión del estado del arte se encuentra recogido en el Capítulo 2, separando los antecedentes en los tres tipos de información con la que se puede contar: sistemática

exclusivamente, regional e histórica.

Los objetivos ii) y iii) se desarrollan en los Capítulos 3, 4 y 5 de este trabajo. El Capítulo 3 resume las ideas centrales y sirve de armazón base para los dos capítulos siguientes.

El Capítulo 4 desarrolla los objetivos para la familia de distribuciones "Extreme Value" (EV), que está formada por las distribuciones Weibull, LogGumbel y Gumbel. Esta última muy importante en España por su amplia utilización. Realmente en este capítulo solo se analiza la distribución LogGumbel, dejando para los Anejos B y C a las otras dos distribuciones de la familia.

El Capítulo 5 repite la misma estructura del Capítulo 4 pero para la distribución "Two Component Extreme Value" (TCEV). De las referencias consultadas y de la experiencia del Departamento de Ingeniería Hidráulica y Medio Ambiente de la Universidad Politécnica de Valencia se puede adelantar que esta distribución se adapta muy bien a los ríos mediterráneos y, además, su uso está recomendado dentro de un análisis regional dado el alto número de parámetros que posee.

El caso práctico analizado es el de tres ríos importantes de la cuenca española del Mediterráneo: los ríos Júcar, Turia y Segura. El Capítulo 6 se ha dedicado a esta aplicación.

Por último, un sumario y las conclusiones generales de este trabajo pueden encontrarse en el Capítulo 7.

2.- ANTECEDENTES

2.1.- INTRODUCCION

En este Capítulo de Antecedentes se va a hacer un repaso de todas las técnicas disponibles en la actualidad para el análisis de la frecuencia de las avenidas. Las diferentes metodologías se pueden clasificar de tres formas distintas:

i) Según el tipo de datos utilizados o disponibles:

- Datos sistemáticos locales. Si se utilizan los datos de aforos del único punto de interés.
- Datos sistemáticos regionales. Si se utilizan los datos de aforos de toda una región.
- Datos sistemáticos e históricos, bien sea locales o regionales.

ii) Según el tipo de variable aleatoria considerada:

- Caudal instantáneo máximo anual.
- Picos de caudal instantáneo máximo diario por encima de un cierto límite.

iii) Según la familia de la función de distribución y del procedimiento de estimación de sus parámetros.

En los Apartados 2.2, 2.3 y 2.4 se van a considerar las diferentes aproximaciones al problema atendiendo a la primera clasificación. Además, se considera que la selección del modelo estadístico es un punto fundamental en el análisis de la frecuencia de las crecidas, por lo que se recogen los diferentes criterios de selección en el Apartado 2.5. El Apartado 2.6 resume las conclusiones fundamentales de esta revisión del estado del arte.

2.2.- EMPLEO DE LOS DATOS SISTEMATICOS LOCALES EXCLUSIVAMENTE

Este sería el enfoque clásico; ajuste a los datos del punto de interés de la función de distribución más adecuada (ya se comentarán más adelante los criterios de selección entre las posibles funciones de distribución).

Si los caudales son máximos anuales, los cuantiles tendrán la expresión siguiente:

$$X_T = \alpha_1 + \alpha_2 Y_T(\alpha_3) \quad (2.1)$$

donde:

α_1 = parámetro de localización

α_2 = parámetro de escala

α_3 = parámetro de forma

Y_T = variable estandarizada de período de retorno T

Si por el contrario son picos por encima de un límite, los cuantiles vendrán expresados por funciones del tipo:

$$X_T = \alpha_1 + \alpha_2 Y_T(\alpha_3, \delta) \quad (2.2)$$

donde δ es el número medio anual de picos por encima del límite.

En esta investigación no va a ser considerado este último tipo de variable. Sobre la primera, solo se va a hacer mención a las distribuciones más utilizadas en Hidrología sin describir sus características estadísticas.

2.2.1.- Modelos de una sola distribución

Los modelos estadísticos clásicos de la Hidrología utilizados en el estudio de avenidas, solo consideran una única población de máximos caudales anuales. Estos modelos clásicos son los siguientes:

i).- Lognormal

Esta distribución puede tener dos o tres parámetros. Con dos parámetros (LN2) su expresión es:

$$f_x(x) = \frac{1}{\sqrt{2\pi x} \sigma_y} \exp \left[-\frac{1}{2} \left(\frac{Lx - \mu_y}{\sigma_y} \right)^2 \right] \quad (2.3)$$

o con tres parámetros (LN3):

$$f_x(x) = \frac{1}{\sqrt{2\pi(x-x_0)} \sigma_y} \exp \left[-\frac{1}{2} \left(\frac{L(x-x_0) - \mu_y}{\sigma_y} \right)^2 \right] \quad (2.4)$$

donde:

$f_x(\cdot)$ = función de densidad de probabilidad de la variable aleatoria X

μ_y = parámetro de escala

x_0 = parámetro de posición

σ_y = parámetro de forma

ii).-Gamma

También llamada distribución de Pearson. Si es de dos parámetros (Pearson tipo II):

$$f_x(x) = \frac{1}{|\alpha| \Gamma(\beta) \left(\frac{x}{\alpha} \right)^{\beta-1}} e^{-\frac{x}{\alpha}} \quad (2.5)$$

Con tres parámetros (Pearson tipo III):

$$f_x(x) = \frac{1}{|\alpha| \Gamma(\beta) x \left(\frac{x-x_0}{\alpha} \right)^{\beta-1}} e^{-\frac{x-x_0}{\alpha}} \quad (2.6)$$

donde:

α = parámetro de escala

x_0 = parámetro de posición

β = parámetro de forma

iii).-LogPearson tipo III

O loggamma de 3 parámetros. Es la recomendada por el "United States Water

Resources Council" (1982). Tiene la expresión:

$$f_x(x) = \frac{1}{|\alpha| \Gamma(\beta)x} \left(\frac{Lx-y_0}{\alpha} \right)^{\beta-1} e^{-\frac{Lx-y_0}{\alpha}} \quad (2.7)$$

donde:

α = parámetro de escala de los logaritmos
 y_0 = parámetro de posición de los logaritmos
 β = parámetro de forma de los logaritmos

iv).-"General Extreme Value" (GEV)

Esta distribución es la recomendada por el "Natural Environment Research Council" británico (1975):

$$F_x(x) = \exp \left[- \left(1 - \beta \frac{x-x_0}{\alpha} \right)^{\frac{1}{\beta}} \right] \quad (2.8)$$

donde:

$F_x(.)$ = función de densidad de probabilidad acumulada
 α = parámetro de escala
 x_0 = parámetro de posición
 β = parámetro de forma

Según el valor del parámetro β , se tendrá:

-GEV 1 o Gumbel o EV1, si $\beta=0$. Tomando límites en la ecuación (2.8) la función de distribución queda:

$$F_x(x) = \exp(-\lambda e^{-\theta x}) \quad (2.9)$$

donde:

λ = parámetro de forma
 θ = parámetro de escala

-GEV 2, si $\beta < 0$.

-GEV 3, si $\beta > 0$.

Si el parámetro de posición es nulo, se tiene la familia "extreme value" (EV). Usualmente se denomina a la EV3 como distribución Weibull:

$$F_x(x) = 1 - e^{-(\theta x)^\gamma} \quad (2.10)$$

y EV2 a la distribución LogGumbel:

$$F_x(x) = e^{-\left(\frac{\theta}{x}\right)^\gamma} \quad (2.11)$$

donde para ambas ecuaciones:

θ = parámetro de escala

γ = parámetro de forma

2.2.2.-Modelos mixtos

Durante los años 70, cobraron especial interés la observación de dos propiedades estadísticas de las poblaciones de caudales máximos anuales, que eran difícilmente representadas por las funciones de distribución clásicas anteriormente descritas. Estas dos propiedades son el "efecto pata de perro" y la "condición de separación".

El "efecto pata de perro" (llamado "dog leg" por Potter en 1958) se presenta en los ríos en los que los registros de crecidas provienen de la combinación o mezcla de dos o más factores desencadenadores. Por ejemplo, la fusión de la nieve, tormentas extratropicales o huracanes. Este fenómeno se refleja en un cambio brusco de la pendiente de la curva de densidad de probabilidad acumulada empírica.

La "condición de separación" ("separation condition") fue estudiada por Matalas et al. (1975) para los EEUU. Si se representa la curva que relaciona la media del coeficiente de sesgo y su desviación típica para una función de distribución de probabilidad y los puntos correspondientes a poblaciones de diversas regiones, se observó que las funciones de probabilidad clásicas se alejaban de los datos reales. Es decir, las funciones disponibles hasta aquel momento no pueden representar la variabilidad muestral que tienen los caudales de crecidas reales.

Conocidos estos problemas, los hidrólogos trataron de solucionarlos desde dos puntos

de vista diferentes:

i) Modelo de Waylen y Woo (1982)

Waylen y Woo consideraron que las avenidas podían provenir de dos mecanismos de producción distintos. En su artículo llamaron R a la variable aleatoria que representa las crecidas anuales debidas a la precipitación y S a las de fusión de la nieve, siendo ambas claramente identificables en la serie de datos de aforos. Se puede definir una nueva variable aleatoria que sea el máximo de ambas:

$$X_i = \max (R_i, S_i) \quad (2.12)$$

siendo i el año. Entonces, si se acepta la hipótesis de que R y S son independientes, la función de probabilidad de no excedencia de X es:

$$F_X(x) = P(X \leq x) = P(R \leq x, S \leq x) = P(R \leq x) P(S \leq x) = F_R(x) F_S(x) \quad (2.13)$$

Al poder separar las máximas avenidas por lluvia de las máximas avenidas por nieve, las funciones $F_R(\cdot)$ y $F_S(\cdot)$ se estiman por separado.

El modelo "Two Component Extreme Value" o TCEV desarrollado por Rossi et al. (1984) llega al mismo resultado final, pero con un origen conceptual diferente al considerar que no es posible separar las avenidas producidas por cada una de los mecanismos generadores. En el Capítulo 5 se describe con mayor detalle la génesis de este modelo. El resultado final es el producto de dos distribuciones de Gumbel:

$$F_X(x) = \exp(-\lambda_1 e^{-\theta_1 x} - \lambda_2 e^{-\theta_2 x}) \quad (2.14)$$

donde:

λ_1, λ_2 = parámetros de forma
 θ_1, θ_2 = parámetros de escala

ii) Distribución Wakeby

Desarrollada por Houghton (1978), es un modelo de variables mixtas. Ahora, la nueva variable aleatoria queda definida como:

$$X_i = R_i + S_i \quad (2.15)$$

Es decir, en este caso la máxima avenida anual se produce por la combinación en el tiempo de dos mecanismos generadores. La función de densidad de probabilidad acumulada llamada de Wakeby queda (en forma explícita):

$$x = -a(1-F_x(x))^b + c(1-F_x(x))^{-d} + e \quad (2.16)$$

donde a, b, c, d y e son parámetros del modelo.

Dado el número de parámetros utilizados por los modelos mixtos (4 en la TCEV y 5 en la Wakeby), éstos hay que emplearlos o bien con series largas (superiores a 75 años), o bien en un contexto regional, si no se quiere contravenir el Principio estadístico de Parsimonia de los parámetros. Este principio recomienda un mínimo de 15 datos por parámetro.

2.3.- EMPLEO DE DATOS LOCALES Y REGIONALES

Uno de los problemas más importantes del análisis de frecuencia de las crecidas es que hay que realizar estimaciones de los caudales de muy alto período de retorno (es decir, estimaciones en la cola derecha de las funciones de densidad de probabilidad), cuando los datos disponibles se agrupan en la parte media o izquierda de la función. Esto es debido a que las series de aforos son usualmente excesivamente cortas para resolver con precisión el problema. Si se pudiera extender de alguna forma la "longitud estadística" de las series de caudales, el error cometido sería menor. Este es el objetivo de las diferentes metodologías que tratan de agrupar los datos de estaciones en una región estadísticamente homogénea. El Apartado 2.3.3 analiza los criterios actuales de obtención de estas regiones homogéneas.

2.3.1.- Métodos de análisis regional

Supongase que se dispone de las series de caudales máximos de varias estaciones con N_i años de longitud cada una. Si se quieren utilizar todos los datos conjuntamente, habrá que realizar algún tipo de hipótesis sobre la homogeneidad de las estaciones respecto de alguno o algunos de los estadísticos de la población de la variable aleatoria. Las principales hipótesis de homogeneidad realizadas hasta la fecha son:

2.3.1.1.-Método "index flood"

Este método fue sugerido por Dalrymple (1960), quedando en el olvido, hasta que fue retomado posteriormente por el "Natural Environment Research Council" británico (1975), Wallis (1980) y Stedinger (1983).

Estima la distribución regional a partir de la variable estadística adimensional Y obtenida como:

$$Y = \frac{X_i}{X_I} \quad (2.17)$$

donde:

X_i = serie de avenidas de la estación i .

X_I es el "index flood", normalmente igual al valor medio de las avenidas de la estación i .

Implícitamente se está haciendo la hipótesis de que los caudales adimensionales de las distintas estaciones siguen la misma distribución estadística. Esto implica que, por ejemplo, para el caso de una distribución regional del tipo Weibull, el parámetro de forma es el mismo en toda la región (Boes et al. 1989).

Si se estima Y_T del análisis regional, el cuantil local de la estación i será:

$$X_{iT} = Y_T X_i \quad (2.18)$$

El "Natural Environment Research Council" británico (1975) ajusta una GEV a los caudales adimensionales, Wallis (1980) mejora la estimación de parámetros utilizando el método de estimación de parámetros por momentos compensados y coeficientes de peso estadístico en función de la longitud de cada serie.

2.3.1.2.-Método de boletín 17B del "U.S. Water Resources Council"

El boletín 17B del "U.S. Water Resources Council" asume que la función de distribución regional de caudales es una LogPearson tipo III. Y esto lo hace por inercia respecto a recomendaciones anteriores norteamericanas más que por razones estadísticas (el boletín 15 del mismo organismo, ya la adoptaba en su metodología en 1967), pues existen artículos anteriores a la publicación de este boletín demostrando la mejor bondad de otras distribuciones para la mayoría de los ríos norteamericanos (Shen et al. 1980).

El método es el siguiente:

- i) Estimar la media, desviación típica y coeficiente de sesgo a partir de los logaritmos de las avenidas locales.
- ii) Bien sea por análisis regional o recurriendo a unos mapas de sesgos (obviamente solo disponibles para los EEUU), obtener el coeficiente de sesgo regional.
- iii) Modificar el sesgo local con la siguiente expresión:

$$\gamma_w = \frac{MSE(\gamma_g) + \gamma_g MSE(\gamma)}{MSE(\gamma_g) + MSE(\gamma)} \quad (2.19)$$

donde:

γ_w = coeficiente de sesgo compensado

γ = coeficiente de sesgo local
 γ_B = coeficiente de sesgo regional
 MSE(.) = error cuadrático medio

iv) Ajustar una distribución LogPearson tipo III tomando como sesgo el coeficiente de sesgo compensado.

El error cuadrático medio del coeficiente de sesgo local, tiene la expresión:

$$MSE(\gamma) = 10^{A - B \log_{10}\left(\frac{N}{10}\right)} \quad (2.20)$$

donde N es la longitud de la serie y:

$$A = -0.33 + 0.08 |\gamma| \quad \text{si } |\gamma| \leq 0.9 \quad (2.21)$$

$$A = -0.52 + 0.30 |\gamma| \quad \text{si } |\gamma| > 0.9 \quad (2.22)$$

$$B = 0.94 - 0.26 |\gamma| \quad \text{si } |\gamma| \leq 1.5 \quad (2.23)$$

$$B = 0.55 \quad \text{si } |\gamma| > 1.5 \quad (2.24)$$

Para los mapas de sesgo americanos, $MSE(\gamma_B) = 0.302$.

2.3.1.3.-Aplicación de la teoría bayesiana de la decisión

Se trata más bien de un método de estimación regional de parámetros. Sea $f_X(x; \underline{\alpha})$ la función de densidad de probabilidad de las máximas avenidas anuales dependiente del vector de parámetros $\underline{\alpha}$.

Sea \underline{C} el vector de características hidrológicas y/o geomorfológicas principales de la cuenca. Por ejemplo: área, pendiente, lluvia media, niveles de segmentación, coeficientes de forma, etc...

Si se consideran a los vectores $\underline{\alpha}$ y \underline{C} como variables estadísticas, la interdependencia entre ellos puede analizarse mediante regresiones regionales del tipo (Nash y Shaw, 1966, Cunnane y Nash, 1971):

$$\underline{\alpha} = \underline{C}' \beta + \underline{\epsilon} \quad (2.25)$$

donde:

β = matriz de la regresión lineal.

$\underline{\epsilon}$ = vector de ruidos independientes e idénticamente distribuidos.

y por tanto se puede definir una distribución "a priori" condicionada de los parámetros $\underline{\alpha}$: $\phi(\underline{\alpha} | \underline{C})$. Esta función expresa la información regional previa acerca de $\underline{\alpha}$.

La información local se incorpora a través de la función de verosimilitud:

$$L(x | \underline{\alpha}) = \prod_{i=1}^N f_{x_i}(x_i; \underline{\alpha}) \quad (2.26)$$

Utilizando el teorema de Bayes para obtener la distribución "a posteriori" de los parámetros se tiene que:

$$\Phi(\underline{\alpha} | x, \underline{C}) = \frac{\phi(\underline{\alpha} | \underline{C}) L(x | \underline{\alpha})}{\int_{\underline{\alpha}} \phi(\underline{\alpha} | \underline{C}) L(x | \underline{\alpha}) d\underline{\alpha}} \quad (2.27)$$

Los estimadores locales de $\underline{\alpha}$ se obtienen tomando los valores esperados de la función Φ . A partir de ellos se pueden calcular los cuantiles X_T .

La última revisión a esta metodología corresponde a Kuczera (1982a, 1982b). Tiene la ventaja de admitir heterogeneidad en las características estadísticas de las series de avenidas.

2.3.1.4.-Método TCEV

Desarrollado por Rossi et al. (1984) y Fiorentino et al. (1985). Se basa en la distribución de cuatro parámetros definida en la ecuación (2.14). Estos parámetros se estiman por máxima verosimilitud sobre series estandarizadas de máximos anuales. La estandarización propuesta es:

$$Y = \theta_1 X - L\lambda_1 \quad (2.28)$$

Con esta estandarización, las avenidas estandarizadas siguen la distribución:

$$F_Y(y) = \exp(-e^{-y} - \lambda e^{-\theta y}) \quad (2.29)$$

La hipótesis de regionalización asumida es que λ y θ son constantes en toda la región.

La metodología propuesta por Rossi et al. (1984) para un análisis regional es:

- i) Localización de los "high outliers" (valores que se salen de la tendencia general por ser excesivamente grandes) en cada estación.
- ii) Estimación de λ_1 y θ_1 por el Método de Máxima Verosimilitud en cada estación, sin considerar los "outliers", ajustando una distribución Gumbel.
- iii) Estandarización de las crecidas regionales, Y , según la ecuación (2.28).
- iv) Estimación por el Método de Máxima Verosimilitud de los parámetros regionales $\hat{\lambda}_1'$, $\hat{\theta}_1'$, $\hat{\lambda}_2'$ y $\hat{\theta}_2'$ ajustando una distribución TCEV.
- v) Obtención de los parámetros de la distribución TCEV local desestandarizando los regionales asumiendo que $\lambda_1'=1$ y $\theta_1'=1$.

Arnell y Gabrielle (1988) proponen un método iterativo de estimación que no tiene en cuenta la presencia de los "outliers". Consiste en:

- i) Ajuste de una distribución TCEV local por el Método de Máxima Verosimilitud.
- ii) Estandarización de las crecidas regionales.
- iii) Ajuste de una distribución TCEV regional por el Método de Máxima Verosimilitud.
- iv) Obtención de los parámetros locales desestandarizando los regionales y volver al paso ii), hasta cumplir algún criterio de convergencia de la solución.

2.3.1.5.-Modelos bivariados y multivariados

Son utilizados para extender las series a la o las estaciones de menor longitud de datos. Por lo tanto en la estación con mayor número de años de registro no se produce ningún

incremento en la información estadística y, en el resto de estaciones, este incremento es función de las correlaciones cruzadas y de las longitudes relativas de las series.

Desafortunadamente no existe amplia bibliografía sobre distribuciones multivariadas de extremos y la mayoría es sobre bivariadas, aunque siempre se puede recurrir a una normalización previa (con todos los problemas que ello conlleva) y tratar el problema como multinormal, que sí que está muy bien estudiado. Por ejemplo, para un modelo binormal, se produce incremento en la información de la media muestral de la serie más corta si el coeficiente de correlación cruzada entre las dos series cumple que:

$$\rho_{x_1x_2}^2 > \frac{1}{N_1 - 2} \quad (2.30)$$

donde N_1 es la longitud de la serie más corta.

En cuanto a modelos bivariados de caudales máximos anuales con distribución normal, se tiene (Raynal 1985):

i) Modelo logístico. Considera que las dos distribuciones marginales son Gumbel:

$$F_{x_1x_2}(x_1, x_2) = \exp\left\{-\left[-LF_{x_1}(x_1)\right]^m + \left[-LF_{x_2}(x_2)\right]^m\right\}^{1/m} \quad (2.31)$$

donde m (≥ 1) es el llamado parámetro asociado. Se relaciona con el coeficiente de correlación cruzada por:

$$m = \frac{1}{1 - \rho_{x_1x_2}} \quad (2.32)$$

ii) Modelo mixto. La distribución conjunta es:

$$F_{x_1x_2}(x_1, x_2) = e^{-k \left[\frac{1}{-LF_{x_1}(x_1)} + \frac{1}{-LF_{x_2}(x_2)} \right]} \quad (2.33)$$

donde k es el parámetro asociado ($0 \leq k \leq 1$). En el caso de que las distribuciones marginales sean del tipo Gumbel, está relacionado con el coeficiente de correlación cruzada por:

$$\rho_{x_1x_2}^2 = \frac{6}{\pi^2} \left[\arccos\left(1 - \frac{k}{2}\right) \right]^2 \quad (2.34)$$

En cuanto a modelos multivariados solo existen posibles extensiones de los casos

anteriores (para un conjunto fijo de distribuciones marginales hay infinitas distribuciones conjuntas que tengan dichas marginales). El problema es que no pueden tener en cuenta todas las correlaciones dentro de la región, ya que recurren a un único parámetro asociado.

2.3.2.- Datos regionales únicamente

En el caso de que se tenga que realizar el análisis de crecidas en una cuenca no aforada, se puede hacer uso de los datos regionales de aforos. Los pasos a seguir serían:

- i) Realizar el análisis regional con alguna de las metodologías anteriormente descritas. Por ejemplo, en el caso del boletín 17B del "U.S. Water Resources Council" bastaría con tomar como coeficiente de sesgo de la cuenca no aforada el regional.
- ii) Obtener los parámetros locales necesarios mediante regresión respecto a las características de la cuenca. En una técnica del tipo "index flood" se estimaría X_1 con alguna ecuación de regresión sobre la superficie, precipitación media, pendiente media, etc... de la cuenca.

Sin embargo, hay que tener en cuenta que el error cometido con los parámetros locales si se obtienen mediante una regresión puede ser muy grande.

2.3.3.- Obtención de las regiones homogéneas

Una hipótesis básica en el análisis regional de crecidas es la de que la variable estadística estandarizada tiene la misma distribución en todas las estaciones de la región considerada. Esto supone que algunos de los parámetros estadísticos dentro de la región sean homogéneos, es decir, respecto a ese o esos parámetros la región debe de ser homogénea.

Evidentemente la forma más sencilla de regionalización es por proximidad geográfica. Con este método no se consideran las características hidrológicas y geomorfológicas de las cuencas vertientes, además de no seguir un criterio objetivo de clasificación. Siempre puede servir como aproximación previa a otra regionalización más depurada.

Los métodos de regionalización se basan en las técnicas de clasificación de datos utilizadas en otros campos de la ciencia. Se puede clasificar un conjunto de cuencas de una región en subregiones homogéneas mediante las siguientes técnicas:

i) Análisis factorial

Empleado por primera vez por White (1975). La clasificación se obtiene al maximizar la varianza de combinaciones lineales de características hidrogeomorfológicas.

ii) Clusterización

Es una clasificación en el espacio de características, bien utilizando una función objetivo o mediante alguna definición de distancia. De Coursey (1972) y Tasker (1982) utilizan técnicas de clusterización definiendo la distancia entre las cuencas mediante la distancia euclídea de las características consideradas: área de drenaje, precipitación media anual e intensidades de precipitación con períodos de retorno 2 y 100 años. La selección de características a tener en cuenta es en base a un análisis factorial previo de las mismas.

iii) Ajuste de una GEV regional

Wiltshire (1985) define un estadístico que mide el ajuste de una distribución GEV a los datos de un conjunto de estaciones con el método del "Natural Environment Research Council" (1975). Ese estadístico es el utilizado para realizar la clasificación de las estaciones en los diferentes grupos.

iv) Clusterización y ajuste de una GEV regional

Acreman y Sinclair (1986) utilizan las dos metodologías anteriores iterativamente.

A partir de las características consideradas en la regionalización utilizada, se podrá obtener la "distancia" al centro de cada región y situarla en la que esté más próxima. Wiltshire (1985) propone para cuencas no aforadas el estimar el cuantil X_T como si perteneciera a todas las regiones y posteriormente obtener el cuantil X_T pesando cada resultado con el inverso de la "distancia" definida.

2.4.- EMPLEO DE DATOS HISTORICOS

En ciertos ríos se conoce la existencia y el valor de crecidas anteriores a la serie de caudales sistemáticos de la estación de aforos. Esta tipo de información puede ser o bien información histórica o información sobre paleoavenidas. La información histórica es definida por todas aquellas medidas previas al registro sistemático. Estas medidas, habitualmente acerca de niveles alcanzados durante las avenidas, se pueden localizar en los periódicos locales, crónicas de los municipios, marcas en el llano de inundación, etc... Por otra parte, la información sobre paleoavenidas es previa a estas medidas históricas y puede ser obtenida por las siguientes técnicas:

- i) Estudios geológicos de los sedimentos del río y de los depósitos erosionados, especialmente los llamados "slackwater deposits".
- ii) Datación mediante carbono 14 de los materiales orgánicos depositados durante las avenidas.
- iii) Ciertas especies arbóreas, cuando se sumergen durante un período prolongado, sus troncos sufren decoloraciones y desarrollos anormales de los anillos.

Más detalles describiendo estas técnicas y su aplicación en el análisis de la frecuencia de las avenidas pueden encontrarse en Costa (1978), Kochel y Baker (1982), Hupp (1982), Baker et al. (1983), Lane (1987) y Stedinger y Baker (1989).

El uso de la información histórica en la predicción de cuantiles de avenidas fue descrito por primera vez por Benson (1950) para el río Susquehanna a su paso por la ciudad de Harrisburg (EEUU). Los datos históricos que utilizó en su estudio fueron censurados por debajo de un cierto nivel de la lámina de agua durante 88 años previos a la instalación de una estación de aforos. La estimación de los cuantiles la realizó mediante la técnica de asignación de probabilidad con la función de densidad de probabilidad acumulada empírica, haciendo uso de los datos históricos y sistemáticos.

Leese (1973) obtuvo los estimadores por Máxima Verosimilitud de una distribución Gumbel para una muestra sin censurar, censurada por encima de un cierto límite (valores perdidos) o censurada por debajo (información histórica). Este tipo de censura en la que se conoce el límite censor se denomina del tipo 1. Censura del tipo 2 es aquella en la que se fija el número de valores censurados. Leese calculó también los Límites Inferiores de Cramer-Rao de los parámetros y cuantiles. Las hipótesis que hace son:

- i) La longitud del período histórico es independiente de la magnitud de las avenidas.
- ii) Todas las avenidas por encima del límite censor son conocidas.
- iii) Existe consistencia estadística entre los periodos histórico y sistemático.

Hace una aplicación práctica para el río Avon a su paso por la ciudad de Bath (Gran Bretaña). El límite histórico adoptado es ligeramente inferior a la menor avenida histórica. La reducción en la varianza asintótica de la media muestral es del orden del 33 %. Sin embargo, apunta que esta ganancia estadística puede ser menor debido a la incertidumbre en la elección del límite censor.

Tasker y Thomas (1978) comparan la bondad de 4 estimadores por el Método de los Momentos Ajustados utilizando información histórica para una distribución LogPearson tipo III. Los resultados de un experimento de Montecarlo se reflejan como función del error cuadrático medio, probabilidad de sobreestimación y valor esperado del coste de oportunidad. Dependiendo del escenario cada método funciona mejor, aunque en la mayoría de los casos los estimadores de los cuantiles mejoran utilizando la información histórica.

Gerard y Karpuz (1979) obtienen la curva de densidad de probabilidad acumulada empírica para máximos niveles de avenidas utilizando diversas fuentes de información histórica. Subrayan que existe un "nivel de percepción" para cada una de las fuentes, que lo definen como "... el nivel por encima del cual se espera que la fuente habrá recogido la información de la avenida en cualquier año".

El "United States Water Resources Council Hydrology Committee" en el boletín 17B (1982, apéndice 6A) recomienda la incorporación de la información histórica por aplicación del teorema de Bayes a la distribución LogPearson tipo III. Obtiene un factor compensador para el recálculo de la media, desviación típica y coeficiente de sesgo que se utilizan en el Método de Estimación de los Momentos clásico. El resultado es el denominado Método de los Momentos Ajustados.

Condie y Lee (1982) comparan mediante simulaciones de Montecarlo la bondad de un pseudo Método por Máxima Verosimilitud y el de los Momentos Ajustados para una distribución Lognormal de 3 parámetros. Describen 3 tipos de información histórica:

- i) Noticias sobre las avenidas que han desbordado el cauce del río. Esta sería información histórica censurada del tipo 1 por debajo del límite.

ii) Noticias sobre la máxima avenida ocurrida, que es información histórica censurada del tipo 2.

iii) Noticias sobre las mayores avenidas extraordinarias, que vuelve a ser una censura del tipo 2 pero con más de una avenida histórica.

Ellos muestran una aplicación para el río Irwell a su paso por Adelphi Weir (Inglaterra). Los cuantiles de 50, 100 y 500 años de período de retorno son muy similares con ambos métodos pero no así los parámetros. Mediante simulaciones por Montecarlo llegan a la conclusión de que el error cuadrático medio de los parámetros es menor con el Método de Máxima Verosimilitud que con el de los Momentos Ajustados, pero muy similares para los cuantiles. Este último resultado entra en contradicción con experiencias posteriores (por ejemplo las de Stedinger y Cohn, 1986) que muestran que en términos del error cuadrático medio siempre el método de Máxima Verosimilitud es mejor que el de los Momentos Ajustados.

Una posible razón de éste resultado engañoso es que Condie y Lee utilizan una función de verosimilitud para información histórica censurada del tipo 1, cuando la información que hacen uso en las simulaciones por Montecarlo es del tipo 2.

Condie (1986) obtiene, aunque no de forma explícita, los estimadores de los parámetros y sus Límites de Cramer-Rao de una distribución Lognormal de tres parámetros, como una extensión de su trabajo junto con Lee descrito anteriormente.

Hosking y Wallis (1986a y b) estudian el valor de la información histórica sujeta a errores en diferentes escenarios. En su primer artículo consideran que se conoce la máxima paleoavenida en un período de 1000 a 10000 años, y en el segundo artículo la máxima avenida histórica en un período de 100 a 200 años.

Según ellos, los diferentes factores que pueden afectar la utilidad de la información histórica son:

- i) Naturaleza de la distribución de la frecuencia de las avenidas.
- ii) Longitud del período de registro sistemático.
- iii) Número de estaciones en un análisis regional.
- iv) Heterogeneidad entre la estaciones.

- v) Correlación en y entre las series de avenidas.
- vi) Errores de medida en los datos sistemáticos.
- vii) Tipo de información histórica.
- viii) Errores en la información histórica.
- ix) No estacionaridad de las series de avenidas.
- x) Elección de la función de distribución a ajustar.
- xi) Método de ajuste.
- xii) Período de retorno del cuantil a estimar.
- xiii) Criterio de bondad de la estimación.

Los diferentes escenarios que Hosking y Wallis simulan por Montecarlo son:

- i) Datos locales y ajuste de una distribución EV1.
- ii) Datos locales y ajuste de una distribución GEV.
- iii) Dos estaciones de aforos y ajuste de una distribución EV1.
- iv) Dos estaciones de aforos y ajuste de una distribución GEV.
- v) Gran número de estaciones de aforos y ajuste de una distribución GEV.

En todos los casos la población es EV1. En sus simulaciones por Montecarlo introducen un error en el valor de la máxima avenida utilizando un coeficiente de error con distribución lognormal. Sus principales conclusiones son:

- i) La ganancia estadística utilizando información histórica puede ser eliminada si los errores en la determinación de la máxima avenida son del orden del 20 al 50 %.
- ii) No hay transferencia de información histórica en análisis regional.

La principal objeción a su estudio es que sus conclusiones son solo válidas para el caso de conocer exclusivamente la máxima avenida histórica, pero esto es solo una forma de información histórica, como se argumentará en el Capítulo siguiente.

Stedinger y Cohn (1986 y 1987), y Cohn y Stedinger (1987) comparan la bondad de tres métodos de estimación para información histórica censurada del tipo I, es decir, con un límite censor. Los métodos son el de los Momentos Ajustados, Máxima Verosimilitud con información censurada y Máxima verosimilitud con información binomial censurada. La distribución de la población es lognormal de dos parámetros. Sus conclusiones son:

- i) "... la información histórica y de paleoavenidas puede tener un tremendo valor en el análisis de la frecuencia de las avenidas".
- ii) Las varianzas asintóticas y los errores cuadráticos medios muestrales son similares.
- iii) Tanto para períodos cortos como largos de información histórica, la información binomial censurada es casi tan buena como la censurada, sobre todo cuanto mayor sea el período de retorno del cuantil.
- iv) El método de Máxima Verosimilitud es más robusto y eficiente que el de los Momentos Ajustados.

Stedinger y Baker (1987) hacen una revisión de todos los procedimientos geobotánicos para el estudio de las paleoavenidas, remarcando que las investigaciones futuras se centrarán en las técnicas de "slackwater deposits" e indicadores de paleoniveles. También hacen una revisión del posible uso de esta información en el estudio de las avenidas, haciendo hincapié en la influencia de los errores en la información histórica y en la posibilidad de incorporación de la misma a un análisis regional.

Wall, Kibler, Newton y Herrin (1987) y Wall, Kibler y Hastings (1987), comparan tres métodos de estimar picos de avenidas en cuencas no aforadas en el valle del Tennessee (EEUU): un procedimiento de estimación regional, estimación no paramétrica con la fórmula de Hazen para la información histórica, y utilización conjunta de la información sistemática regional e histórica local mediante una varianza compensada. Asumen diferentes poblaciones y comparan los resultados con los obtenidos con la serie sistemática local. La incorporación de la información histórica a partir de la varianza compensada incrementa la fiabilidad de los estimadores de los cuantiles, e incluso los estimadores con la fórmula de Hazen son mejores que los del análisis regional.

Lane (1987), estudiando el río Pecos (EEUU) encuentra que el método de los Momentos Ajustados (recomendado por el "U. S. Water Resources Council") para la incorporación de la información histórica empeora los estimadores. Propone una alternativa que es la de añadir información histórica sintética.

Hirsh y Stedinger (1987) comparan diferentes fórmulas de asignación empírica de probabilidad con la presencia de información histórica. Introducen dos nuevas fórmulas que llaman de excedencia y bayesiana. La precisión de las nuevas expresiones son solo algo mayor que las clásicas fórmulas de Hazen y Weibull, y solo se muestran claramente superiores en un contexto estadístico bayesiano. En tres apéndices estudian las influencias de los errores en la longitud histórica y en el límite censor, y el caso de presencia de múltiples límites censores. La utilización de fórmulas de asignación empírica también es objeto de los trabajos de Adamowski y Feluch (1990) y Guo y Cunnane (1991).

Phien y Fang (1989) analizan las ventajas de incorporar información censurada de tipo I para la distribución GEV, utilizando el método de Máxima Verosimilitud. Sus conclusiones son muy favorables al uso de este tipo de información adicional.

Por último, Pilon y Adamowski (1993) realizan un estudio similar pero para la distribución LogPearson tipo III, llegando al mismo tipo de resultados.

2.5.- CRITERIOS DE SELECCION DEL MODELO

Uno de los puntos cruciales en el análisis de la frecuencia de las crecidas es la selección del modelo estadístico utilizado. Los criterios a seguir en esta elección que se pueden encontrar en la literatura son:

i) Personales o institucionales

La experiencia personal o de otros técnicos puede indicar cual es la metodología que mejor se adapta a los problemas habituales en la región. Por otra parte, en países como los EEUU o el Reino Unido existen recomendaciones a seguir en el análisis de crecidas, lo cual no quiere decir que sean la única solución. En lo que respecta a España, la vieja Instrucción de Grandes Presas (1967) menciona este tipo de estudio, recomendando (art. 14.6) "... establecer una función entre caudales y períodos de recurrencia bien por extrapolación estadística ..., o bien por deducción mediante la aplicación de coeficientes adecuados a la superficie de la cuenca receptora, ...". Dado que las técnicas de regionalización son relativamente recientes, la Instrucción española no hace mención de ellas.

ii) Gráficos

La utilización de los gráficos se puede hacer desde dos enfoques distintos:

- a) Utilizando papel probabilístico. Si al representar la distribución de frecuencias empírica se obtiene una línea recta, los datos siguen la distribución correspondiente al tipo de papel.
- b) Comparando la distribución de frecuencias empírica con las distintas curvas teóricas consideradas. La curva de distribución de frecuencias empírica se obtiene asignando una probabilidad de no excedencia a las crecidas clasificadas (ordenadas de menor a mayor o viceversa). Es la técnica más empleada en Hidrología, por lo que más detalles se pueden encontrar en el Anejo A.

iii) Tests estadísticos de ajuste

Se pueden utilizar los clásicos tests de ajuste, como el Chi cuadrado o el Smirnov-Kolmogorov. Sin embargo, estos tests no son sensibles al desarrollo del modelo en la cola

derecha, que es precisamente la zona de interés con las avenidas. Esto da lugar a que casi todas las distribuciones de extremos pasen estos tests, siendo que los resultados en cuanto a cuantiles de alto período de retorno son bien diferentes de una familia de distribuciones a otra.

Existe un test de verosimilitud modificado desarrollado por Hosking (1984) que tiene en cuenta este efecto, pero que solo es valido para la familia de distribuciones GEV.

iv) Tests basados en las propiedades estadísticas de las poblaciones

Los momentos de orden superior al segundo son más representativos de la forma de la cola derecha de la distribución. Por ello es conveniente el utilizar la curva teórica de los coeficientes de sesgo y kurtosis y comprobar que no se aleja del punto correspondiente a nuestra serie de datos, tanto a nivel local como regional. También conviene analizar si se presentan los efectos de "pata de perro" y condición de separación en nuestros datos. En el caso de que sea así, habrá que considerar distribuciones del tipo mixto.

v) Análisis de robustez del modelo

Se considera como modelo el conjunto tipo de distribución y método de estimación (D/E). Una D/E es robusta si estima el cuantil con pequeños sesgo y error cuadrático medio, incluso si la muestra sobre la que se realiza la estimación está extraída de otra distribución. La robustez es por tanto una característica intrínseca del modelo.

El análisis de robustez requiere el uso de técnicas de generación del tipo Montecarlo. Existen diversos estudios sobre este tema como los de Landwehr et al. (1980), Kuczera (1982b) y Lettenmaier et al. (1985), aunque sin llegar a conclusiones generales. Si se puede decir que para series cortas son más robustos los modelos de pocos parámetros estimados por el método de los Momentos Compensados y para series largas o en un contexto regional el mejor método de estimación es el de Máxima Verosimilitud. El método de los Momentos Compensados fue introducido por primera vez por Greenwood et al. (1979) y Landwehr et al. (1979).

2.6.- CONCLUSIONES

En el análisis de la frecuencia de las crecidas intervienen los factores de identificación de la distribución adecuada, determinación del mejor método de estimación, elección entre un análisis local o regional, definición de una región homogénea y tratamiento apropiado de los problemas especiales (series censuradas, años con valores nulos, etc...). Durante los últimos 10 años se han publicado gran cantidad de artículos estudiando estos temas. Es por ello de esperar que los libros de Hidrología, manuales y recomendaciones institucionales sean revisados y puestos al día.

El método de análisis por ajuste de una distribución Gumbel (en Europa) o una Log-Pearson tipo III (en EEUU) a los datos de una única estación está desfasado. El análisis regional por "index flood" (y ajuste de una "vieja" GEV o Wakeby) o por TCEV obtiene en cualquier caso mejores estimadores.

En regiones donde exista más de un mecanismo de producción de avenidas (como por ejemplo los ríos de la cuenca mediterránea española) la condición de separación será mucho más fuerte y solo las distribuciones mixtas pueden representarla. El único problema que presentan estas distribuciones es el alto número de parámetros, que condiciona su uso a los análisis regionales exclusivamente. La distribución TCEV se considera que es la más adecuada para las características de los ríos mediterráneos.

El método de estimación más robusto es el de Máxima Verosimilitud, aunque para series cortas funciona algo mejor el de los Momentos Compensados. En cualquier caso, ambos resultan superiores al más conocido Método de los Momentos. Los métodos de estimación regional obtenidos por aplicación de la Teoría de la Decisión de Bayes, aún siendo muy atractivos, han sido abandonados en los últimos años, dado el gran error cometido en la distribución "a priori" de los parámetros, condicionada a ciertas características geomorfológicas de las cuencas.

La utilización de la información histórica existente aumenta en gran medida la fiabilidad de los estimadores de los cuantiles, aunque puede estar sometida a mayores errores que la información sistemática. Sin embargo, modelos que incorporen la información histórica disponible dentro de un análisis regional y un estudio riguroso de las ventajas que se producen con ello, no han sido realizados hasta la actualidad para la distribución TCEV. Este va a ser el principal objetivo del presente trabajo.

3.- LA INCORPORACION DE LA INFORMACION HISTORICA

3.1.- JUSTIFICACION DE LA METODOLOGIA ADOPTADA

3.1.1.- Introducción

Los ríos mediterráneos se caracterizan por presentar el efecto "pata de perro" descrito por Potter (1958). Este efecto se puede observar claramente en las curvas de densidad de probabilidad acumulada de los ríos Júcar, Turia y Segura presentadas en el Capítulo 6. Esto hace que distribuciones clásicas como la distribución de Gumbel sean inadecuadas en el análisis de la frecuencia de las avenidas de estos ríos. La exhaustiva revisión de literatura y la experiencia en el Departamento de Ingeniería Hidráulica de la Universidad Politécnica de Valencia apuntaron al uso de la distribución "Two Component Extreme Value" (o TCEV) como representativa de los ríos de la cuenca mediterránea española.

Aunque el objetivo fundamental de esta investigación es el analizar en profundidad la distribución TCEV, se estudian previamente las tres distribuciones correspondientes a la familia "Extreme Value" de dos parámetros, que son: la LogGumbel, la Weibull y la Gumbel. Las razones de ello son, por una parte, el que son más sencillas que la TCEV, y por tanto de su análisis se pueden extraer conclusiones que faciliten el estudio de la TCEV. Y por otra parte, son de amplio uso en Hidrología, especialmente las funciones de distribución Gumbel en España y LogGumbel en Francia.

El análisis se va a realizar primero creando los modelos estadísticos que sean capaces de tener en cuenta la información histórica disponible, y segundo explorando los aumentos de fiabilidad de los estimadores de los cuantiles mediante dicha utilización. El análisis regional solo se va a estudiar y aplicar para la distribución TCEV, que por su número de parámetros debe de ser utilizada con abundancia de datos. En este trabajo, no se va a distinguir entre información histórica e información sobre paleoavenidas, tal como hacen algunos autores en el Capítulo de Antecedentes. Aunque tienen un origen bien distinto, desde el punto de vista estadístico son exactamente el mismo tipo de información.

Dos aspectos muy importantes en Hidrología no van a ser considerados aquí: la posible no estacionaridad de las series de avenidas, y los errores de la información, especialmente en la información histórica.

La constancia en el tiempo de las características de la cuenca y de los mecanismos que afectan a la producción de avenidas, es algo que no se puede garantizar, sobre todo durante períodos largos de tiempo, que es la situación cuando se emplea información histórica.

Normalmente, la información histórica original es acerca de niveles alcanzados durante las crecidas, y muy raramente se dispone del dato de caudal. La conversión correcta de niveles a caudales supone un conocimiento exacto de las características hidráulicas del tramo de río en cuestión, y esto se hace tanto más difícil cuanto más se intente retrotraerse en el tiempo. Esta situación también se presenta en las grandes avenidas de la serie sistemática, que pueden situarse fuera del límite de aplicación de la curva de gastos de la estación de aforos, o que pueden haber destruido la propia estación.

El método de estimación de parámetros y cuantiles que vamos a emplear es exclusivamente el Método de Máxima Verosimilitud. Se trata de uno de los métodos más robustos y, además, en este método la incorporación de la información histórica es "más natural", como se verá en el Apartado 3.3. El método de Bayes, tan atractivo a priori como éste, no se ha utilizado porque requiere el conocer la distribución de los parámetros en función de algunas de las características geomorfológicas de la cuenca, cuestión no resuelta hasta el momento y que parece muy difícil que se resuelva en el futuro.

3.1.2.- Tipos de información histórica considerados

Para poder desarrollar modelos que incorporen la información histórica hay que conocer sus diferentes formas de presentación. De la experiencia adquirida en la búsqueda de información histórica, ésta se va a clasificar en tres tipos (figura 3.1):

i) Información histórica binomial censurada

Este tipo de información consiste en conocer que durante el período histórico previo a la serie sistemática hubo un cierto número de avenidas que superaron un determinado umbral de percepción histórico, de tal forma que las avenidas inferiores a dicho umbral no han dejado ningún tipo de constancia. Este umbral puede ser real o ficticio, sin que afecte teóricamente al modelo estadístico. Ejemplo de un umbral ficticio es la información histórica utilizada en este trabajo para el río Turia a su paso por la ciudad de Valencia. El umbral es el correspondiente al cauce antiguo del río a sección llena, que cuando se supera provoca inundaciones en la ciudad. Evidentemente, el umbral real es mucho menor, ya que hay constancia de muchas avenidas por debajo de ese nivel. Sin embargo, el umbral real de percepción es difícil de conocer, entre otras cosas porque puede cambiar en el tiempo. Por ejemplo, se produce una modificación del umbral de percepción al variar la sensibilidad de la ciudad frente a las avenidas, o por modificaciones en el lecho del río.

ii) Información histórica censurada

Es similar a la anterior, pero además no solo se conoce el número sino también el valor de las avenidas que superaron el límite de percepción histórico. Es una información más abundante que en el caso de binomial censurada. Existe una cierta polémica sobre si esta información es censurada de tipo 1 o 2, es decir, si el elemento censor es el límite de percepción o no existe tal límite y se está seleccionando a priori el número de avenidas históricas. Nos inclinamos por que exista siempre un límite de percepción real de la información histórica. Por poner un ejemplo, los periódicos no recogen en sus primeras planas "las 3 avenidas más importantes del siglo", sino que recogen "3 avenidas de gran magnitud, que son las 3 más importantes del siglo". El elemento censor no es el número, sino el límite de percepción del periódico. Otra cosa es que sea difícil determinar ese límite censor, pero como en el caso de la información histórica binomial censurada, siempre se puede recurrir a un límite ficticio mayor más fácil de obtener.

iii) Información histórica sobre la máxima avenida

En este caso, la información que se tiene solamente es el valor de la máxima avenida ocurrida durante el período histórico. Este es el único caso que se considera información censurada de tipo 2. Se puede presentar como resultado de un estudio sobre paleoavenidas, en el que la máxima a borrado las huellas de las avenidas menores, o bien de una información histórica confusa, en la que artificialmente se desprecia la existencia del límite de percepción.

3.1.3.- Medida de la bondad en la estimación de los cuantiles

La fiabilidad de la estimación de un cuantil X_T se puede medir por el error cuadrático medio muestral (MSE) del estimador, definido como:

$$MSE(\hat{X}) = E[(\hat{X}_T - X_T)^2] \quad (3.1)$$

donde $E[.]$ es el operador valor esperado.

Este error es muy complejo de obtener analíticamente, por lo que se recurre a

simulaciones del tipo Montecarlo para su obtención aproximada. Otra forma de resolver el problema es hacer 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 de Máxima Verosimilitud es asintóticamente insesgado, por lo que el CRLB es el límite inferior de su MSE cuando la información utilizada tiende a infinito. El CRLB sí que es más asequible de obtener analíticamente. Si se demostrase que es una buena aproximación del verdadero MSE muestral, podría ser utilizado en lugar de éste.

En este trabajo realmente se analizarán no las varianzas asintóticas y error cuadrático medio, sino las ganancias estadísticas asintóticas SG y muestrales SG*, definidas como:

$$SG_i = 1 - \frac{CRLB_i}{CRLB_0} \quad (3.2)$$

$$SG^*_i = 1 - \frac{MSE_i}{MSE_0} \quad (3.3)$$

donde el subíndice 0 indica que solo se utiliza la información sistemática y el subíndice i corresponde a los diferentes usos de información histórica.

La mayor parte de los estudios previos solo consideran simulaciones de Montecarlo, olvidando que, si los valores asintóticos son lo suficientemente aproximados, se pueden descubrir propiedades interesantes acerca del uso de la información histórica. Incluso puede servir para simplificar las simulaciones si se está empeñado en hacerlas.

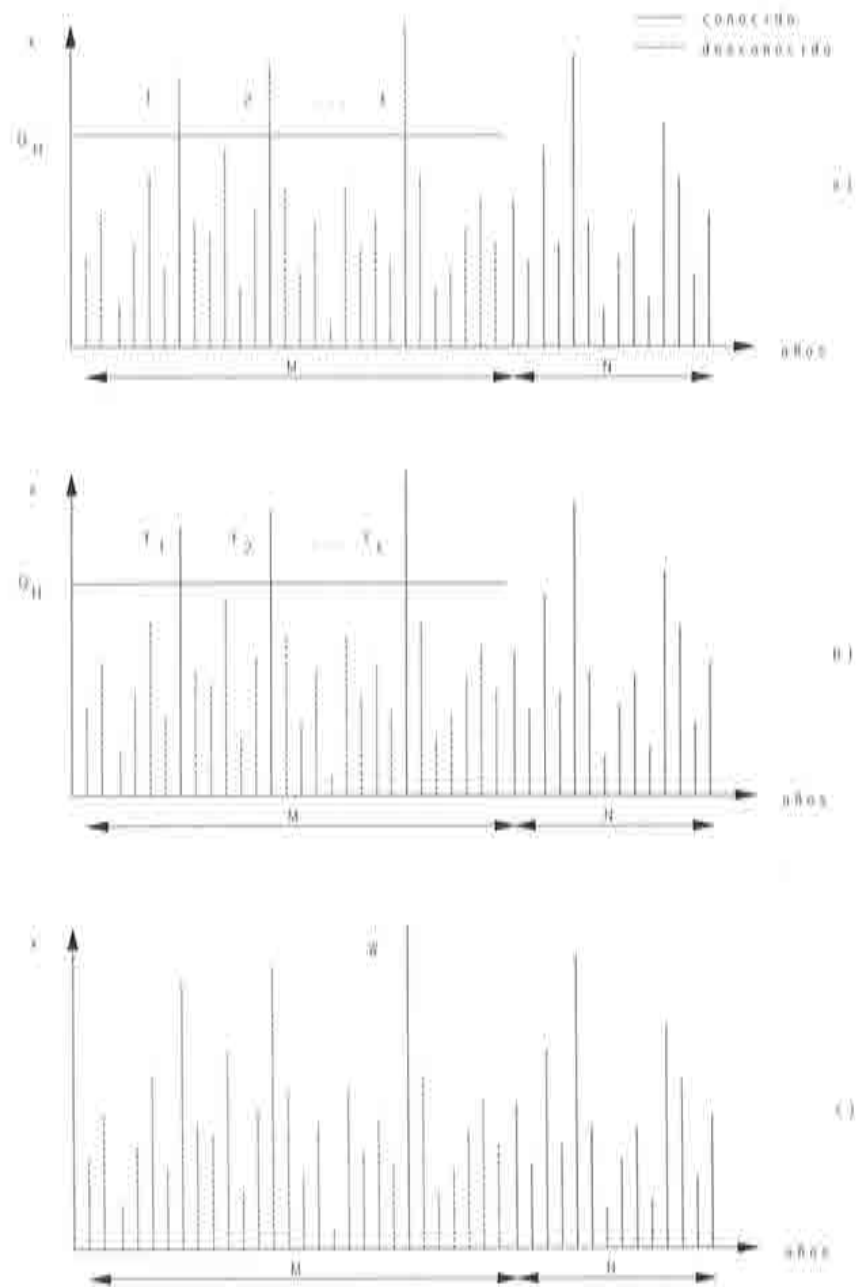


Figura 3.1. Tipos de información histórica o paleohistórica: a) binomial censurada, b) censurada y c) máxima avenida.

3.2.- HIPOTESIS CONSIDERADAS

Los modelos estadísticos esquematizados en el siguiente apartado y desarrollados en los Capítulos 4 y 5, presentan en todo momento unas hipótesis básicas, que al mismo tiempo los limitan. Estas hipótesis son las adoptadas habitualmente en Hidrología, y son:

i) La avenida anual (bien sea máximo instantáneo o máximo diario) es una variable estadística continua. Los posibles valores negativos de la variable aleatoria se asimilan a la probabilidad de un año completamente seco.

ii) La población de avenidas sigue exactamente la función de distribución que en cada momento se analice.

iii) Las realizaciones de la variable aleatoria, cuyo resultado es la serie de aforos, son independientes e idénticamente distribuidas.

iv) La información sistemática e histórica no presenta errores.

v) Los tipos de información histórica a considerar son solo binomial censurada (BC), censurada (CE) y de máxima avenida (MF), según se ha descrito en el Apartado 3.1.

vi) En análisis regional, las estaciones de aforos consideradas cumplen la hipótesis de homogeneización.

vii) El método de estimación será siempre el de Máxima Verosimilitud.

3.3.- PROCEDIMIENTO ESTADISTICO

3.3.1.- Función de verosimilitud

Uno de los principales objetivos de este trabajo es crear un modelo estadístico que pueda hacer uso de la información histórica. El modelo escogido es el que tiene como método de estimación de parámetros el de Máxima Verosimilitud. Este método 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. Para 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:

$$f_X(x) = \prod_{i=1}^N f_X(x_i) \quad (3.4)$$

donde:

\underline{x}, x_i = registro sistemático

$f_X(\cdot)$ = función de densidad de probabilidad de la variable aleatoria X , común para toda la serie de aforos

N = longitud del registro sistemático

Por tanto, una función de verosimilitud sería:

$$L_{SY}(\Theta) = \prod_{i=1}^N f_X(x_i; \Theta) \quad (3.5)$$

donde Θ es el conjunto de parámetros de la distribución considerada.

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

$$f_k(k) = \binom{M}{k} (1-p)^{M-k} p^k \quad k=0,1,\dots,M \quad (3.6)$$

con lo que la función de distribución conjunta quedaría como:

$$f_{X,k}(x,k) = \binom{M}{k} (1-p)^{M-k} p^k \prod_{i=1}^N f_X(x_i) \quad (3.7)$$

donde:

p = probabilidad de excedencia de Q_{tp} , función de los parámetros del modelo

M = longitud del período histórico

Y una función de verosimilitud sería:

$$L_{BC}(\Theta) = (1-p)^{M-k} p^k \prod_{i=1}^N f_X(x_i; \Theta) \quad (3.8)$$

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

$$f_Y(y) = \frac{f_X(y)}{p} \quad (3.9)$$

La función de distribución conjunta es ahora:

$$f_{X,Y,k}(x,y,k) = \binom{M}{k} (1-p)^{M-k} p^k \prod_{i=1}^N f_X(x_i) \prod_{j=1}^k f_Y(y_j) \quad (3.10)$$

donde:

y_j = avenidas históricas

$f_Y(\cdot)$ = función de densidad de probabilidad de Y

Y la función de verosimilitud a emplear, después de sustituir la ecuación (3.9) en la ecuación (3.10), es:

$$L_{CE}(\Theta) = (1-p)^{M-k} \prod_{i=1}^N f_X(x_i; \Theta) \prod_{j=1}^k f_X(y_j; \Theta) \quad (3.11)$$

Si la única información disponible durante el período histórico es la máxima avenida W , su función de densidad de probabilidad es:

$$f_W(w) = M f_X(w) F_X^{M-1}(w) \quad (3.12)$$

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

$$f_{X,W}(x,w) = \prod_{i=1}^N f_X(x_i) f_W(w) \quad (3.13)$$

Que, sustituyendo la ecuación (3.12), da lugar a la función de verosimilitud:

$$L_{MF}(\Theta) = \prod_{i=1}^N f_X(x_i; \Theta) f_X(w; \Theta) F_X^{M-1}(w; \Theta) \quad (3.14)$$

En la mayor parte de los casos es mucho más ventajoso maximizar el logaritmo de la función de verosimilitud. Con ello, los productos de la ecuaciones anteriores se convierten en sumatorios y las exponenciales desaparecen.

En el caso de tener más de un tipo de información histórica, sería inmediato obtener la función de verosimilitud por combinación de las anteriores.

3.3.2.- Obtención de las ganancias estadísticas asintóticas de los cuantiles

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 I. Los elementos de esta matriz se obtienen como:

$$I_i(j,k) = E \left[- \frac{\partial^2 LL_i(\Theta)}{\partial \Theta_j \partial \Theta_k} \right] \quad (3.15)$$

donde:

$LL_i(\cdot)$ = logaritmo de la función de verosimilitud

Θ_j, Θ_k = parámetros del modelo

i = tipo de información utilizada: sistemática, binomial censurada, censurada o de máxima avenida

La varianza asintótica del estimador de cualquier función $g(\cdot)$ de los parámetros es función de la matriz de Fisher y el vector de derivadas parciales de la función respecto de los parámetros. En particular, para el cuantil de período de retorno T viene dada por:

$$CRLB_i(\hat{X}_T) = \underline{V} I^{-1} \underline{V}' \quad (3.16)$$

donde el elemento j del vector \underline{V} viene dado por la ecuación:

$$V(j) = \frac{\partial X_T}{\partial \Theta_j} \quad (3.17)$$

y la ganancia estadística del estimador del cuantil quedaría según la ecuación (3.2).

Las demostraciones de estas expresiones para el caso de distribuciones de un único parámetro, pueden encontrarse en Mood et al. (1974), y para múltiples parámetros en Kendall y Stuart (1967).

3.3.3.- Obtención de las ganancias estadísticas muestrales

Para obtener la ganancia estadística muestral se requiere el realizar un número grande de simulaciones de Montecarlo con la siguiente estructura:

- i) Generar sintéticamente una serie sistemática y, si fuera necesaria, la información histórica.
- ii) Obtener el cuantil por el método de Máxima Verosimilitud.
- iii) Repetir i) y ii) el número de veces considerado.
- iv) Calcular el error cuadrático medio muestral del estimador del cuantil como:

$$MSE_i(\hat{X}_T) = \frac{1}{S} \sum_{j=1}^S (X_{Tj} - X_T)^2 \quad (3.18)$$

donde:

X_T = verdadero valor del cuantil de período de retorno T

X_{Tj} = 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 (3.18) 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. La ganancia estadística muestral se obtiene mediante la ecuación (3.3).

El núcleo central de la generación de series sintéticas es siempre un generador de números aleatorios con distribución uniforme entre 0 y 1. El algoritmo empleado es el siguiente:

$$S_i = \text{mod} (16807 S_{i-1}, 2^{31}-1)$$

$$U_i = \frac{S_i}{2^{31}-1} \quad (3.19)$$

donde:

S_i = semilla, número aleatorio uniforme entre 0 y $2^{31}-1$

U_i = número aleatorio uniforme entre 0 y 1

Este algoritmo pertenece a la familia de los generadores congruenciales lineales. Es de los más simples y al mismo tiempo de los más rápidos. Se estudió el uso de otros algoritmos más complejos: el de L'Ecuyer, descrito por Bartley et al. (1987), y combinaciones de congruenciales, descritos en Press et al. (1989). A pesar de su simplicidad cumplió las condiciones de uniformidad, no correlación y recirculación de alto período necesarias en esta investigación, por lo que fue el algoritmo utilizado.

Otro elemento crítico en la simulación por Montecarlo es la subrutina de obtención de los estimadores de los parámetros. Estos estimadores son los que maximizan las diferentes funciones de verosimilitud, definidas en el Apartado 3.3.1, y los posibles algoritmos son siempre iterativos, es decir, necesitan repetir varias veces los cálculos. El número de repeticiones será función del criterio de convergencia elegido. Como la estimación se va a realizar un número muy elevado de veces, la casuística es muy grande y la probabilidad de que en alguna simulación el algoritmo no converga debe de ser lo más pequeña posible. Todo ello, sin incrementar en exceso el tiempo de computación. Con estos criterios se ha seleccionado para cada distribución y nivel de información el algoritmo más conveniente.

**4.- ANALISIS DE LA INCORPORACION DE LA INFORMACION
HISTORICA A LAS DISTRIBUCIONES "EXTREME VALUE"**

4.1.- INTRODUCCION

La familia de funciones de distribución "Extreme Value" (EV) está formada por las distribuciones Gumbel o EV1, LogGumbel o EV2 y Weibull o EV3. Aunque ninguna de ellas se ajusta a los ríos Júcar, Turia y Segura analizados en el Capítulo 6, se consideró su estudio como un paso previo de aprendizaje al estudio de la distribución más completa que es la "Two Component Extreme Value". En cualquier caso, la familia EV ha tenido y tiene un amplio uso en todo el mundo.

En este capítulo solo se va a analizar en detalle la distribución LogGumbel, mientras que de las otras dos solo se dejará constancia de los resultados finales en los Anejos B y C, ya que las expresiones son similares y las conclusiones idénticas.

Esta similitud de resultados era de esperar, ya que mediante sencillos cambios de variable se puede pasar de una distribución a otra. Si X_1 es una variable aleatoria con distribución LogGumbel, entonces la variable aleatoria X_2 definida como:

$$X_2 = LX_1 \quad (4.1)$$

se podría demostrar que sigue una distribución Gumbel. Si se define otra variable aleatoria X_3 como:

$$X_3 = \frac{1}{X_1} \quad (4.2)$$

entonces sigue una distribución Weibull.

La elección de la distribución LogGumbel como representativa de esta familia, es debida a su amplio uso en el resto de Europa (Francia especialmente) y su buen ajuste a ríos muy caudalosos (como el río Paraná, estudiado en Francés et al., 1994). En el Apartado 4.2 se analiza en profundidad la incorporación de la información histórica a esta distribución, mientras que en el Apartado 4.3 se reflejan las principales conclusiones de este análisis, que, como ya se ha mencionado previamente, son extrapolables a toda la familia de distribuciones EV.

4.2.- ANALISIS DE LA DISTRIBUCION LOGGUMBEL

4.2.1.- Características estadísticas

Como paso previo al análisis de la distribución LogGumbel, se van a obtener algunos de los estadísticos más importantes de la misma. Su función de densidad de probabilidad acumulada viene dada por la ecuación:

$$F_X(x) = e^{-(\theta/x)^\gamma} \quad x \geq 0 \quad (4.3)$$

donde:

θ = parámetro de escala

γ = parámetro de forma

Derivando la ecuación (4.3) se obtiene la función de densidad de probabilidad:

$$f_X(x) = \frac{\gamma}{\theta} \left(\frac{\theta}{x}\right)^{\gamma+1} e^{-(\theta/x)^\gamma} \quad x \geq 0 \quad (4.4)$$

A partir de la función de densidad de probabilidad, es posible obtener la media de la distribución mediante integración, siendo el resultado final:

$$\mu = E[X] = \theta \Gamma\left(1 - \frac{1}{\gamma}\right) \quad (4.5)$$

donde $\Gamma(\cdot)$ es la función gamma.

La varianza de la distribución LogGumbel resulta ser:

$$\sigma^2 = E[(X-\mu)^2] = \theta^2 \left[\Gamma\left(1 - \frac{2}{\gamma}\right) - \Gamma^2\left(1 - \frac{1}{\gamma}\right) \right] \quad (4.6)$$

Y el sesgo:

$$\mu_3 = E[(X-\mu)^3] = \theta^3 \left[\Gamma\left(1 - \frac{3}{\gamma}\right) - 3\Gamma\left(1 - \frac{1}{\gamma}\right)\Gamma\left(1 - \frac{2}{\gamma}\right) + 2\Gamma^3\left(1 - \frac{1}{\gamma}\right) \right] \quad (4.7)$$

Con lo que el coeficiente de sesgo, más utilizado que el sesgo, es:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\Gamma\left(1-\frac{3}{\gamma}\right) - 3\Gamma\left(1-\frac{1}{\gamma}\right)\Gamma\left(1-\frac{2}{\gamma}\right) + 2\Gamma^3\left(1-\frac{1}{\gamma}\right)}{\left[\Gamma\left(1-\frac{2}{\gamma}\right) - \Gamma^2\left(1-\frac{1}{\gamma}\right)\right]^{3/2}} \quad (4.8)$$

que es función exclusiva del parámetro de forma γ . Esta independencia respecto del parámetro de escala también se reproduce en los coeficientes adimensionales de variación (función del momento central de segundo orden) y de kurtosis (función del momento central de cuarto orden).

4.2.2.- Estimadores de los parámetros por Máxima Verosimilitud

Según se vio en el Capítulo 3, los estimadores de los parámetros por Máxima Verosimilitud son aquellos que maximizan el logaritmo de la función de verosimilitud. Para el caso de disponer exclusivamente de los datos sistemáticos, la particularización de la ecuación (3.5) para la distribución LogGumbel es:

$$LL_{gy}(\Theta) = N\gamma L\theta + NL\gamma - (\gamma + 1) \sum_{i=1}^N Lx_i - \theta^\gamma \sum_{i=1}^N x_i^{-\gamma} \quad (4.9)$$

donde:

x_i = serie sistemática

$LL(\cdot)$ = logaritmo de la función de verosimilitud

Θ = conjunto de parámetros

N = longitud de la serie sistemática

En este caso, la forma más rápida y efectiva de obtener el máximo de la función es anulando las derivadas parciales de $LL(\cdot)$, y resolviendo el sistema de ecuaciones resultante. Operando en el mismo se obtiene:

$$\theta^\gamma = \frac{N}{\sum_{i=1}^N x_i^{-\gamma}} \quad (4.10)$$

$$\gamma = \frac{N}{\sum_{i=1}^N L\left(\frac{\theta}{x_i}\right) \left[\left(\frac{\theta}{x_i}\right)^\gamma - 1\right]} \quad (4.11)$$

Dado que estas ecuaciones se iban a resolver un gran número de veces en la fase de simulación por Montecarlo, era imprescindible obtener un método que fuera lo más rápido posible, pero que la convergencia estuviera garantizada. Se propusieron tres métodos: uno iterativo, el de Newton-Raphson y el de la secante. Después de un estudio de bondad de cada uno de estos métodos, el más eficaz en velocidad y convergencia fue el método iterativo, por lo que éste fue el incluido en los programas de simulación.

Si además de la serie sistemática se dispone de información histórica del tipo binomial censurada, la función de verosimilitud vendría dada por la ecuación (3.8), y su logaritmo para la distribución LogGumbel es:

$$LL_{BC}(\Theta) = -(M-k) \left(\frac{\theta}{Q_H}\right)^\gamma + kLp + N\gamma L\theta + NL\gamma - (\gamma+1) \sum_{i=1}^N Lx_i - \theta^\gamma \sum_{i=1}^N x_i^{-\gamma} \quad (4.12)$$

dónde:

M= longitud del período histórico

Q_H = límite de percepción histórico, de período de retorno H

p= probabilidad de excedencia de Q_H

k= número de avenidas históricas por encima de Q_H

El sistema de ecuaciones resultante es ahora:

$$\theta^\gamma = \frac{N}{\sum_{i=1}^N x_i^{-\gamma} + \left(M - \frac{k}{p}\right) Q_H^{-\gamma}} \quad (4.13)$$

$$\gamma = \frac{N}{\sum_{i=1}^N L\left(\frac{\theta}{x_i}\right) \left[\left(\frac{\theta}{x_i}\right)^\gamma - 1\right] + \left(M - \frac{k}{p}\right) \left(\frac{\theta}{Q_H}\right)^\gamma L\left(\frac{\theta}{Q_H}\right)} \quad (4.14)$$

El método iterativo resultó inestable para este sistema de ecuaciones, con lo que el

algoritmo utilizado fue el método de la secante, que es incondicionalmente convergente.

Si la información histórica es del tipo censurada, el logaritmo de la función de verosimilitud es:

$$LL_{CE}(\Theta) = -(M-k)\left(\frac{\theta}{Q_H}\right)^\gamma + (N+k)L\theta - (N+k)L\gamma + (\gamma+1)\sum_{i=1}^N L\left(\frac{\theta}{x_i}\right) - \sum_{i=1}^N \left(\frac{\theta}{x_i}\right)^\gamma + (\gamma+1)\sum_{j=1}^k L\left(\frac{\theta}{y_j}\right) - \sum_{j=1}^k \left(\frac{\theta}{y_j}\right)^\gamma \quad (4.15)$$

donde y_j es la serie de avenidas históricas por encima de Q_H

El sistema resultante está formado por las ecuaciones:

$$\theta^\gamma = \frac{N+k}{\sum_{i=1}^N x_i^{-\gamma} + \sum_{j=1}^k y_j^{-\gamma} + (M-k)Q_H^{-\gamma}} \quad (4.16)$$

$$\gamma = \frac{N+k}{\sum_{i=1}^N L\left(\frac{\theta}{x_i}\right)\left[\left(\frac{\theta}{x_i}\right)^\gamma - 1\right] + \sum_{j=1}^k L\left(\frac{\theta}{y_j}\right)\left[\left(\frac{\theta}{y_j}\right)^\gamma - 1\right] + (M-k)\left(\frac{\theta}{Q_H}\right)^\gamma L\left(\frac{\theta}{Q_H}\right)} \quad (4.17)$$

El método de la secante es el utilizado para resolver este sistema de ecuaciones.

Si la información adicional es acerca de la máxima avenida histórica, el logaritmo de la correspondiente función de verosimilitud es:

$$LL_{MF}(\Theta) = (N+1)L\left(\frac{\gamma}{\theta}\right) + (\gamma+1)\sum_{i=1}^N L\left(\frac{\theta}{x_i}\right) - \sum_{i=1}^N \left(\frac{\theta}{x_i}\right)^\gamma + (\gamma+1)L\left(\frac{\theta}{w}\right) - M\left(\frac{\theta}{w}\right)^\gamma \quad (4.18)$$

donde w es la máxima avenida durante el período histórico.

El sistema de ecuaciones resultante es ahora:

$$\theta^\gamma = \frac{N+1}{\sum_{i=1}^N x_i^{-\gamma} + Mw^{-\gamma}} \quad (4.19)$$

$$\gamma = \frac{N+1}{\sum_{i=1}^N L\left(\frac{\theta}{x_1}\right)\left[\left(\frac{\theta}{x_i}\right)^\gamma - 1\right] + ML\left(\frac{\theta}{w}\right)\left[\left(\frac{\theta}{w}\right)^\gamma - 1\right]} \quad (4.20)$$

También en este caso, el algoritmo utilizado es el método de la secante.

4.2.3.- Varianzas asintóticas de los estimadores de los cuantiles

La matriz de información de Fisher es el paso previo para la obtención de las varianzas y ganancias estadísticas asintóticas de los estimadores de los cuantiles. Sus elementos se obtienen a partir de los valores esperados de las segundas derivadas del logaritmo de la función de verosimilitud. Afortunadamente, las integrales que aparecen al tomar valores esperados son funciones gamma, gamma incompletas o sus derivadas, con lo que las expresiones quedan mucho más "manejables". A continuación se presentan los resultados finales de los elementos de las matrices de Fisher para los distintos niveles de información. Con información sistemática exclusivamente son:

$$I_0(1,1) = -E\left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta^2}\right] = \frac{N\gamma^2}{\theta^2} \quad (4.21)$$

$$I_0(2,2) = -E\left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \gamma^2}\right] = \frac{N[1+\Gamma''(2)]}{\gamma^2} \quad (4.22)$$

$$I_0(1,2) = I_0(2,1) = -E\left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta \partial \gamma}\right] = \frac{N\Gamma'(2)}{\theta} \quad (4.23)$$

Si la información disponible es binomial censurada, el resultado es:

$$I_1(1,1) = -E\left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \theta^2}\right] = \left(\frac{\gamma}{\theta}\right)^2 \left[N + M \left(\frac{\theta}{Q_H}\right)^{2\gamma} \frac{q}{p} \right] \quad (4.24)$$

$$I_1(1,2) = I_1(2,1) = -E\left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \theta \partial \gamma}\right] = \frac{N\Gamma'(2)}{\theta} + M \frac{\gamma}{\theta} \left(\frac{\theta}{Q_H}\right)^{2\gamma} \frac{q}{p} L\left(\frac{\theta}{Q_H}\right) \quad (4.26)$$

$$I_1(2,2) = -E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \gamma^2} \right] = \frac{N[1+\Gamma''(2)]}{\gamma^2} + M \left[\left(\frac{\theta}{Q_H} \right)^\gamma L \left(\frac{\theta}{Q_H} \right) \right]^2 \frac{q}{P} \quad (4.25)$$

donde $q = 1-p$.

Si la información disponible es censurada, el resultado es:

$$I_2(1,1) = -E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \theta^2} \right] = \left(\frac{\gamma}{\theta} \right)^2 (N+Mp) \quad (4.27)$$

$$I_2(2,2) = -E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \gamma^2} \right] = \frac{N[1+\Gamma''(2)] + M[p+\Gamma''(2,(\theta/Q_H)^\gamma)]}{\gamma^2} + M \left(\frac{\theta}{Q_H} \right)^\gamma L^2 \left(\frac{\theta}{Q_H} \right) q \quad (4.28)$$

$$I_2(1,2) = I_2(2,1) = -E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \theta \partial \gamma} \right] = \frac{N\Gamma'(2) + M[\Gamma'(2) - \Gamma'(2,D)]}{\theta} + M \frac{\gamma}{\theta} \left(\frac{\theta}{Q_H} \right)^\gamma L \left(\frac{\theta}{Q_H} \right) q \quad (4.29)$$

donde $\Gamma(\dots)$ es la función gamma incompleta, con el límite inferior de la integral igual a cero.

Y por último, si la información histórica es exclusivamente acerca de la máxima avenida:

$$I_3(1,1) = -E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \theta^2} \right] = \left(\frac{\gamma}{\theta} \right)^2 (N+1) \quad (4.30)$$

$$I_3(2,2) = -E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \gamma^2} \right] = \frac{(N+1)[1+\Gamma''(2)] + L^2 M - 2\Gamma'(2)LM}{\gamma^2} \quad (4.31)$$

$$I_3(1,2) = I_3(2,1) = -E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \theta \partial \gamma} \right] = \frac{(N+1)\Gamma'(2) - LM}{\theta} \quad (4.32)$$

Las varianzas asintóticas de los estimadores de los cuantiles se obtiene haciendo uso de la ecuación (3.16), en la que el vector \underline{V} de derivadas parciales del cuantil respecto de los

parámetros para una distribución LogGumbel es:

$$V(1) = (-LF)^{-1/\gamma} \quad (4.33)$$

$$V(2) = \frac{\theta}{\gamma^2} (-LF)^{-1/\gamma} L(-LF) \quad (4.34)$$

donde $F = 1 - 1/T$, es la probabilidad de no excedencia del cuantil X_T .

No se acompañan los resultados finales de las varianzas asintóticas de los cuantiles por no ser expresiones compactas, y realmente servir de enlace para el siguiente apartado sobre las ganancias estadísticas, mucho más interesantes.

4.2.4.- Ganancias estadísticas asintóticas de los estimadores de los cuantiles

Las ganancias estadísticas asintóticas de los estimadores de los cuantiles para los diferentes niveles de información utilizados en este trabajo, SG_1 , se definió en la ecuación (3.2) a partir de las varianzas asintóticas, que después de muchas operaciones se llegan a simplificar en las expresiones que se acompañan en este apartado. Si la información histórica es del tipo binomial censurada, el resultado final es:

$$SG_1 = 1 - R_1 \left[1 + \frac{rD^2L^2 \left(\frac{D}{K} \right) \frac{q}{p}}{a - 2bLK + L^2K} \right] \quad (4.35)$$

donde:

$$\frac{1}{R_1} = 1 + \frac{rD^2(L^2D - 2bLD + a) \frac{q}{p}}{a - b^2} \quad (4.36)$$

con:

$$a = 1 + \Gamma''(2) = 1.8237$$

$$b = \Gamma'(2) = 0.4228$$

$$D = -L(1 - 1/H)$$

$$K = -L(1 - 1/T)$$

$$r = M/N$$

Si la información histórica es censurada, la ganancia estadística es:

$$SG_2 = 1 - R_2 \left[1 + r \frac{a' + p + qDL^2D - 2(b' + qDLD)LK + pL^2K}{a - 2bLK + L^2K} \right] \quad (4.37)$$

donde:

$$\frac{1}{R_2} = 1 + \frac{r}{a - b^2} [(a' + p + qDL^2D)(rp + 1) - 2(b' + qDLD)(rb' + b) + r(b'^2 - (qDLD)^2) + ap] \quad (4.38)$$

con:

$$a' = \Gamma''(2, D)$$

$$b' = \Gamma'(2, D)$$

Y por último, si la información histórica es acerca de la máxima avenida:

$$SG_3 = 1 - R_3 \left[N + 1 + \frac{LM(LM - 2b + 2LK)}{a - 2bLK + L^2K} \right] \quad (4.39)$$

donde:

$$\frac{1}{R_3} = \frac{(N+1)^2}{N} + \frac{L^2M}{a - b^2} \quad (4.40)$$

4.2.5.- Propiedades de las ganancias estadísticas asintóticas

Casi más importante que conocer las expresiones de las ganancias estadísticas asintóticas, es el analizar las propiedades de las mismas. A través de estas propiedades se podrían (y se pueden) obtener conclusiones que servirían para orientar el uso de la información histórica disponible. La principal de estas propiedades es que las ganancias correspondientes al uso de información binomial censurada y censurada (SG_1 y SG_2) son exclusivamente función de:

- i) El ratio r , que es la relación entre las longitudes de los períodos histórico y sistemático.

ii) El periodo de retorno H del limite de percepción (a través de las variables p , q y D).

iii) El periodo de retorno T del cuantil (a través de la variable K).

Y la ganancia estadística utilizando el valor de la máxima avenida (SG_1) es función de:

i) La longitud del periodo sistemático N .

ii) La longitud del periodo histórico M .

iii) El periodo de retorno T del cuantil.

En los tres casos, **la ganancia estadística asintótica es independiente de los parámetros de la distribución**. Es decir, es independiente de las características estadísticas de la población: media, varianza, sesgo, etc...

A continuación se analizan con mayor detalle las propiedades respecto a los tres factores dependientes.

4.2.5.1.- Variación de las ganancias estadísticas en función del periodo de retorno del cuantil

En las Figuras (4.1) y (4.2) se representan las tres ganancias estadísticas asintóticas SG_i en función del periodo de retorno T del cuantil, para dos diferentes limites de percepción histórica (de 10 y 100 años de periodo de retorno) y para un ratio r de 5. Las principales características observables en dichas figuras son:

i) Existe un cuantil para el que la información histórica maximiza su ganancia estadística. Este cuantil puede ser obtenido analíticamente y se sitúa, para el caso de información binomial censurada (SG_1) en:

$$T=H \tag{4.41}$$

para el caso de información censurada, SG_2 (normalmente T cercano y mayor que H):

$$LK = \frac{-A + (A^2 + 4BC)^{\frac{1}{2}}}{2B} \quad (4.42)$$

donde:

$$A = a' + (1-a)p + qDL^2D$$

$$B = pb - b' - qDLD$$

$$C = b(a' + p + qDL^2D) - a(b' + qDLD)$$

y para la información de máxima avenida, SG_3 (T mayor que H e independiente de la longitud sistemática N):

$$LK = b - \frac{LM + [L^2M + 4(a - b^2)]^{\frac{1}{2}}}{2} \quad (4.43)$$

ii) Existe un cuantil para el que la ganancia estadística es nula, si se utiliza información binomial censurada o máxima avenida, o muy pequeña si se utiliza información censurada. Este cuantil es un mínimo de SG_3 , y para la información binomial censurada es dado por:

$$LK = \frac{bLD - a}{LD - b} \quad (4.44)$$

para la información censurada:

$$LK = \frac{-A + (A^2 + 4BC)^{1/2}}{2B} \quad (4.45)$$

y para la información de máxima avenida:

$$LK = b - \frac{LM - [L^2M + 4(a - b^2)]^{1/2}}{2} \quad (4.46)$$

Este cuantil normalmente es de muy bajo periodo de retorno (menor de 3 años) y por tanto no es de interés en Hidrología.

iii) Para límites de percepción mayores de 50 años de periodo de retorno aproximadamente, las SG_3 son prácticamente constantes para cuantiles mayores de 10 años de periodo de retorno.

iv) Los valores límites, tanto para cuando T tiende a 1 como para T tendiendo a infinito, son iguales y valen:

$$\lim_{T \rightarrow 1} SG_1 = \lim_{T \rightarrow \infty} SG_1 = 1 - R_1 \left(1 + rD^2 \frac{q}{p} \right) \quad (4.47)$$

$$\lim_{T \rightarrow 1} SG_2 = \lim_{T \rightarrow \infty} SG_2 = 1 - R_2(1 + rp) \quad (4.48)$$

$$\lim_{T \rightarrow 1} SG_3 = \lim_{T \rightarrow \infty} SG_3 = 1 - R_3(N+1) \quad (4.49)$$

v) La ganancia estadística utilizando información censurada siempre es mayor que en el caso de ser binomial censurada, aunque en general son muy similares. Por poner un ejemplo, para T igual a infinito, donde la diferencia es máxima, y con H igual a 10 y r igual a 5, es decir con abundante información histórica, la diferencia entre SG_1 y SG_2 es de solo un 8%.

vi) El uso de la información de máxima avenida normalmente proporciona menor ganancia estadística, excepto cuando la información histórica es pobre (alto H y pequeño r).

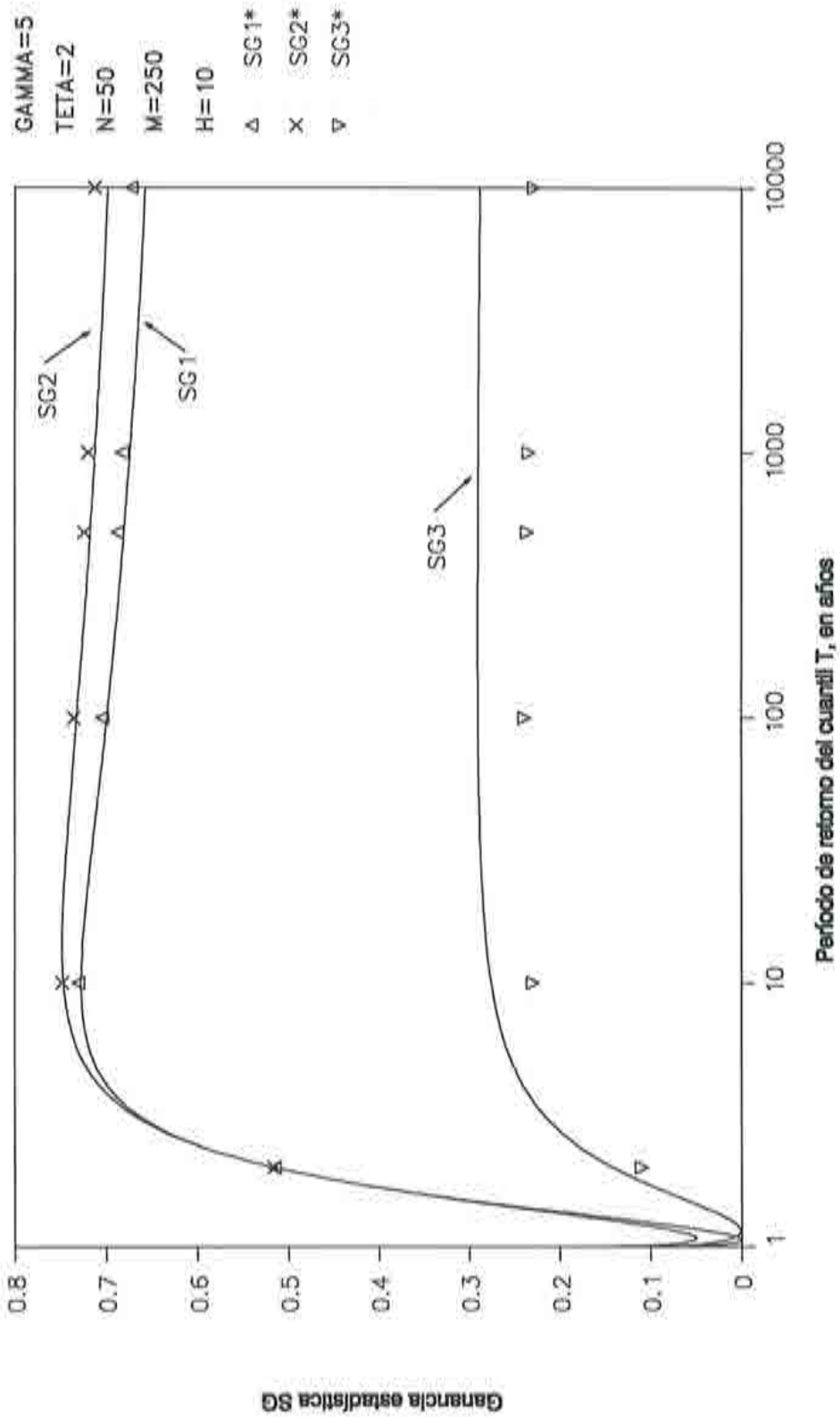


Figura 4.1. Ganancias estadísticas asintótica y media muestral en función del período de retorno del cuantil, para un límite de percepción histórico de 10 años.

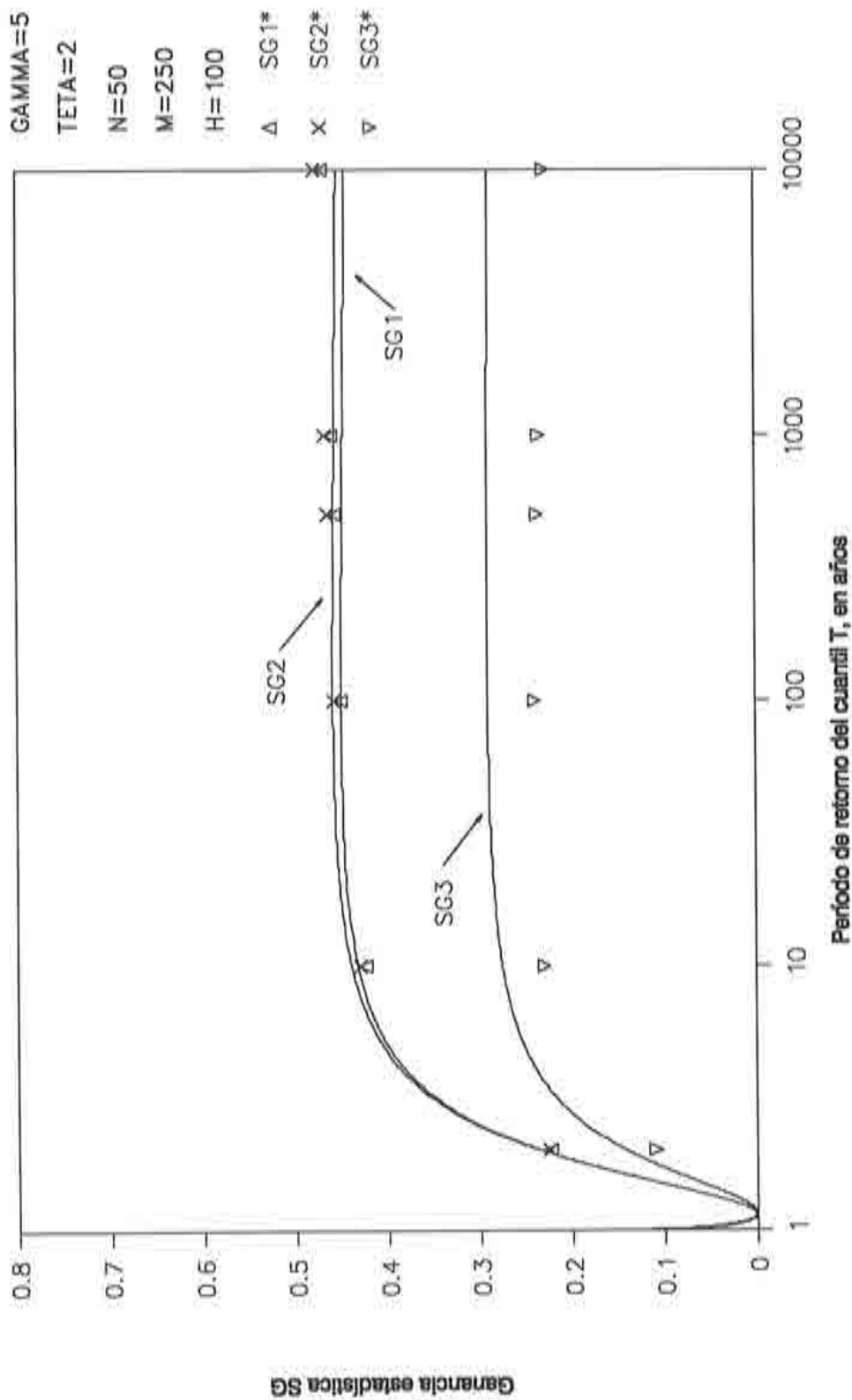


Figura 4.2. Ganancias estadísticas asintótica y media muestral en función del período de retorno del cuantil, para un límite de percepción histórico de 100 años.

4.2.5.2.- Variación de las ganancias estadísticas con información binomial censurada y censurada en función del límite de percepción

En las Figuras (4.3) y (4.4) se representan esas ganancias asintóticas para el cuantil de 500 años y dos diferentes ratios r (5 y 1). Las principales propiedades de esta relación son:

- i) Existe un límite de percepción que maximiza la ganancia estadística utilizando información binomial censurada, pero no pudo ser obtenido de forma explícita.
- ii) Para límites de percepción históricos muy grandes, H mayor que 1000 años, la ganancia estadística tiende a cero.
- iii) Existe un punto de ganancia estadística nula con información binomial censurada, que se corresponde con bajos períodos de retorno del límite de percepción (menor de 2 años). Este punto tampoco ha podido ser obtenido de forma explícita.
- iv) Siempre la ganancia estadística utilizando información censurada es mayor que la binomial censurada, aunque, al aumentar el período de retorno H del límite de percepción, la diferencia va disminuyendo.
- v) Para un H completamente distinto del período de retorno del cuantil (muy pequeño o muy grande) puede ocurrir que la ganancia estadística con máxima avenida sea mayor que con información binomial censurada.
- vi) Solo para H muy elevados, la ganancia estadística utilizando información censurada se reduce tanto que es inferior al caso de utilizar la máxima avenida.

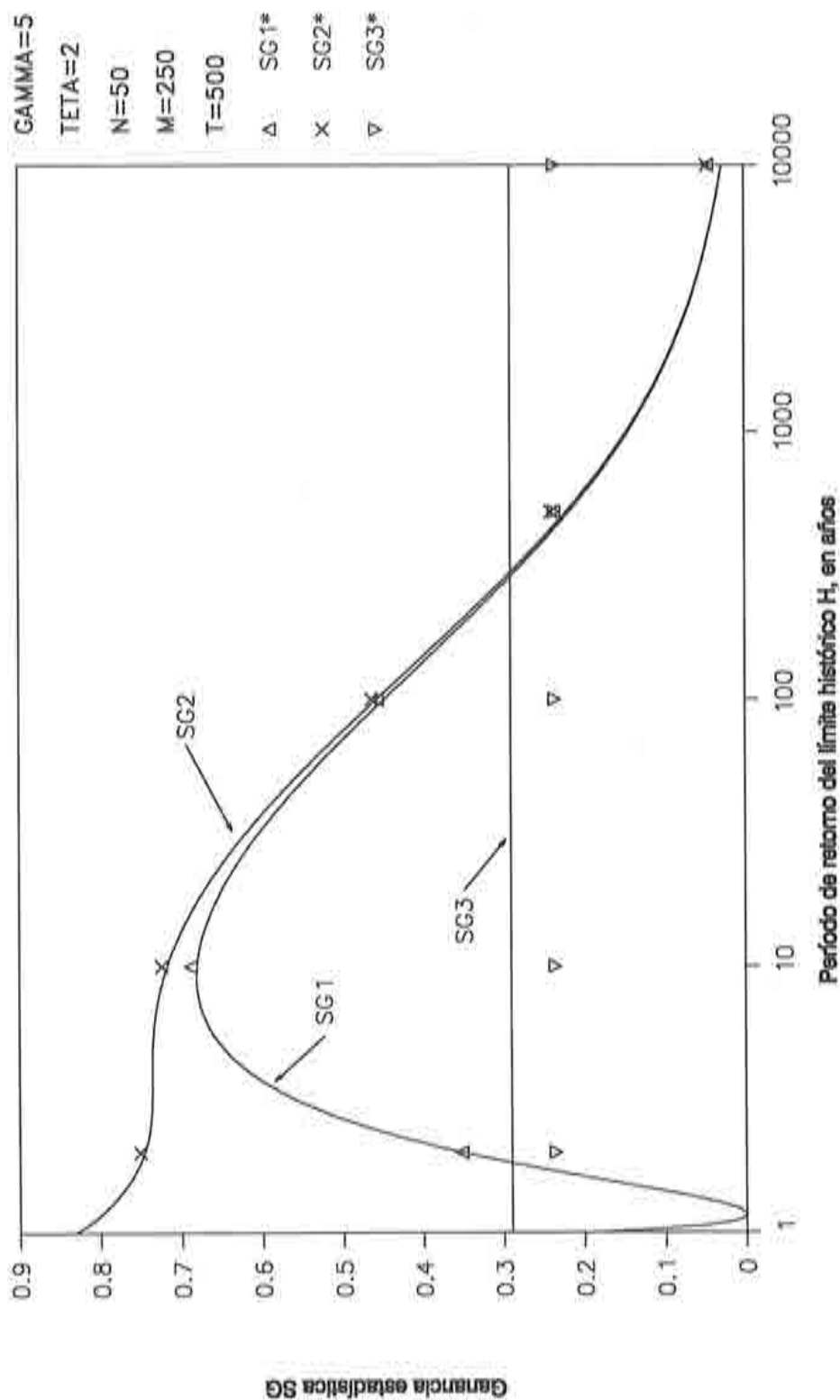


Figura 4.3. Ganancias estadísticas asintótica y media muestral en función del límite de percepción histórico, para un ratio de longitudes entre períodos histórico y sistemático de 5.

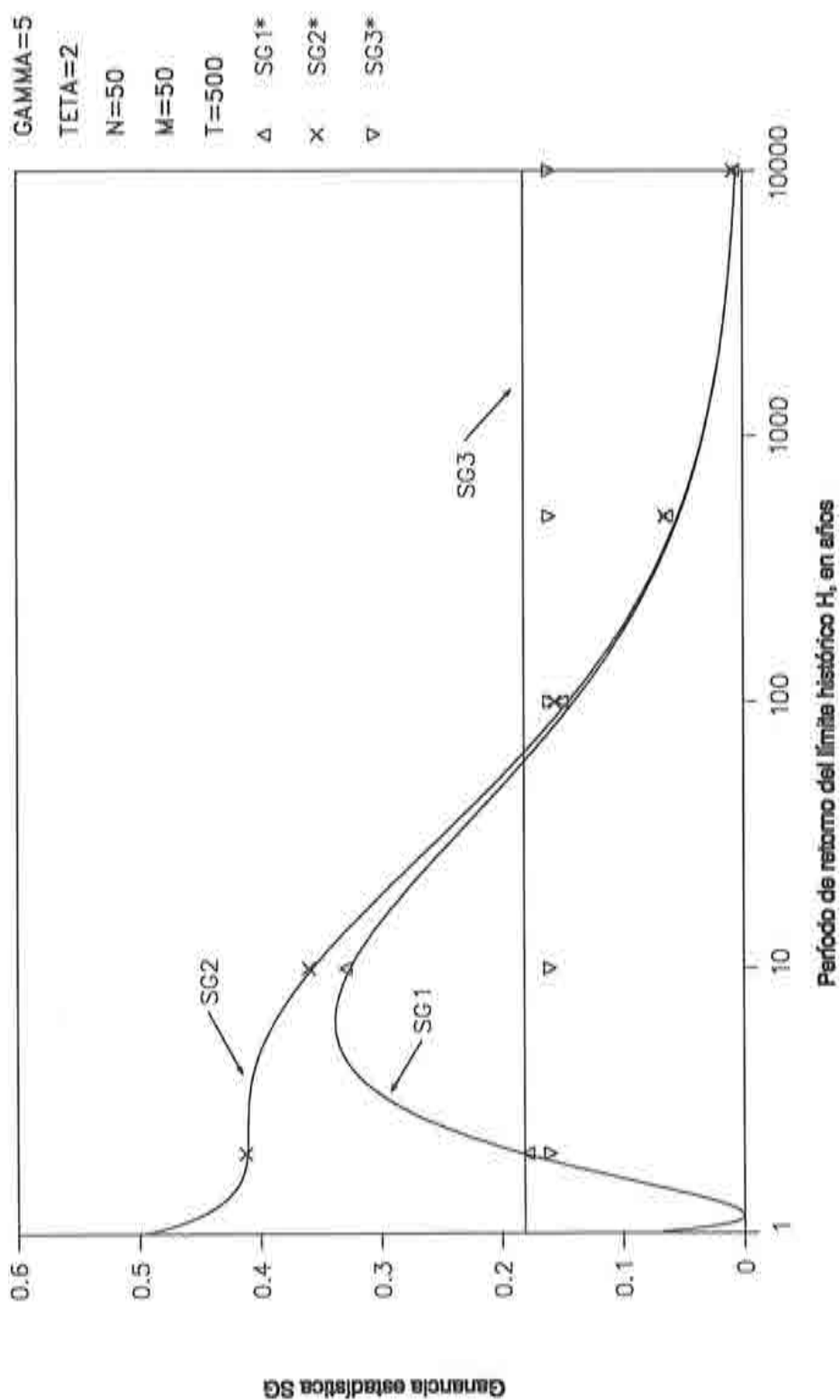


Figura 4.4. Ganancias estadísticas asintótica y media muestral en función del límite de percepción histórica, para un ratio de longitudes entre los períodos histórico y sistemático de 1.

4.2.5.3.- Variación de las ganancias estadísticas en función del ratio de longitudes entre período histórico y sistemático

Las representaciones gráficas se encuentran en las Figuras (4.5) y (4.6) para el cuantil de 500 años y dos diferentes límites de percepción (H igual a 10 y 100 años de período de retorno). Algunas propiedades que pueden observarse en dichas figuras son:

- i) Como era de esperar, para una determinada longitud de registro sistemático, la ganancia estadística aumenta a medida que aumenta la longitud del período histórico considerado.
- ii) Cuando la información tiende a infinita, los valores límites de las ganancias estadísticas son:

$$\lim_{r \rightarrow \infty} SG_1 = 1 - \frac{(a-b^2)L^2 \left(\frac{D}{K}\right)}{(L^2K - 2bLK + a)(L^2D - 2bLD + a)} \quad (4.50)$$

que valdrá 1 si H=T.

$$\lim_{r \rightarrow \infty} SG_2 = 1 \quad (4.51)$$

$$\lim_{r \rightarrow \infty} SG_3 = 1 - \frac{a-b^2}{L^2K - 2bLK + a} \quad (4.52)$$

Las ganancias estadísticas se aproximan muy lentamente a estos límites, como se observa en las Figuras (4.5) y (4.6).

- iii) El orden de mayor a menor ganancia habitual es utilizando información censurada, binomial censurada y máxima avenida. Sin embargo, para T muy diferente de H puede ocurrir que SG_3 sea mayor que SG_1 y, por tanto, también sus límites.

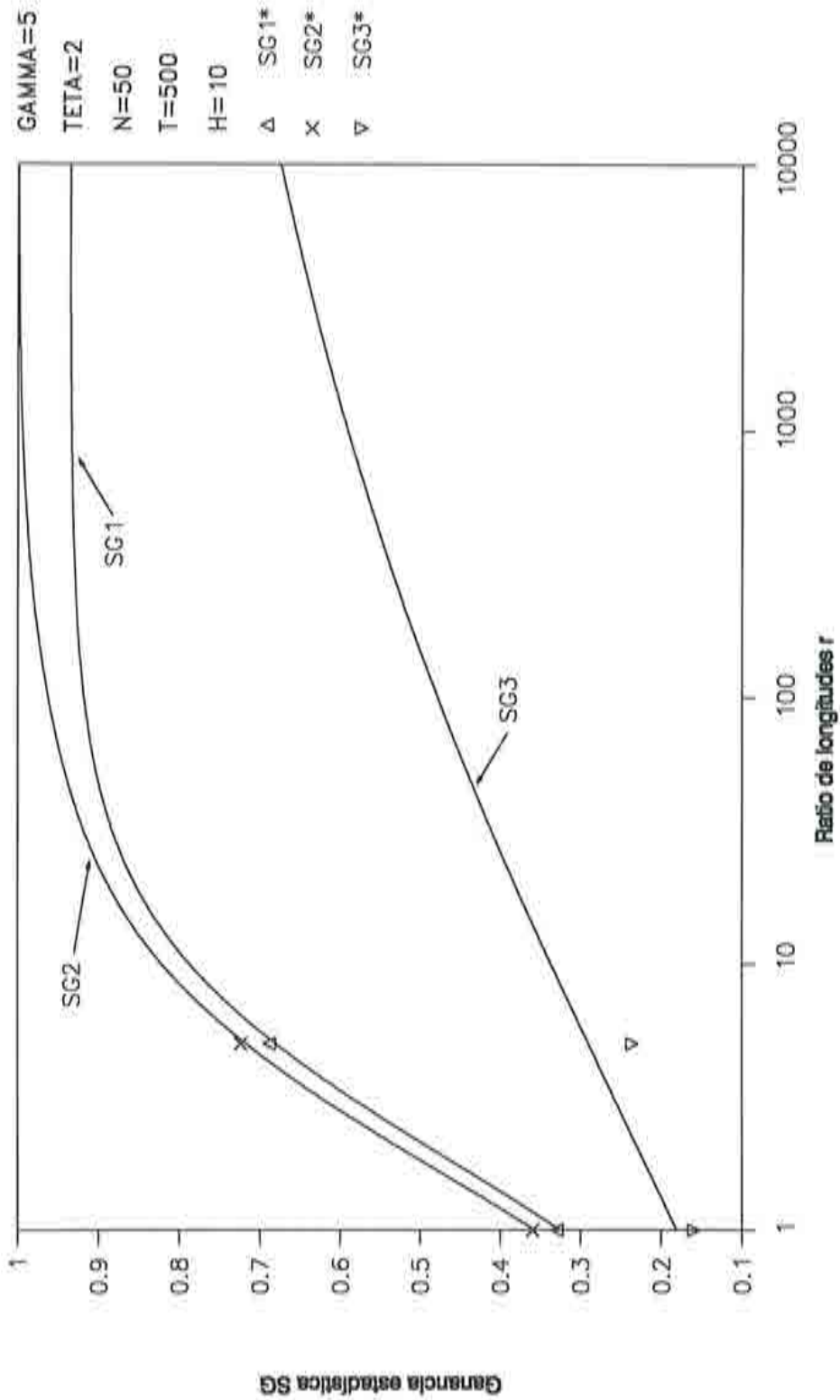


Figura 4.5. Ganancias estadísticas asintótica y media muestral en función del ratio entre las longitudes de los periodos histórico y sistemático, para un límite de percepción histórico de 10 años.

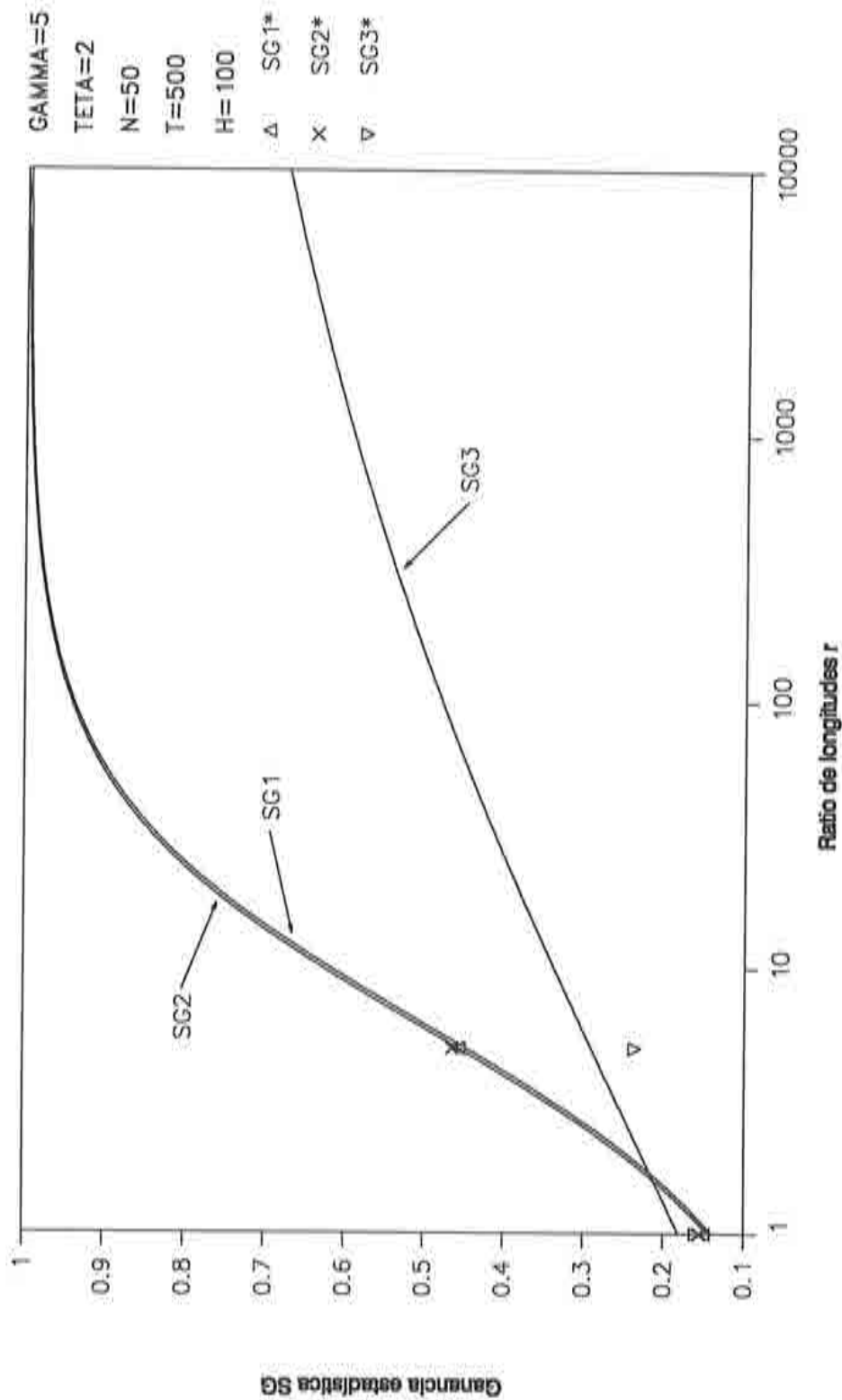


Figura 4.6. Ganancias estadísticas asintótica y media muestral en función del ratio entre las longitudes de los periodos histórico y sistemático, para un límite de percepción histórico de 100 años.

4.2.6.- Errores cuadráticos medios y ganancias estadísticas muestrales de los cuantiles

Se realizaron simulaciones de tipo Montecarlo intentando contestar a la pregunta de si la ganancia estadística asintótica obtenida y estudiada en el Apartado anterior era una buena aproximación de la ganancia estadística muestral. Si esto era así, no solo se podrían utilizar las expresiones asintóticas sino que también serían de aplicación las propiedades estudiadas anteriormente.

Las simulaciones se realizaron según se describe en el Apartado 3.3.4 generando cada vez 5000 series sintéticas, haciendo uso del programa SIMLG. A pesar de este gran número de simulaciones, no hay que olvidar que el resultado es a su vez una variable aleatoria, de la que se quiere conocer su valor medio. Por esta razón fueron consideradas 5 semillas iniciales, representándose gráficamente los resultados medios de las mismas junto a los valores asintóticos.

Las distintas simulaciones se realizaron de tal forma que en cada simulación se modificó algún factor del escenario inicial, definido por:

$$\theta = 2$$

$$\gamma = 5$$

$$N = 50$$

$$M = 250$$

$$H = 100 \text{ años}$$

$$T = \text{diversos valores, desde 2 a 10000 años}$$

Este escenario se consideró como representativo de un caso general. De hecho, es similar a la situación encontrada en el estudio del río Paraná (Francés et al., 1994).

De las simulaciones, algunos resultados interesantes son:

i) Como era de esperar, siempre los errores cuadráticos medios muestrales utilizando información histórica (MSE_1 , MSE_2 y MSE_3) fueron menores que sin utilizarla (MSE_0).

ii) Las varianzas asintóticas (CRLB) son una buena aproximación de los errores cuadráticos medios muestrales (MSE). En todos los escenarios, la diferencia fue menor del 5% para los casos de no usar información histórica, uso de información histórica binomial censurada y uso de información histórica censurada. Con información histórica de máxima avenida la diferencia fue menor del 10%.

iii) Según se observa en las Figuras (4.1) a (4.6) las ganancias estadísticas muestrales con información binomial censurada (SG_1^*) y censurada (SG_2^*) son ligeramente superiores que sus correspondientes valores asintóticos (SG_1 y SG_2). La razón es que MSE_1 y MSE_2 se aproximan más rápido a sus valores asintóticos que MSE_0 . Por otra parte, la ganancia estadística muestral con máxima avenida (SG_3^*) es menor que su valor asintótico (SG_3), ya que, en este caso, MSE_3 se aproxima más lentamente a su valor asintótico que MSE_0 .

iv) Las ganancias estadísticas muestrales parecen ser independientes del parámetro γ , como se observa en la Figura (4.7). Para el rango del coeficiente de sesgo entre 9.5 y 1.2 la variación es del orden de la tercera cifra significativa, y solo se producen variaciones mayores para sesgos muy altos (γ muy bajo).

v) Un efecto similar se reproduce para la independencia de SG_1^* y SG_2^* frente a la longitud del registro sistemático N , como se refleja en la Figura (4.9).

vi) En la Figura (4.8) se observa que las ganancias estadísticas muestrales son claramente independientes del parámetro θ .

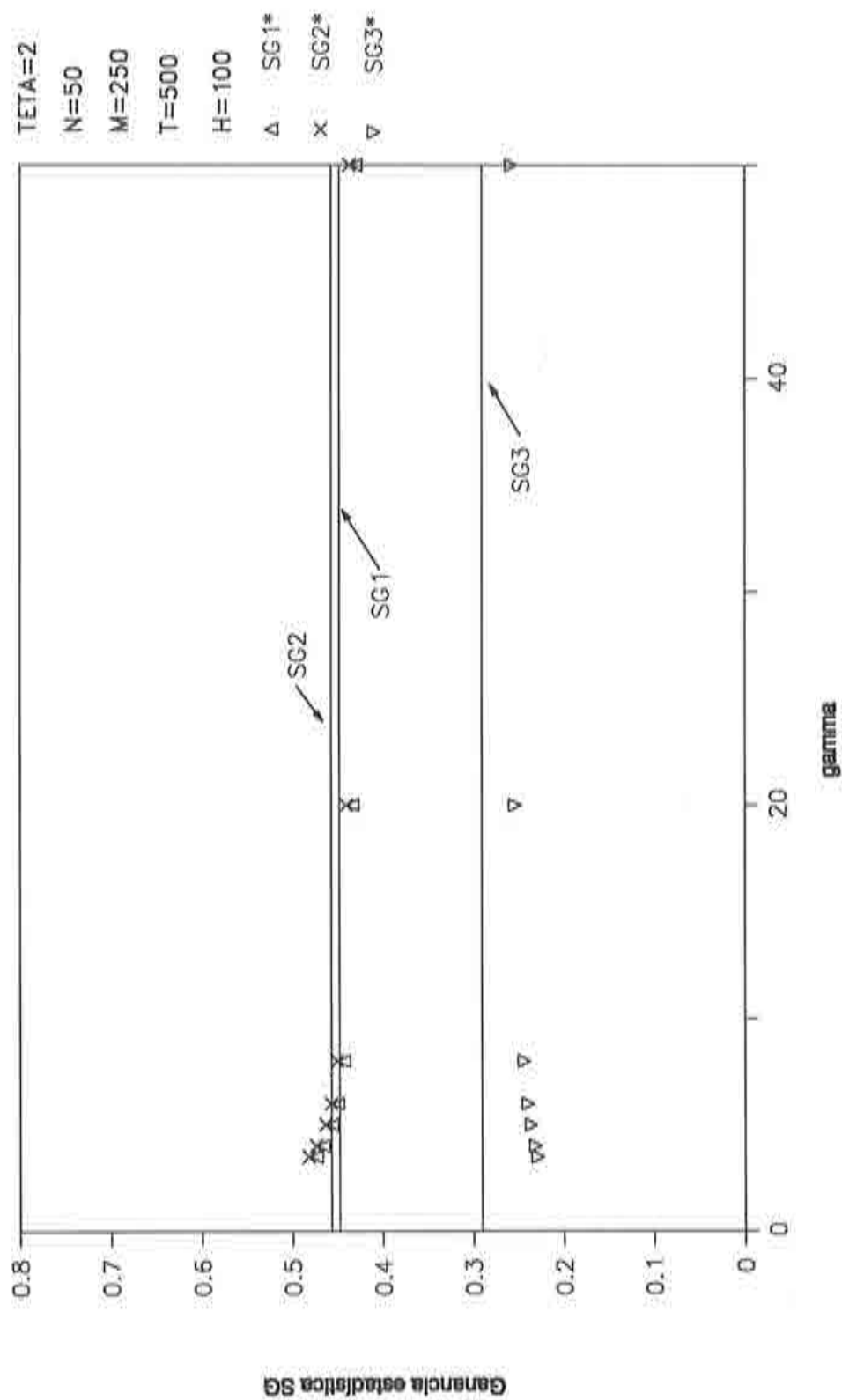


Figura 4.7. Ganancias estadísticas asintótica y media muestral en función del parámetro γ .

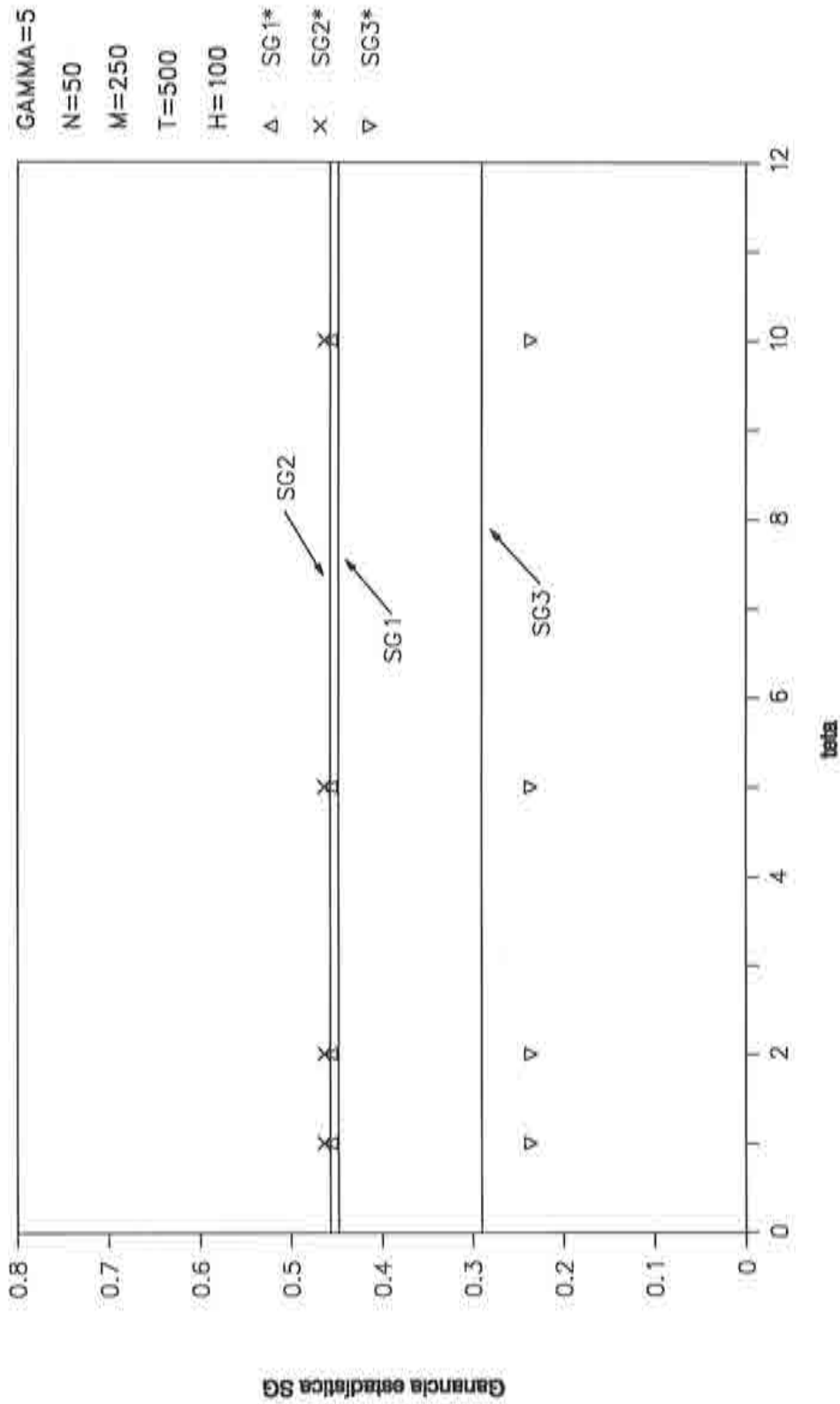


Figura 4.8. Ganancias estadísticas asintótica y media muestral en función del parámetro θ .

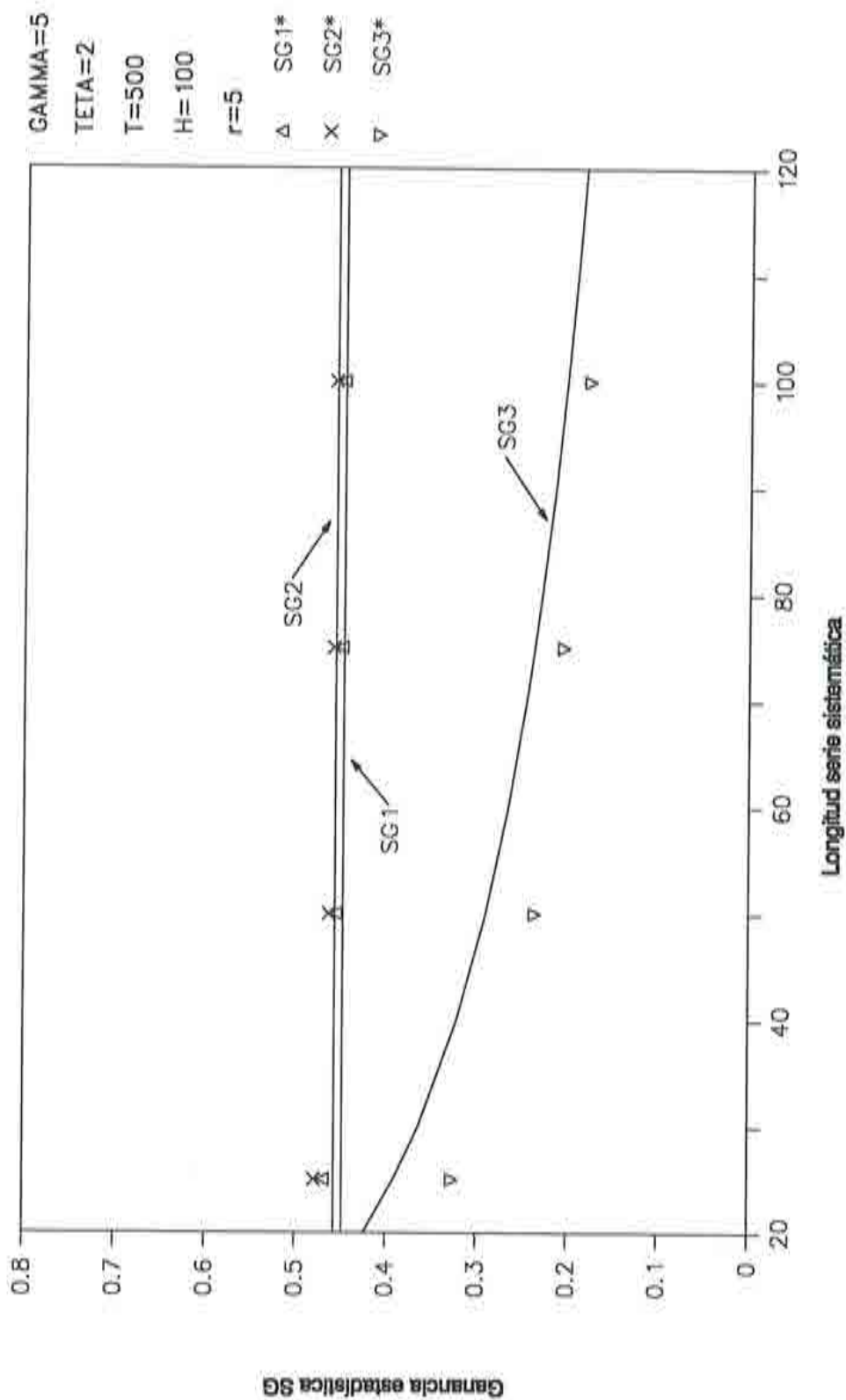


Figura 4.9. Ganancias estadísticas asintótica y media muestral en función de la longitud de la serie sistemática.

4.3.- CONCLUSIONES

En este capítulo se han desarrollado modelos estadísticos para la distribución LogGumbel que son capaces de incorporar al análisis de la frecuencia de las avenidas tres clase de información histórica: binomial censurada, censurada y de máxima avenida. Esta distribución pertenece a la familia de distribuciones "Extreme Value" o EV, de la que también forman parte las distribuciones Gumbel y Weibull. Estas últimas se estudian en los Anejos B y C.

Se han analizado las ventajas de utilizar la información histórica obteniendo las ganancias estadísticas asintóticas y, mediante simulaciones de Montecarlo, una aproximación de las ganancias estadísticas muestrales. Los principales resultados, aplicables en el rango habitual de la Hidrología, de la familia de distribuciones EV son (se considera como rango posible de información una longitud de registro sistemático de 10 a 200 años, longitud del período histórico de 25 a 50000 años, límite de percepción histórico inferior a 1000 años de período de retorno y cuantiles de 10 a 10000 años):

- i) Siempre se obtienen mejores estimadores utilizando información histórica que no haciéndolo.
- ii) Las expresiones asintóticas tanto de las varianzas como de las ganancias asintóticas de los estimadores de los cuantiles son buenas aproximaciones de sus correspondientes valores poblacionales.
- iii) Las ganancias estadísticas son independientes de los parámetros de las distribuciones de la familia EV. Los valores asintóticos se ha demostrado analíticamente que son así exactamente, y los muestrales lo son desde un punto de vista práctico. Al ser las características estadísticas de la población función de los parámetros de la distribución, se convierte en absurdo tratar de analizar las ventajas de uso de la información histórica en función de dichas características.
- iv) La ganancia estadística es en media del orden del 50% utilizando información histórica binomial censurada o censurada. Es decir, el aumento de fiabilidad en los estimadores de los cuantiles es lo suficientemente grande como para justificar en la mayoría de los casos el uso de información histórica, si ésta está disponible. Esta conclusión está plenamente de acuerdo con las ideas de Stedinger y Cohn (1986).
- v) Debido al comportamiento asintótico de las ganancias estadísticas utilizando

información binomial censurada y censurada, con grandes aumentos del período histórico se producen muy pequeñas mejoras de las mismas. Por tanto, para estos tipos de información no será "rentable estadísticamente" aumentar desmesuradamente el período de investigación sobre avenidas históricas o paleoavenidas. Con la información sobre la máxima avenida ocurre todo lo contrario.

vi) El factor más influyente en las ganancias estadísticas con información binomial censurada y censurada es el límite de percepción histórico, y para el caso de información de máxima avenida el ratio entre las longitudes de los períodos histórico y sistemático.

vii) Salvo que exista abundancia de información histórica (bajo límite de percepción) hay escasas diferencias utilizando información binomial censurada o censurada. Así, ante la posibilidad de errores en la determinación de las avenidas históricas es mejor hacer empleo de la información como binomial censurada.

viii) Normalmente, la información sobre la máxima avenida (censurada tipo 2) es más pobre que la binomial censurada y censurada (censurada tipo 1). Esto explica los resultados contrarios al uso de la información histórica obtenidos por Hosking y Wallis (1986a y 1986b) y es coincidente con las conclusiones de Guo y Cunnane (1991).

**5.- ANALISIS DE LA INCORPORACION DE LA INFORMACION
HISTORICA A LA DISTRIBUCION "TWO COMPONENT
EXTREME VALUE"**

5.1.- INTRODUCCION

Como se comprobará en el Capítulo 6 de Aplicación a los ríos Júcar, Turia y Segura, la distribución "Two Component Extreme Value" o TCEV es la que mejor se adapta a las características de las avenidas de estos ríos y, por extensión, a los ríos de la cuenca mediterránea española. Por ello se dedica este capítulo, central del trabajo, al análisis de esta distribución.

En primer lugar se presenta la génesis de la TCEV, para a continuación explorar sus características estadísticas. Estas características se relacionarán fundamentalmente con los parámetros de la distribución, relaciones que ayudarán en la comprensión de la dependencia de las ganancias estadísticas respecto de los mismos.

El modelo estadístico que es capaz de incorporar la información histórica se desarrolla en el Apartado 5.4, obteniéndose las funciones de verosimilitud para los cuatro niveles de información considerados en esta investigación. Esto niveles son: información sistemática exclusivamente, información histórica binomial censurada, información histórica censurada e información histórica de máxima avenida. Al final del apartado se selecciona el mejor método de búsqueda del máximo de estas funciones, para utilizar el Método de Máxima Verosimilitud en la estimación de los cuantiles.

El siguiente paso es analizar las ventajas de uso de los diferentes tipos de información histórica a través de las ganancias estadísticas asintóticas y muestrales, definidas en las ecuaciones (3.2) y (3.3) y descrita su obtención en el Apartado 3.3 de capítulo anterior. Desafortunadamente, las ganancias estadísticas asintóticas de la distribución TCEV no han podido ser obtenidas analíticamente, por lo que se han dejado en función de las matrices de Fisher. Lo que si ha podido obtenerse son sus propiedades fundamentales, que servirán para un uso más racional de la información histórica, siempre y cuando sean una buena aproximación de las ganancias estadísticas muestrales. Esta comparación se ha realizado mediante simulaciones de Montecarlo.

Por último, el Apartado 5.8 presenta las modificaciones de los apartados anteriores necesarias en un estudio regional.

5.2.- GENESIS DE LA DISTRIBUCION "TWO COMPONENT EXTREME VALUE"

Como se vio el Capítulo de Antecedentes, la función de distribución "Two Component Extreme Value", o TCEV, trata de tener en cuenta que las avenidas de un río pueden estar provocadas por dos mecanismos diferentes. Cada uno de estos mecanismos da lugar a una secuencia de avenidas en un año que se consideran independientes e idénticamente distribuidas (iid):

$$Z_{1i} ; \quad i=1,2,\dots,K_1 \quad (5.1)$$

$$Z_{2i} ; \quad i=1,2,\dots,K_2 \quad (5.2)$$

Donde K_1 y K_2 son el número de avenidas de un año. Se hace la hipótesis de que siguen una distribución de Poisson, con valores medios λ_1 y λ_2 respectivamente. Si Z_{1i} genera las crecidas ordinarias y Z_{2i} las extraordinarias se tendrá que $\lambda_1 > \lambda_2$.

El número total de crecidas en un año es $K = K_1 + K_2$. La variable aleatoria K , como suma de dos variables aleatorias de Poisson, es también de Poisson con media $\lambda = \lambda_1 + \lambda_2$.

Por tanto, la función de distribución de las crecidas de un río a lo largo de un año es:

$$F_Z(z) = p_1 F_{Z_1}(z) + (1-p_1) F_{Z_2}(z) \quad (5.3)$$

donde $p_1 = \lambda_1/\lambda$ es la probabilidad de que una crecida sea de tipo ordinaria.

Si se define la avenida ordinaria anual como:

$$X_1 = \max Z_{1i} \quad 1 \leq i \leq K_1 \quad (5.4)$$

Su función de distribución (máximo de X_1), es el valor máximo de un proceso de Poisson y su función de distribución es:

$$F_{X_1}(x_1) = \exp \{-\lambda_1 [1 - F_{Z_1}(x_1)]\} \quad (5.5)$$

Donde $F_{Z_1}(\cdot)$ es la función de distribución común de todas las Z_{1i} .

Si se asume que sigue una distribución exponencial de media $1/\theta_1$, se tiene que:

$$F_{Z_i}(z) = 1 - e^{-\theta_1 z} \quad z > 0 \quad (5.6)$$

Y por tanto, la ecuación (5.5) queda:

$$F_{X_1} = \exp(-\lambda e^{-\theta_1 x}) \quad x > 0 \quad (5.7)$$

Esta función de distribución tiene una discontinuidad en $X_1=0$; donde presenta una probabilidad finita e igual a:

$$P(X_1=0) = e^{-\lambda_1} \quad (5.8)$$

Intentando simplificar el resultado final, se puede asumir que esta probabilidad concentrada queda representada en los valores negativos de X_1 , con lo que la variable aleatoria X_1 pasa a ser continua:

$$F_{X_1}(x) = \exp(-\lambda_1 e^{-\theta_1 x}) \quad (5.9)$$

Análogamente, para las avenidas extraordinarias, se tiene que la avenida anual extraordinaria:

$$X_2 = \max Z_{2i} \quad 1 \leq i \leq K_2 \quad (5.10)$$

sigue la función de distribución:

$$F_{X_2}(x) = \exp(-\lambda_2 e^{-\theta_2 x}) \quad (5.11)$$

Es decir, las avenidas ordinarias y extraordinarias siguen sendas distribuciones de Gumbel o EVI. La máxima avenida anual estará definida como:

$$X = \max (X_1, X_2) \quad (5.12)$$

Si se considera que X_1 y X_2 son independientes, la función de distribución de X es:

$$F_X(x) = \exp[-\lambda_1 e^{-\theta_1 x} - \lambda_2 e^{-\theta_2 x}] \quad (5.13)$$

Que es la expresión de la TCEV. Esta función es una generalización de la función Gumbel y es la función de distribución exacta de la máxima avenida anual si las series de avenidas dentro de un año siguen una distribución exponencial. En caso contrario se trata de

un resultado asintótico.

La función de densidad de probabilidad (f.d.p.) es la derivada de la ecuación (5.13):

$$f_X(x) = F_X(x) \psi(x) \quad (5.14)$$

donde:

$$\psi(x) = \lambda_1 \theta_1 e^{-\theta_1 x} + \lambda_2 \theta_2 e^{-\theta_2 x} \quad (5.15)$$

El empleo de la TCEV en un análisis regional exige hacer algún tipo de homogeneización de las variables aleatorias correspondientes a las diferentes estaciones de aforos. Esta homogeneización consiste en una estandarización local previa (Rossi et al., 1984) que es:

$$Y = \theta_1 X - L\lambda_1 \quad (5.16)$$

Con lo que la función de distribución de la variable estandarizada es:

$$F_Y(y) = \exp[-e^{-y} - \lambda e^{-\theta y}] \quad (5.17)$$

donde:

$$\theta = \frac{\theta_2}{\theta_1} \quad (5.18)$$

$$\lambda = \lambda_2 \lambda_1^{-\theta} \quad (5.19)$$

La hipótesis de regionalización es que los parámetros λ y θ permanecen constantes dentro de una misma región y, por tanto, podrán utilizarse conjuntamente para la obtención de la distribución regional.

5.3.- CARACTERÍSTICAS ESTADÍSTICAS DE LA TCEV

Dado que se trata de una distribución de reciente creación, no existe en la literatura una recopilación de las propiedades estadísticas de la TCEV. Por ello se analizan en este apartado sus características estadísticas, haciendo especial hincapié en la relación con los parámetros de la distribución.

La variable aleatoria de una población TCEV es continua y está definida en toda la recta real. Los valores negativos están representando en su conjunto la probabilidad de que se produzca una avenida de valor nulo, probabilidad que será muy pequeña excepto en climas áridos.

Las propiedades que debe cumplir la ecuación (5.13) para ser una función de distribución de probabilidad acumulada (f.d.a.) son:

- i) $F_X(-\infty) = 0$
- ii) $F_X(+\infty) = 1$
- iii) $F_X(\cdot)$ tiene que ser monótona no decreciente
- iv) $F_X(\cdot)$ tiene que ser continua por la derecha

Las propiedades i) y ii) definen el espacio posible de los parámetros.

La propiedad i) implica que:

$$\lim_{x \rightarrow -\infty} (\lambda_1 e^{-\theta_1 x} + \lambda_2 e^{-\theta_2 x}) = 0 \quad (5.20)$$

Lo cual solo es posible en los casos:

$$\text{Si } \theta_1 > 0 \Leftrightarrow \lambda_1 > 0$$

$$\text{Si } \theta_2 > 0 \Leftrightarrow \lambda_2 > 0$$

$$\text{Si } \theta_1 \leq 0 \Leftrightarrow \theta_2 > 0$$

$$\text{Si } \theta_2 \leq 0 \Leftrightarrow \theta_1 > 0$$

La propiedad ii) implica que:

$$\lim_{x \rightarrow \infty} \left(\lambda_1 e^{-\theta_1 x} + \lambda_2 e^{-\theta_2 x} \right) = 1 \quad \text{donde:} \quad (5.21)$$

Lo cual solo es posible si:

$$\theta_1 > 0 \text{ y } \theta_2 > 0$$

Luego el espacio de parámetros es realmente:

$$\theta_1, \theta_2 > 0$$

$$\lambda_1, \lambda_2 \geq 0$$

La propiedad iii) se cumple automáticamente al ser la derivada de $F_X(\cdot)$, obtenida en la ecuación (5.14), siempre positiva en el espacio de parámetros.

La propiedad iv) se cumple por definición. Al cumplirse las cuatro condiciones, la ecuación (5.13) representa efectivamente una f.d.a.

A continuación se estudian los estadísticos más importantes de la distribución TCEV.

5.3.1.- Probabilidad de presentación de una avenida extraordinaria

La probabilidad de presentación de una avenida extraordinaria o "high outlier" define la proporción de avenidas extraordinarias en un río. Vendrá definida por:

$$p_2 = P[X_1 < X_2] = \int_{-\infty}^{\infty} F_{X_1}(x) f_{X_2}(x) dx \quad (5.22)$$

Las distribuciones del miembro de derecho de la ecuación (5.22) están definidas en las ecuaciones (5.9) y (5.10). Con el cambio de variable:

$$z = \lambda_2 e^{-\theta_2 x} \quad (5.23)$$

y operando queda:

$$p_2 = \int_0^{\infty} e^{-t-z} dz \quad (5.24)$$

$$t = \lambda z^\theta \quad (5.25)$$

Es decir, la probabilidad de un "outlier" es función exclusiva de los parámetros estandarizados λ y θ .

Como se observa en la Figura (5.1), a efectos prácticos, se puede considerar que p_2 es función exclusiva de λ , especialmente para valores bajos del parámetro. Cuanto mayor sea λ , mayor es la proporción de avenidas extraordinarias.

5.3.2.- Media poblacional

Las avenidas extraordinarias y ordinarias siguen independientemente distribuciones de Gumbel, con lo que sus medias serán (ver Anejo C):

$$\mu_1 = \frac{0.5772 + L\lambda_1}{\theta_1} \quad (5.26)$$

$$\mu_2 = \frac{0.5772 + L\lambda_2}{\theta_2} \quad (5.27)$$

Luego la relación entre las avenidas extraordinarias y ordinarias es:

$$\frac{\mu_1}{\mu_2} = \frac{0.5772 + L\lambda_1}{0.5772 + L\lambda_2} \theta \quad (5.28)$$

Como los parámetros λ_1 y λ_2 están amortiguados por el logaritmo, el factor más importante es θ . Es decir, θ es un índice aproximado de la relación en magnitud entre las avenidas ordinarias y extraordinarias.

Por otra parte, la media poblacional de la TCEV viene definida por:

$$\mu = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx \quad (5.29)$$

Realizando el cambio definido por la ecuación (5.23) se llega a que:

$$\mu = \frac{1}{\theta_1} [L\lambda_1 - I_1] \quad (5.30)$$

donde:

$$I_1 = \int_0^{\infty} L z e^{-z-t} \left(\theta \frac{t}{z} + 1 \right) dz \quad (5.31)$$

En la Figura (5.2) se representa la media poblacional en función de los parámetros estandarizados ($\theta_1 = \lambda_1 = 1$). Esta media estandarizada crece cuanto mayor sea la proporción de avenidas extraordinarias (λ alto) y mayores sean éstas comparadas con las ordinarias (θ bajo).

5.3.3.- Varianza poblacional

La varianza poblacional es una medida de la dispersión de la variable aleatoria y viene dada por:

$$\sigma^2 = E[(X-\mu)^2] = E[X^2] - \mu^2 \quad (5.32)$$

Para obtenerla, en primer lugar se calcula el valor esperado de los cuadrados, que para la TCEV es:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_x(x) dx = \frac{1}{\theta_1^2} (L^2\lambda_1 - 2L\lambda_1 I_1 + I_2)$$

donde:

$$I_2 = \int_0^{\infty} L^2 z^2 e^{-z-t} \left(\theta \frac{t}{z} + 1 \right) dz \quad (5.34)$$

Sustituyendo las ecuaciones (5.30) y (5.33) en (5.32), queda:

$$\sigma^2 = \frac{1}{\theta_1^2} (I_2 - I_1^2) \quad (5.35)$$

Es decir, la varianza poblacional es función de θ_1 , θ_2 y del parámetro estandarizado λ .

El coeficiente de variación, adimensional, resulta ser:

$$\eta = \frac{\sigma}{\mu} = \frac{\sqrt{I_2 - I_1^2}}{L\lambda_1 - I_1} \quad (5.36)$$

que es función exclusiva de λ_1 , λ_2 y el parámetro estandarizado θ . El coeficiente de variación en función de λ y θ para $\lambda_1 = 1$ queda representado en la Figura (5.3). La dispersión es tanto mayor cuanto menores sean λ y θ , o lo que es equivalente, cuanto mayores y más raras sean las avenidas extraordinarias. Además, el coeficiente de variación crece cuando disminuye λ_1 .

5.3.4. Sesgo poblacional

El sesgo es una medida de la simetría de la distribución de la población. Viene definido por:

$$\mu_3 = E[(X - \mu)^3] = E[X^3] - 3\mu E[X^2] + 2\mu^3 \quad (5.37)$$

Para la TCEV, el valor esperado de los cubos de la variable aleatoria es:

$$E[X^3] = \int_{-\infty}^{\infty} x^3 f_x(z) dx = \frac{1}{\theta^3} (L^3 \lambda_1 - 3L^2 \lambda_1 I_1 + 3L \lambda_1 I_2 - I_3) \quad (5.38)$$

donde:

$$I_3 = \int_0^{\infty} L^3 z e^{-z/\theta} \left(\theta \frac{L}{z} + 1 \right) dz \quad (5.39)$$

Sustituyendo en la ecuación (5.37) queda:

$$\mu_3 = \frac{1}{\theta^3} (3I_1 I_2 - 2I_1^3 - I_3) \quad (5.40)$$

El coeficiente de sesgo es un estadístico adimensional y vale:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{3I_1I_2 - 2I_1^3 - I_3}{(I_2 - I_1^2)^{3/2}} \quad (5.41)$$

Es decir, el coeficiente de sesgo es función exclusiva de los parámetros estandarizados. Como valores límite se obtiene el coeficiente de sesgo de una Gumbel:

$$\lim_{\lambda \rightarrow 0} \gamma_1 = \lim_{\lambda \rightarrow 1} \gamma_1 = 1.29857 \quad (5.42)$$

Según se observa en la Figura (5.4) el coeficiente de sesgo crece cuando disminuyen θ y λ en el rango habitual de los mismos. En ese rango, cuanto mayores y menos frecuentes sean las avenidas extraordinarias el sesgo será mayor.

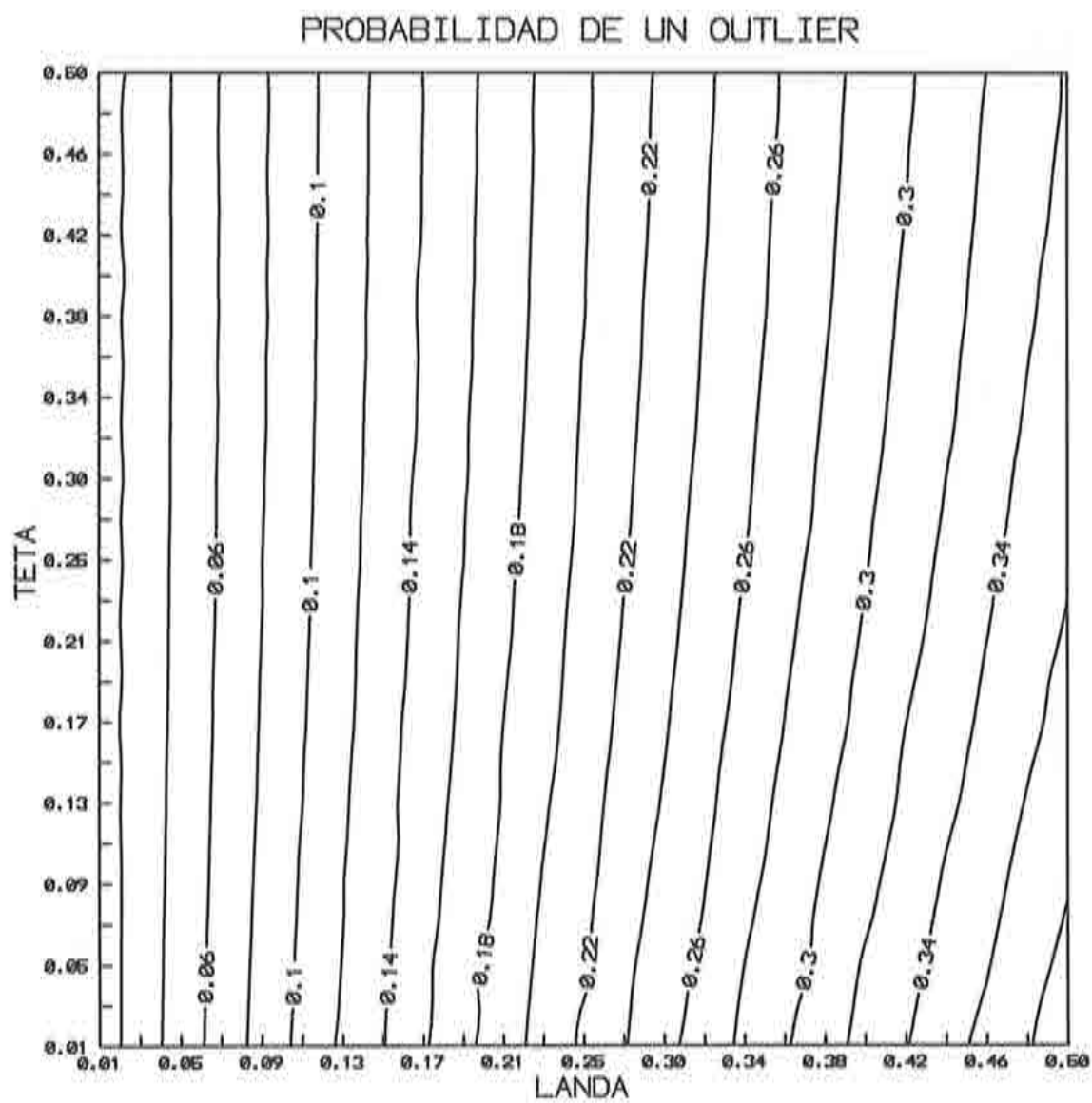


Figura 5.1. Probabilidad de una avenida extraordinaria de una distribución TCEV, en función de los parámetros estandarizados.

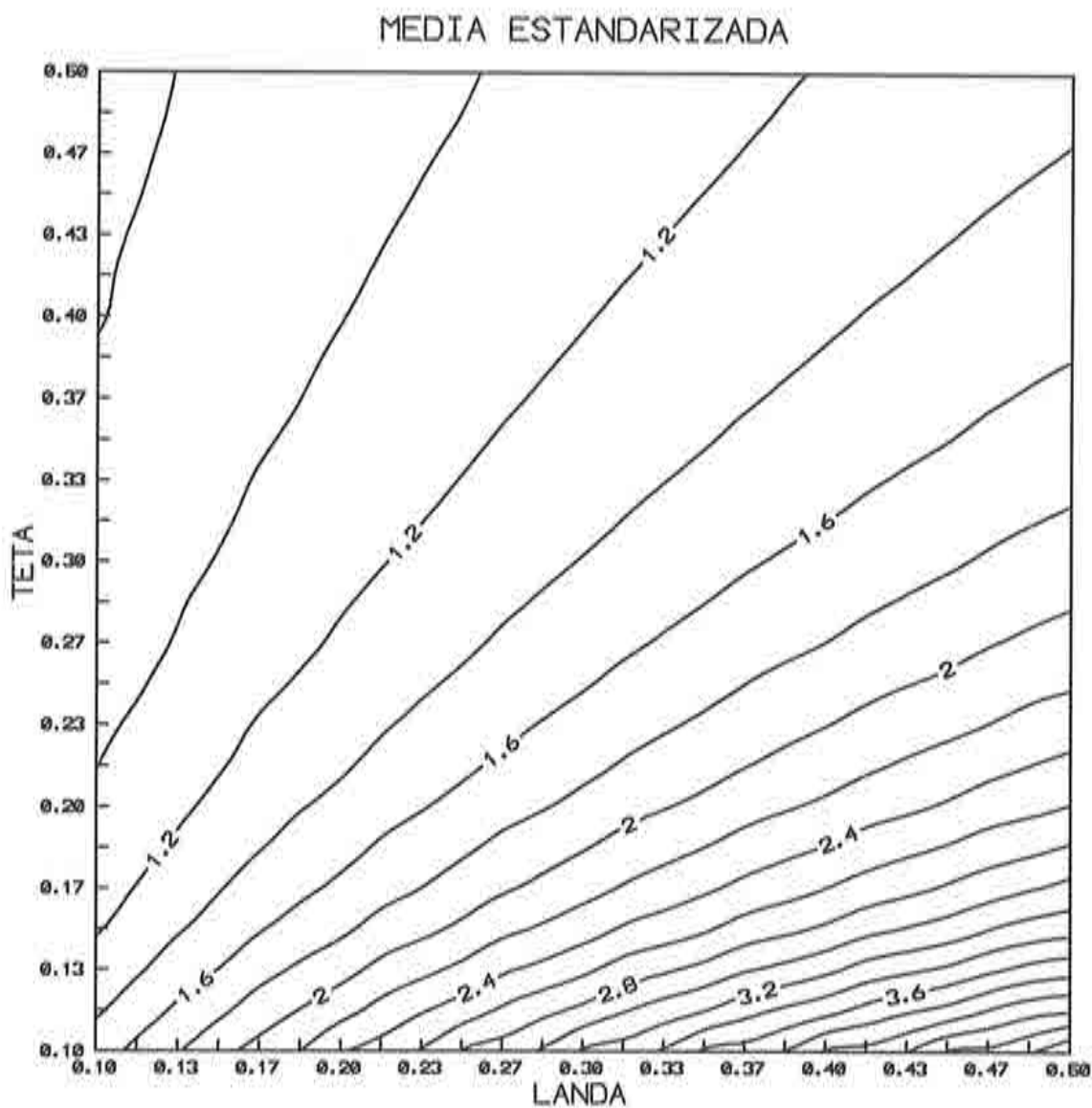


Figura 5.2. Media poblacional de una distribución TCEV con los parámetros λ_1 y θ_1 unidad, en función de los parámetros estandarizados.

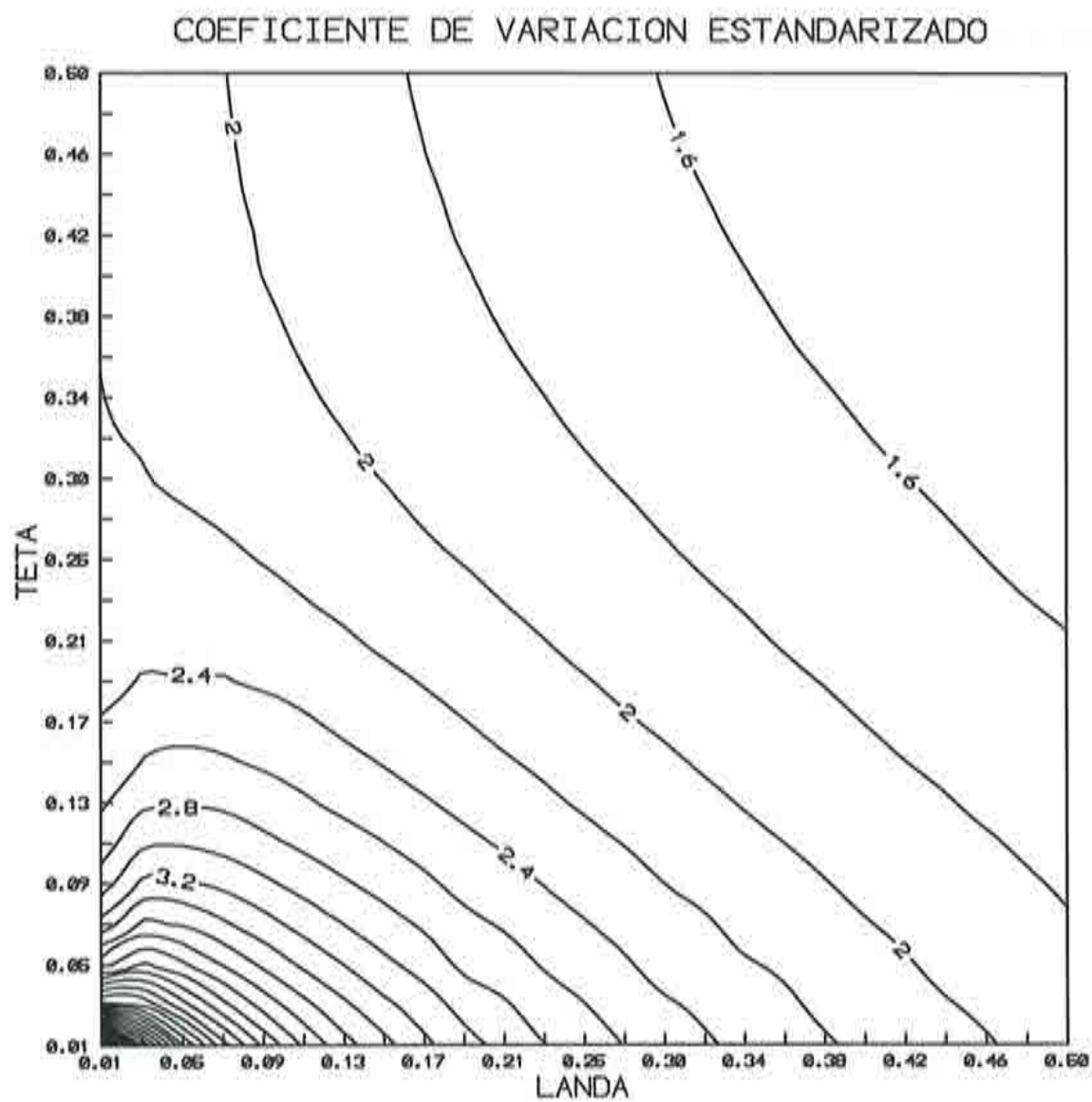


Figura 5.3. Coeficiente de variación de una distribución TCEV con los parámetros λ_1 y θ_1 unidad, en función de los parámetros estandarizados.

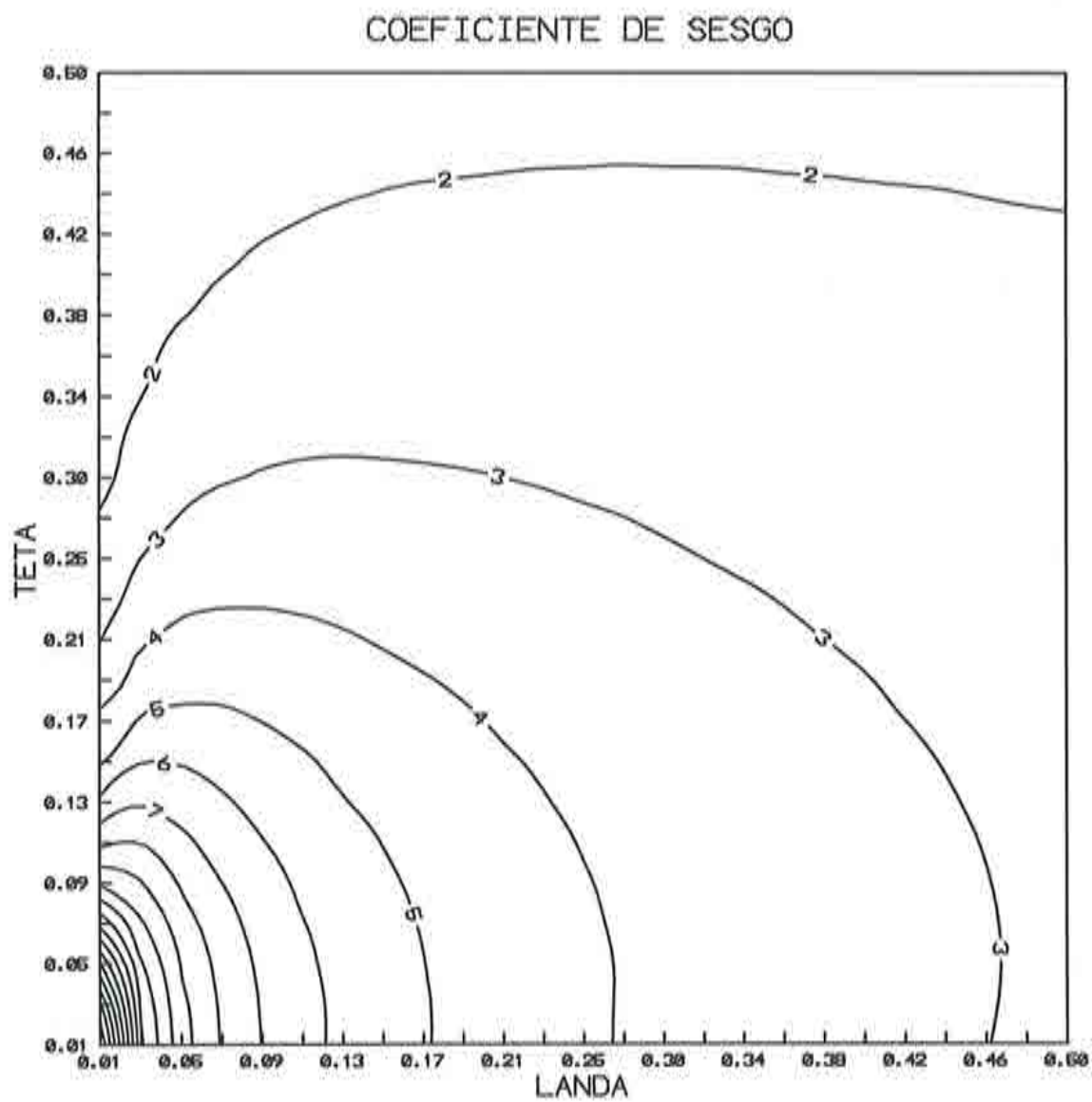


Figura 5.4. Coeficiente de sesgo de una distribución TCEV, en función de los parámetros estandarizados.

5.4. ESTIMADORES DE LOS PARAMETROS POR MÁXIMA VEROSIMILITUD

5.4.1.- Funciones de verosimilitud

Según se ha visto en el Capítulo de Antecedentes, uno de los métodos de estimación más robustos es el Método de la Máxima Verosimilitud. Además, este método permite la incorporación de la información histórica de una forma inmediata, según se vio en el Capítulo 3 sobre el Procedimiento Estadístico. Para el caso de no disponer de información histórica, y solo contar los N datos del registro sistemático, una función de verosimilitud sería:

$$L_{sy}(\Theta) = \prod_{i=1}^N \exp \left[-\lambda_1 e^{-\theta_1 x_i} - \lambda_2 e^{-\theta_2 x_i} \right] \psi(x_i) \quad (5.43)$$

Realmente es mucho más cómodo maximizar el logaritmo de la función de verosimilitud, que es en este caso:

$$LL_{sy}(\Theta) = -\lambda_1 \sum_{i=1}^N e^{-\theta_1 x_i} - \lambda_2 \sum_{i=1}^N e^{-\theta_2 x_i} + \sum_{i=1}^N L\psi(x_i) \quad (5.44)$$

Si la información histórica es del tipo binomial censurada, es decir, se conoce solo el número de avenidas K que superaron durante el período histórico M el límite de percepción Q_H , hay que incorporar a la función de verosimilitud la función de probabilidad de la distribución binomial que sigue este tipo de información:

$$f_K(k) = \binom{M}{k} q^{M-k} p^k \quad (5.45)$$

donde:

q = probabilidad de no excedencia del límite Q_H

p = probabilidad de excedencia del límite Q_H

Por lo que :

$$q = \exp \left[-\lambda_1 e^{-\theta_1 Q_H} - \lambda_2 e^{-\theta_2 Q_H} \right] \quad (5.46)$$

$$p = 1 - q \quad (5.47)$$

Y el logaritmo de la función de verosimilitud sería:

$$LL_{BC}(\Theta) = (M-k) \left(-\lambda_1 e^{-\theta_1 Q_H} - \lambda_2 e^{-\theta_2 Q_H} \right) + kLp + LL_{SY}(\Theta) \quad (5.48)$$

Si además se conoce el valor de las avenidas históricas y_j , es decir, se está utilizando información histórica censurada, según se vio en el Capítulo 3 el logaritmo de la función de verosimilitud quedaría:

$$LL_{CE}(\Theta) = (M-k) \left(-\lambda_1 e^{-\theta_1 Q_H} - \lambda_2 e^{-\theta_2 Q_H} \right) - \lambda_1 \sum_{j=1}^k e^{-\theta_1 y_j} - \lambda_2 \sum_{j=1}^k e^{-\theta_2 y_j} + \sum_{j=1}^k L\psi(y_j) + LL_{SY}(\Theta) \quad (5.49)$$

Y por último, si la única información histórica es acerca de la máxima avenida, el logaritmo de la función de verosimilitud particularizada para la TCEV quedaría:

$$LL_{MF}(\Theta) = -M\lambda_1 e^{-\theta_1 w} - M\lambda_2 e^{-\theta_2 w} + L\psi(w) + LL_{SY}(\Theta) \quad (5.50)$$

Los parámetros por Máxima Verosimilitud serán aquellos que maximicen las ecuaciones (5.44), (5.48), (5.49) o (5.50) según el tipo de información que se esté utilizando.

5.4.2.- Obtención de los máximos de las funciones de verosimilitud

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. Desafortunadamente como se observa en la Figura (5.5), éste es el caso. En dicha figura, se ha representado la superficie LL_{SY} en dos dimensiones haciendo $\lambda_1 = \theta_1 = 1$. La serie sistemática \underline{x} se ha generado sintéticamente. La situación es que el máximo de la función se encuentra en una superficie extensa con forma de meseta, donde la variación de la función es mínima, presentando por otra parte unos bordes muy pendientes a su alrededor.

Como el algoritmo de búsqueda del máximo iba a ser utilizado decenas de miles de veces en la fase de simulación (Capítulo 5.6) era conveniente el que fuera rápido y a la vez efectivo sobre una superficie compleja, para una casuística casi infinita. Los algoritmos testeados fueron 3 que se describen a continuación:

5.4.2.1.- Método iterativo

Este método es el propuesto por Rossi et al. (1984). El máximo del logaritmo de la función de verosimilitud es un punto donde todas las derivadas parciales se anulan. Es decir,

los estimadores de los parámetros por Máxima Verosimilitud son solución del siguiente sistema de ecuaciones:

$$\begin{aligned}
 \frac{\partial LL_i(\Theta)}{\partial \lambda_1} &= 0 \\
 \frac{\partial LL_i(\Theta)}{\partial \theta_1} &= 0 \\
 \frac{\partial LL_i(\Theta)}{\partial \lambda_2} &= 0 \\
 \frac{\partial LL_i(\Theta)}{\partial \theta_2} &= 0
 \end{aligned}
 \tag{5.51}$$

donde i puede ser SY, BC, CE o MF.

Como las ecuaciones son altamente no lineales la única forma de resolverlas es mediante un método iterativo, que para el caso de utilizar exclusivamente la información sistemática vendría dado por las siguientes ecuaciones recursivas:

$$\theta_1 = \frac{\sum_{i=1}^N e^{-\theta_1 x_i}}{\sum_{i=1}^N \frac{e^{-\theta_1 x_i}}{\psi(x_i)}}
 \tag{5.52}$$

$$\lambda_1 = \lambda_1 \frac{\sum_{i=1}^N x_i e^{-\theta_1 x_i}}{\sum_{i=1}^N \frac{\theta_1 x_i e^{-\theta_1 x_i}}{\psi(x_i)} - \sum_{i=1}^N \frac{e^{-\theta_1 x_i}}{\psi(x_i)}}
 \tag{5.53}$$

$$\theta_2 = \frac{\sum_{i=1}^N e^{-\theta_2 x_i}}{\sum_{i=1}^N \frac{e^{-\theta_2 x_i}}{\psi(x_i)}}
 \tag{5.54}$$

$$\lambda_2 = \lambda_2 \frac{\sum_{i=1}^N x_i e^{-\theta_2 x_i}}{\sum_{i=1}^N \frac{\theta_2 x_i e^{-\theta_2 x_i}}{\psi(x_i)} - \sum_{i=1}^N \frac{e^{-\theta_2 x_i}}{\psi(x_i)}} \quad (5.55)$$

Este método sencillo de programar es bastante rápido pero, en ciertos casos, no converge. Por este motivo fue rechazado.

5.4.2.2- Método del Simplex

Este método se basa en el movimiento sobre la superficie en la dirección de la máxima pendiente de una figura geométrica llamada simplex. El simplex en un espacio de 4 parámetros posee 5 vértices y no tiene porqué englobar inicialmente el máximo de la función.

El algoritmo central de este método es el desarrollado en el paquete de subrutinas en Fortran de Press et al. (1989). Su principal ventaja es que converge siempre. Sin embargo, si se quiere evitar que el algoritmo se pare en los alrededores de la planicie, el criterio de parada debe ser muy fino de tal forma que el algoritmo se ralentiza muchísimo.

5.4.2.3 - Método de Powell de las direcciones conjugadas

Se basa en una serie de maximizaciones unidimensionales. Las direcciones de maximización viene definidas por el algoritmo de Polak-Riviere (Press et al., 1989). Las maximizaciones unidimensionales engloban al punto óptimo, por lo que el algoritmo no tenderá a pararse en los bordes de la planicie de la Figura (5.5).

Además es casi tan rápido, a igualdad de precisión, que el método iterativo. Por todo ello es el algoritmo utilizado.

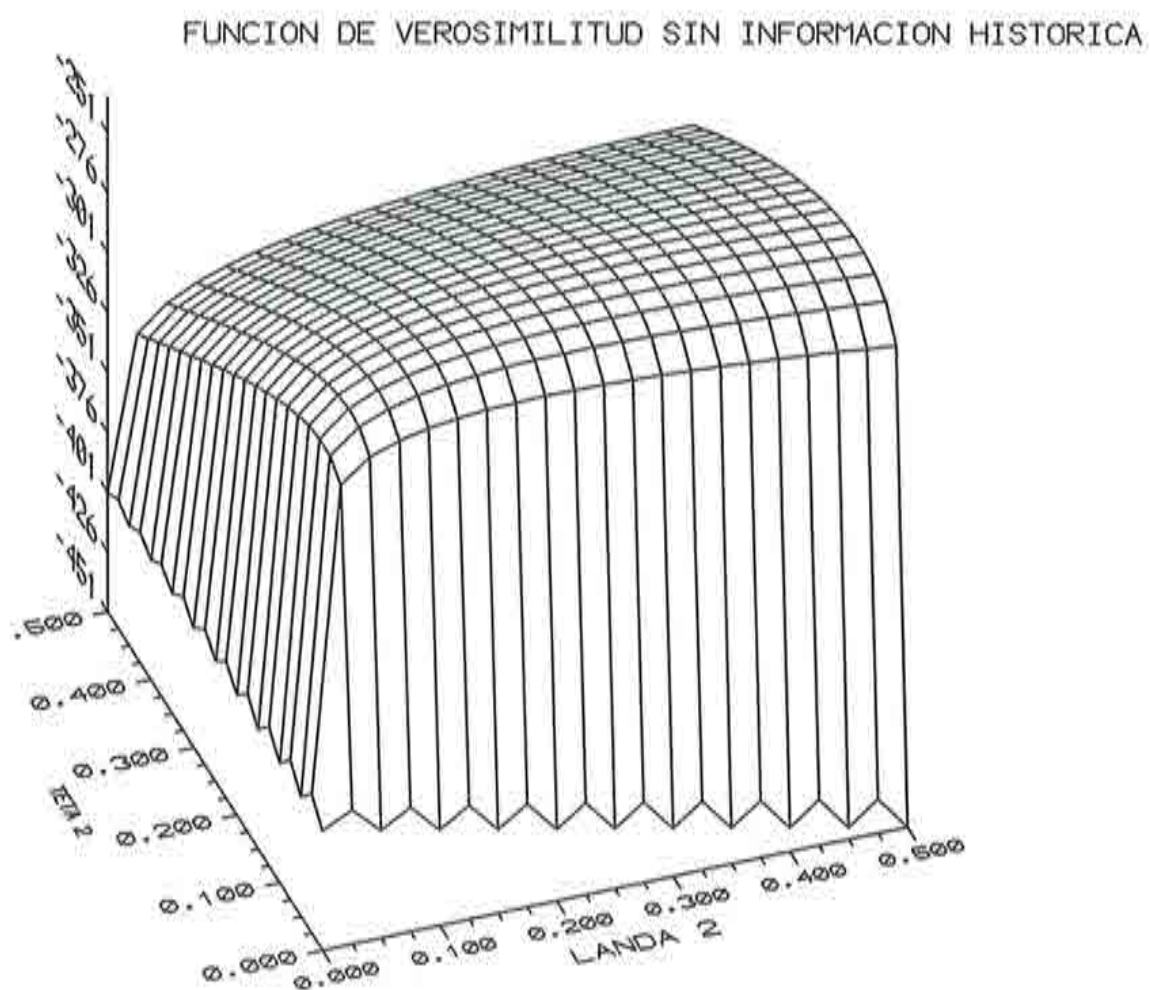


Figura 5.5. Función de verosimilitud de una distribución TCEV con λ_1 y θ_1 unidad, para una serie de datos de 100 años generada sintéticamente, en función de los parámetros estandarizados.

5.5.- MATRICES DE INFORMACION DE FISHER PARA LOS DISTINTOS TIPOS DE INFORMACION

Para dejar constancia de la metodología seguida en la obtención de los elementos de esta matriz, se va a seguir con cierto detalle los pasos operacionales para el caso con información sistemática exclusivamente, mientras que del resto de matrices solo se van a adjuntar los resultados finales.

5.5.1.- Matriz de información de Fisher con información sistemática exclusivamente

Como se consignó en el Capítulo 3, los elementos de esta matriz son los valores esperados de la derivadas segundas de la función de verosimilitud, cambiadas de signo, respecto de los parámetros.

Para el parámetro λ_1 , la primera derivada de LL_{SY} es, derivando en la ecuación (5.44):

$$\frac{\partial LL_{SY}(\Theta)}{\partial \lambda_1} = -\sum_{i=1}^N e^{-\theta_1 x_i} + \sum_{i=1}^N \frac{\theta_1 e^{-\theta_1 x_i}}{\psi(x_i)} \quad (5.56)$$

Y la segunda derivada es:

$$\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1^2} = -\sum_{i=1}^N \frac{\theta_1^2 e^{-2\theta_1 x_i}}{\psi^2(x_i)} \quad (5.57)$$

Al ser X_i una serie de datos independientes e idénticamente distribuidos, el valor esperado del sumatorio es N veces el valor esperado de uno de sus elementos, de tal forma que queda:

$$-E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1^2} \right] = N \theta_1^2 E \left[\frac{e^{-2\theta_1 x}}{\psi^2(x)} \right] \quad (5.58)$$

donde:

$$E \left[\frac{e^{-2\theta_1 x}}{\psi^2(x)} \right] = \int_{-\infty}^{\infty} \frac{e^{-2\theta_1 x}}{\psi^2(x)} F_x(x) \psi(x) dx = \int_{-\infty}^{\infty} \frac{e^{-2\theta_1 x}}{\psi(x)} F_x(x) dx \quad (5.59)$$

que con el cambio de variable:

$$z = \lambda_1 e^{-\theta_1 x} \quad (5.60)$$

resulta ser:

$$E \left[\frac{e^{-2\theta_1 x}}{\psi^2(x)} \right] = \frac{1}{\theta_1 \lambda_1^2} \int_0^{\infty} e^{-z-t_1} \frac{z}{\theta_1 z + \theta_2 t_1} dz = \frac{S_{31}}{\theta_1 \lambda_1^2} \quad (5.61)$$

donde:

$$t_1 = \lambda_1 \left(\frac{z}{\lambda_1} \right)^{\theta_2/\theta_1} \quad (5.62)$$

La integral S_{31} de la ecuación anterior no es inmediata, por lo que hay que resolverla numéricamente. Sustituyendo queda que:

$$-E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1^2} \right] = N \frac{\theta_1}{\lambda_1^2} S_{31} \quad (5.63)$$

La segunda derivada de LL_{SY} respecto a los parámetros λ_1 y θ_1 resulta:

$$\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_1} = \sum_{i=1}^N x_i e^{-\theta_1 x_i} + \lambda_2 \theta_2 \sum_{i=1}^N \frac{e^{-\theta_1 x_i} e^{-\theta_2 x_i} - \theta_1 x_i e^{-\theta_1 x_i} e^{-\theta_2 x_i}}{\psi^2(x_i)} \quad (5.64)$$

Y su valor esperado, cambiado de signo:

$$-E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_1} \right] = -NE[xe^{-\theta_1 x}] - N\lambda_2 \theta_2 E \left[\frac{e^{-\theta_1 x} e^{-\theta_2 x}}{\psi^2(x)} \right] + N\lambda_2 \theta_2 \theta_1 E \left[x \frac{e^{-\theta_1 x} e^{-\theta_2 x}}{\psi^2(x)} \right] \quad (5.65)$$

El primer valor esperado del miembro derecho de la ecuación queda, utilizando el cambio de variable definido en la ecuación (5.66):

$$E [xe^{-\theta_1 x}] = \frac{S_{11}}{\theta_1^2 \lambda_1} \quad (5.66)$$

con:

$$S_{11} = \int_0^{\infty} L\left(\frac{\lambda_1}{z}\right) e^{-z-t_1} (\theta_1 z + \theta_2 t_1) dz \quad (5.67)$$

Para el segundo valor esperado, por problemas de convergencia numérica, se realiza el cambio:

$$z = \lambda_2 e^{-\theta_2 x} \quad (5.68)$$

y el resultado es:

$$E\left[\frac{e^{-\theta_1 x} e^{-\theta_2 x}}{\psi^2(x)}\right] = \frac{S_4}{\theta_2 \lambda_1 \lambda_2} \quad (5.69)$$

con:

$$S_4 = \int_0^{\infty} e^{-z-t_2} \frac{t_2}{\theta_1 z + \theta_2 t_2} dz \quad (5.70)$$

siendo:

$$t_2 = \lambda_1 \left(\frac{z}{\lambda_2}\right)^{\theta_1/\theta_2} \quad (5.71)$$

El tercer valor esperado con el mismo cambio de variable queda:

$$E\left[x \frac{e^{-\theta_1 x} e^{-\theta_2 x}}{\psi^2(x)}\right] = \frac{S_5}{\theta_2^2 \lambda_1 \lambda_2} \quad (5.72)$$

con:

$$S_5 = \int_0^{\infty} L\left(\frac{\lambda_2}{z}\right) e^{-z-t_2} \frac{t_2}{\theta_2 z + \theta_1 t_2} dz \quad (5.73)$$

Y sustituyendo resulta:

$$-E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_1} \right] = N \left(-\frac{1}{\theta_1^2 \lambda_1} S_{11} - \frac{1}{\lambda_1} S_4 + \frac{\theta_1}{\lambda_1 \theta_2} S_5 \right) \quad (5.74)$$

La segunda derivada con respecto a λ_1 y λ_2 es:

$$\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \lambda_2} = -\sum_{i=1}^N \frac{\theta_1 e^{-\theta_1 x_i} \theta_2 e^{-\theta_2 x_i}}{\psi^2(x_i)} \quad (5.75)$$

Y su valor esperado:

$$-E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \lambda_2} \right] = N \theta_1 \theta_2 E \left[\frac{e^{-\theta_1 x} e^{-\theta_2 x}}{\psi^2(x)} \right] = N \frac{\theta_1}{\lambda_1 \lambda_2} S_4 \quad (5.76)$$

La segunda derivada de LL_{SY} respecto a λ_1 y θ_2 es:

$$\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_2} = -\sum_{i=1}^N \frac{\theta_1 e^{-\theta_1 x_i} (\lambda_2 e^{-\theta_2 x_i} - \theta_2 \lambda_2 x_i e^{-\theta_2 x_i})}{\psi^2(x_i)} \quad (5.77)$$

Y su valor esperado:

$$\begin{aligned} -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_2} \right] &= N \theta_1 \lambda_2 E \left[\frac{e^{-\theta_1 x} e^{-\theta_2 x}}{\psi^2(x)} \right] - N \theta_1 \theta_2 \lambda_2 E \left[x \frac{e^{-\theta_1 x} e^{-\theta_2 x}}{\psi^2(x)} \right] = \\ &= N \left(\frac{\theta_1}{\lambda_1 \theta_2} S_4 - \frac{\theta_1}{\theta_2 \lambda_1} S_5 \right) \end{aligned} \quad (5.78)$$

Dada la simetría del operador segunda derivada, no es necesario obtener las 16 derivadas parciales, sino solo 10. Las 3 siguientes son las correspondientes al parámetro θ_1 . La primera derivada parcial es:

$$\frac{\partial LL_{SY}(\Theta)}{\partial \theta_1} = \lambda_1 \sum_{i=1}^N e^{-\theta_1 x_i} + \sum_{i=1}^N \frac{\lambda_1 e^{-\theta_1 x_i} - \theta_1 \lambda_1 e^{-\theta_1 x_i}}{\psi(x_i)} \quad (5.79)$$

La segunda derivada respecto a θ_1 dos veces es:

$$\begin{aligned} \frac{\partial LL_{SY}(\Theta)}{\partial \theta_1^2} &= -\lambda_1 \sum_{i=1}^N x_i e^{-\theta_1 x_i} + \\ &+ \sum_{i=1}^N \frac{-\lambda_1^2 e^{-2\theta_1 x_i} - 2\lambda_2 \theta_2 \lambda_1 x_i e^{-\theta_1 x_i} e^{-\theta_2 x_i} + \lambda_2 \theta_2 \lambda_1 \theta_1 x_i^2 e^{-\theta_1 x_i} e^{-\theta_2 x_i}}{\psi^2(x_i)} \end{aligned} \quad (5.80)$$

Que tomando los valores esperados, queda:

$$-E \left[\frac{\partial LL_{SY}(\Theta)}{\partial \theta_1^2} \right] = N \left(\frac{1}{\theta_1^3} S_{21} + \frac{1}{\theta_1} S_{31} + \frac{2}{\theta_2} S_5 - \frac{\theta_1}{\theta_2^2} S_6 \right) \quad (5.81)$$

donde:

$$S_{21} = \int_0^{\infty} L^2 \left(\frac{\lambda_1}{z} \right) e^{-z-t_1} (\theta_1 z + \theta_2 t_1) dz \quad (5.82)$$

$$S_{31} = \int_0^{\infty} e^{-z-t_1} \frac{z}{\theta_1 z + \theta_2 t_1} dz \quad (5.83)$$

$$S_6 = \int_0^{\infty} L^2 \left(\frac{\lambda_2}{z} \right) e^{-z-t_2} \frac{t_2}{\theta_2 z + \theta_1 t_2} dz \quad (5.84)$$

La segunda derivada respecto a θ_1 y λ_2 es:

$$\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1 \partial \lambda_2} = -\sum_{i=1}^N \frac{(\lambda_1 e^{-\theta_1 x_i} - \theta_1 \lambda_1 x_i e^{\theta_1 x_i}) \theta_2 e^{-\theta_2 x_i}}{\psi^2(x_i)} \quad (5.85)$$

Que tomando valores esperados resulta:

$$-E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1 \partial \lambda_2} \right] = N \left(\frac{1}{\lambda_2} S_4 - \frac{\theta_1}{\lambda_2 \theta_2} S_5 \right) \quad (5.86)$$

La segunda derivada respecto a θ_1 y θ_2 es:

$$\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1 \partial \theta_2} = - \sum_{i=1}^N \frac{(\lambda_1 e^{-\theta_1 x_i} - \theta_1 \lambda_1 x_i e^{-\theta_1 x_i}) (\lambda_2 e^{-\theta_2 x_i} - \theta_2 \lambda_2 x_i e^{-\theta_2 x_i})}{\psi^2(x_i)} \quad (5.87)$$

Que tomando los valores esperados queda:

$$-E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1 \partial \theta_2} \right] = N \left(\frac{1}{\theta_2} S_4 - \frac{\theta_1 + \theta_2}{\theta_2^2} S_5 + \frac{\theta_1}{\theta_2^2} S_6 \right) \quad (5.88)$$

Los tres valores esperados de las segundas derivadas tomando como base λ_2 y θ_2 se pueden obtener sin más que intercambiar los subíndices 1 por 2. El resultado es:

$$-E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1^2} \right] = N \frac{\theta_2}{\lambda_2^2} S_{32} \quad (5.89)$$

donde:

$$S_{32} = \int_0^{\infty} e^{-z-t_2} \frac{z}{\theta_2 z + \theta_1 t_2} dz \quad (5.90)$$

$$-E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_2 \partial \theta_2} \right] = N \left(-\frac{1}{\theta_2^2 \lambda_2} S_{12} - \frac{\theta_1}{\lambda_2 \theta_2} S_4 + \frac{\theta_1}{\theta_2 \lambda_2} S_5 \right) \quad (5.91)$$

donde:

$$S_{12} = \int_0^{\infty} L \left(\frac{\lambda_2}{z} \right) e^{-z-t_2} (\theta_2 z + \theta_1 t_2) dz \quad (5.92)$$

$$-E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_2^2} \right] = N \left(\frac{1}{\theta_2^3} S_{22} + \frac{1}{\theta_2} S_{32} + \frac{2\theta_1}{\theta_2^2} S_5 - \frac{\theta_1}{\theta_2^2} S_6 \right) \quad (5.93)$$

donde:

$$S_{22} = \int_0^{\infty} L^2 \left(\frac{\lambda_2}{z} \right) e^{-z-t_2} (\theta_2 z + \theta_1 t_2) dz \quad (5.94)$$

Con todo esto, la matriz de información de Fisher, para el caso de emplear exclusivamente la serie sistemática de aforos, queda totalmente definida en función de 9 integrales S, las cuales solo se pueden resolver de forma aproximada numéricamente.

5.5.2.- Matriz de información de Fisher con información histórica binomial censurada

Las cuatro derivadas parciales del logaritmo de la función de verosimilitud definida en la ecuación (5.48) son:

$$\frac{\partial LL_{BC}(\Theta)}{\partial \lambda_1} = e^{-\theta_1 Q_H} \left(-M + \frac{k}{p} \right) + \frac{\partial LL_{SY}(\Theta)}{\partial \lambda_1} \quad (5.95)$$

$$\frac{\partial LL_{BC}(\Theta)}{\partial \theta_1} = \lambda_1 Q_H e^{-\theta_1 Q_H} \left(M - \frac{k}{p} \right) + \frac{\partial LL_{SY}(\Theta)}{\partial \theta_1} \quad (5.96)$$

$$\frac{\partial LL_{BC}(\Theta)}{\partial \lambda_2} = e^{-\theta_2 Q_H} \left(-M + \frac{k}{p} \right) + \frac{\partial LL_{SY}(\Theta)}{\partial \lambda_2} \quad (5.97)$$

$$\frac{\partial LL_{BC}(\Theta)}{\partial \theta_2} = \lambda_2 Q_H e^{-\theta_2 Q_H} \left(M - \frac{k}{p} \right) + \frac{\partial LL_{SY}(\Theta)}{\partial \theta_2} \quad (5.98)$$

Tomando segundas derivadas parciales y teniendo en cuenta que $E[K] = M p$, los elementos de la matriz de información de Fisher son:

$$-E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \lambda_1^2} \right] = M e^{-2\theta_1 Q_H} \frac{q}{p} - E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1^2} \right] \quad (5.99)$$

$$-E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \lambda_1 \partial \theta_1} \right] = -M \lambda_1 Q_H e^{-2\theta_1 Q_H} \frac{q}{p} - E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_1} \right] \quad (5.100)$$

$$-E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \lambda_1 \partial \lambda_2} \right] = M e^{-(\theta_1 + \theta_2) Q_H} \frac{q}{p} - E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \lambda_2} \right] \quad (5.101)$$

$$-E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \lambda_1 \partial \theta_2} \right] = -M \lambda_2 Q_H e^{-(\theta_1 + \theta_2) Q_H} \frac{q}{p} - E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_2} \right] \quad (5.102)$$

$$-E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \theta_1^2} \right] = M \lambda_1^2 Q_H^2 e^{-2\theta_1 Q_H} \frac{q}{p} - E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1^2} \right] \quad (5.103)$$

$$-E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \theta_1 \partial \lambda_2} \right] = -M \lambda_1 Q_H e^{-(\theta_1 + \theta_2) Q_H} \frac{q}{p} - E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1 \partial \lambda_2} \right] \quad (5.104)$$

$$-E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \theta_1 \partial \theta_2} \right] = M \lambda_1 \lambda_2 Q_H^2 e^{-(\theta_1 + \theta_2) Q_H} \frac{q}{p} - E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1 \partial \theta_2} \right] \quad (5.105)$$

$$-E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \lambda_2^2} \right] = M e^{-2\theta_2 Q_H} \frac{q}{p} - E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_2^2} \right] \quad (5.106)$$

$$-E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \lambda_2 \partial \theta_2} \right] = -M \lambda_2 Q_H e^{-2\theta_2 Q_H} \frac{q}{p} - E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_2 \partial \theta_2} \right] \quad (5.107)$$

$$-E \left[\frac{\partial^2 LL_{BC}(\Theta)}{\partial \theta_2^2} \right] = M \lambda_2^2 Q_H^2 e^{-2\theta_2 Q_H} \frac{q}{p} - E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_2^2} \right] \quad (5.108)$$

5.5.3.- Matriz de información de Fisher con información histórica censurada

Las derivadas parciales de la ecuación (5.49) son:

$$\frac{\partial LL_{CE}(\Theta)}{\partial \lambda_1} = - (M-k) e^{-\theta_1 Q_H} - \sum_{j=1}^k e^{-\theta_1 y_j} + \sum_{j=1}^k \frac{\theta_1 e^{-\theta_1 y_j}}{\psi(y_j)} + \frac{\partial LL_{SY}(\Theta)}{\partial \lambda_1} \quad (5.109)$$

$$\begin{aligned} \frac{\partial LL_{CE}(\Theta)}{\partial \theta_1} &= (M-k) \lambda_1 e^{-\theta_1 Q_H} + \lambda_1 \sum_{j=1}^k y_j e^{-\theta_1 y_j} + \sum_{j=1}^k \frac{\lambda_1 e^{-\theta_1 y_j} - \theta_1 \lambda_1 e^{-\theta_1 y_j}}{\Psi(y_j)} + \\ &+ \frac{\partial LL_{SY}(\Theta)}{\partial \theta_1} \end{aligned} \quad (5.110)$$

$$\frac{\partial LL_{CE}(\Theta)}{\partial \lambda_2} = - (M-k) e^{-\theta_2 Q_H} - \sum_{j=1}^k e^{-\theta_2 y_j} + \sum_{j=1}^k \frac{\theta_2 e^{-\theta_2 y_j}}{\Psi(y_j)} + \frac{\partial LL_{SY}(\Theta)}{\partial \lambda_2} \quad (5.111)$$

$$\begin{aligned} \frac{\partial LL_{CE}(\Theta)}{\partial \theta_2} &= (M-k) \lambda_2 e^{-\theta_2 Q_H} + \lambda_2 \sum_{j=1}^k y_j e^{-\theta_2 y_j} + \sum_{j=1}^k \frac{\lambda_2 e^{-\theta_2 y_j} - \theta_2 \lambda_2 e^{-\theta_2 y_j}}{\Psi(y_j)} + \\ &+ \frac{\partial LL_{SY}(\Theta)}{\partial \theta_2} \end{aligned} \quad (5.112)$$

Tomando los valores esperados de las segundas derivadas y simplificando, el resultado es:

$$-E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \lambda_1^2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1^2} \right] + M \frac{\theta_1}{\lambda_1^2} S_{31}^* \quad (5.113)$$

$$-E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \lambda_1 \partial \theta_1} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_1} \right] - M q Q_H e^{-\theta_1 Q_H} + M \left(-\frac{S_{11}^*}{\theta_1^2 \lambda_1} - \frac{S_4^*}{\lambda_1} + \frac{\theta_1}{\lambda_1 \theta_2} S_5^* \right) \quad (5.114)$$

$$-E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \lambda_1 \partial \lambda_2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \lambda_2} \right] + M \frac{\theta_1}{\lambda_1 \lambda_2} S_4^* \quad (5.115)$$

$$-E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \lambda_1 \partial \theta_2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_2} \right] + M \left(\frac{\theta_1}{\lambda_1 \theta_2} S_4^* - \frac{\theta_1}{\lambda_1 \theta_2} S_5^* \right) \quad (5.116)$$

$$\begin{aligned}
 -E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \theta_1^2} \right] &= -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1^2} \right] + M q \lambda_1 Q_H^2 e^{-\theta_1 Q_H} + \\
 &+ M \left(\frac{S_{21}^*}{\theta_1^3} + \frac{S_{31}^*}{\theta_1} + 2 \frac{S_5^*}{\theta_2} - \frac{\theta_1}{\theta_2^2} S_6^* \right)
 \end{aligned} \tag{5.117}$$

$$-E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \theta_1 \partial \lambda_2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1 \partial \lambda_2} \right] + M \left(\frac{S_4^*}{\lambda_2} - \frac{\theta_1}{\lambda_2 \theta_2} S_5^* \right) \tag{5.118}$$

$$-E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \theta_1 \partial \theta_2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1 \partial \theta_2} \right] + M \left(\frac{S_4^*}{\theta_2} - \frac{\theta_1 + \theta_2}{\theta_2^2} S_5^* + \frac{\theta_1}{\theta_2^2} S_6^* \right) \tag{5.119}$$

$$-E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \lambda_2^2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_2^2} \right] + M \frac{\theta_2^2}{\lambda_2^2} S_{32}^* \tag{5.120}$$

$$\begin{aligned}
 -E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \lambda_2 \partial \theta_2} \right] &= -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_2 \partial \theta_2} \right] - M q Q_H e^{-\theta_2 Q_H} + \\
 &+ M \left(-\frac{S_{12}^*}{\theta_2^2 \lambda_2} - \frac{\theta_1}{\lambda_2 \theta_2} S_4^* + \frac{\theta_1}{\theta_2 \lambda_2} S_5^* \right)
 \end{aligned} \tag{5.121}$$

$$\begin{aligned}
 -E \left[\frac{\partial^2 LL_{CE}(\Theta)}{\partial \theta_2^2} \right] &= -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_2^2} \right] + M q \lambda_2 Q_H e^{-\theta_2 Q_H} + \\
 &+ M \left(-\frac{S_{22}^*}{\theta_2^3} - \frac{S_{32}^*}{\theta_2} + \frac{2\theta_1}{\theta_2^2} S_5^* - \frac{\theta_1}{\theta_2^2} S_6^* \right)
 \end{aligned} \tag{5.122}$$

donde los elementos S^* vienen definidos por las integrales:

$$S_{11}^* = \int_0^{z_{1H}} L \frac{\lambda_1}{z} e^{-z-t_1} (\theta_1 z - \theta_2 t_1) dz \quad (5.123)$$

$$S_{21}^* = \int_0^{z_{1H}} L^2 \frac{\lambda_1}{z} e^{-z-t_1} (\theta_1 z + \theta_2 t_1) dz \quad (5.124)$$

$$S_{31}^* = \int_0^{z_{1H}} e^{-z-t_1} \frac{z}{\theta_1 z + \theta_2 t_1} dz \quad (5.125)$$

con el límite superior:

$$z_{1H} = \lambda_1 e^{-\theta_1 Q_H} \quad (5.126)$$

mientras que las integrales:

$$S_4^* = \int_0^{z_{2H}} e^{-z-t_2} \frac{t_2}{\theta_1 z + \theta_2 t_2} dz \quad (5.127)$$

$$S_5^* = \int_0^{z_{2H}} L \frac{\lambda_2}{z} e^{-z-t_2} \frac{t_2}{\theta_2 z + \theta_1 t_2} dz \quad (5.128)$$

$$S_6^* = \int_0^{z_{2H}} L^2 \frac{\lambda_2}{z} e^{-z-t_2} \frac{t_2}{\theta_2 z + \theta_1 t_2} dz \quad (5.129)$$

$$S_{12}^* = \int_0^{z_{2H}} L \frac{\lambda_2}{z} e^{-z-t_2} \theta_2 z + \theta_1 t_2 dz \quad (5.130)$$

$$S_{22}^* = \int_0^{z_{2H}} L^2 \frac{\lambda_2}{z} e^{-z-t_2} \theta_2 z + \theta_1 t_2 dz \quad (5.131)$$

$$S_{32}^* = \int_0^{z_{2H}} e^{-z-t_2} \frac{z}{\theta_2 z + \theta_1 t_2} dz \quad (5.132)$$

tienen como límite superior:

$$z_{2H} = \lambda_2 e^{-\theta_2 Q_H} \quad (5.133)$$

5.5.4.- Matriz de información de Fisher con información histórica de máxima avenida

Las derivadas parciales son:

$$\frac{\partial LL_{MF}(\Theta)}{\partial \lambda_1} = -M e^{-\theta_1 \omega} + \frac{\theta_1 e^{-\theta_1 \omega}}{\psi(\omega)} + \frac{\partial LL_{SY}(\Theta)}{\partial \lambda_1} \quad (5.134)$$

$$\frac{\partial LL_{MF}(\Theta)}{\partial \theta_1} = M \lambda_1 \omega e^{-\theta_1 \omega} + \frac{\lambda_1 e^{-\theta_1 \omega} - \theta_1 \lambda_1 \omega e^{-\theta_1 \omega}}{\psi(\omega)} + \frac{\partial LL_{SY}(\Theta)}{\partial \theta_1} \quad (5.135)$$

$$\frac{\partial LL_{MF}(\Theta)}{\partial \lambda_2} = -M e^{-\theta_2 \omega} + \frac{\theta_2 e^{-\theta_2 \omega}}{\psi(\omega)} + \frac{\partial LL_{SY}(\Theta)}{\partial \lambda_2} \quad (5.136)$$

$$\frac{\partial LL_{MF}(\Theta)}{\partial \theta_2} = M \lambda_2 \omega e^{-\theta_2 \omega} + \frac{\lambda_2 e^{-\theta_2 \omega} - \theta_2 \lambda_2 \omega e^{-\theta_2 \omega}}{\psi(\omega)} + \frac{\partial LL_{SY}(\Theta)}{\partial \theta_2} \quad (5.137)$$

En este caso, al tomar valores esperados de las segundas derivadas por motivos de integración numérica el cambio de variable realizado en todas las integrales es:

$$z = M \lambda_2 e^{-\theta_2 \omega} \quad (5.138)$$

con lo que la variable auxiliar t queda:

$$t = M \lambda_1 \left(\frac{z}{M\lambda_2} \right)^{\theta_1/\theta_2} \quad (5.139)$$

El resultado final es:

$$-E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \lambda_1^2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1^2} \right] + \frac{\theta_1^2}{\lambda_1^2 \theta_2} S_9^{**} \quad (5.140)$$

$$-E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \lambda_1 \partial \theta_1} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_1} \right] - \frac{S_7^{**}}{\theta_2^2 \lambda_1} - \frac{S_4^{**}}{\lambda_1} + \frac{\theta_1}{\lambda_1 \theta_2} S_5^{**} \quad (5.141)$$

$$-E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \lambda_1 \partial \lambda_2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \lambda_2} \right] + \frac{\theta_1}{\lambda_1 \lambda_2} S_4^{**} \quad (5.142)$$

$$-E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \lambda_1 \partial \theta_2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_1 \partial \theta_2} \right] + \frac{\theta_1}{\lambda_1 \theta_2} S_4^{**} - \frac{\theta_1}{\theta_2 \lambda_1} S_5^{**} \quad (5.143)$$

$$-E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \theta_1^2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1^2} \right] + \frac{S_8^{**}}{\theta_2^3} + \frac{S_9^{**}}{\theta_2} + \frac{2S_5^{**}}{\theta_2} - \frac{\theta_1}{\theta_2^2} S_6^{**} \quad (5.144)$$

$$-E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \theta_1 \partial \lambda_2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1 \partial \lambda_2} \right] + \frac{S_4^{**}}{\lambda_2} - \frac{\theta_1}{\lambda_2 \theta_2} S_5^{**} \quad (5.145)$$

$$-E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \theta_1 \partial \theta_2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_1 \partial \theta_2} \right] + \frac{S_4^{**}}{\theta_2} - \frac{\theta_1 + \theta_2}{\theta_2^2} S_5^{**} + \frac{\theta_1}{\theta_2^2} S_6^{**} \quad (5.146)$$

$$-E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \lambda_2^2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_2^2} \right] + \frac{\theta_2}{\lambda_2^2} S_{32}^{**} \quad (5.147)$$

$$-E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \lambda_2 \theta_2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda_2 \theta_2} \right] - \frac{S_{12}^{**}}{\theta_2^2 \lambda_2} - \frac{\theta_1}{\lambda_2 \theta_2} S_4^{**} + \frac{\theta_1}{\lambda_2 \theta_2} S_5^{**} \quad (5.148)$$

$$-E \left[\frac{\partial^2 LL_{MF}(\Theta)}{\partial \theta_2^2} \right] = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta_2^2} \right] + \frac{S_{22}^{**}}{\theta_2^3} - \frac{S_{32}^{**}}{\theta_2} + \frac{2\theta_1}{\theta_2^2} S_5^{**} - \frac{\theta_1}{\theta_2^2} S_6^{**} \quad (5.149)$$

donde:

$$S_7^{**} = \int_0^{\infty} L \frac{M\lambda_2}{z} \frac{e^{-z-t}}{\theta_2 z + \theta_1 t} dz \quad (5.150)$$

$$S_8^{**} = \int_0^{\infty} L^2 \frac{M\lambda_2}{z} \frac{t}{z} \frac{e^{-z-t}}{\theta_2 z + \theta_1 t} dz \quad (5.151)$$

$$S_9^{**} = \int_0^{\infty} \frac{t^2}{z} \frac{e^{-z-t}}{\theta_2 z + \theta_1 t} dz \quad (5.152)$$

$$S_4^{**} = \int_0^{\infty} e^{-z-t} \frac{t}{\theta_1 z + \theta_2 t} dz \quad (5.153)$$

$$S_5^{**} = \int_0^{\infty} L \frac{\lambda_2}{z} e^{-z-t} \frac{t}{\theta_2 z + \theta_1 t} dz \quad (5.154)$$

$$S_6^{**} = \int_0^{\infty} L^2 \frac{\lambda_2}{z} e^{-z-t} \frac{t}{\theta_2 z + \theta_1 t} dz \quad (5.155)$$

$$S_{12}^{**} = \int_0^{\infty} L \frac{\lambda_2}{z} e^{-z-t} \theta_2 z + \theta_1 t \, dz \quad (5.156)$$

$$S_{22}^{**} = \int_0^{\infty} L^2 \frac{\lambda_2}{z} e^{-z-t} \theta_2 z + \theta_1 t \, dz \quad (5.157)$$

$$S_{32}^{**} = \int_0^{\infty} e^{-z-t} \frac{z}{\theta_2 z + \theta_1 t} \, dz \quad (5.158)$$

5.6.- VARIANZAS Y GANANCIAS ESTADÍSTICAS ASINTÓTICAS DE LOS CUANTILES

5.6.1.- Obtención de la varianza y ganancia estadística asintótica

Según se vio en el Capítulo 3, la inversa de la matriz de Fisher da lugar a la matriz de varianzas y covarianzas asintóticas de los parámetros. Desafortunadamente para la función de distribución TCEV esta nueva matriz es excesivamente compleja (como se observa en los resultados del apartado anterior), por lo que en ningún momento pudo ser obtenida analíticamente de una forma correcta.

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

$$CRLB_i(\hat{X}_T) = \underline{V} I_i^{-1} \underline{V}' \quad i = 0,1,2,3 \quad (5.159)$$

donde:

$$\underline{V} = \left(\frac{\partial X_T}{\partial \lambda_1}, \frac{\partial X_T}{\partial \theta_1}, \frac{\partial X_T}{\partial \lambda_2}, \frac{\partial X_T}{\partial \theta_2} \right) \quad (5.160)$$

es decir, \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:

$$\frac{\partial X_T}{\partial \lambda_1} = \frac{e^{-\theta_1 X_T}}{\psi(X_T)} \quad (5.161)$$

$$\frac{\partial X_T}{\partial \theta_1} = -\frac{\lambda_1 X_T e^{-\theta_1 X_T}}{\psi(X_T)} \quad (5.162)$$

$$\frac{\partial X_T}{\partial \lambda_2} = \frac{e^{-\theta_2 X_T}}{\psi(X_T)} \quad (5.163)$$

Las ganancias estadísticas asintóticas de los cuantiles representan la disminución en la varianza asintótica al utilizar algún tipo de información histórica. Estas ganancias vienen dadas por la ecuación:

$$\frac{\partial X_T}{\partial \theta_2} = -\frac{\lambda_2 X_T e^{-\theta_2 X_T}}{\Psi(X_T)} \quad (5.164)$$

$$SG_i(\hat{X}_T) = 1 - \frac{CRLB_i(\hat{X}_T)}{CRLB_0(\hat{X}_T)} \quad i = 1,2,3 \quad (5.165)$$

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

- i) La relación r entre las longitudes de los periodos histórico y sistemático.
- ii) El período de retorno H del límite de percepción histórico.
- iii) El período de retorno T del cuantil considerado.
- iv) Los parámetros estandarizados λ y θ , que a su vez son función de los cuatro parámetros de la distribución TCEV según las ecuaciones (5.18) y (5.19).

Mientras que la ganancia estadística asintótica utilizando la información histórica de máxima avenida es función exclusiva de:

- i) La longitud N del registro sistemático.
- ii) La longitud M del periodo histórico.
- iii) El período de retorno T del cuantil considerado.
- iv) Los parámetros estandarizados λ y θ .

En la comprobación numérica de estos factores dependientes, los resultados se repetían al variar el resto de posibles factores con cuatro cifras significativas. Teniendo en cuenta la gran cantidad de integrales resueltas de forma aproximada, se puede considerar que dichos resultados eran realmente constantes.

Mientras que en la familia de distribuciones "Extreme Value", analizadas en el Capítulo 4, la ganancia estadística asintótica de los cuantiles era independiente de los dos parámetros posibles, en la distribución TCEV no ocurre lo mismo. Aunque eso sí, no es función de los 4 parámetros, sino solo de los 2 parámetros estandarizados.

Al no disponer de resultados analíticos, solo se puede analizar la influencia de los factores dependientes desde un punto cualitativo, que es lo que se refleja a continuación.

5.6.2.- Influencia de los parámetros estandarizados en la ganancia estadística asintótica

Mediante el programa PLOTLA se han obtenido las ganancias estadísticas asintóticas para 400 puntos del plano λ - θ para los cuantiles de 100 y 500 años de periodo de retorno. Estos periodos de retorno se han escogido por ser representativos de las zonas medias y altas de los cuantiles, que es donde las ganancias estadísticas van a ser mayores por corresponder a periodos de retorno superiores a la longitud del registro sistemático. El ratio de longitudes es de 5, con una longitud del registro sistemático de 100 años para el caso de emplear información histórica de máxima avenida. Cuando la información histórica es binomial censurada y censurada, el límite de percepción es de 50 años de periodo de retorno.

Los resultados se han transformado en líneas de isoganancia estadística utilizando el paquete informático SURFER. Los resultados se presentan en las Figuras (5.6) a (5.11). Las irregularidades en las líneas son debidas a los pequeños errores de las integrales que aparecen en las expresiones de los elementos de las matrices de Fisher y al método de interpolación de la red inicial para la obtención de la superficie.

Como se observa en las Figuras (5.6) y (5.7), la ganancia estadística utilizando información histórica binomial censurada aumenta con el parámetro λ , o lo que es lo mismo, cuanto mayor es la proporción de avenidas extraordinarias. Los cuantiles de 100 y 500 años son originados por las avenidas extraordinarias, por lo que la ganancia será tanto mayor cuanto mayor sea la información sobre las mismas. Para λ muy pequeños existe un valle, donde la ganancia estadística disminuye con θ , de tal forma que decrece a medida que aumenta el coeficiente de sesgo de la población.

En el valle, la ganancia estadística es muy pequeña para periodos de retorno altos, mientras que para retornos medios las diferencias en todo el plano son de solo el 15%. Además aparece una ligera cumbre paralela al valle y para valores de λ mayores.

En lo que respecta a la ganancia estadística utilizando información histórica censurada

(Figuras 5.8 y 5.9) para el rango habitual en Hidrología de los parámetros, disminuye con λ y θ . Aunque al ser las diferencias mínimas (del orden del 10%) se puede considerar a efectos prácticos que es independiente de los 4 parámetros.

De todas formas, aparece una cumbre donde con la información binomial censurada había un valle. Es decir, si las avenidas extraordinarias son grandes comparadas con las ordinarias y además poco frecuentes, la información binomial censurada se deteriora, mientras que la información censurada aumenta de valor. La razón se encuentra en que el número de avenidas que superan el límite de percepción sigue una distribución binomial cuya varianza aumenta a medida que disminuye la probabilidad de superación del límite. Por tanto, la información binomial censurada se empobrece por ese aumento de variabilidad. Sin embargo, la información censurada además hace uso del valor de las avenidas históricas que probablemente son extraordinarias y no se encuentran en la serie sistemática.

Las Figuras (5.10) y (5.11) representan la ganancia estadística asintótica utilizando la información histórica de máxima avenida. Para periodos de retorno altos ($T=500$ años) presenta una cumbre para λ bajo. En la zona alta del parámetro λ la ganancia estadística crece con el parámetro θ , es decir, cuanto más similares en magnitud son las avenidas extraordinarias y ordinarias. En cualquier caso, la tendencia general es a aumentar a medida que disminuye λ y las avenidas extraordinarias se hacen menos frecuentes.

Para periodos de retorno medios ($T=100$ años) la cumbre se desplaza hacia λ mayores y aparece un poco profundo valle para λ bajos, siendo las tendencias generales similares a las anteriores.

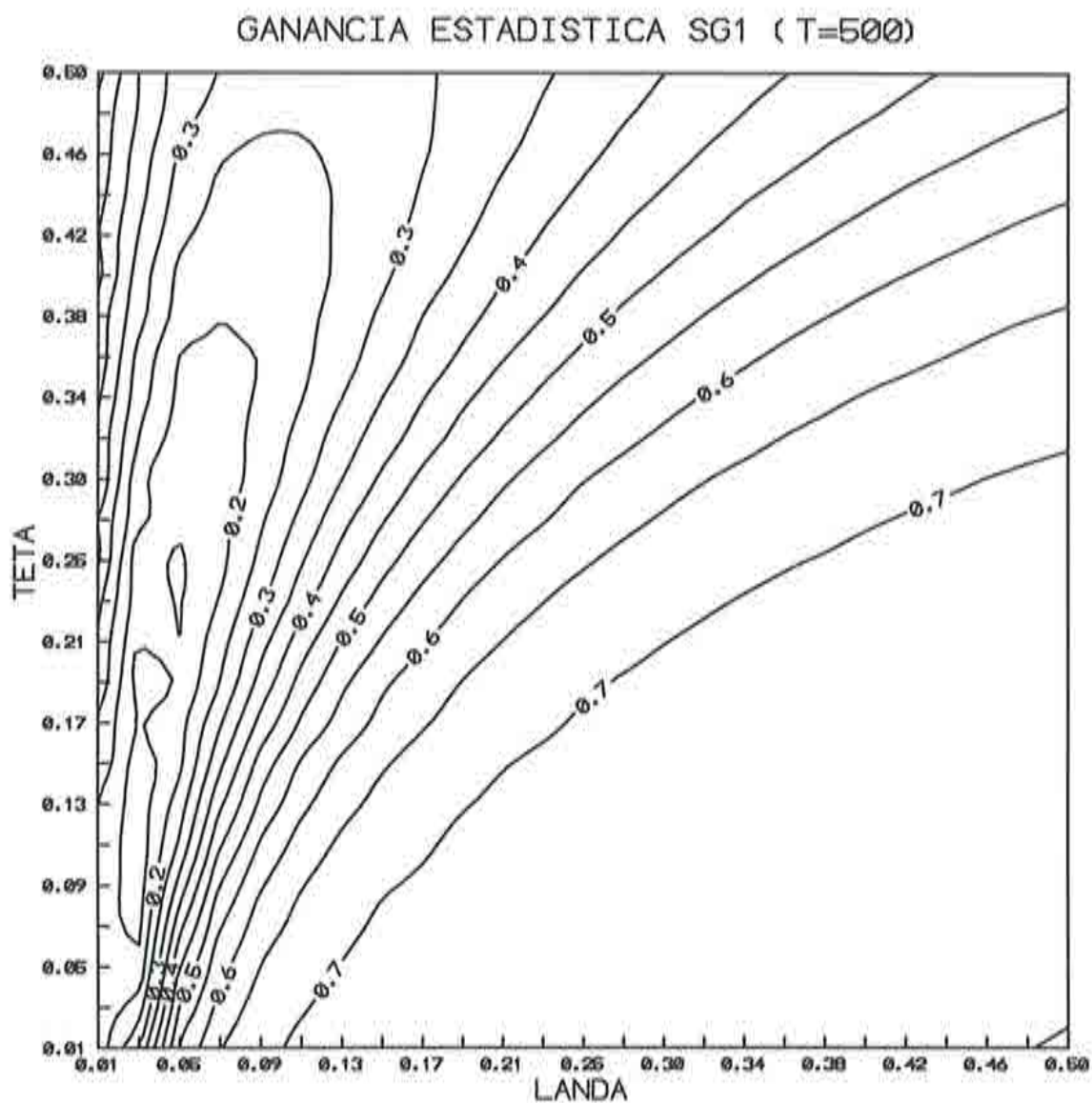


Figura 5.6. Ganancia estadística asintótica con información histórica binomial censurada en función de los parámetros estandarizados, para el cuantil de 500 años y escenario inicial.

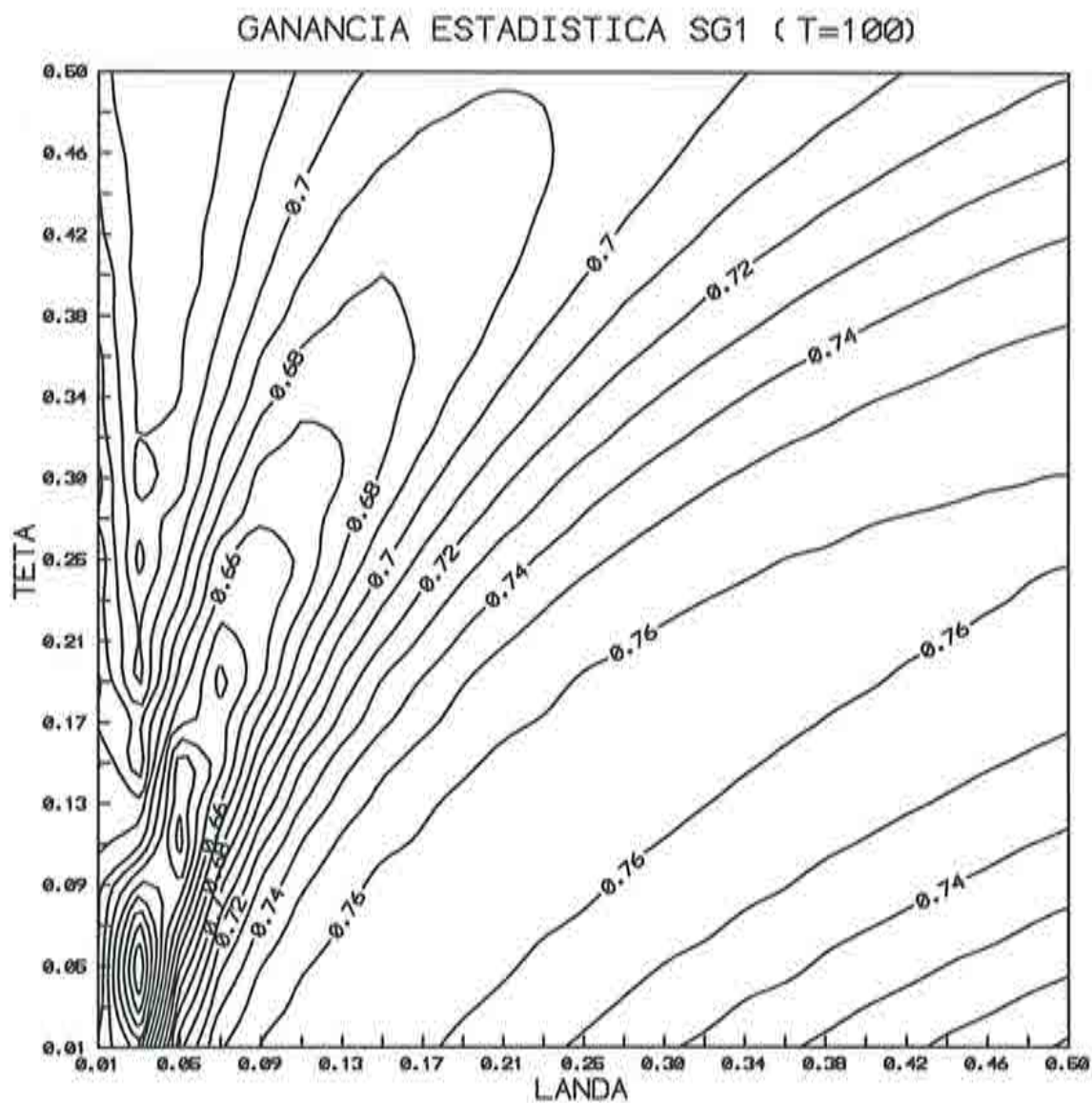


Figura 5.7. Ganancia estadística asintótica con información histórica binomial censurada en función de los parámetros estandarizados, para el cuantil de 100 años y escenario inicial.

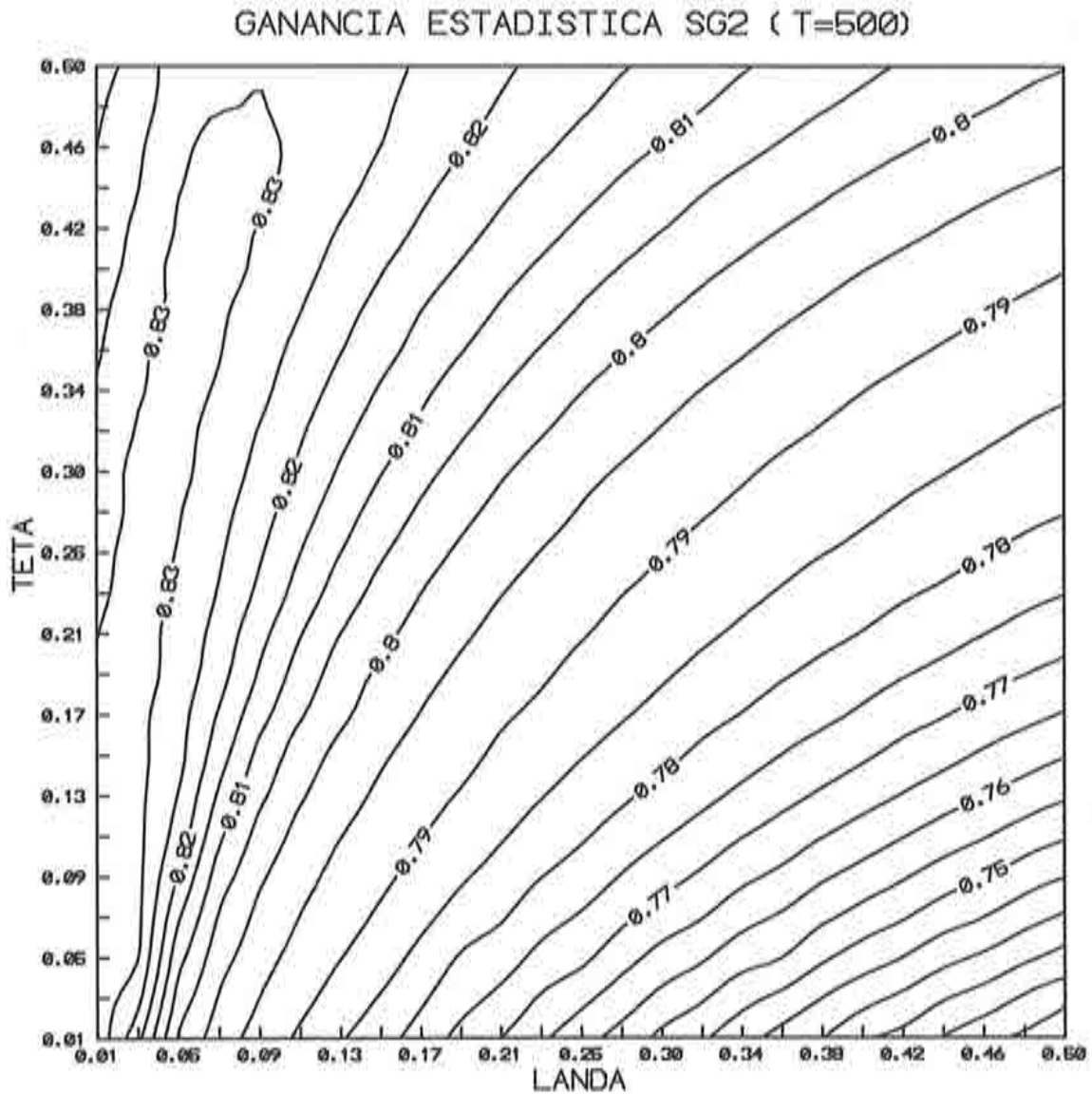


Figura 5.8. Ganancia estadística asintótica con información histórica censurada en función de los parámetros estandarizados, para el cuantil de 500 años y escenario inicial.

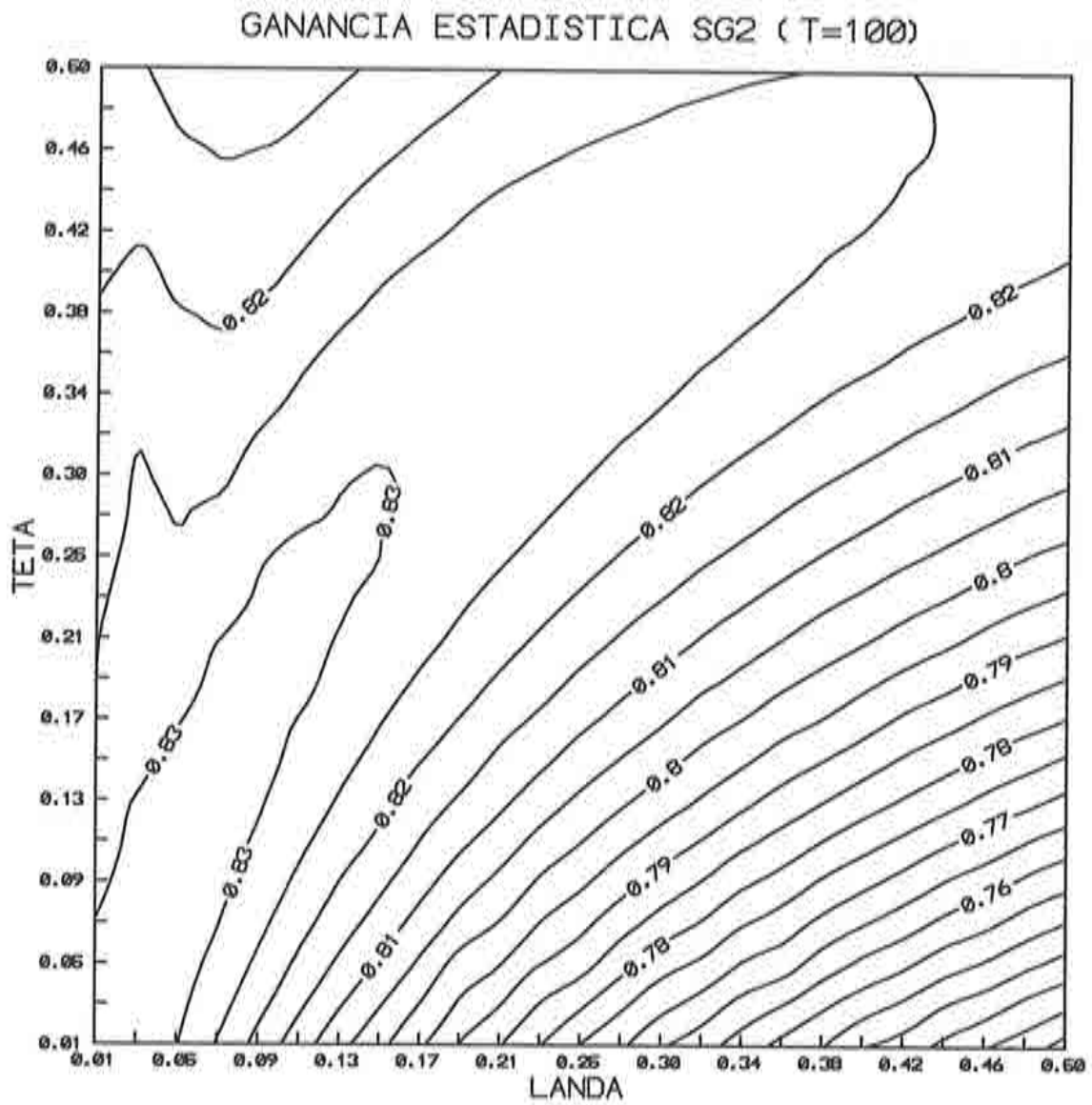


Figura 5.9. Ganancia estadística asintótica con información histórica censurada en función de los parámetros estandarizados, para el cuantil de 100 años y escenario inicial.

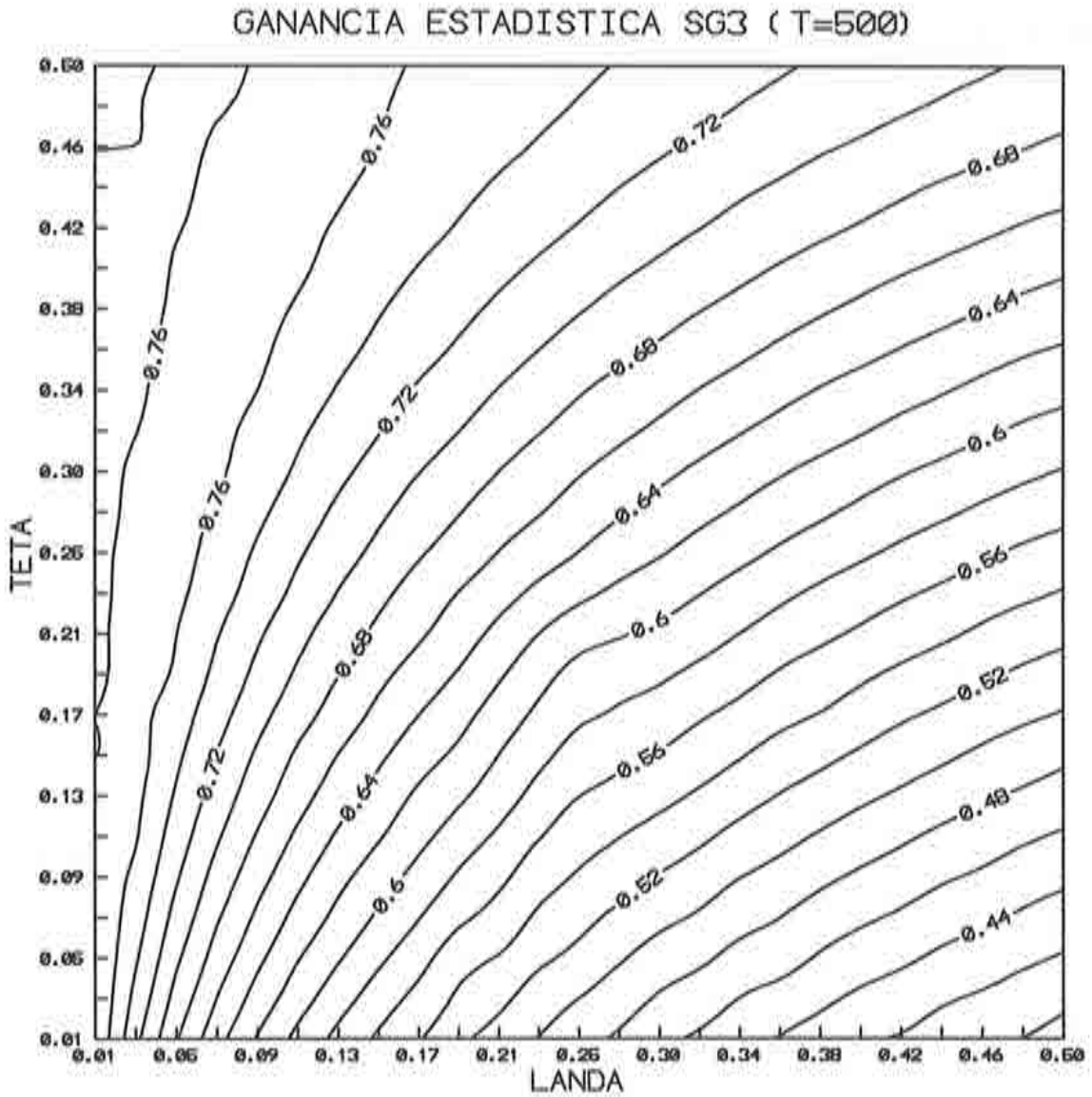


Figura 5.10. Ganancia estadística asintótica con información histórica de máxima avenida en función de los parámetros estandarizados, para el cuantil de 500 años y escenario inicial.

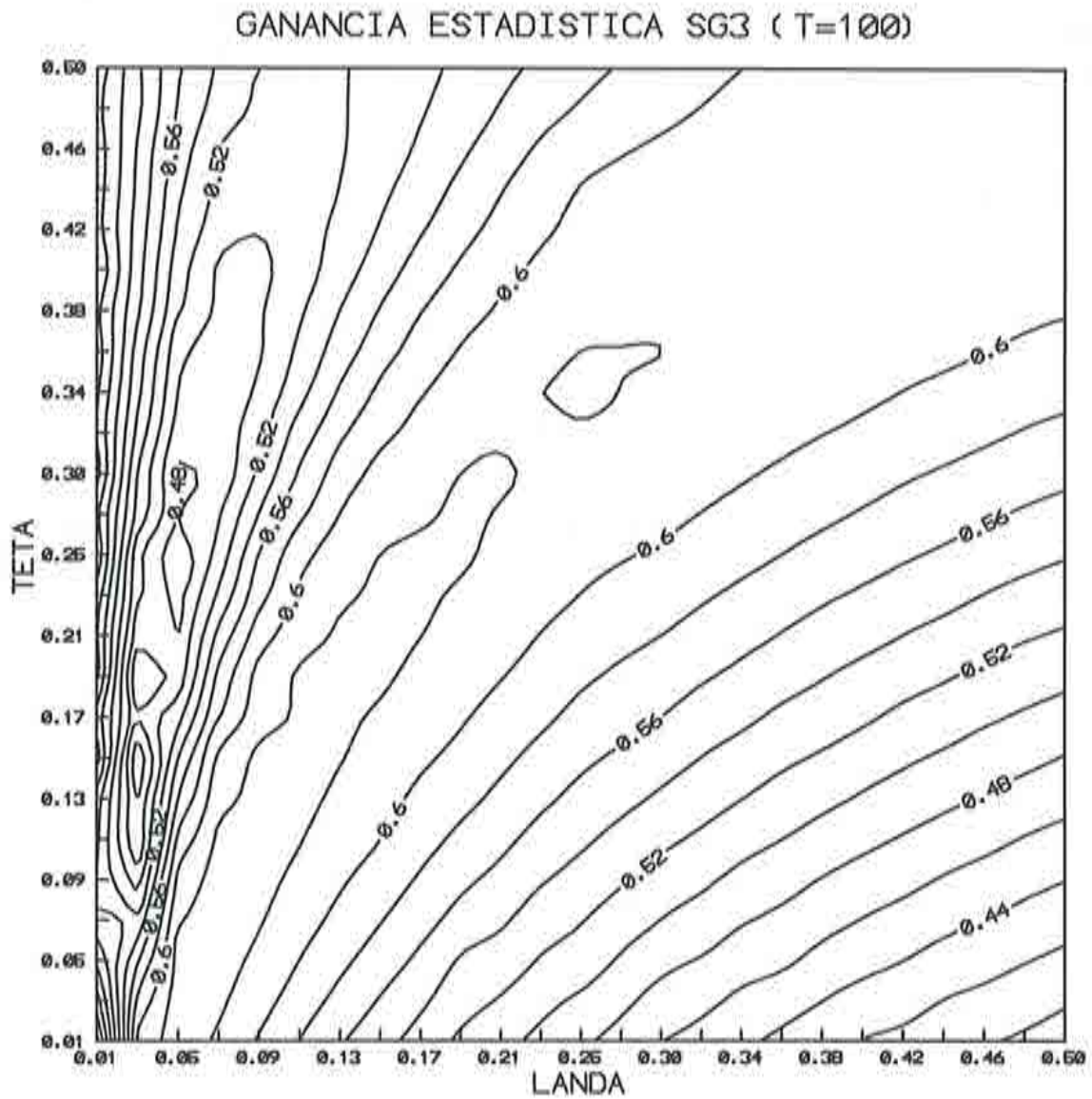


Figura 5.11. Ganancia estadística asintótica con información histórica de máxima avenida en función de los parámetros estandarizados, para el cuantil de 100 años y escenario inicial.

5.6.3.- Variación de las ganancias estadísticas asintóticas en función del período de retorno del cuantil

En este apartado se analiza la variación de la ganancia estadística asintótica en función del período de retorno del cuantil en diferentes escenarios. Estos escenarios se forman a partir de uno inicial cuyas características son:

- Parámetro estandarizado $\lambda = 0.2$. Es decir, una probabilidad de presentación de la avenida extraordinaria de aproximadamente 0.18, según se puede observar en la Figura (5.1).
- Parámetro estandarizado $\theta = 0.1$. Es decir, una relación entre las avenidas ordinarias y extraordinarias de un orden de magnitud, como se comentó en el Apartado 5.3.2.
- ratio entre las longitudes del período histórico y sistemático de 5.
- Para SG_1 y SG_2 un límite de percepción de 50 años.
- Para SG_3 una longitud de la serie sistemática de 100 años.

El resto de escenarios se han formado variando estas características iniciales. El programa utilizado es PLOTT que genera las ganancias estadísticas de 200 cuantiles. Estos resultados se han condensado en las Figuras (5.12) a (5.20). Las propiedades cualitativas que pueden extraerse de estos resultados son:

- i) En todos los casos y para toda la gama de cuantiles, la información histórica censurada SG_2 proporciona la mayor ganancia estadística asintótica.
- ii) En general, la información histórica binomial censurada SG_1 proporciona mayores ganancias estadísticas asintóticas que la de máxima avenida SG_3 . Las excepciones son:
 - Que las avenidas extraordinarias y las ordinarias estén dentro del mismo orden de magnitud (θ alto, como se observa en la Figura 5.14, frente a las Figuras 5.12 y 5.13). Esta situación no debe ser la habitual en climas mediterráneos.
 - Que la frecuencia de las avenidas extraordinarias sea baja (λ bajo, Figura 5.15 en contraposición con las Figuras 5.12 y 5.16). En esta situación, SG_1 cae por debajo de SG_3 para cuantiles medios y altos.

- Que el límite de percepción histórica sea pequeño (Figura 5.17 frente a las Figuras 5.12 y 5.18), y por tanto el número de avenidas que superen este límite van a ser ordinarias y extraordinarias. Esta mezcla de información aporta muy poco a los cuantiles medios y altos que están generados, probabilísticamente hablando, por las avenidas extraordinarias.

iii) La ganancia estadística asintótica máxima con información binomial censurada SG_1 se alcanza con el cuantil de período de retorno igual al límite de percepción. Esto solo ha podido ser probado numéricamente.

iv) A partir del cuantil de período de retorno igual al límite de percepción, se observa en todos los casos, que SG_2 sufre muy poca variación, aunque alcanza un máximo inmediatamente después del punto mencionado. La posición de esta máximo es función exclusiva de los parámetros estandarizados λ , θ y de H . No lo es del ratio r .

v) Para cuantiles medios y altos, SG_3 sufre pocas variaciones, aunque alcanza un máximo en los alrededores del cuantil de período de retorno igual a la longitud histórica. Esta posición del máximo no es función de r y solo lo es de M y de λ y θ .

vi) Como se observa en las Figuras (5.19) y (5.20), a medida que aumenta r , las ganancias estadísticas en todo el rango de cuantiles aumentan: en la Figura (5.19) con un ratio de 10 las ganancias estadísticas son mucho mayores que en la Figura (5.20) donde el ratio es de 2.5.

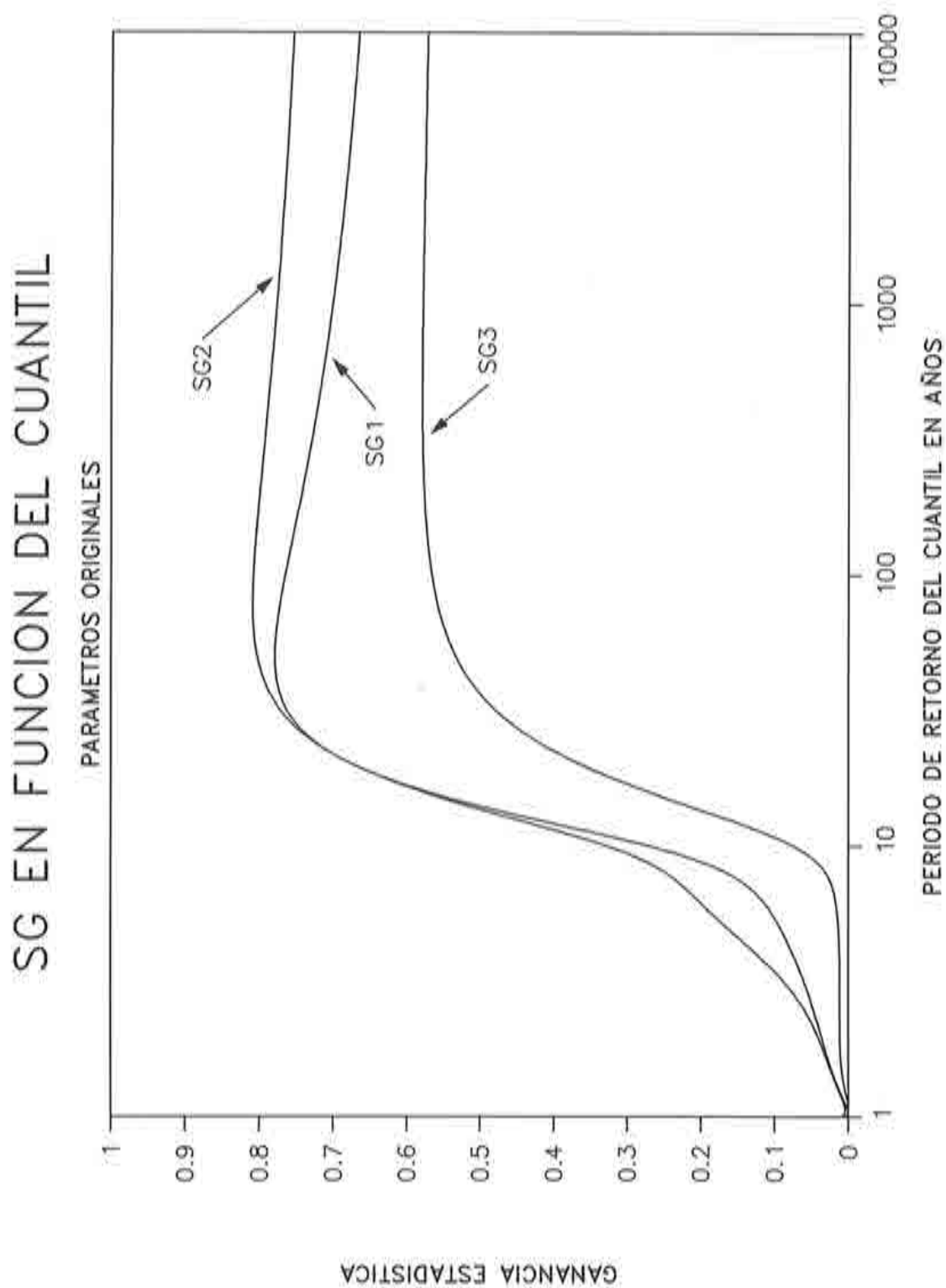


Figura 5.12. Ganancias estadísticas asintóticas en función del período de retorno del cuantil, para el escenario original de factores.

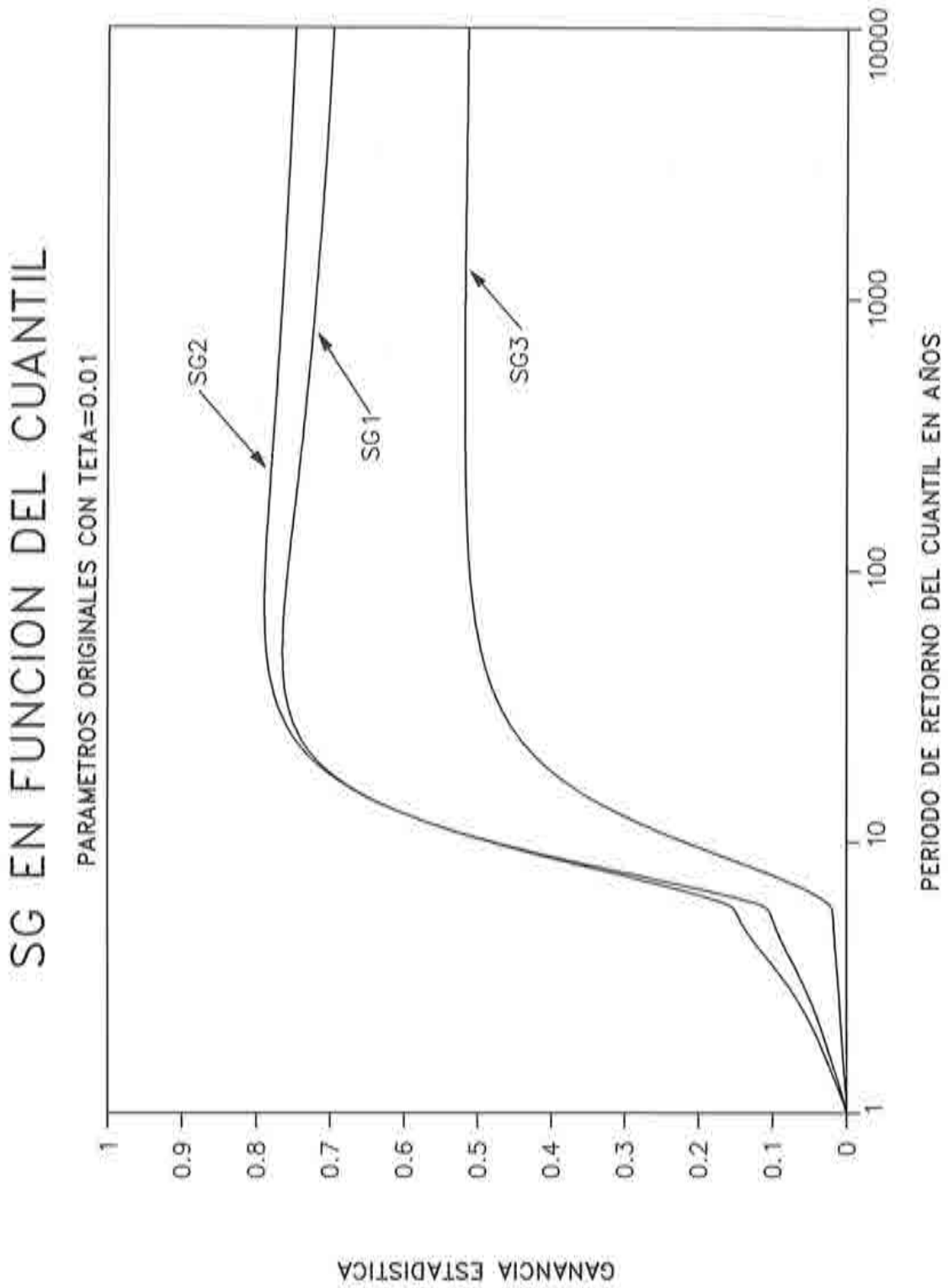


Figura 5.13. Ganancias estadísticas asintóticas en función del periodo de retorno del cuantil, con el parámetro $\theta=0.01$.

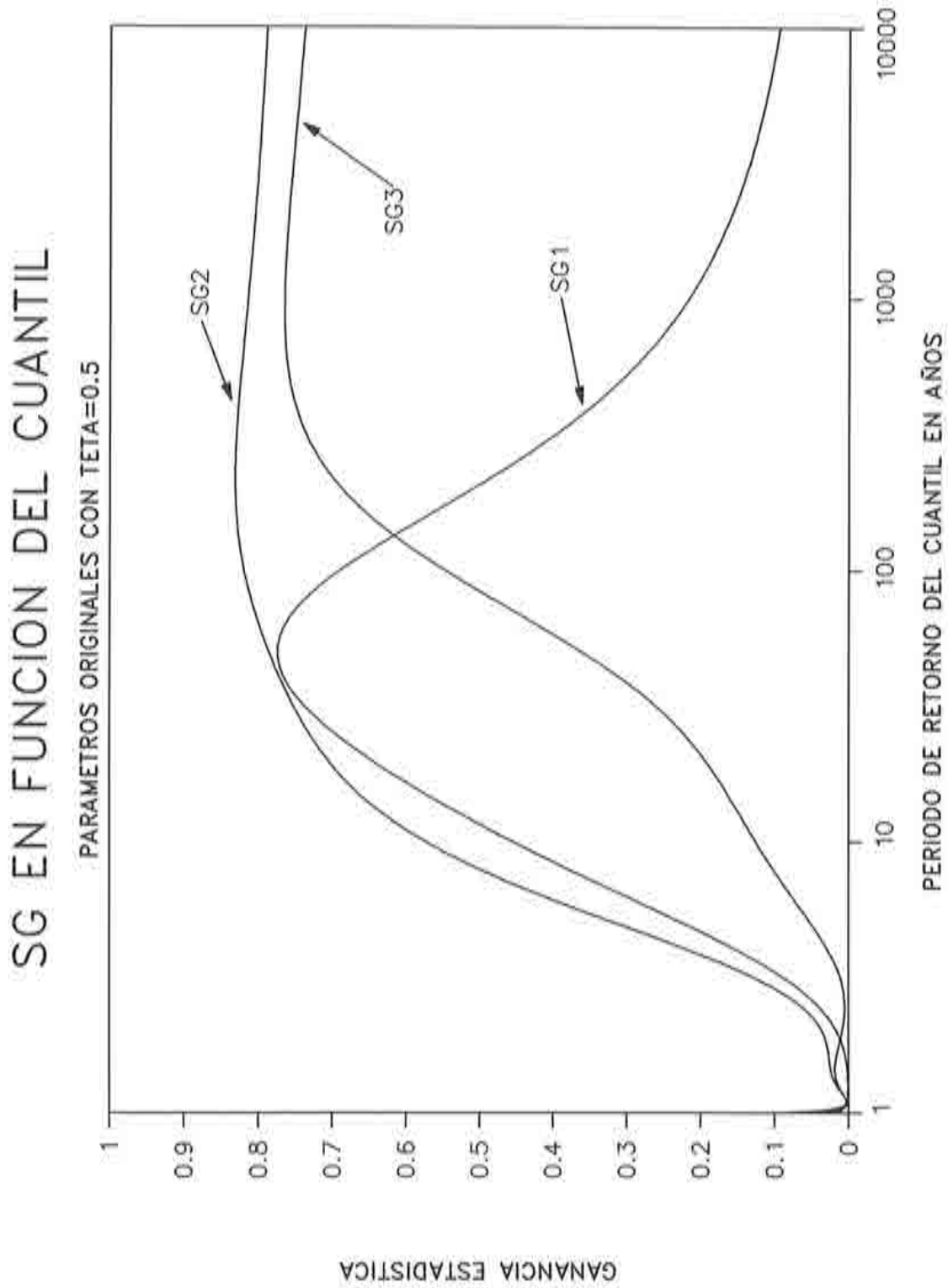


Figura 5.14. Ganancias estadísticas asintóticas en función del período de retorno del cuantil, con el parámetro $\theta=0.5$.

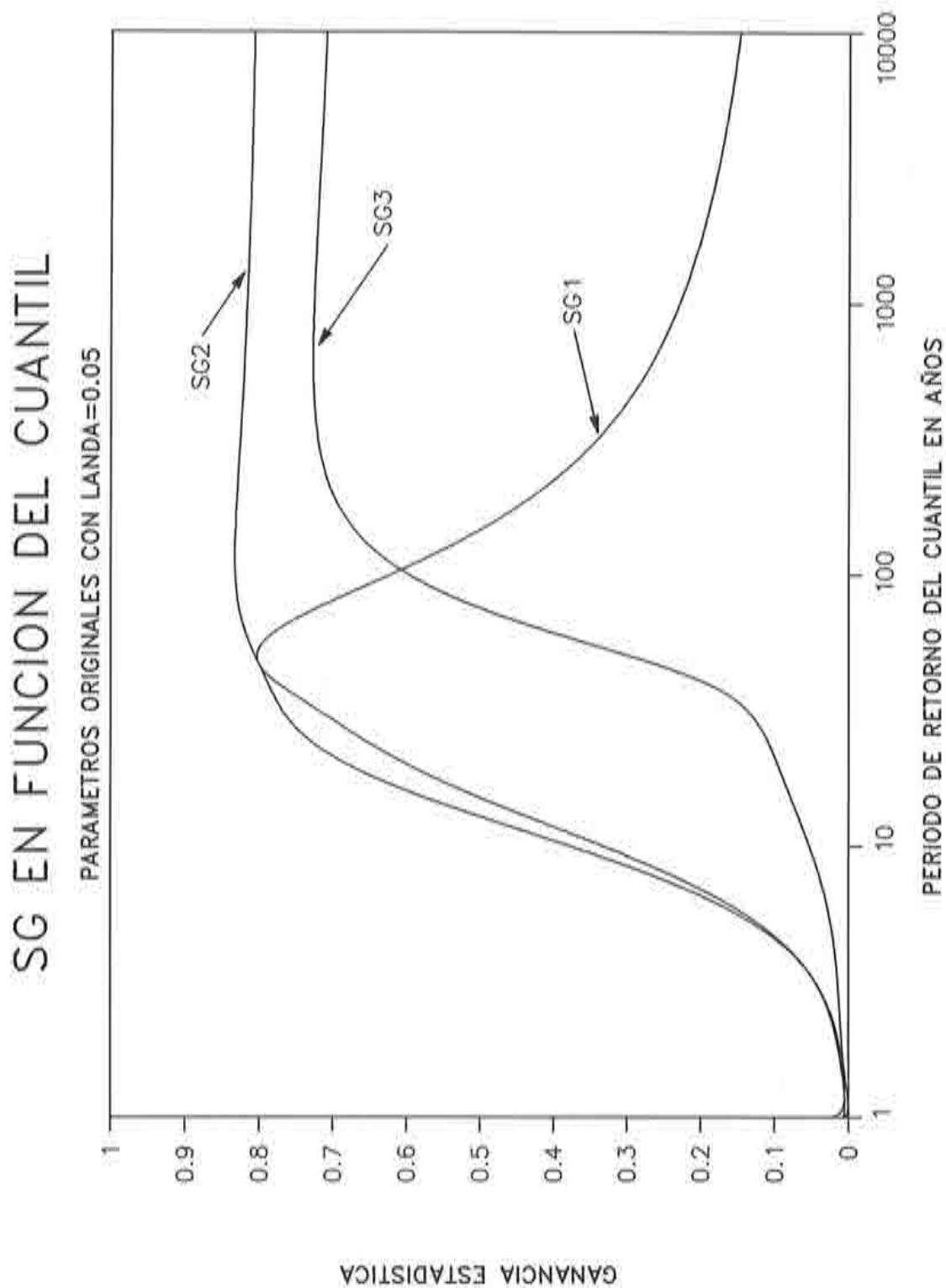


Figura 5.15. Ganancias estadísticas asintóticas en función del período de retorno del cuantil, con el parámetro $\lambda= 0.05$.

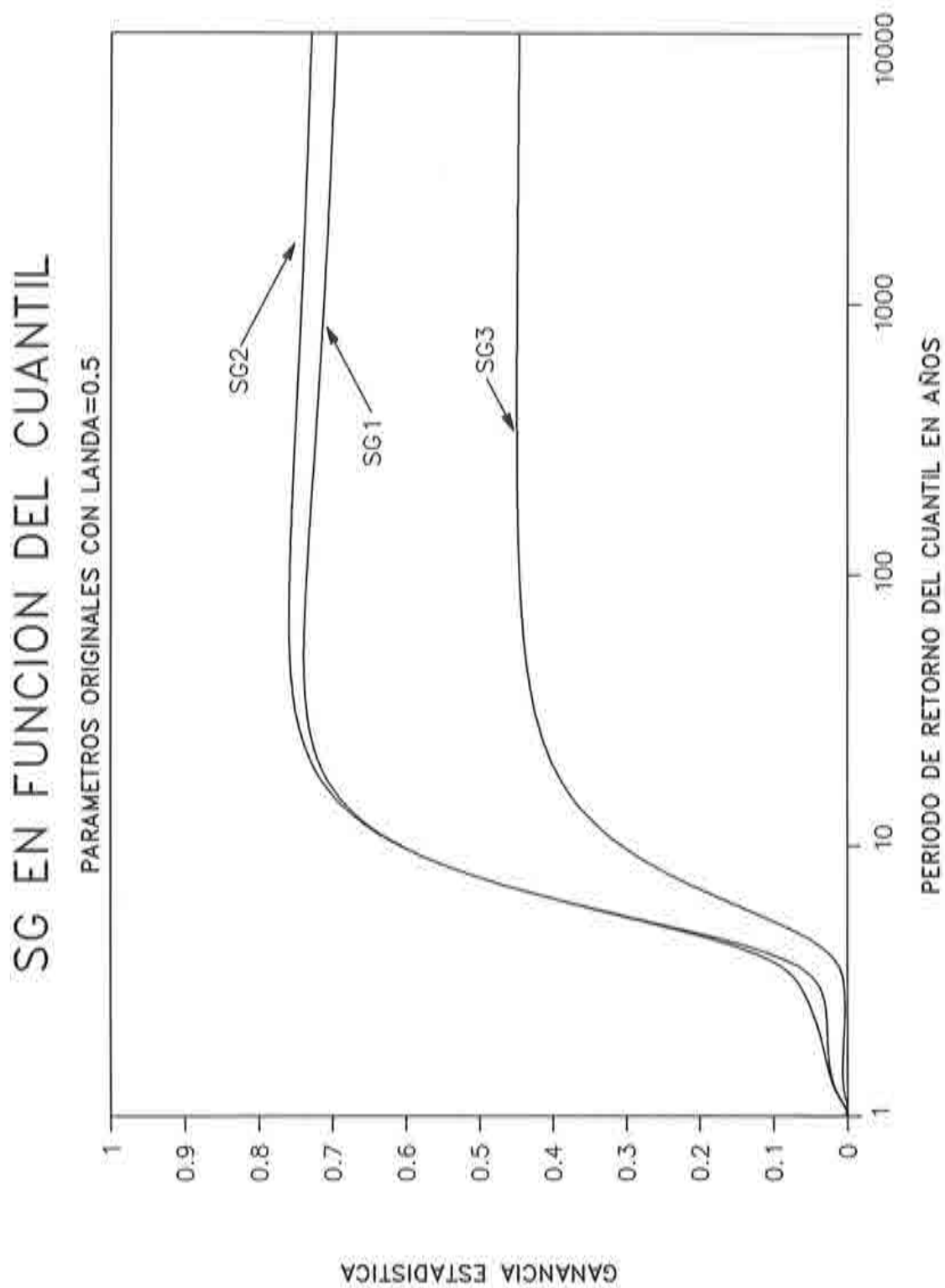


Figura 5.16. Ganancias estadísticas asintóticas en función del período de retorno del cuantil, con el parámetro $\lambda = 0.5$.

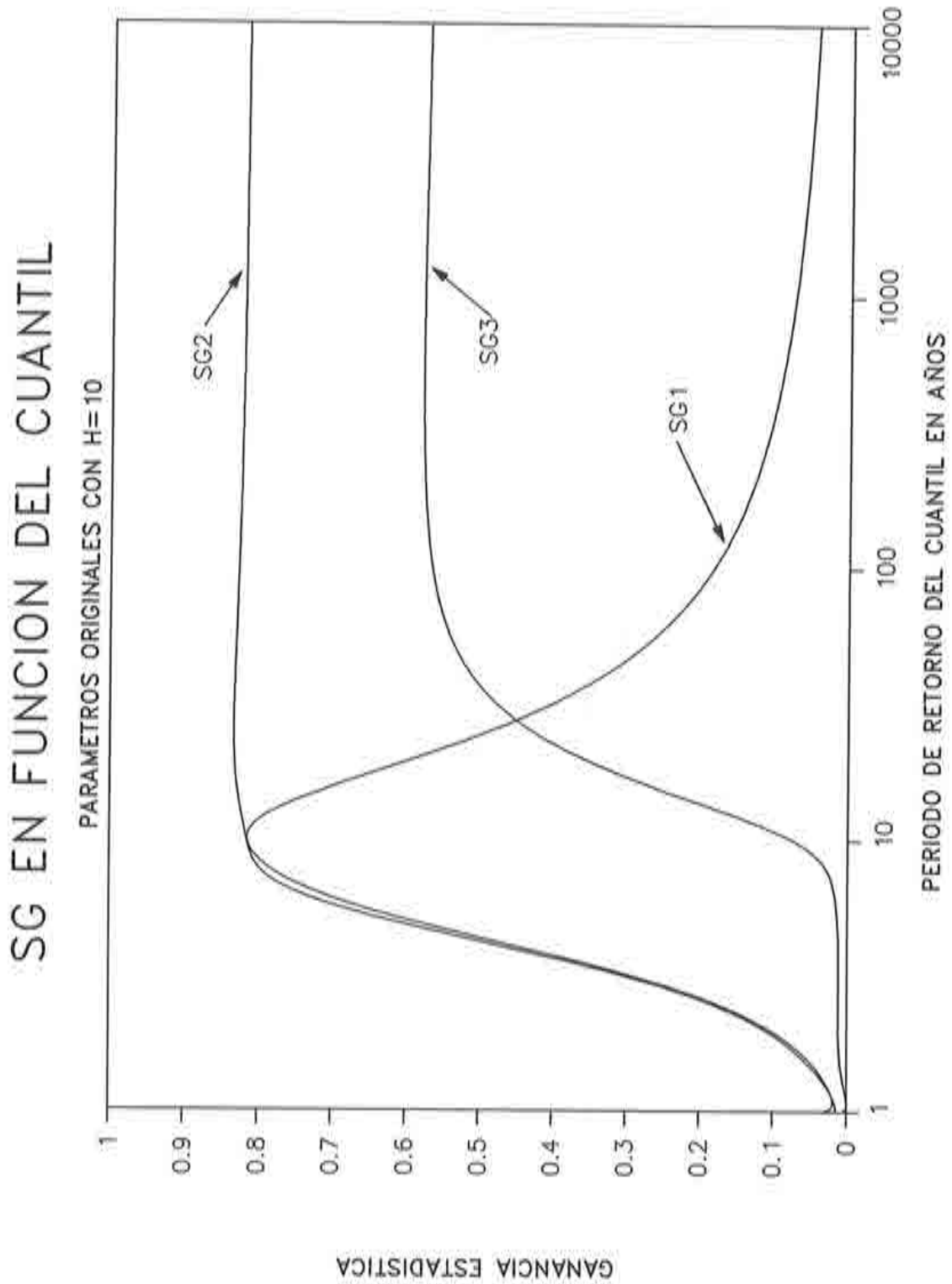


Figura 5.17. Ganancias estadísticas asintóticas en función del período de retorno del cuantil, para un límite de percepción histórico de 10 años.

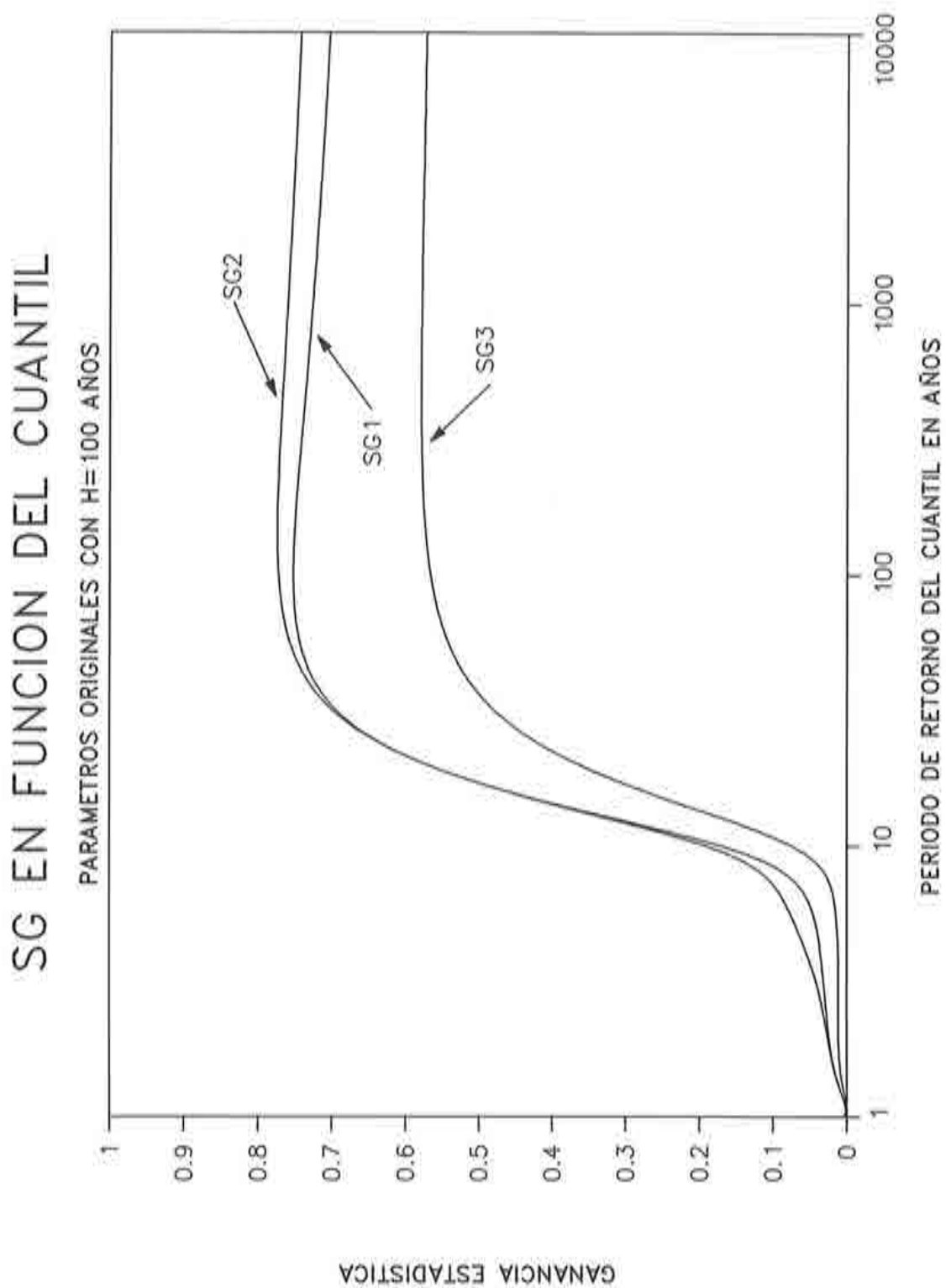


Figura 5.18. Ganancias estadísticas asintóticas en función del periodo de retorno del cuantil, para un límite de percepción histórico de 100 años.

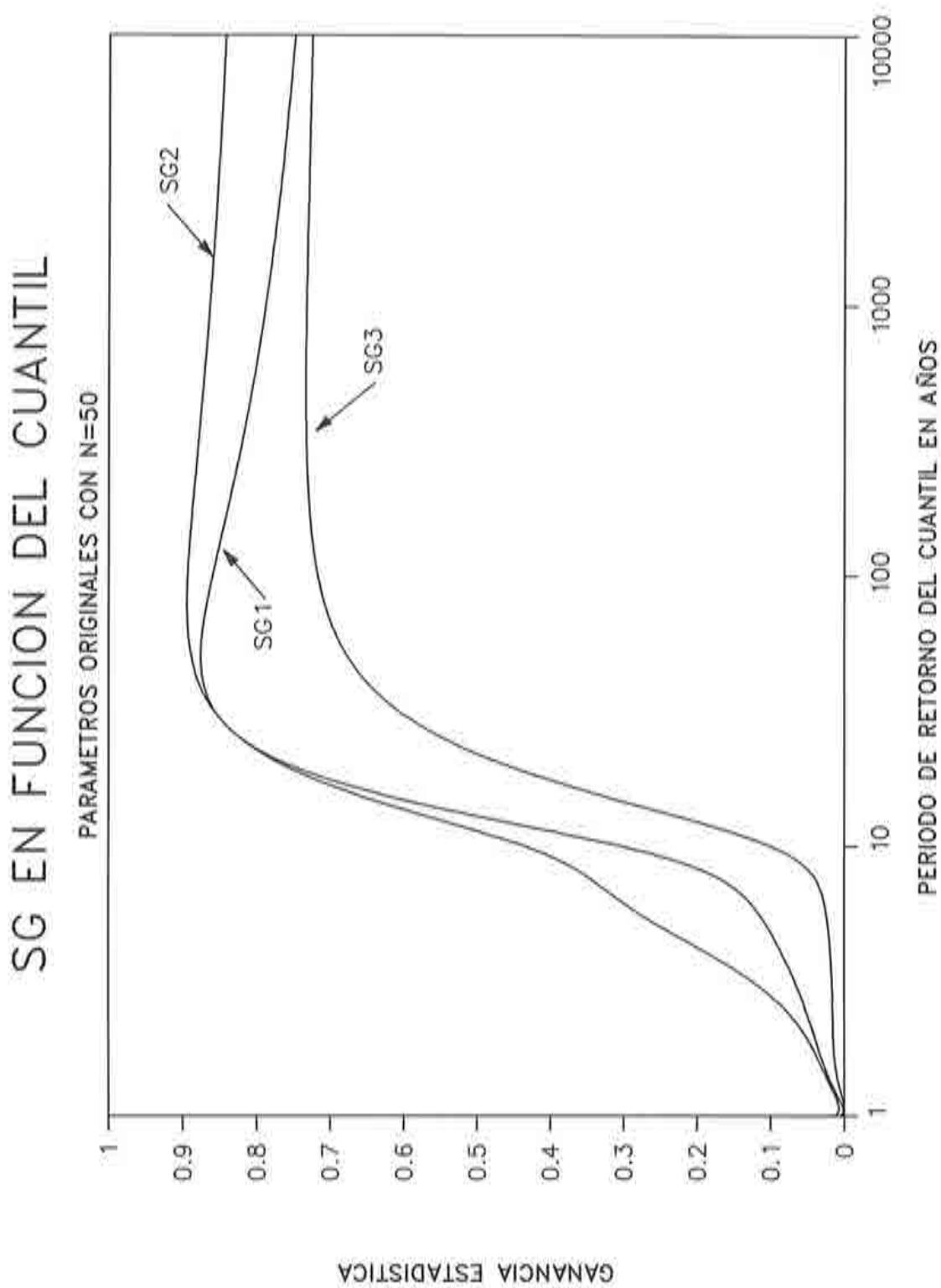


Figura 5.19. Ganancias estadísticas asintóticas en función del período de retorno del cuantil, para una longitud de la serie sistemática de 50 años.

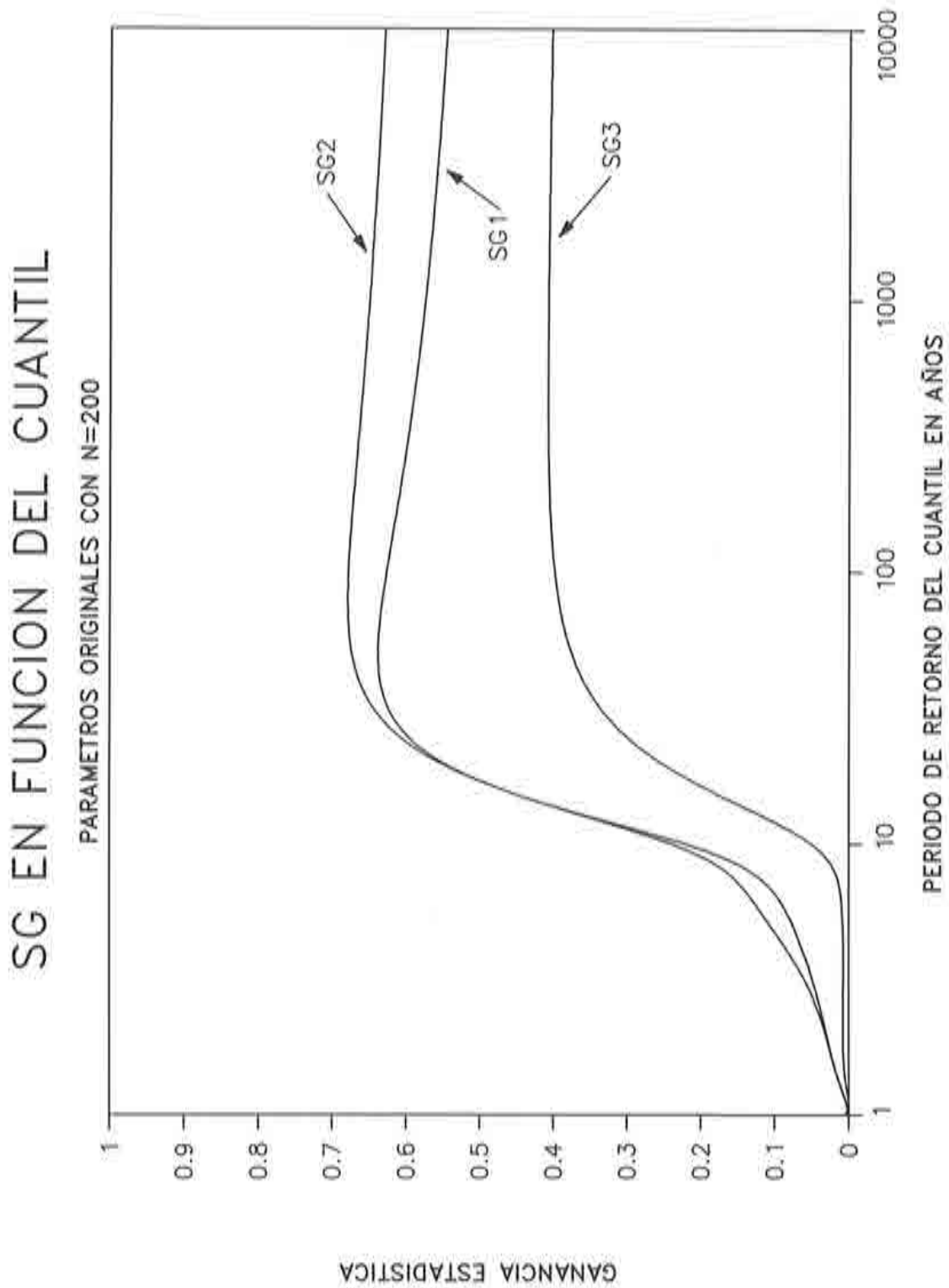


Figura 5.20. Ganancias estadísticas asintóticas en función del período de retorno del cuantil, para una longitud de la serie sistemática de 200 años.

5.6.4.- Variación de la ganancia estadística asintótica en función del límite de percepción histórico

Evidentemente, el límite de percepción histórico solo es un factor influyente en las ganancias estadísticas asintóticas con información histórica binomial censurada SG_1 y censurada SG_2 y no lo es con la máxima avenida. De todas formas, se ha incluido el valor constante de SG_3 en los gráficos para poder comparar las diferencias entre las tres ganancias asintóticas. Como en el apartado anterior se parte del mismo escenario inicial y se han obtenidos los diferentes resultados variando algún factor de ese escenario.

El programa utilizado es el PLOTH y los resultados se condensan en las Figuras (5.21) a (5.29). De los mismos, se pueden deducir las propiedades siguientes:

i) Para límites de percepción bajos y medios, que son los que habitualmente se dan en la Hidrología Histórica, la ganancia estadística asintótica mayor se obtiene con la información histórica censurada. Solo para límites de percepción superiores a 500 años de período de retorno la información de máxima avenida es superior.

ii) Para el cuantil de 100 años, SG_2 permanece prácticamente constante para límites inferiores a 100 años aproximadamente, que es el rango habitual. El valor de la ganancia estadística asintótica en esta zona es, a efectos prácticos, independiente de los parámetros estandarizados. (Figuras 5.21, 5.22, 5.23, 5.24 y 5.25). También es independiente del cuantil (Figuras 5.28 y 5.29) en magnitud, aunque para cuantiles inferiores (Figura 5.28) el rango disminuye en la misma medida que el cuantil. Para cuantiles altos (Figura 5.29) el rango donde es constante SG_2 aumenta ligeramente.

iii) La ganancia estadística asintótica utilizando la información binomial censurada siempre es inferior a la correspondiente a información censurada. La diferencia es muy pequeña para límites de percepción altos en todos los casos. Sin embargo, la diferencia es muy grande para límites de percepción bajos, donde SG_1 es muy pequeña.

iv) Para cuantiles medios y bajos (Figuras 5.21 y 5.28) SG_1 alcanza un máximo en el límite de percepción similar al cuantil. El máximo se adelanta para las situaciones de θ y λ bajos (Figura 6.22 y 6.24), que se corresponden con las avenidas extraordinarias muy grandes y poco frecuentes.

v) Para cuantiles altos (Figura 5.29) el máximo se adelanta en mayor medida.

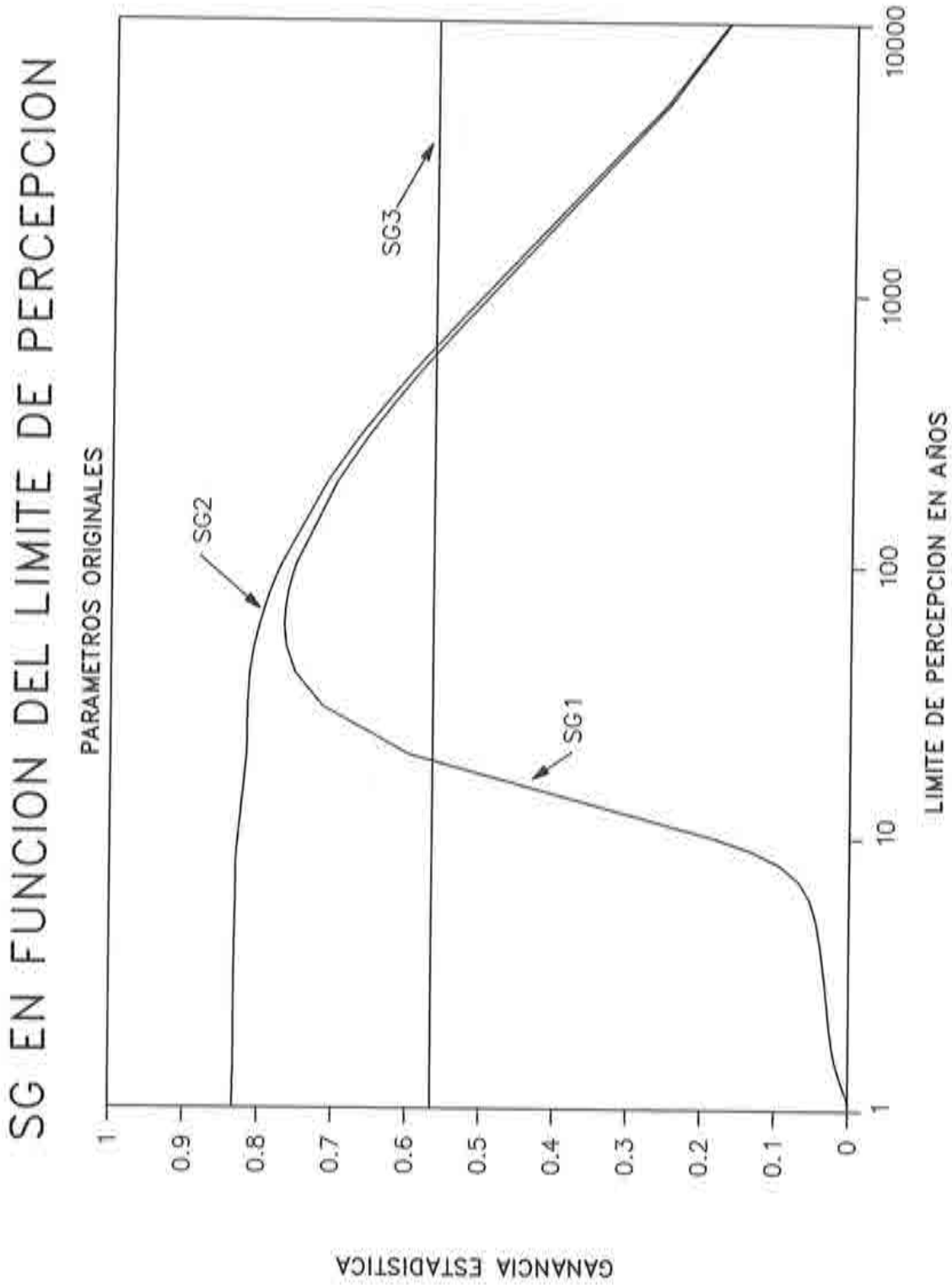


Figura 5.21. Ganancias estadísticas asintóticas en función del límite de percepción histórico, con el escenario de factores inicial.

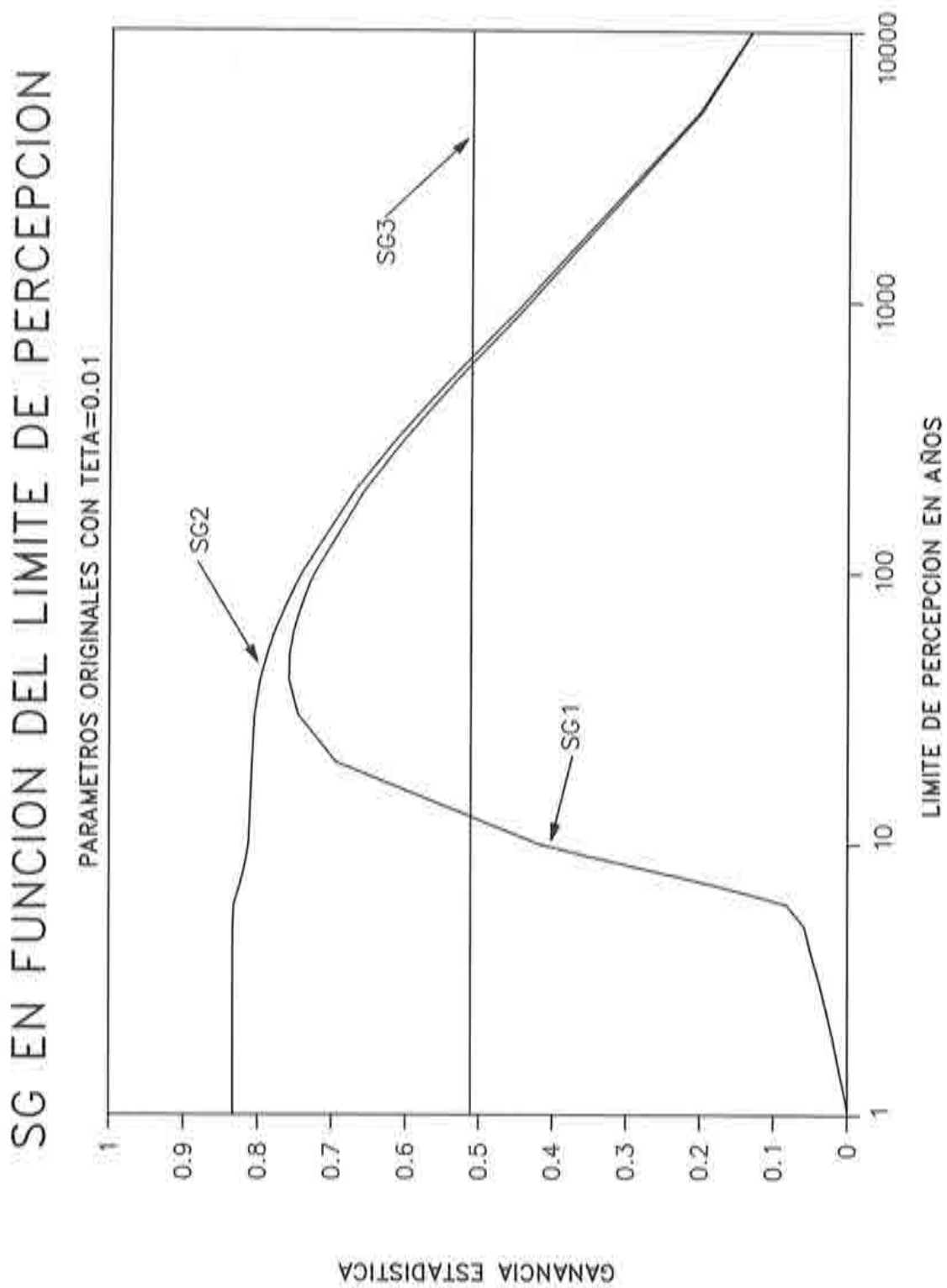


Figura 5.22. Ganancias estadísticas asintóticas en función del límite de percepción histórico, con el parámetro $\theta=0.01$.

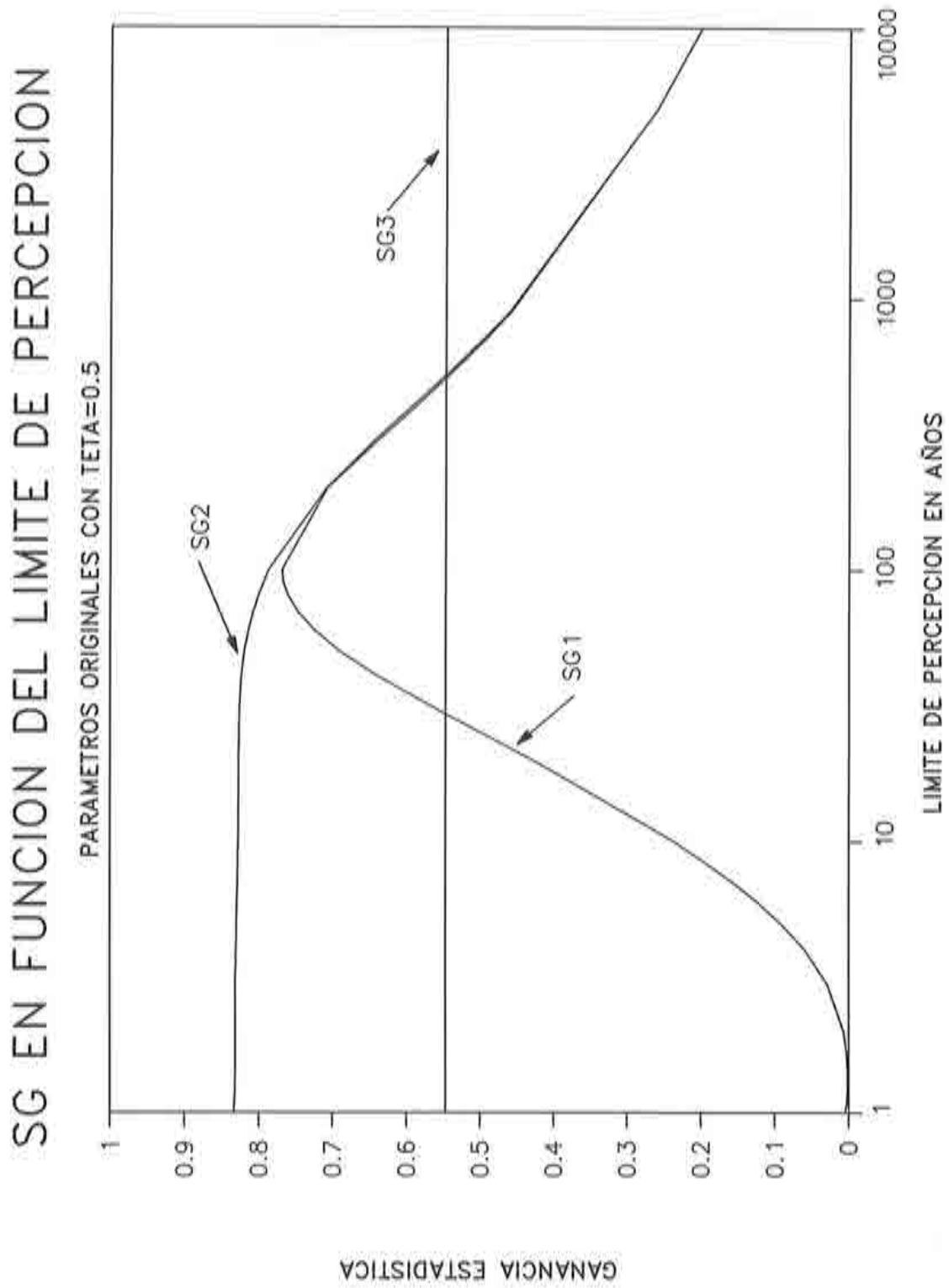


Figura 5.23. Ganancias estadísticas asintóticas en función del límite de percepción histórico, con el parámetro $\theta=0.5$.

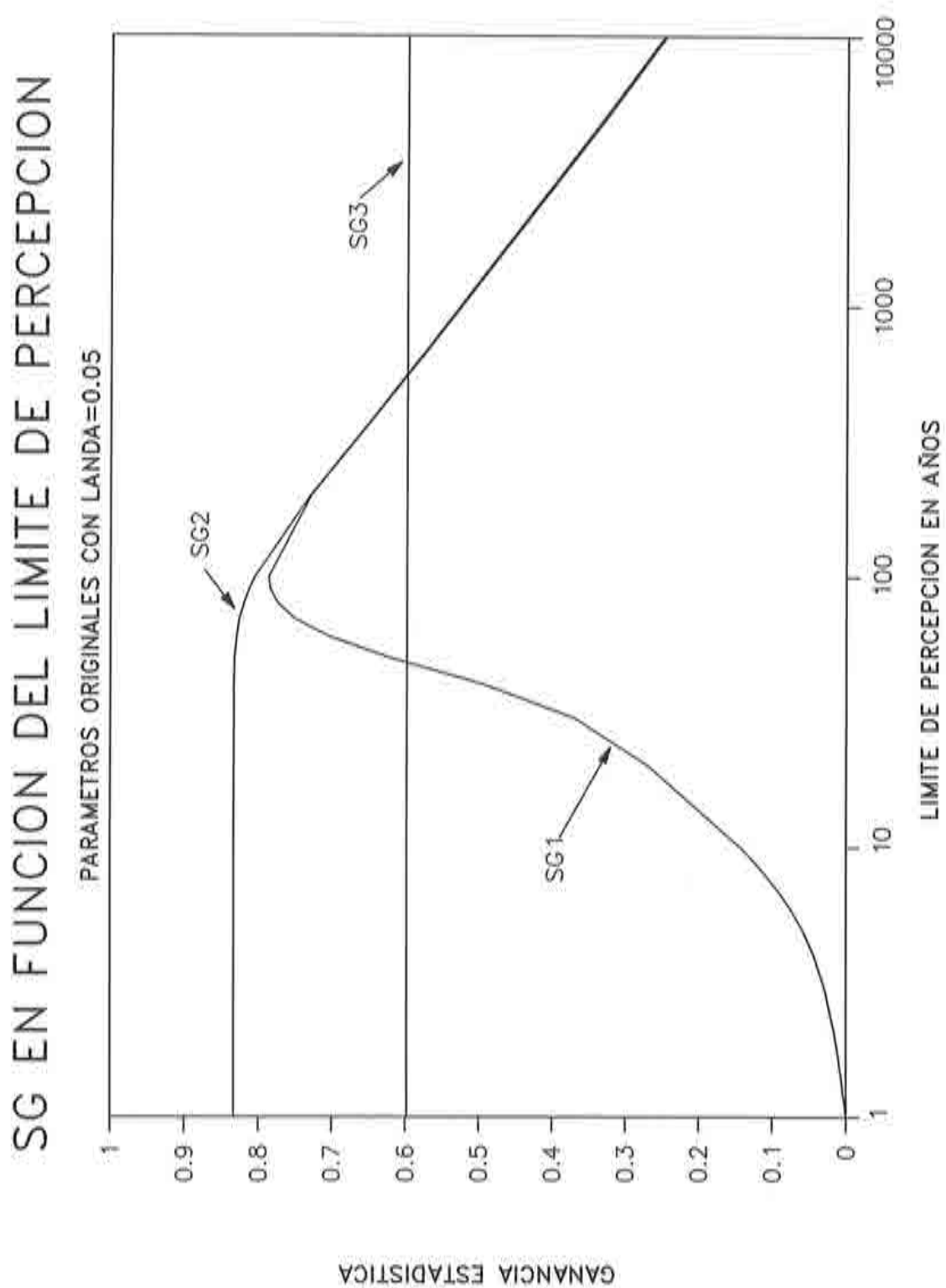


Figura 5.24. Ganancias estadísticas asintóticas en función del límite de percepción histórico, con el parámetro $\lambda=0.05$.

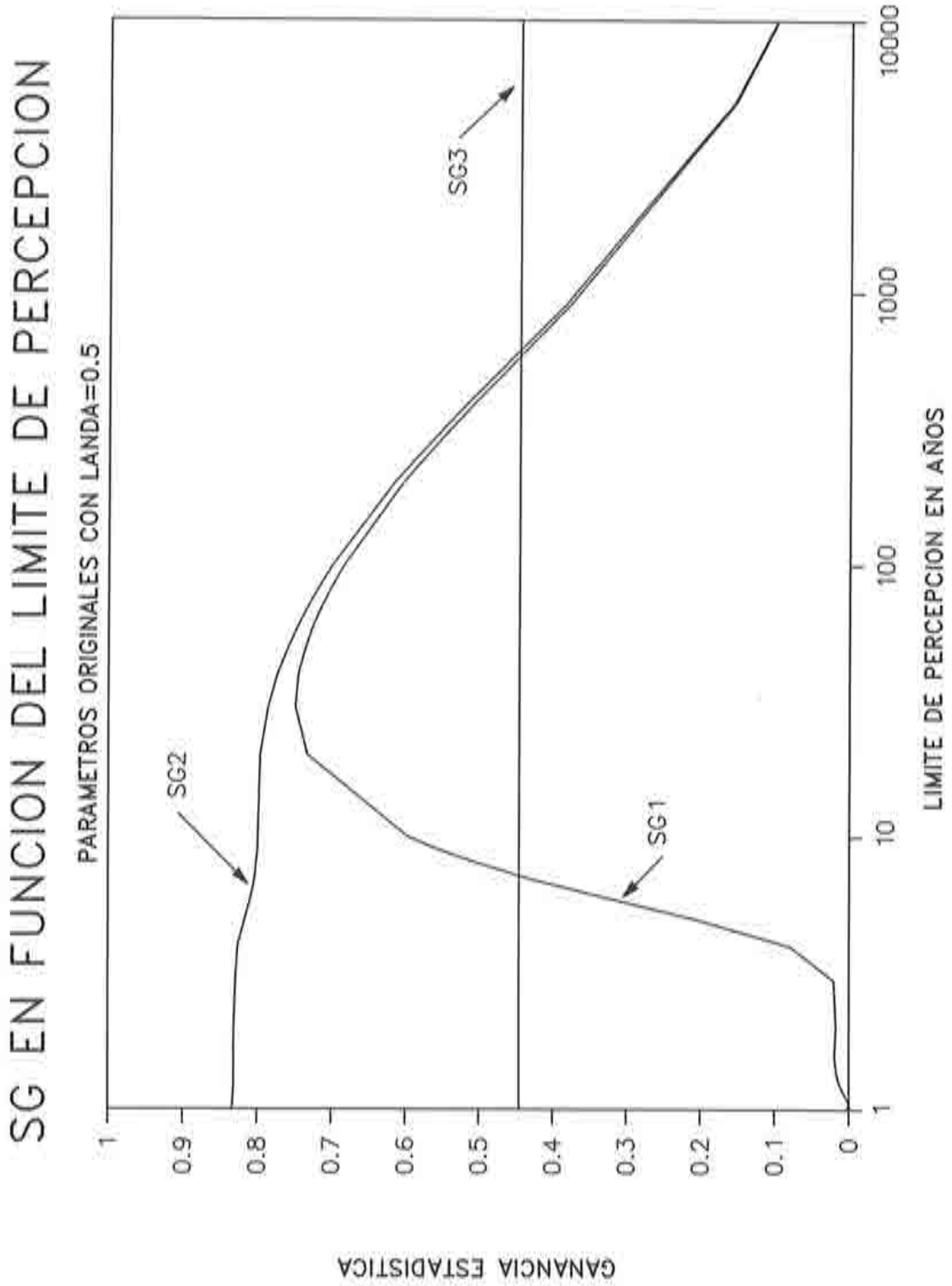


Figura 5.25. Ganancias estadísticas asintóticas en función del límite de percepción histórico, con el parámetro $\lambda=0.5$.

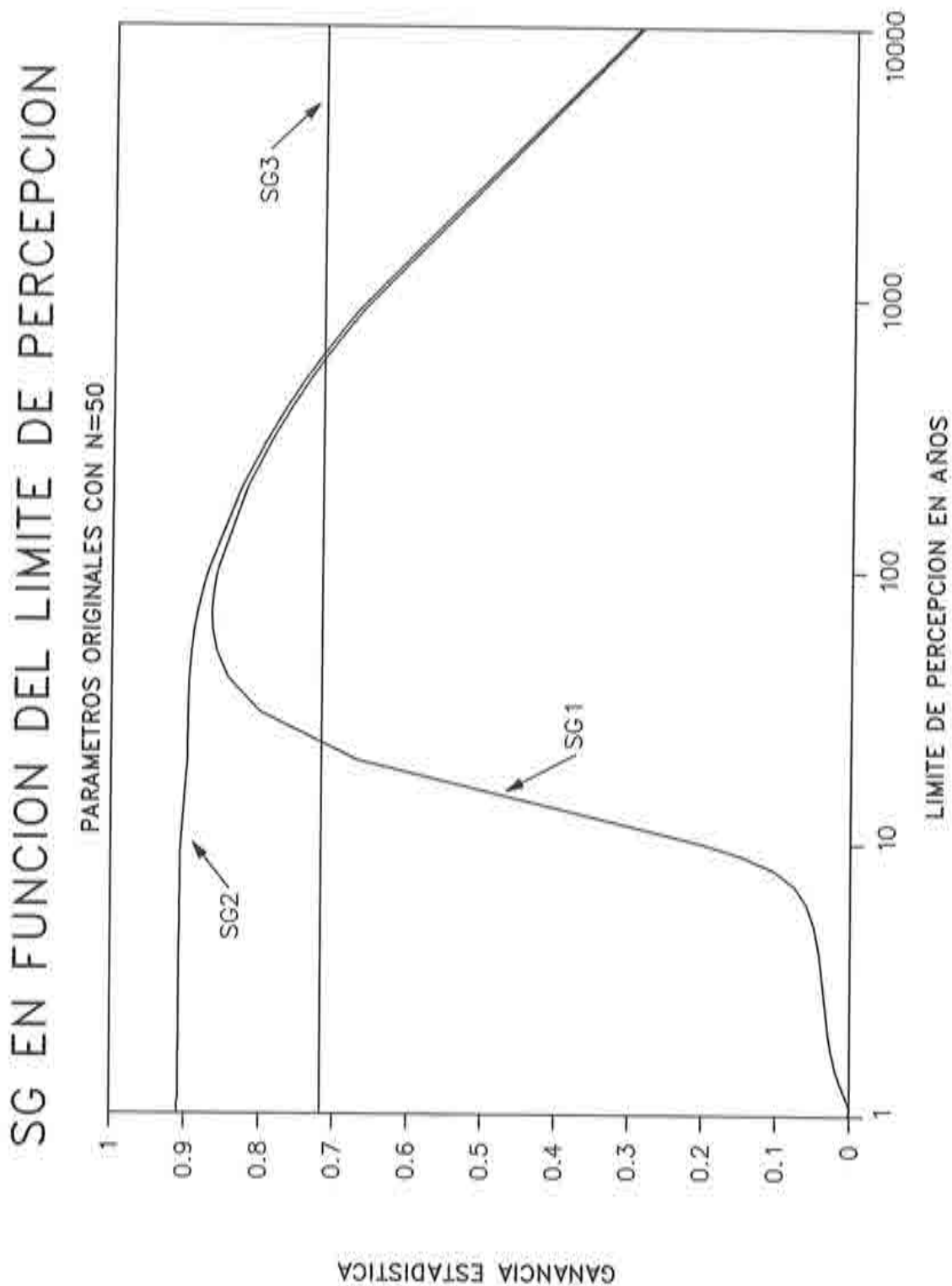


Figura 5.26. Ganancias estadísticas asintóticas en función del límite de percepción histórico, para una longitud de la serie sistemática de 50 años.

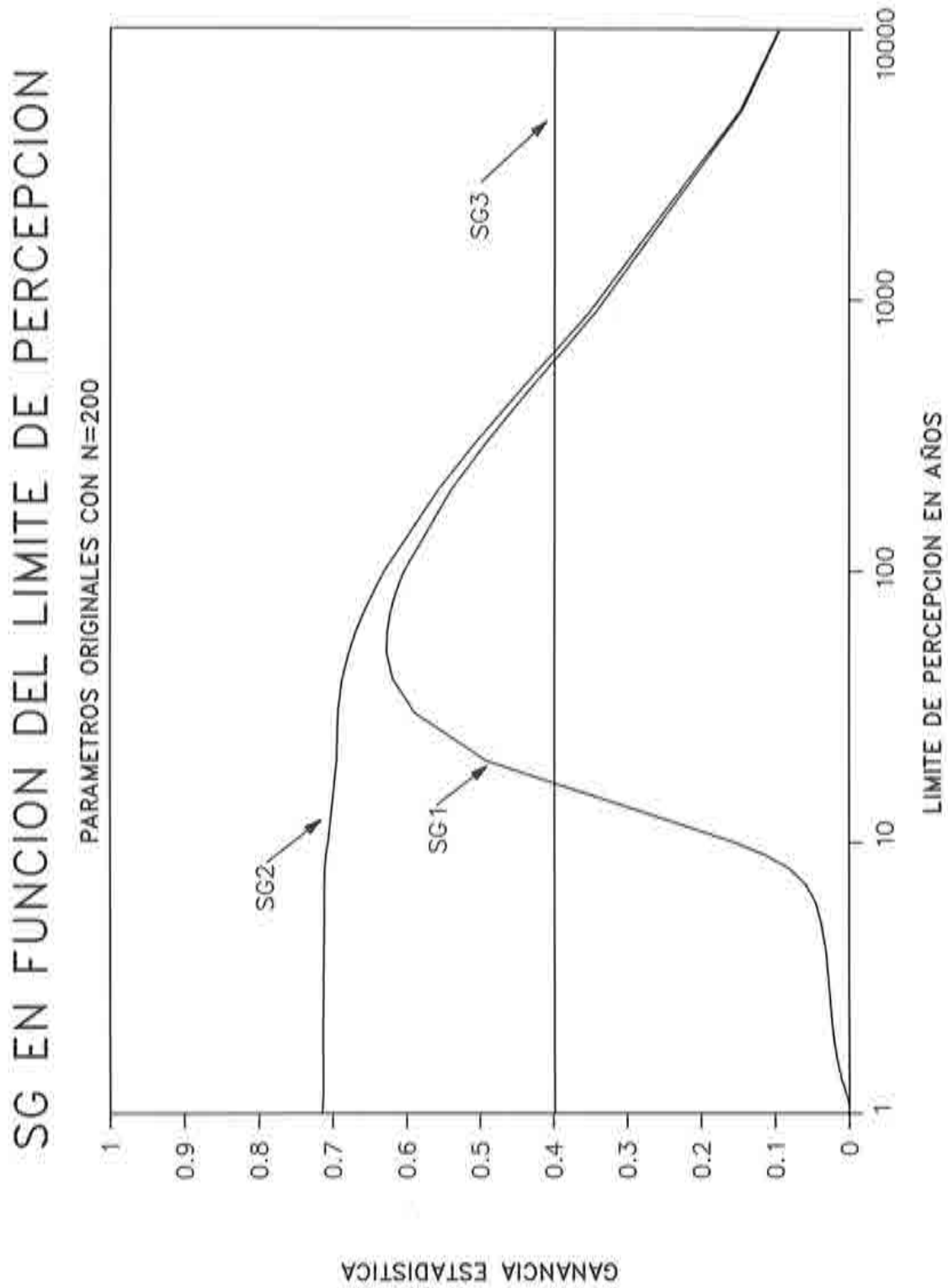


Figura 5.27. Ganancias estadísticas asintóticas en función del límite de percepción histórico, para una longitud de la serie sistemática de 200 años.

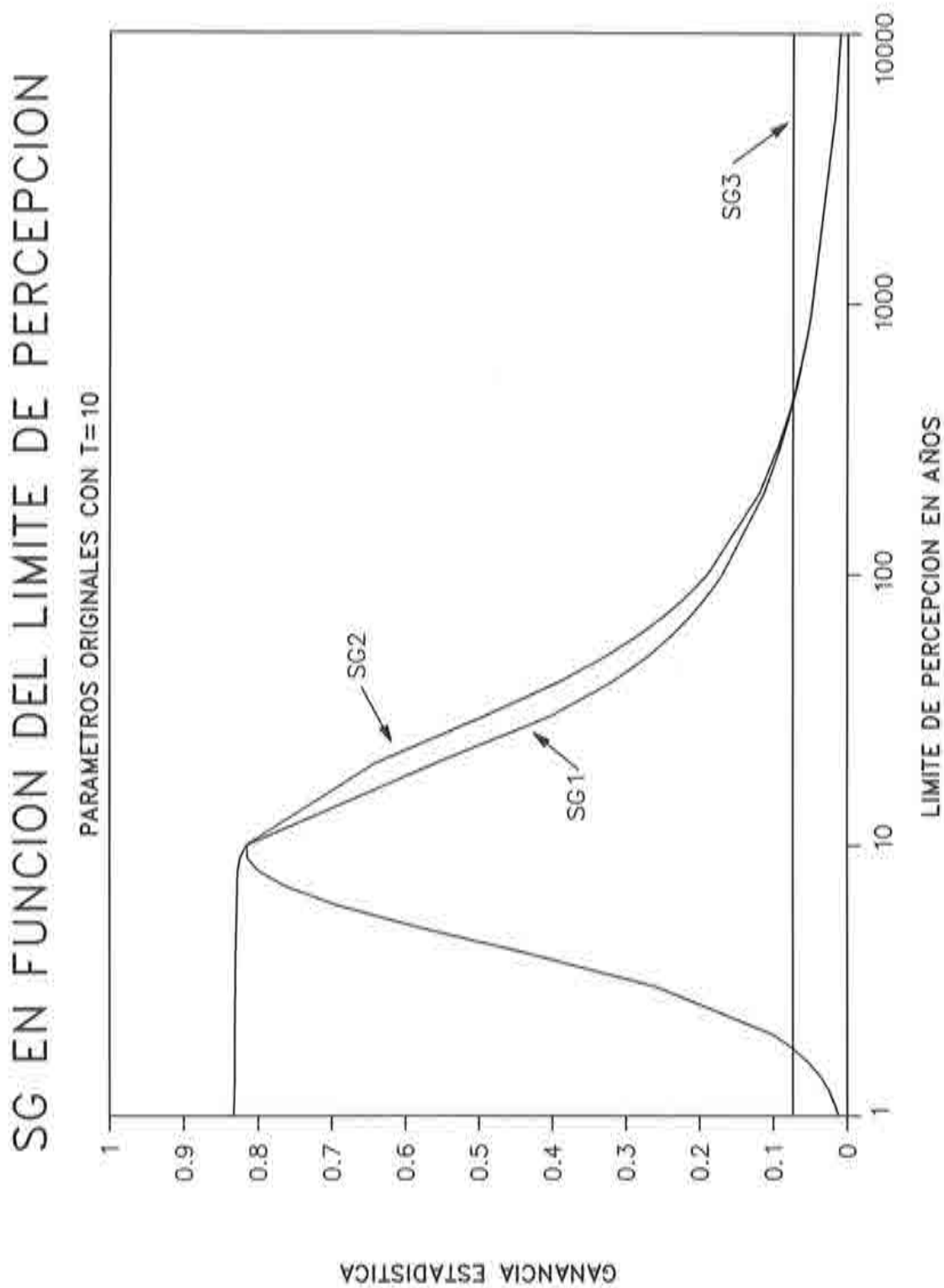


Figura 5.28. Ganancias estadísticas asintóticas en función del límite de percepción histórico, para el cuantil de 10 años.

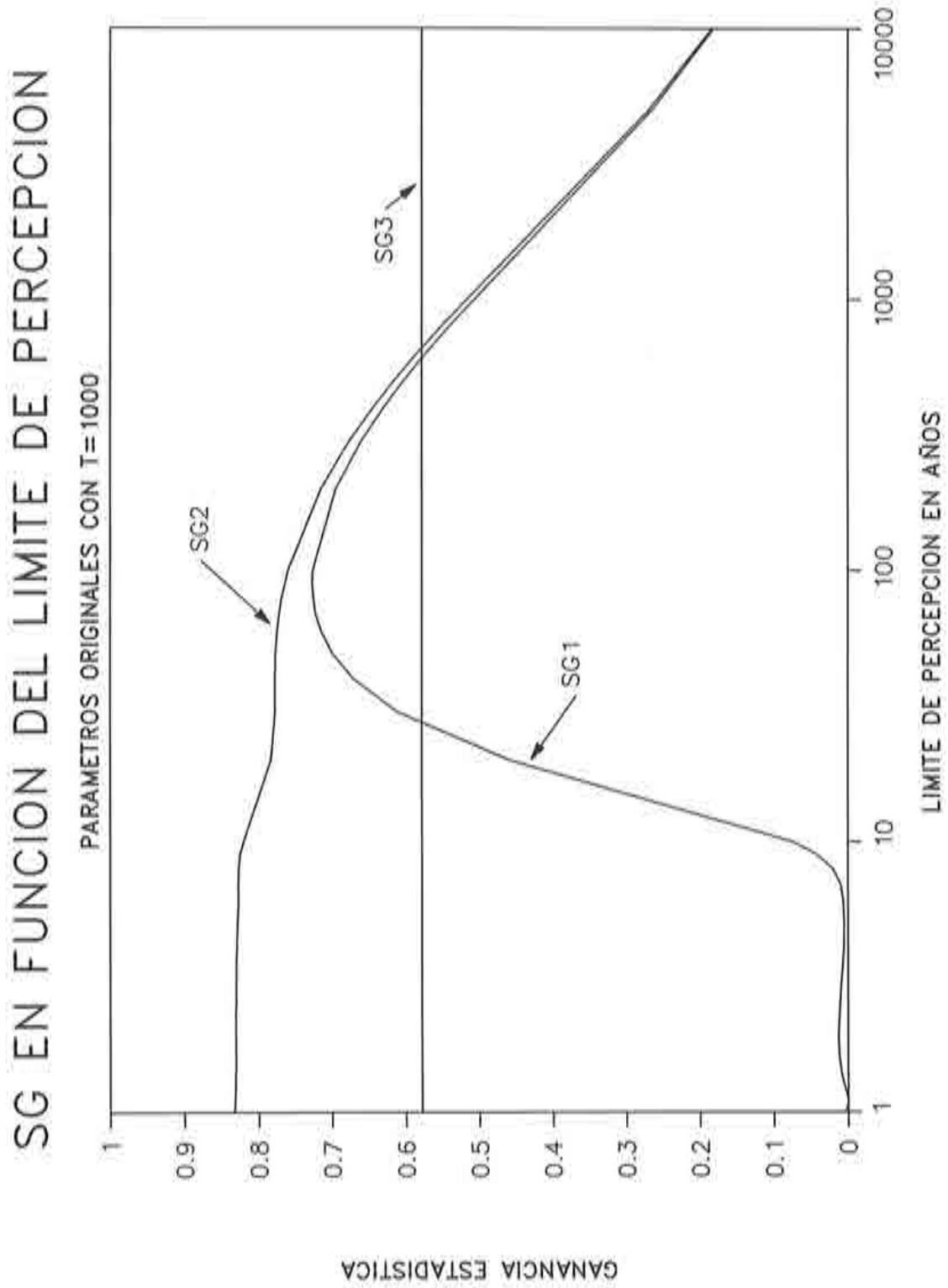


Figura 5.29. Ganancias estadísticas asintóticas en función del límite de percepción histórico, para el cuantil de 1000 años.

5.6.5.- Variación de la ganancia estadística asintótica en función de la relación de longitudes entre el período histórico y sistemático

Como en los casos anteriores, se parte de un escenario original y se han obtenido las ganancias estadísticas asintóticas modificando alguno de los factores. Esto se ha realizado con el programa PLOTM y los resultados se reflejan en las Figuras (5.30) a (5.38). Las propiedades que se deducen de estos resultados son:

- i) Las tres ganancias estadísticas aumentan con el ratio r . La excepción es SG_3 que presenta un máximo para r muy alto, aunque la presentación del máximo se adelanta con θ alto o λ bajo.
- ii) Parece que tiendan para $r \rightarrow \infty$ a valores asintóticos que para SG_2 es 1; es decir, si la longitud histórica es infinita la varianza de los estimadores de los cuantiles utilizando la información censurada es nula.
- iii) Para r menor que 1 (en las figuras para M menor de 100 años), las ganancias estadísticas asintóticas en los tres casos son similares y muy pobres.
- iv) Para r muy grandes (en las figuras para M mayor que 1000 años), el orden de mayor a menor ganancia estadística es con información censurada SG_2 , con información binomial censurada SG_1 y con información de máxima avenida SG_3 .
- v) Para ratios intermedios puede ocurrir que el orden se trastoque en las mismas situaciones descritas en los apartados anteriores (Figuras 5.33, 5.35 y 5.36).
- vi) Para límites de percepción muy bajos, como en la Figura (5.35), la información binomial censurada es muy pobre, incluso si el período histórico se hace infinito.
- vii) Si el límite de percepción es muy alto (Figura 5.36) las informaciones binomial censurada y censurada producen ganancias estadísticas muy similares.
- viii) Los estimadores de los cuantiles bajos (Figura 5.37) apenas mejoran al utilizar información binomial censurada o de máxima avenida.

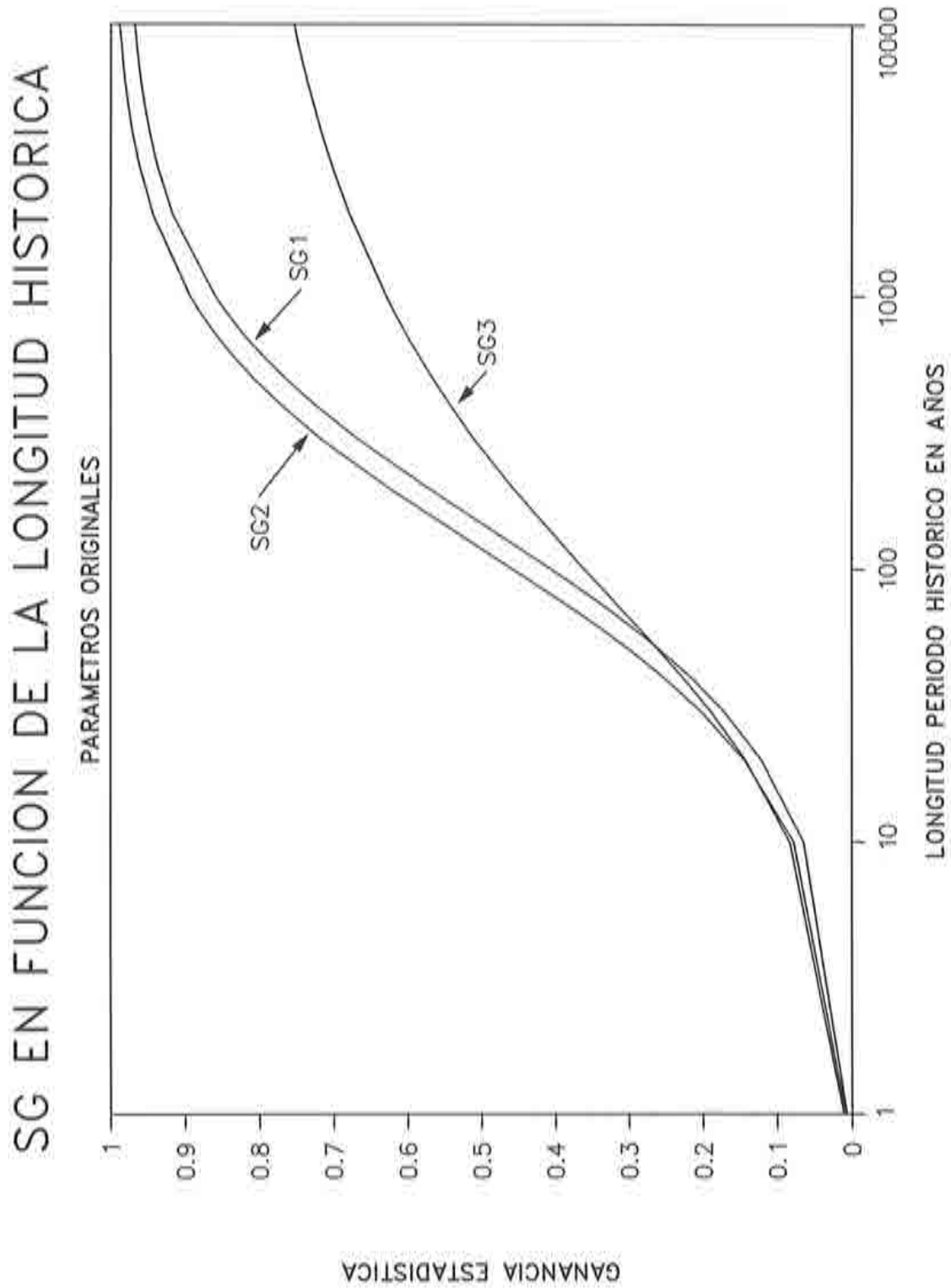


Figura 5.30. Ganancias estadísticas asintóticas en función de la longitud del período histórico, con los factores del escenario original.

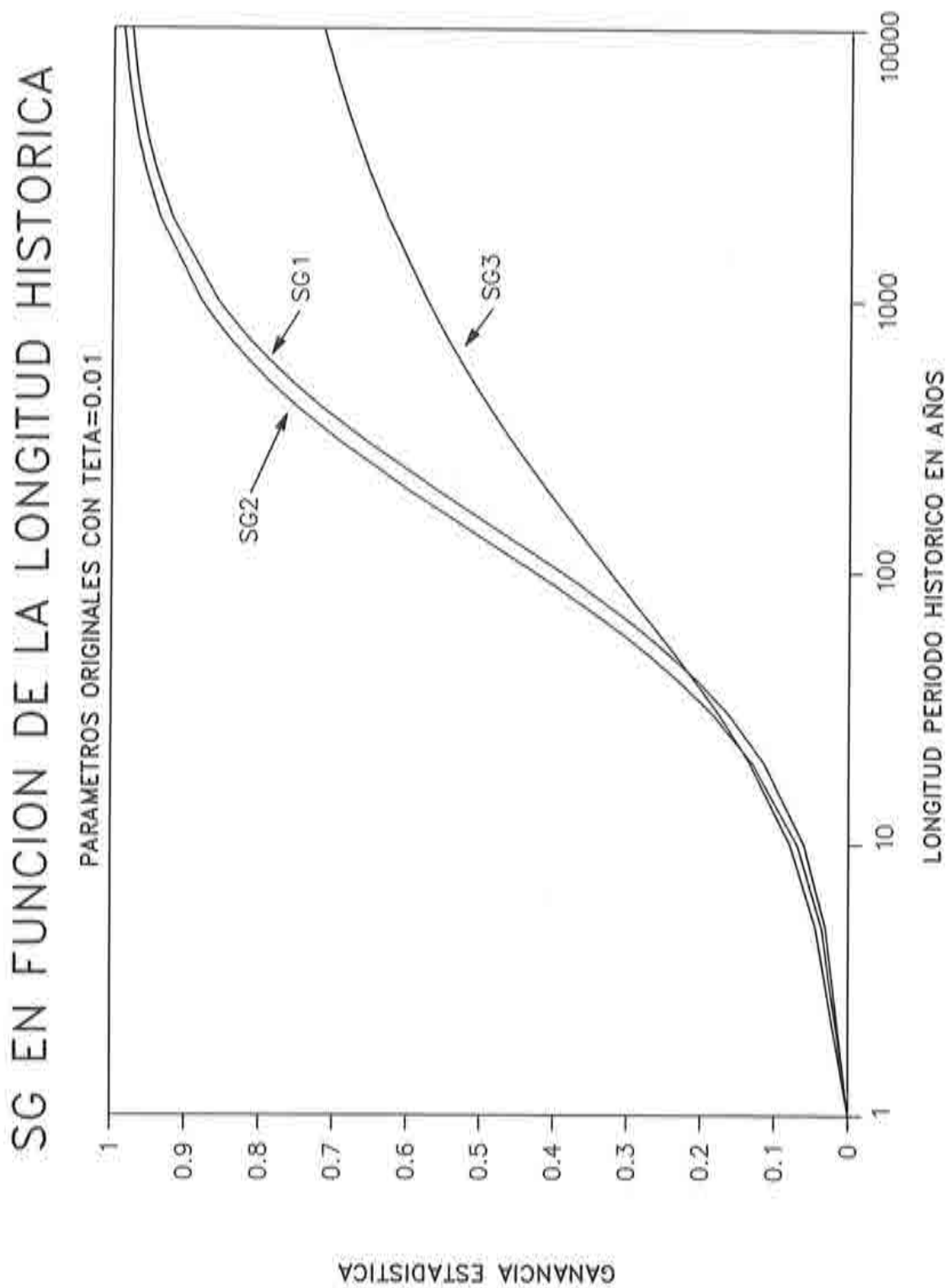


Figura 5.31. Ganancias estadísticas asintóticas en función de la longitud del período histórico, con el parámetro $\theta = 0.01$.

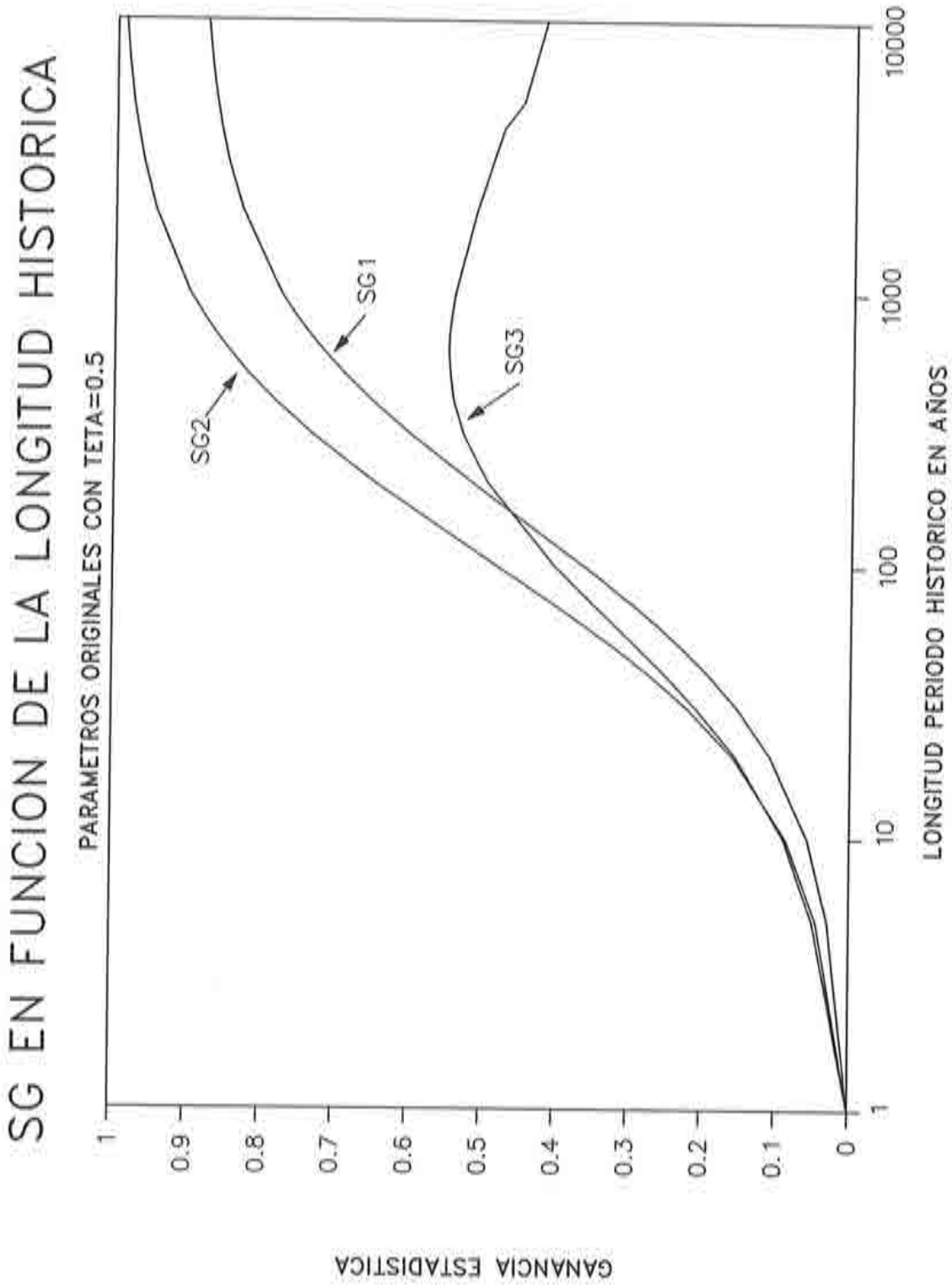


Figura 5.32. Ganancias estadísticas asintóticas en función de la longitud del período histórico, con el parámetro $\theta=0.5$.

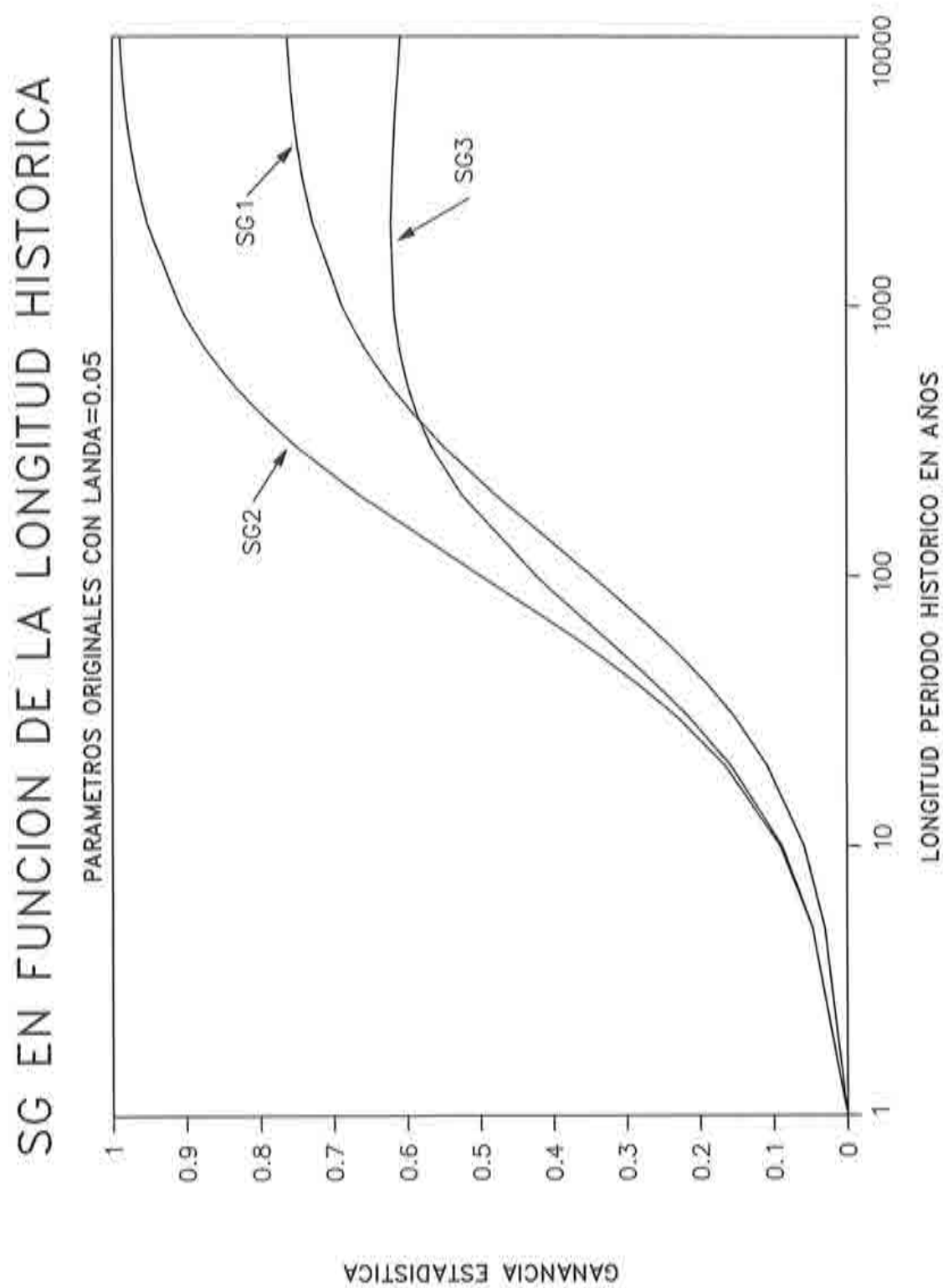


Figura 5.33. Ganancias estadísticas asintóticas en función de la longitud del período histórico, con el parámetro $\lambda=0.05$.

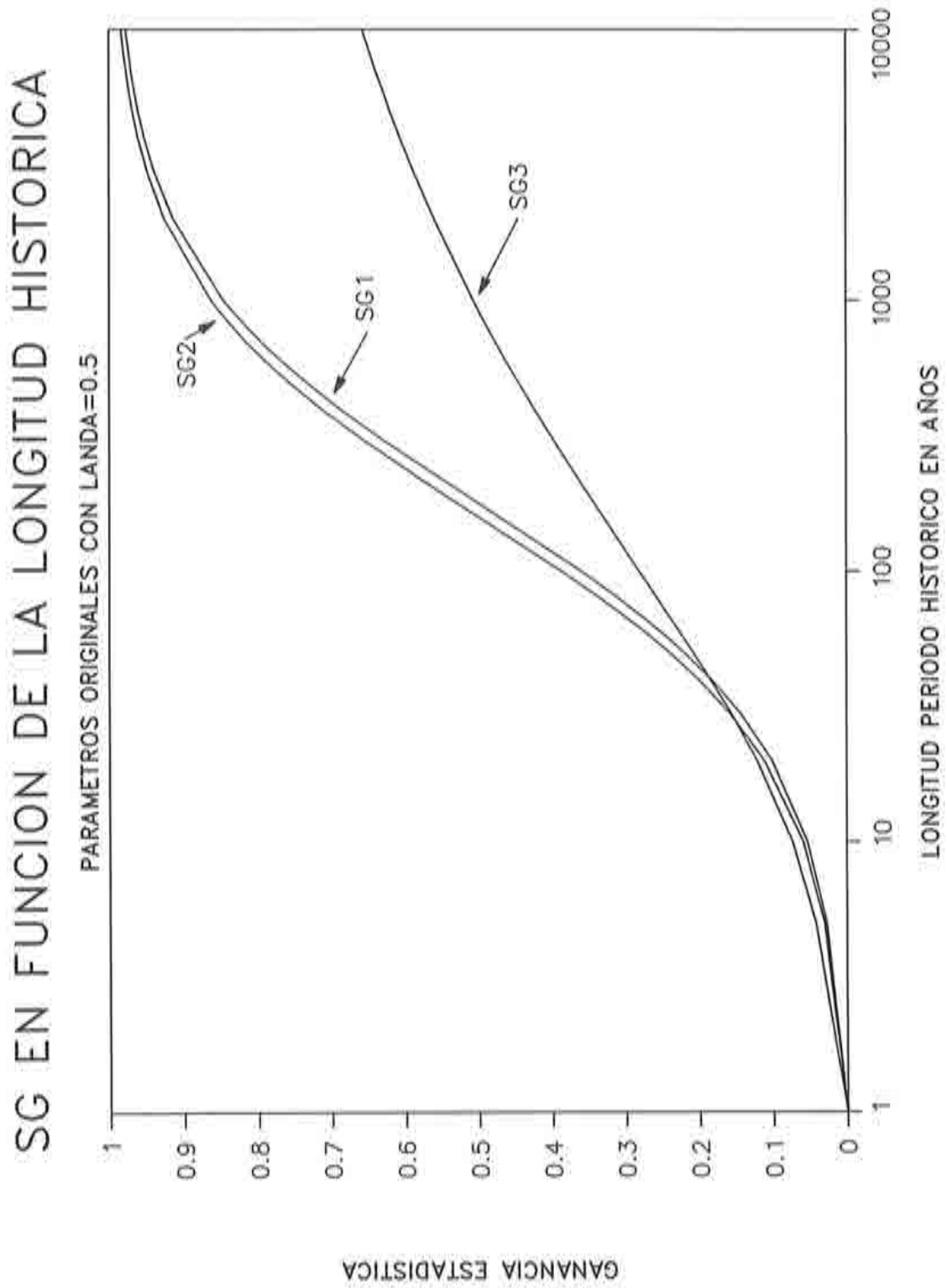


Figura 5.34. Ganancias estadísticas asintóticas en función de la longitud del período histórico, con el parámetro $\lambda=0.5$.

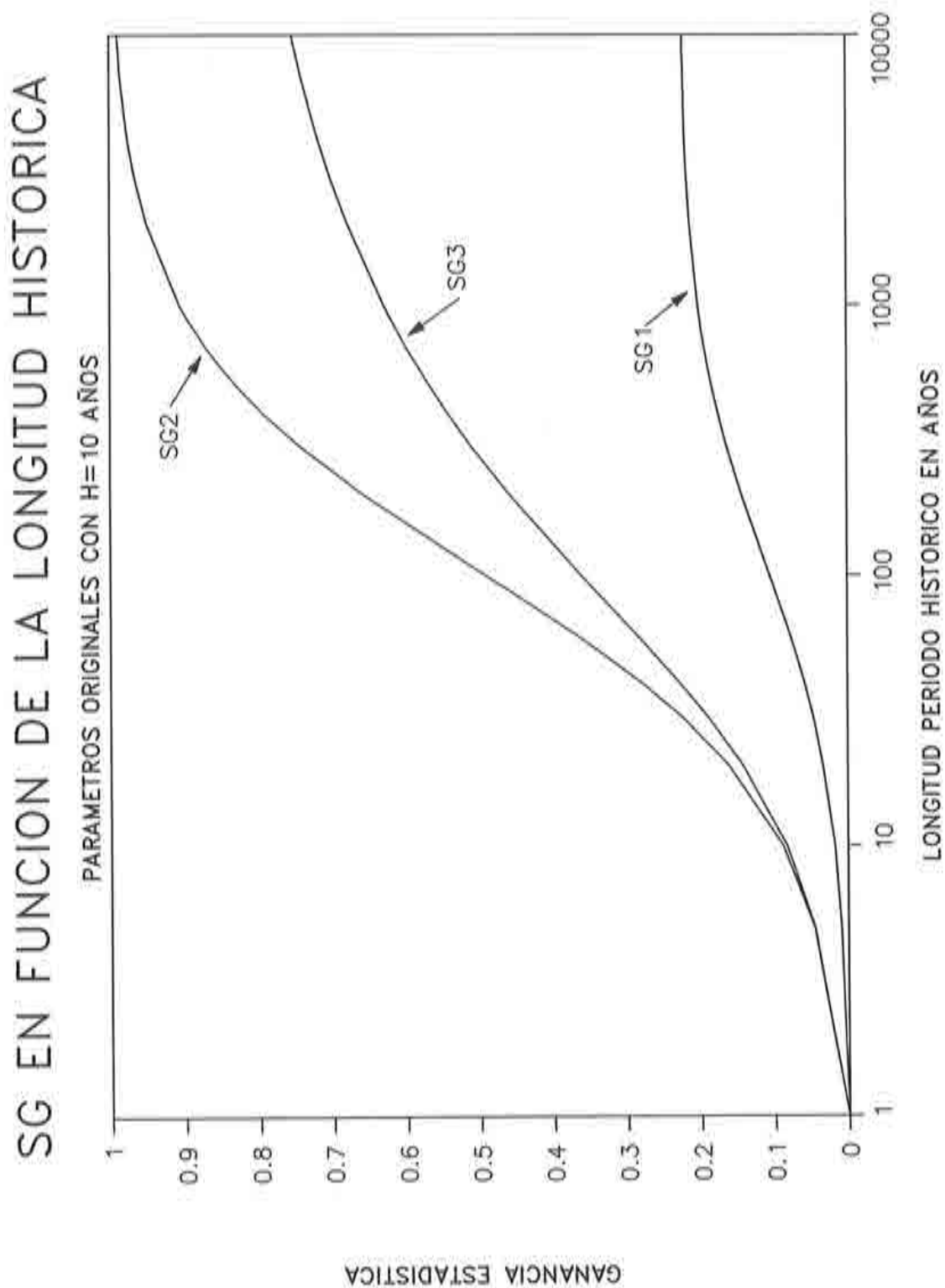


Figura 5.35. Ganancias estadísticas asintóticas en función de la longitud del período histórico, para un límite de percepción histórico de 10 años.

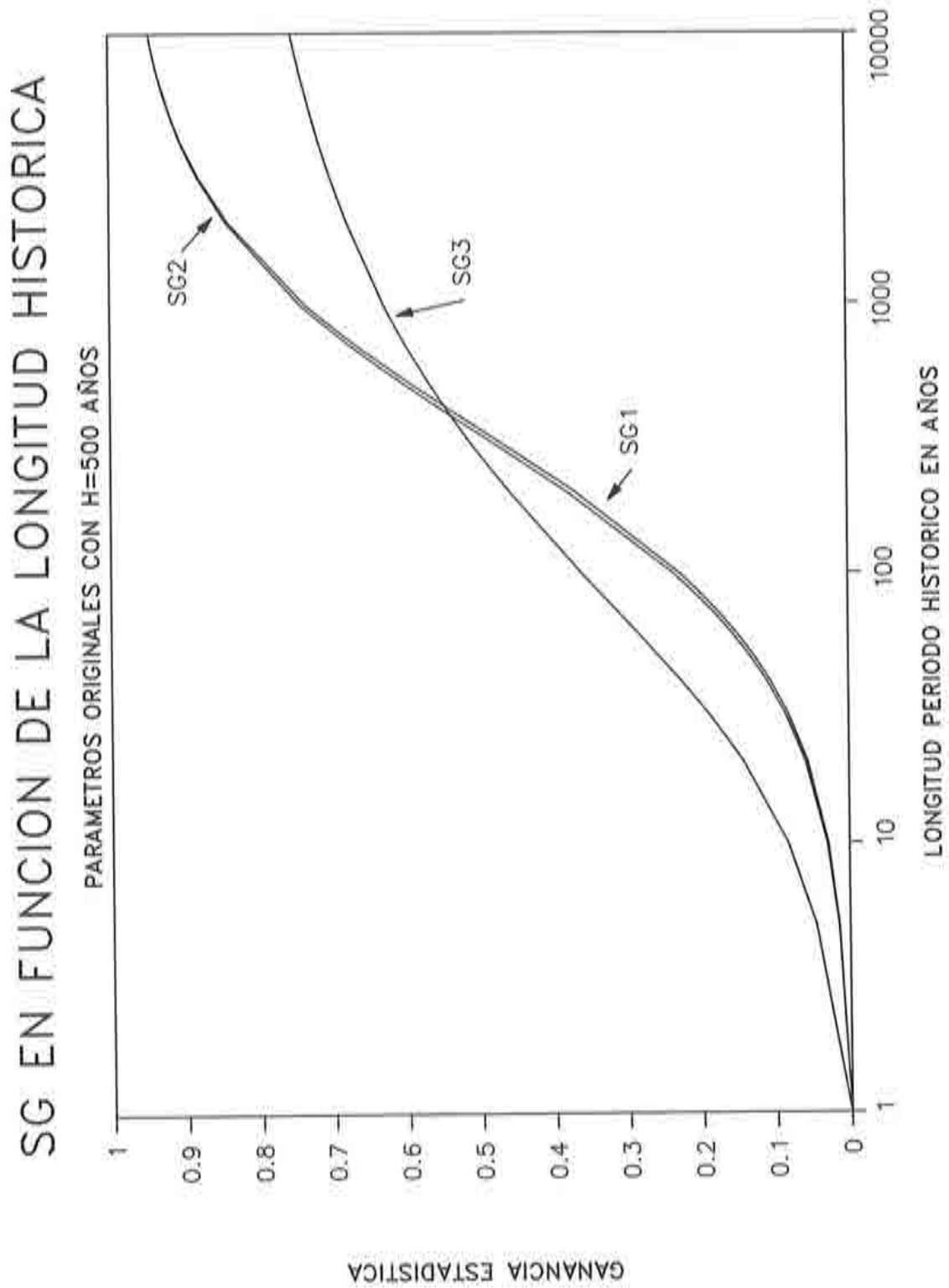


Figura 5.36. Ganancias estadísticas asintóticas en función de la longitud del período histórico, para un límite de percepción histórico de 500 años.

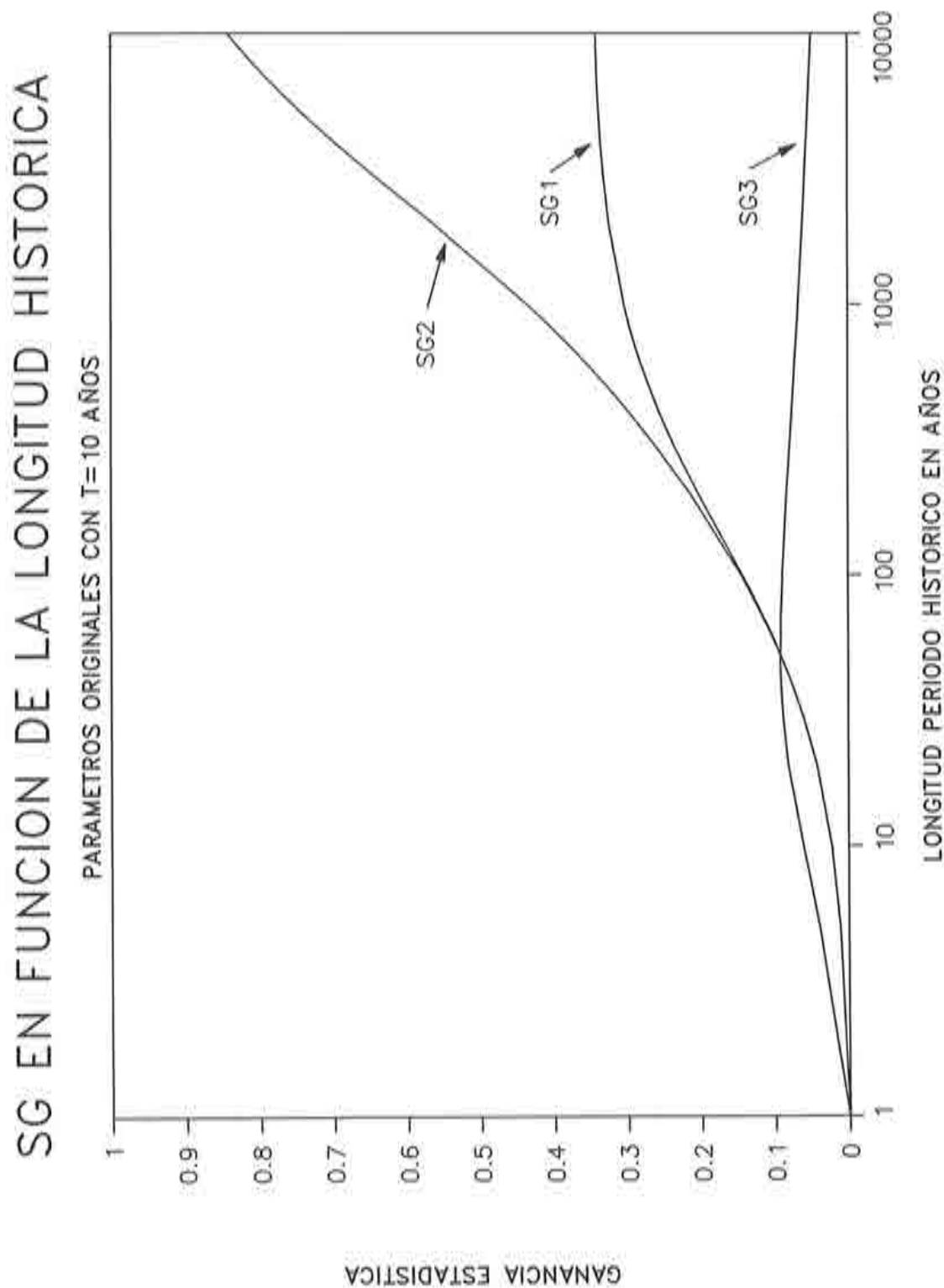


Figura 5.37. Ganancias estadísticas asintóticas en función de la longitud del período histórico, para el cuantil de 10 años de período de retorno.

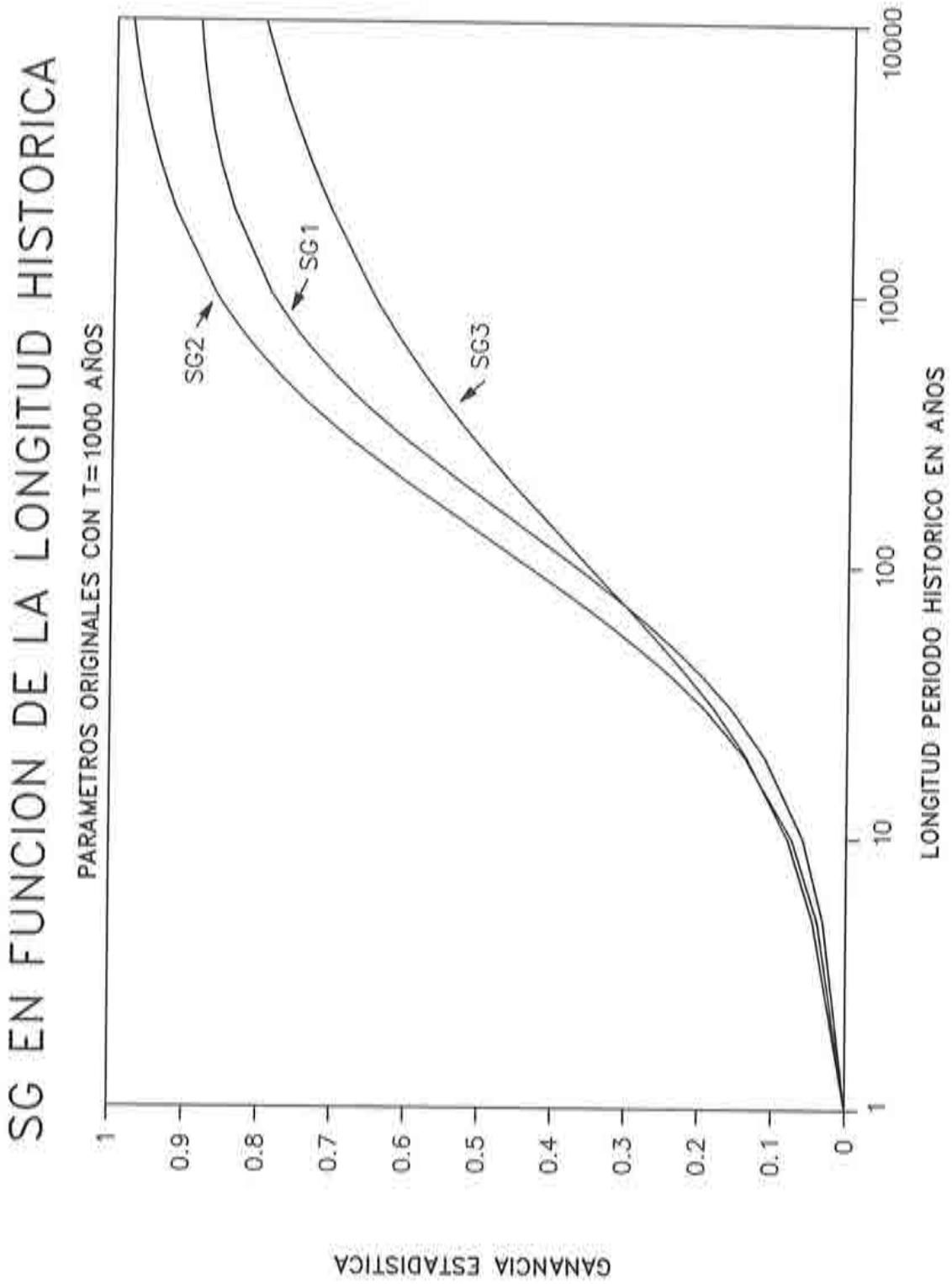


Figura 5.38. Ganancias estadísticas asintóticas en función de la longitud del período histórico, para el cuantil de 1000 años de período de retorno.

5.7.- VARIANZAS Y GANANCIAS ESTADÍSTICAS MUESTRALES DE LOS CUANTILES

Las varianzas asintóticas (CRLB) obtenidas en el Apartado 5.6.1 son un límite inferior de la varianza muestral (MSE) de cualquier estimador insesgado de los cuantiles: representa lo mejor que se puede hacer en la estimación de cuantiles en términos de la varianza.

El Método de Máxima Verosimilitud es un método de estimación asintóticamente insesgado; es decir, el sesgo obtenido en la estimación con una muestra infinita es nulo. Por ello, se puede considerar la varianza asintótica como un límite inferior del Método de Máxima Verosimilitud.

La varianza muestral es muy compleja de obtener analíticamente por lo que en la mayoría de los casos hay que recurrir a simulaciones de Montecarlo para obtenerla. El tiempo de computación y la variabilidad muestral de las simulaciones hacen que se pretenda sustituir éste por su valor asintótico (obtenido y analizado en el apartado anterior). Por ello surge la pregunta: ¿en qué medida las varianzas, o lo que es más interesante las ganancias estadísticas, asintóticas y muestrales son similares?.

Para resolver estas cuestiones se ha desarrollado el programa SIMTC4P, que obtiene el error cuadrático medio y ganancia estadística muestral para diversos cuantiles, generando series sintéticas con la estructura descrita en el Apartado 3.3.4. Además calcula, para comparación, la varianza y ganancia estadística asintótica. Para cada serie generada, obtiene los estimadores de los cuantiles mediante el Método de Máxima Verosimilitud con el algoritmo de Powell, descrito en el Apartado 5.4.2 de este mismo Capítulo.

Para disminuir el tiempo de computación se realizaron las mínimas simulaciones partiendo de un escenario inicial, de tal forma que en cada simulación se varía alguno de los factores de dicho escenario.

El escenario inicial es:

- Parámetros

$$\lambda_1 = 1$$

$$\theta_1 = 1$$

$$\lambda_2 = 0.2$$

$$\theta_2 = 0.1$$

- Longitud período sistemático de 100 años.

- Longitud período histórico de 500 años.
- Límite de percepción histórico de 50 años de período de retorno.

Las conclusiones obtenidas son las siguientes:

i) Como se observa en las Figuras (5.39), (5.40), (5.41) y (5.42) las varianzas asintóticas son una buena aproximación de los errores cuadráticos medios muestrales para casos en los que se utiliza la información sistemática solo, con información histórica censurada y con información histórica de máxima avenida. La diferencia es en todas las situaciones inferior al 10%, en términos de la varianza. Sin embargo, cuando se utiliza la información histórica binomial censurada, la diferencia asciende al 30%. En dichas figuras, realmente está representado el coeficiente de variación del estimador del cuantil, cuya expresión es:

$$CV_i(\bar{X}_T) = \frac{\sqrt{CRLB_i(\bar{X}_T)}}{\bar{X}_T} \quad (5.1)$$

ii) Las ganancias estadísticas asintóticas (SG) con información censurada y de máxima avenida son una buena aproximación de sus correspondientes valores muestrales (SG^{*}). Esto se puede observar en las Figuras (5.43) y (5.44) donde se comparan valores muestrales y asintóticos. Como era de esperar, para el caso de información histórica binomial censurada las diferencias son importantes, aunque se han reducido al 20%. La reducción es debida a que el error cuadrático medio sin información histórica está por encima de la varianza asintótica correspondiente, con lo que se compensa parcialmente la diferencia con información histórica binomial censurada. Esta diferencia disminuye a medida que aumenta la longitud sistemática, según se puede comprobar en la Tabla (5.1), de tal forma que al duplicarla se divide por dos la diferencia aproximadamente. La razón es que MSE₀ está muy cerca de su valor asintótico CRLB₀ desde tamaños de muestra muy pequeños, mientras MSE₁ se aproxima a su valor asintótico CRLB₁ mucho más lentamente.

Tabla 5.1. Ganancia estadística muestral con información binomial censurada (SG_1^*) y censurada (SG_2^*) del cuantil de 100 años en el escenario inicial para diferentes longitudes del registro sistemático.

	N = 50	N = 100	N = 200	Valor asintótico
SG_1^*	.6270	.7191	.7370	.7646
SG_2^*	.7929	.8030	.8104	.8079

iii) Manteniendo r constante, SG_2^* es independiente de N como se observa en la Tabla (5.1). Se observa un ligero aumento con N debido a que MSE_2^* se acerca más rápidamente a su valor asintótico que MSE_0^* . No ocurre lo mismo con SG_1^* por la razón comentada anteriormente.

iv) Manteniendo los parámetros estandarizados constantes, las tres ganancias estadísticas muestrales se muestran independientes respecto a λ_1 y θ_1 , como se refleja en la Tabla (5.2).

Tabla 5.2. Ganancias estadísticas muestrales del cuantil de 100 años para diferentes parámetros λ_1 y θ_1 , con $\lambda = 0.2$ y $\theta = 0.1$.

λ_1	θ_1	SG_1^*	SG_2^*	SG_3^*
1	1	.7191	.8030	.4743
3	0.05	.7212	.8050	.4744
3	0.1	.7205	.8044	.4739
2	0.1	.7255	.8055	.4774
Valor asintótico		.7646	.8079	.5658

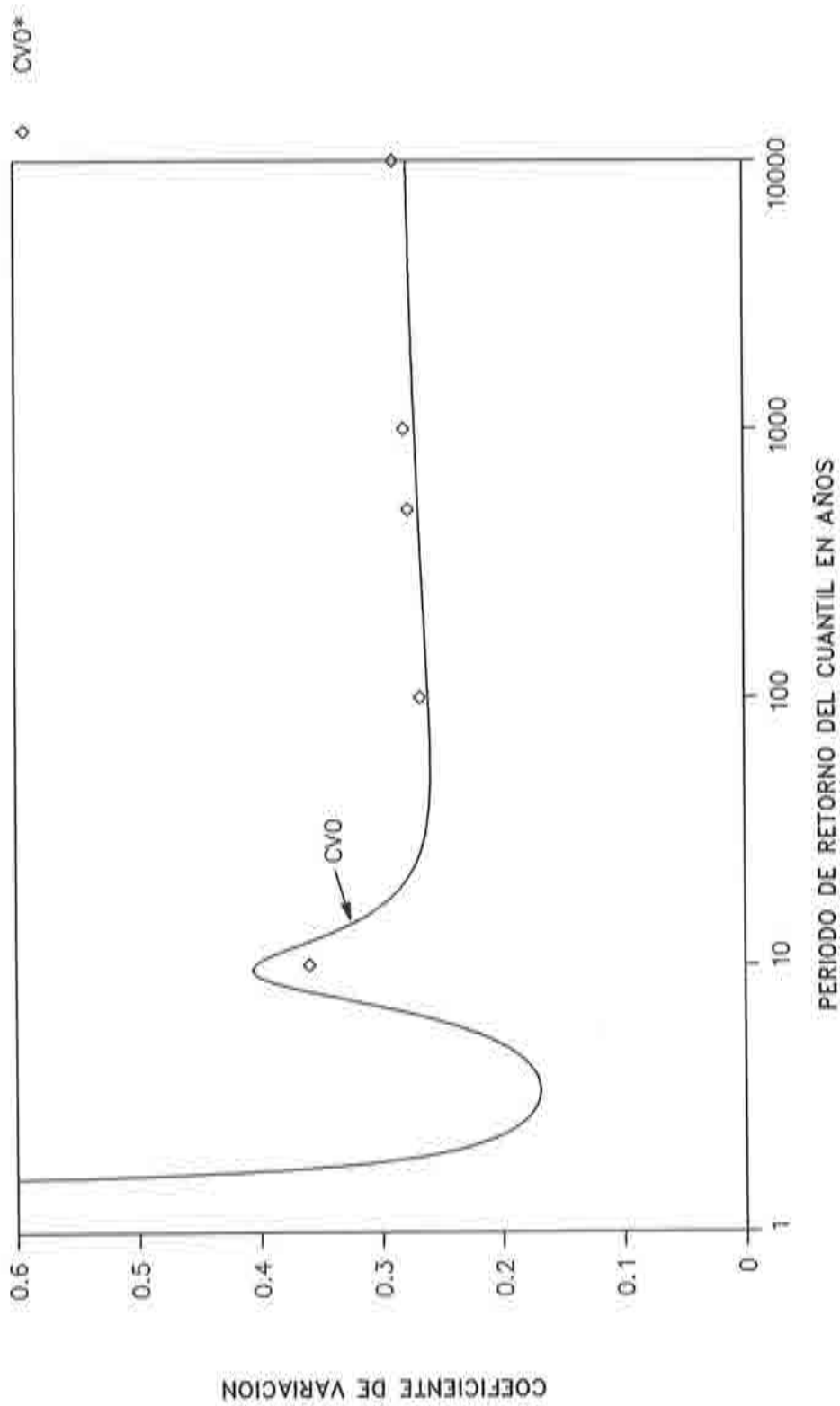


Figura 5.39. Coeficientes de variación asíntotico y muestral utilizando la información sistemática exclusivamente en función del cuantil, para el escenario de factores inicial.

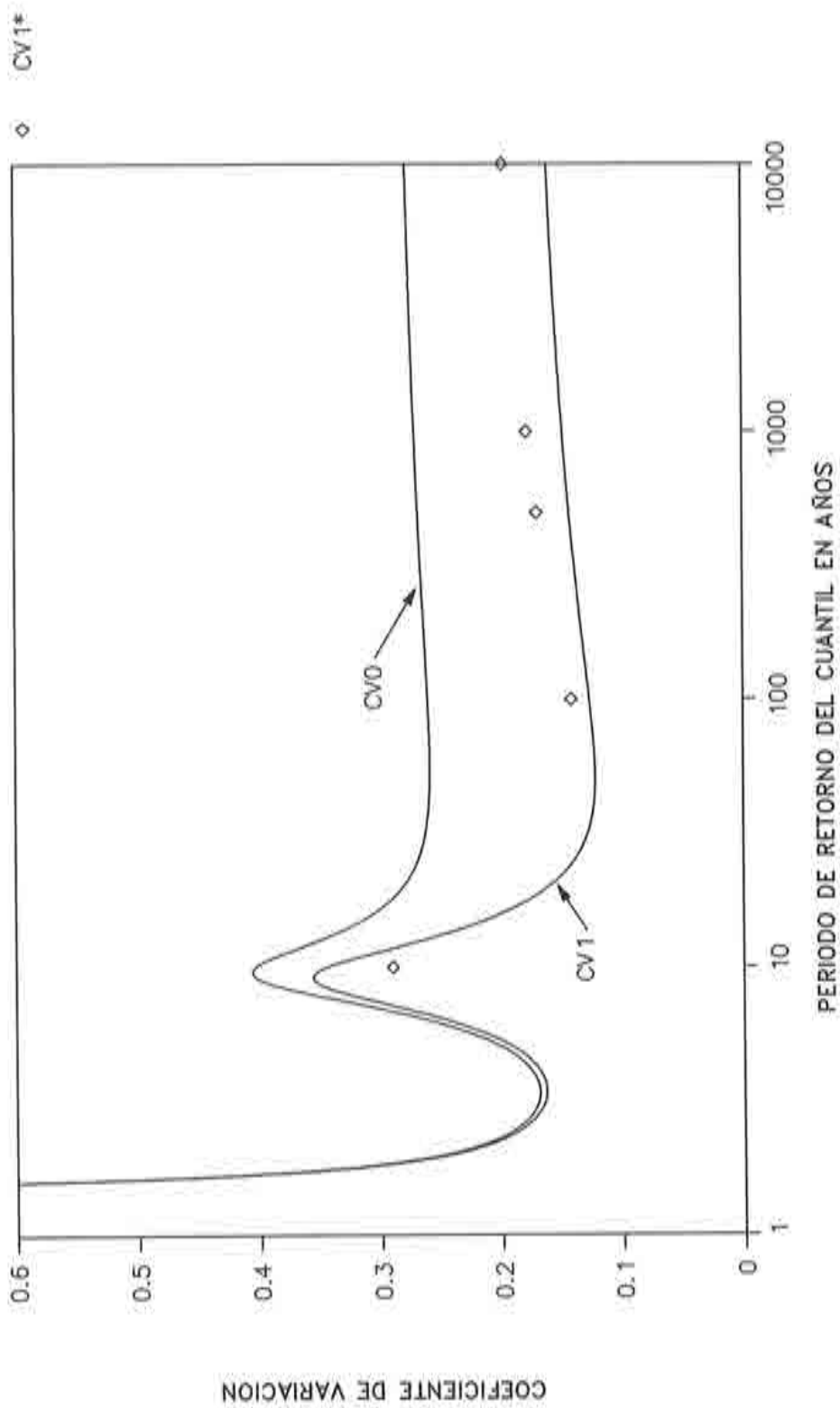


Figura 5.40. Coeficientes de variación asintótico y muestral utilizando información histórica binomial censurada en función del cuantil, para el escenario de factores inicial.

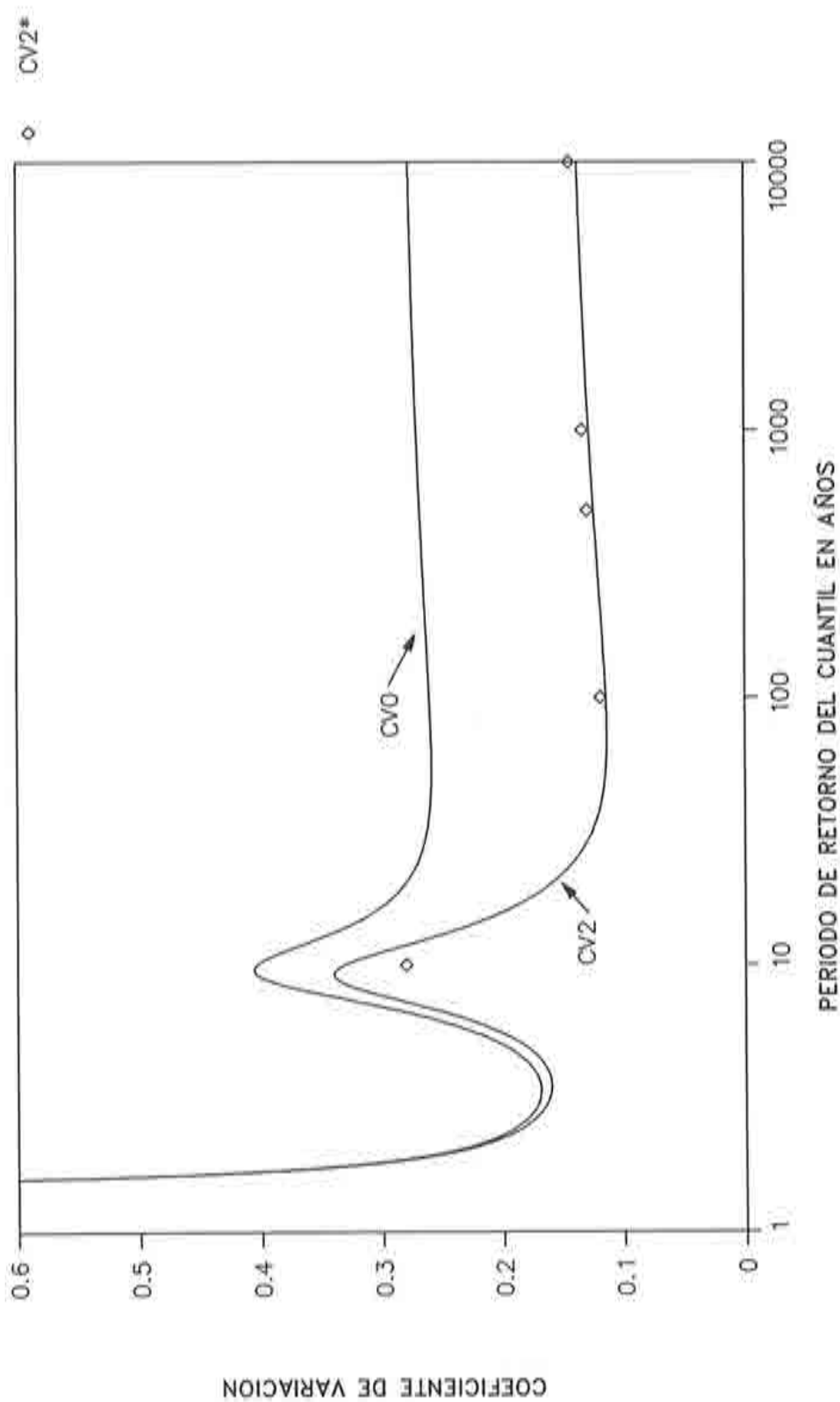


Figura 5.41. Coeficientes de variación asintótico y muestral con información histórica censurada en función del cuantil, para el escenario de factores inicial.

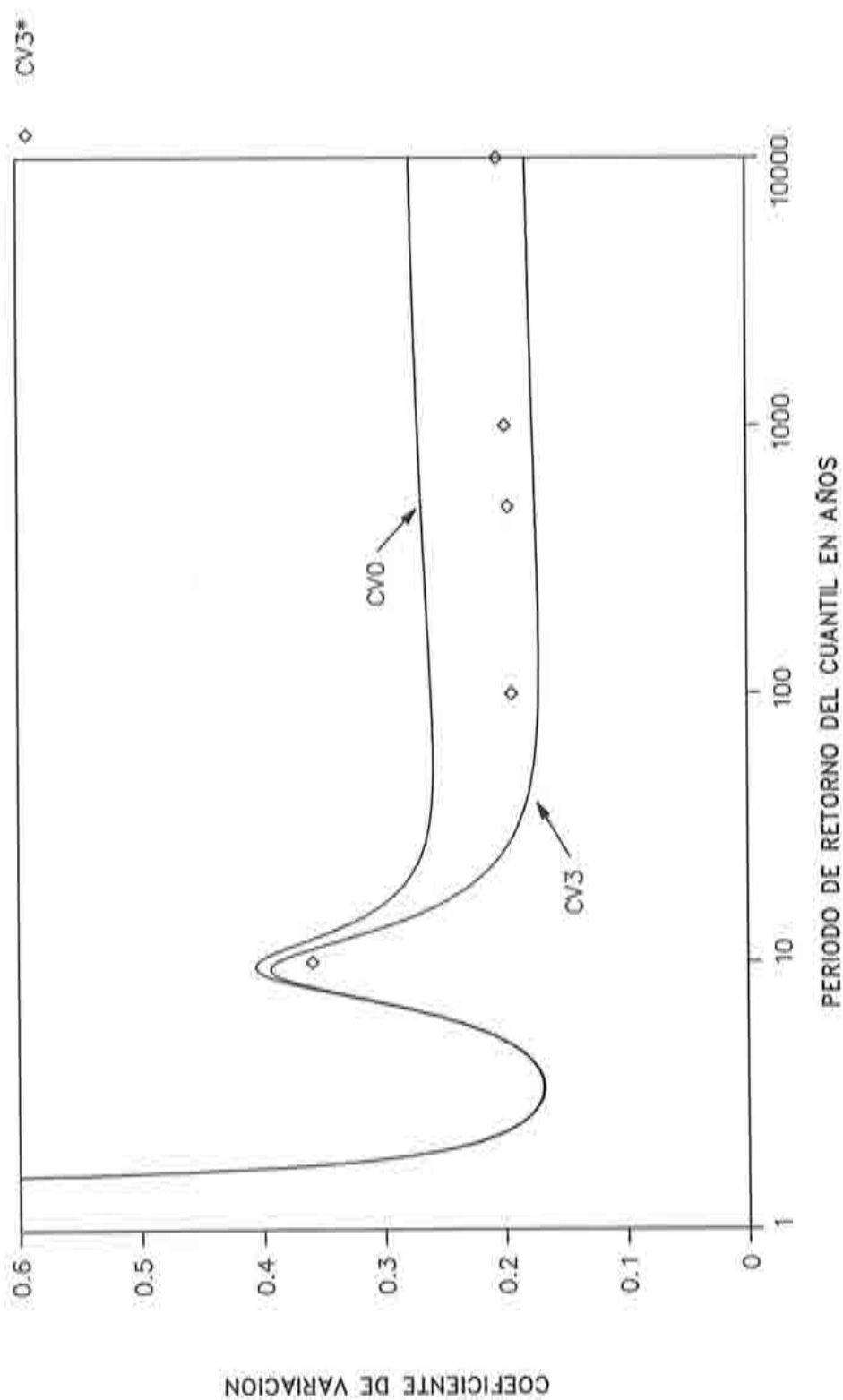


Figura 5.42. Coeficientes de variación asintótico y muestral con información histórica de máxima avenida en función del cuantil, para el escenario de factores inicial.

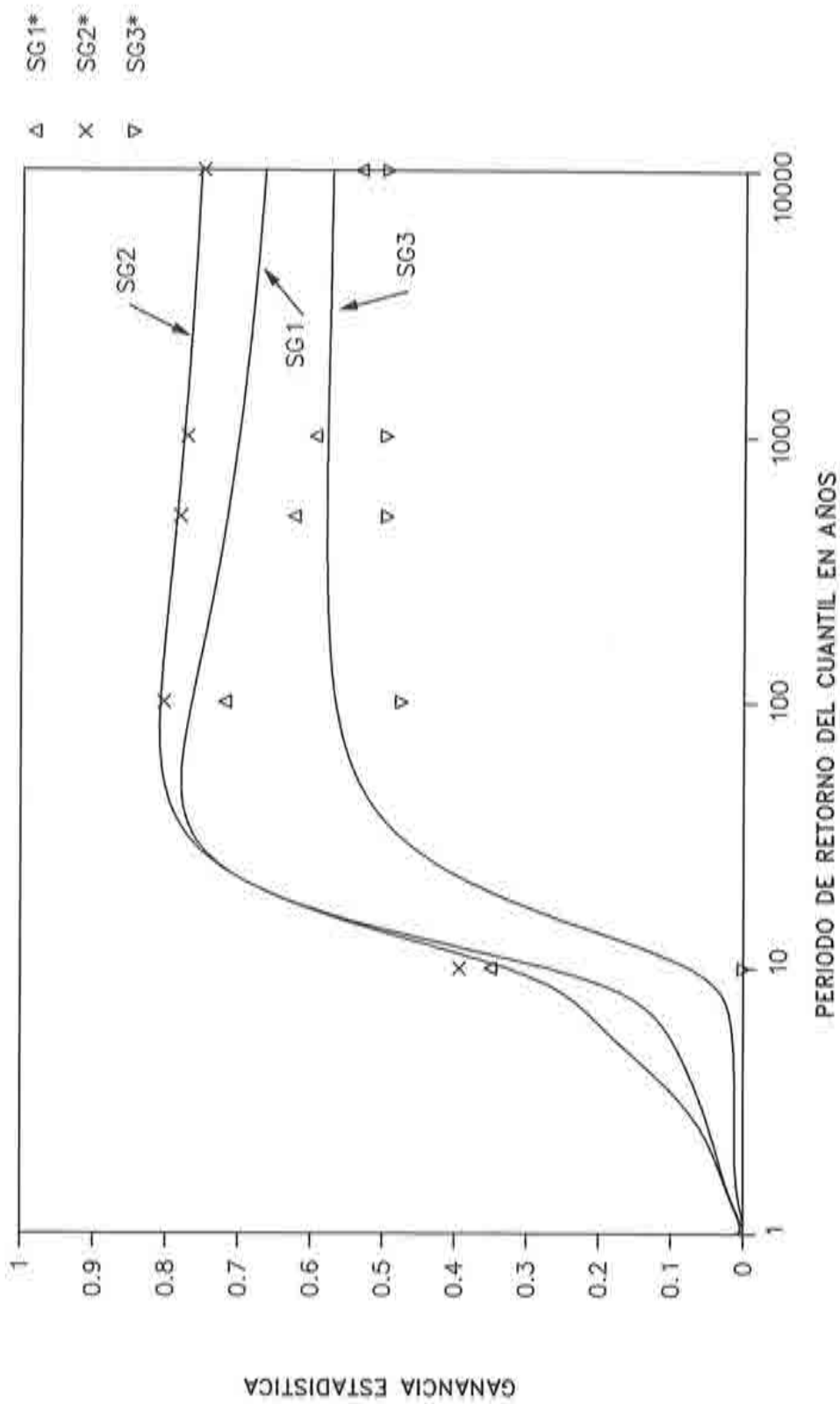


Figura 5.43. Ganancias estadísticas asintótica y muestral en función del cuantil, para el escenario de factores inicial.

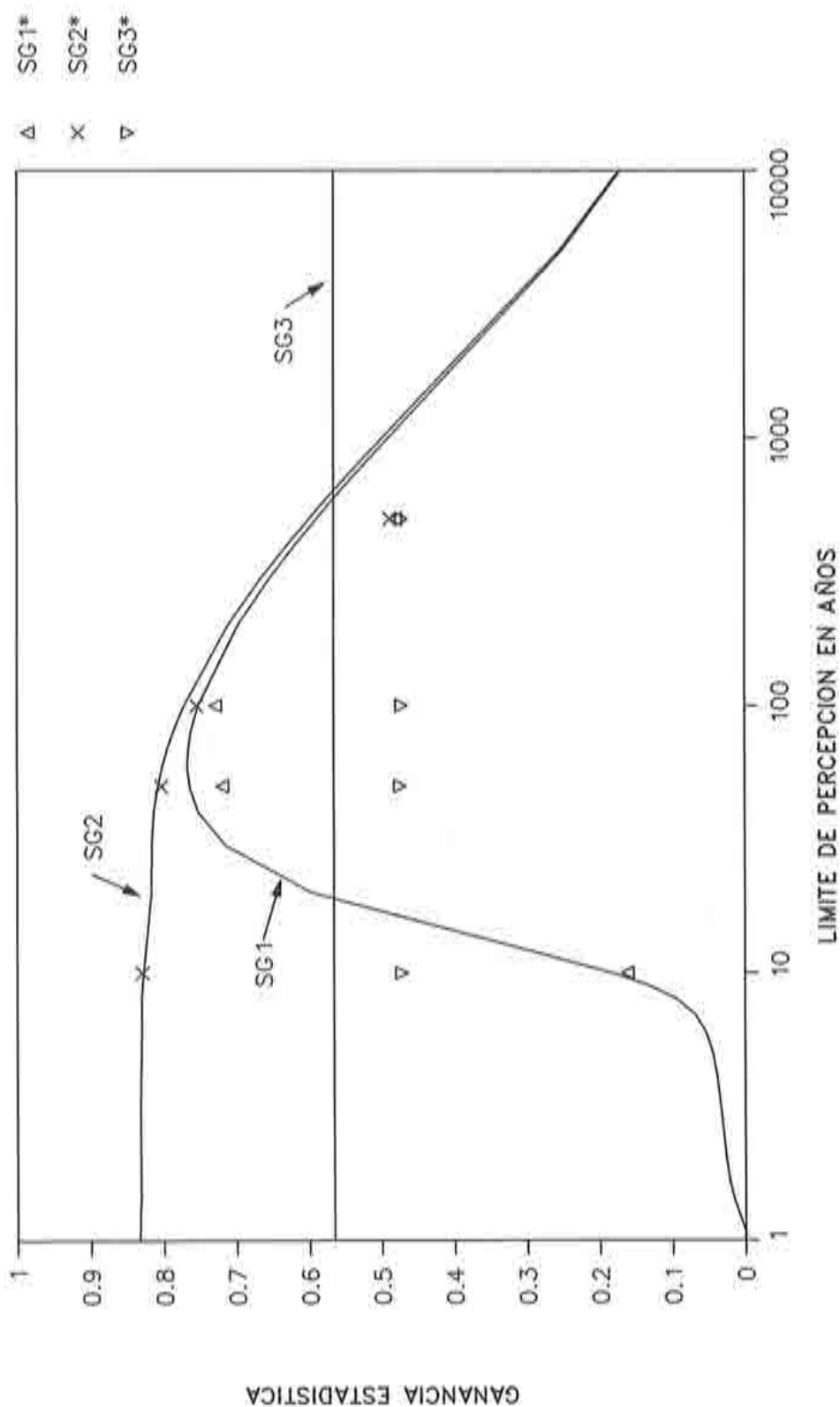


Figura 5.44. Ganancias estadísticas asintótica y muestral en función del límite de percepción histórico, para el escenario de factores inicial.

5.8.- UTILIZACION DE LA TCEV EN UN ANALISIS REGIONAL

5.8.1.- Fases del análisis regional

El campo de utilización habitual de la TCEV es dentro de un análisis regional. La hipótesis de regionalización es que las series estandarizadas por la ecuación (5.167) de diferentes estaciones proviene de una misma población TCEV.

Las fases adoptadas en este trabajo para un análisis regional son una mezcla del procedimiento seguido por Rossi et al. (1984) y el de Arnell y Gabrielle (1988). Estas fases son:

i) Detección de los "high outliers" o datos que se salen por exceso de la tendencia general de la serie. Mediante el test SKEW que se describe más adelante, es posible detectar las avenidas extraordinarias de la serie. Con estas avenidas no hay que contar en la obtención de los parámetros de estandarización (Rossi et al. 1984).

ii) Ajuste de una distribución de Gumbel por Máxima Verosimilitud a las avenidas ordinarias. Para ello se utiliza un método iterativo resultante de anular las derivadas parciales de la función de verosimilitud de cada nivel de información, que se encuentran en el Anejo C.

iii) Estandarización de la serie sistemática y de la información histórica mediante:

$$Y_i = \theta X_i - L\lambda \quad (5.2)$$

donde:

X_i = serie original

Y_i = serie estandarizada

λ, θ = parámetros de la Gumbel local

iv) Ajuste de una TCEV por Máxima Verosimilitud a los datos regionales. Aunque la hipótesis de regionalización supone que los parámetros correspondientes a las avenidas ordinarias son unidad, se relaja esta hipótesis obteniendo los 4 parámetros regionales λ_1' , θ_1' , λ_2' y θ_2' .

v) Obtención de los parámetros locales, mediante la desestandarización de los

regionales, dando por resultado:

$$\lambda_1 = \lambda_1' \lambda^{0_1'} \quad (5.3)$$

$$\theta_1 = \theta_1' \theta \quad (5.4)$$

$$\lambda_2 = \lambda_2' \lambda^{0_2'} \quad (5.5)$$

$$\theta_2 = \theta_2' \theta \quad (5.6)$$

vi) Obtención del cuantil X_T del período de retorno T, resolviendo la ecuación:

$$1 - \frac{1}{T} = \exp(-\lambda_1 e^{-\theta_1 X_T} - \lambda_2 e^{-\theta_2 X_T}) \quad (5.7)$$

El programa ROSSI3 realiza este tipo de análisis con solo la información sistemática. La obtención de la TCEV regional utilizando la información histórica se realiza mediante el programa ANTC4.

El análisis descrito varía respecto al de Rossi et al. (1984) en que éstos asumen que $\theta_1' = \lambda_1' = 1$ a la hora de estandarizar. Arnell y Gabrielle (1988) proponen un método iterativo sin detección de "outliers", pero utilizando como parámetros de estandarización los parámetros λ_1 y θ_1 de la TCEV local. El método utilizado es por tanto una mezcla de estas dos propuestas.

5.8.2.- Test SKEW de detección de "high outliers"

En un conjunto de datos reales existe la posibilidad de que una o más de las observaciones tengan valores mucho mayores o mucho menores que el resto de la muestra. Estas poco probables observaciones se denominan "outliers". Los "outliers" pueden ser debidos a errores de medición o evaluación, error en el modelo estadístico asumido o simplemente realizaciones poco probables.

En esta investigación son solo de interés los "outliers" por exceso o "high outliers". Si se asume que la población de las crecidas de un río sigue una distribución TCEV, se

podrán detectar las avenidas extraordinarias como "high outliers" de una distribución Gumbel.

Kottegoda (1984) describe diversos estadísticos para la detección de "outliers", uno de los cuales se basa en el sesgo muestral. Rossi et al. (1984) también escoge este test en su análisis, pero sin detallarlo.

Si se ordena de menor a mayor la serie sistemática, se puede definir un estadístico G_j :

$$G_j = \sqrt{N-j+1} \frac{\sum_{i=1}^{N-j+1} (x'_i - \bar{X}(N-j+1))^3}{\left[\sum_{i=1}^{N-j+1} (x'_i - \bar{X}(N-j+1))^2 \right]^{3/2}} \quad (5.8)$$

donde:

x'_i = serie ordenada de menor a mayor

$\bar{X}(k)$ = media muestral de las primeras k observaciones

Es decir, G_j es un estimador del sesgo muestral sin considerar los j-1 valores mayores.

Matalas et al. (1975) obtuvieron mediante simulación por Montecarlo la media y desviación típica del sesgo muestral de diversas poblaciones, para muestras de longitudes entre 10 y 90 observaciones. Sus resultados se acompañan en la Tabla (5.3) para una población Gumbel. Dado que el sesgo de una Gumbel tiene un valor constante de aproximadamente 1.140, es evidente que el estadístico G_j es altamente sesgado. En el mismo artículo, los autores asumen que el sesgo muestral de una población Gumbel, tiene una distribución Gumbel. Si se realiza un test estadístico "one sided", los valores críticos del 95% vendrán dados por:

$$G_{0.95} = \frac{L\lambda_G}{\theta_G} - \frac{1}{\theta_G} L(-L0.95)$$

donde θ_G y λ_G son los parámetros de la distribución Gumbel.

Los parámetros se obtienen por el método de los momentos (ver Anejo C), con lo que valen:

$$\theta_G = \frac{\pi}{\sqrt{6}\sigma(G)} \quad (5.10)$$

donde:

$$\lambda_G = \theta_G \mu(G) - 0.5772 \quad (5.11)$$

$\mu(G)$ = media muestral del sesgo muestral G

$\sigma(G)$ = desviación típica muestral del sesgo muestral G

Los resultados para diversas longitudes de muestra se acompañan en la Tabla (5.3). Como comprobación, Fiorentino et al. (1985) dan el valor crítico para una longitud de 35 observaciones de 1.87, que coincide aproximadamente con la tabla. Además, la experiencia con este test siempre ha sido muy buena.

Tabla 5.3, Media, desviación típica y valor crítico del 95% del sesgo muestral.

Longitud de muestra	$\mu(G)$	$\sigma(G)$	$G_{0.95}$
10	0.525	0.625	1.690
20	0.740	0.586	1.833
30	0.841	0.555	1.877
40	0.898	0.56	1.879
50	0.937	0.504	1.877
60	0.964	0.485	1.869
70	0.986	0.468	1.859
80	1.002	0.452	1.845
90	1.015	0.437	1.831
∞	1.140	0.000	1.140

5.8.3.- Varianza asintótica de los cuantiles

La desestandarización de los parámetros equivale a la relación entre el cuantil local y regional:

$$X_T = \frac{Y_T + L\lambda}{\theta} \quad (5.12)$$

Por tanto, si no existiera ningún error en la determinación de los parámetros de estandarización, la varianza del cuantil local vendría dada por:

$$Var(X_T) = \frac{1}{\theta^2} Var(Y_T) \quad (5.13)$$

Realmente la varianza de X_T es mayor y, tomando los primeros términos de un desarrollo de Taylor, se puede aproximar por:

$$Var(X_T) = \left(\frac{\partial X_T}{\partial Y_T}, \frac{\partial X_T}{\partial \lambda}, \frac{\partial X_T}{\partial \theta} \right) S \begin{pmatrix} \frac{\partial X_T}{\partial Y_T} \\ \frac{\partial X_T}{\partial \lambda} \\ \frac{\partial X_T}{\partial \theta} \end{pmatrix} \quad (5.14)$$

Siendo S la matriz de varianzas y covarianzas de las variables estadísticas Y_T , λ y θ . Si se hace la hipótesis de independencia entre el cuantil regional y los parámetros de estandarización, esta matriz tendrá la forma:

$$S = \begin{pmatrix} Var(Y_T) & 0 & 0 \\ 0 & Var(\lambda) & Cov(\lambda, \theta) \\ 0 & Cov(\lambda, \theta) & Var(\theta) \end{pmatrix} \quad (5.15)$$

Si se aproximan los elementos de la matriz por los valores asintóticos y se obtienen los términos de la ecuación (5.179) derivando la ecuación (5.177), la varianza de X_T quedaría:

$$Var(X_T) \approx \frac{1}{\theta^2} CRLB(Y_T) + \frac{1}{\lambda^2 \theta^2} CRLB(\lambda) - 2 \frac{Y_T + L\lambda}{\lambda \theta^3} CRLB(\lambda, \theta) + \frac{(Y_T + L\lambda)^2}{\theta^4} CRLB(\theta) \quad (5.16)$$

donde $CRLB(Y_T)$ es la varianza asintótica del cuantil de una distribución TCEV (visto en esta capítulo) y $CRLB(\lambda)$, $CRLB(\theta)$ y $CRLB(\lambda, \theta)$ 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 se pueden deducir a partir de la función de verosimilitud que se encuentra en el Anejo C, y cuya expresión es:

$$I_0(1,1) = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \lambda^2} \right] = \frac{N}{\lambda^2} \quad (5.17)$$

$$I_0(2,2) = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta^2} \right] = \frac{N[1 + \Gamma''(2) + L^2 \lambda - 2\Gamma'(2)L\lambda]}{\theta^2} \quad (5.18)$$

$$I_0(1,2) = I_0(2,1) = -E \left[\frac{\partial^2 LL_{SY}(\Theta)}{\partial \theta \partial \lambda} \right] = \frac{N[\Gamma'(2) - L\lambda]}{\lambda \theta} \quad (5.19)$$

La hipótesis de no existencia de error en la determinación de los parámetros locales hace que la varianza obtenida por la ecuación (5.178) sea una cota inferior de la varianza real. La hipótesis de independencia en la matriz S da lugar a que una cota superior de $\text{Var}(X_T)$ sea la ecuación (5.181). Afortunadamente, para períodos de retorno altos las dos cotas se hacen aproximadamente iguales. Esto es debido a que los tres últimos sumandos de la ecuación (5.181) se hacen despreciables frente al primero.

Parece bastante evidente, a la vista de las ecuaciones (5.178) y (5.181), que cualquier mejora en la estimación de los cuantiles regionales Y_T repercute automáticamente en los cuantiles locales X_T . Es decir, que la información histórica de cualquier estación de aforos se transmite dentro de un análisis regional al resto de estaciones.

5.9.- CONCLUSIONES

Este capítulo realmente se considera el más importante del trabajo, por que analiza una distribución, la TCEV, que posee unas propiedades estadísticas adaptables a los ríos mediterráneos. Los estadísticos más importantes se han estudiado en el Apartado 5.3 obteniéndose como conclusión general que son fundamentalmente función de los parámetros estandarizados λ y θ . Para poder investigar las posibilidades de uso de los diferentes tipos de información histórica, se ha desarrollado en primer lugar un modelo estadístico que es capaz de incorporar combinaciones de dicha información adicional dentro de un análisis regional. Este modelo se materializa en los programas ROSSI3 y ANTC4 desarrollados en el Departamento de Ingeniería Hidráulica y Medio Ambiente de la Universidad Politécnica de Valencia.

Las principales conclusiones de este capítulo referentes al uso de información histórica son:

- i) Siempre se mejoran los estimadores de los cuantiles si se utiliza algún tipo de información histórica frente al uso exclusivo de la serie sistemática. Las ganancias estadísticas son del orden del 50%, en las situaciones que se consideran habituales en Hidrología.
- ii) Las expresiones asintóticas son representativas de sus correspondientes valores muestrales, por tanto, son generalizables las propiedades obtenidas de aquellas.
- iii) Las ganancias estadísticas no son función de los cuatro parámetros de la distribución, si no solamente de los dos parámetros estandarizados. El parámetro λ es representativo de la proporción de las avenidas extraordinarias en el total de máximos anuales, mientras que el parámetro θ refleja la relación entre los órdenes de magnitud de las avenidas ordinarias y extraordinarias.
- iv) La ganancia estadística utilizando información histórica binomial censurada (SG_1) aumenta a medida que aumenta la proporción de avenidas extraordinarias, es decir, a medida que aumenta λ . A medida que se parecen en magnitud las dos poblaciones de avenidas (incremento de θ) la ganancia estadística crece si la proporción de las avenidas extraordinarias es baja y disminuye para altas proporciones de las mismas.
- v) La ganancia estadística utilizando información histórica censurada (SG_2) a efectos prácticos es independiente de todos los parámetros.

vi) La ganancia estadística utilizando información histórica de máxima avenida (SG_3) disminuye a medida que aumenta la proporción de avenidas extraordinarias. Cuando dicha proporción es pequeña prácticamente es independiente del parámetro θ , mientras que para λ altos aumenta con θ .

vii) Si el límite de percepción es inferior a 20 años es mejor utilizar la información censurada o de máxima avenida que la binomial censurada, si es posible. Si el límite se encuentra entre 20 y 500 años la información censurada es la mejor de todas, aunque a medida que aumenta el límite la diferencia frente a la binomial censurada se reduce. Para límites superiores a 500 años la máxima avenida es la información a utilizar. Como lo habitual será tener siempre límites de percepción históricos inferiores a los 100 años de período de retorno, a diferencia de lo que ocurre en la familia de distribuciones "Extreme Value" (EV), sí que existen diferencias apreciables entre SG_1 y SG_2 .

viii) El comportamiento asintótico de SG_1 y SG_2 hace que no interese incrementar en exceso el período histórico. Por ejemplo, si existe información de tiempos históricos no resultará rentable estadísticamente un estudio sobre paleoavenidas. Este resultado es similar en la familia EV.

ix) Las ganancias estadísticas producidas con el uso de la información histórica en una estación de aforos, se transmite en un análisis regional al resto de estaciones, especialmente para cuantiles de período de retorno alto.

6.- APLICACION A LOS RIOS JUCAR, TURIA Y SEGURA

6.1.- INTRODUCCION

En este capítulo se aplican los modelos estadísticos desarrollados por en los capítulos anteriores a tres ríos importantes de la cuenca mediterránea: Júcar, Turia y Segura.

Las características de los ríos del Levante español dan lugar a que éstos tengan avenidas muy peligrosas para las poblaciones que se asientan en sus márgenes. Son ríos con caudales medios bajos, tanto de aportaciones como de crecidas, pero que de "vez en cuando" originan avenidas catastróficas. Baste como ejemplo recordar la avenida del río Júcar de 1982 que provocó 34 muertos, 200.000 personas sin vivienda y daños por valor de 250.000 millones de pesetas de 1982 (Las Provincias, 1988). Dentro del análisis de la frecuencia de las avenidas precisamente el analizar ese "de vez en cuando" de las extraordinarias es el punto clave del análisis.

Las avenidas de los ríos de la franja costera mediterránea española, se puede considerar que proceden de dos poblaciones estadísticas distintas. La primera de ellas, producida por lluvias de tipo ciclónico, da lugar a las avenidas ordinarias de los ríos. La segunda de ellas es generada por precipitaciones de tipo convectivo, originando las avenidas de tipo extraordinario.

Las más extremas de las avenidas extraordinarias se presentan fundamentalmente en Otoño, cuando se produce el fenómeno llamado de "la gota fría". Los factores necesarios para la presentación de este fenómeno son una elevada presión de vapor en la atmósfera, una orografía favorable a la elevación del aire, vientos que soplen desde el mar a la tierra (vientos de Levante), una bolsa de aire frío en altura y temperatura alta del agua de mar. Los dos últimos factores solo son posibles durante el Otoño, de ahí la concentración del fenómeno en esta época del año.

El ajuste de una función de distribución de probabilidad que no tenga en cuenta esta dualidad en el origen de las avenidas puede dar lugar a una infravaloración en los resultados para períodos de retorno superiores a los 50 años. Por ejemplo, basta observar el ajuste de una distribución Gumbel clásica a los datos sistemáticos del río Júcar en la Figura (6.6). Con este pobre ajuste se produce una clara infravaloración de los cuantiles de alto período de retorno, es decir, fijado un nivel de riesgo, el cuantil dado por la distribución Gumbel es mucho menor de lo que realmente es.

La función TCEV, como se vio en el Capítulo 2 de Antecedentes, sí que puede reflejar un proceso mixto de dos poblaciones estadísticas y además, para los casos analizados, el ajuste es bastante bueno. Sin embargo, esta función de distribución tiene 4 parámetros, con lo que

lo que la incertidumbre de los estimadores de los cuantiles es alta si se tratan de estimar con series de datos sistemáticos del orden de 40 años: se estaría contraviniendo el Principio de Parsimonia estadístico.

Una forma de aumentar la fiabilidad de los estimadores, es aumentar la longitud estadística de la información, bien sea mediante la utilización de la información histórica disponible o mediante un análisis regional. En este estudio de las avenidas de los ríos Júcar, Turia y Segura se han utilizado ambas técnicas.

Las fases del análisis han sido:

- 1.- Estudio individual de cada río, utilizando la información sistemática e histórica disponible.
- 2.- Análisis regional de los ríos Júcar y Turia.

Dada la incompatibilidad regional entre el Segura y los ríos Turia y Júcar, no ha sido posible un análisis conjunto de los tres.

Como criterio de comprobación del ajuste de la distribución estimada, se ha utilizado la comparación gráfica de la función de densidad de probabilidad acumulada (f.d.a.) ajustada con la f.d.a. empírica, cuya obtención se describe en el Anejo A. Los tests de ajuste generales desarrollados hasta la actualidad no consideran especialmente la cola derecha de la distribución, resultando que la mayor parte de las distribuciones pasan estos tests. Sin embargo, la cola derecha es la principal zona de interés en el análisis de la frecuencia de las crecidas. Por ello, estos tipos de tests no son habituales en Hidrología.

La metodología utilizada para el ajuste de una distribución TCEV es la descrita en el Capítulo 5, para lo que el Departamento de Ingeniería Hidráulica y Medio Ambiente de la Universidad Politécnica de Valencia desarrolló los programas específicos ROSSI3 y ANTC4.

6.2.- INFORMACION DISPONIBLE

6.2.1.- Río Júcar

La región más problemática, desde el punto de vista de las avenidas, es el llano de inundación ocupado por las comarcas de La Ribera Alta y de La Ribera Baja. La superficie total de la cuenca es de 22.000 km² aproximadamente, aunque las partes de la cuenca que contribuye a las avenidas en estas comarcas, son las cuencas media y baja.

Las estaciones de aforos disponibles en esta zona son la E-42 en el embalse de Tous, la E-89 en el Huerto Mulet y la E-45 en Alcira, esta última abandonada en 1974.

Por la longitud de la serie y por encontrarse aguas abajo de la confluencia del río Magro, la serie sistemática seleccionada es la E-89. Los datos de caudales máximos instantáneos anuales disponibles abarcan el período desde el año hidrológico 1946-47 hasta el 1988-89 (Tabla 6.1 y Figura 6.1), aunque no de forma continua, por lo que se tienen solo 42 años.

La variable estadística seleccionada es la de máximos instantáneos anuales, a pesar de contar con menor número de datos. Las razones han sido:

- Ajuste defectuoso de los máximos diarios.
- La información histórica que se va a utilizar se refiere a los máximos instantáneos de las avenidas históricas.

En cuanto a la información histórica se tienen noticias de 70 avenidas importantes desde el año 1388 (F. Calvo, 1989, Comisión Nacional de Protección Civil, 1983). En otras fuentes se citan 217 desde el año 1088 (Comisaría de Aguas C.H.J., 1983). Sin embargo, el Centro de Estudios Hidrográficos (1983), en su estudio sobre la avenida de Octubre de 1982, y utilizando diversas fuentes, llega a cuantificar las avenidas más importantes acaecidas desde el siglo XVII, a excepción de la avenida de 1632. Sus resultados se adjuntan en la Tabla (6.2).

La existencia de este valioso estudio hace inmediato el uso de esta información como de tipo censurada, siendo el umbral de percepción de 6.200 m³/s. Este umbral fue superado 5 veces desde el año 1600. A pesar de esta aparente abundancia de información, el período histórico considerado es el del año 1792 (año intermedio entre 1778 y 1805) al año 1945 (comienzo de la serie sistemática). Las razones han sido básicamente dos:

i) El desconocimiento del valor de la avenida de 1632, lo que obligaría a utilizar la información histórica como binomial censurada o de máxima avenida. Si la distribución a ajustar fuese de la familia EV no habría mucha diferencia entre la ganancia estadística con información histórica censurada y binomial censurada, con lo que se podría prolongar el período histórico. Sin embargo, para la TCEV no ocurre lo mismo, ya que la diferencia es importante.

ii) El hecho de que la información acerca de los niveles alcanzados en las avenidas de la Tabla (6.2), tiene su origen en los datos recogidos por el párroco de Carcagente a mediados del siglo XIX (Gómez et al. 1865). Evidentemente, esta información es más detallada para las avenidas vividas por él mismo y tanto más difusa cuanto más se remontan en el tiempo. Por lo tanto, es posible que avenidas importantes de los siglos XVII y XVIII no hayan sido recogidas.

Tabla 6.1. Datos de aforos de la estación 89 del río Júcar en Huerto Mulet.

Año hidrológico	Máximo diario (m ³ /s)	Máximo instantáneo (m ³ /s)
1946-47	429	464
1947	407	447
1948	735	880
1949	144	181
1950	426	860
1951	231	315
1952	84	93
1953	368	375
1954	77	134
1955	137	177
1956	184	954
1957	382	440
1958	505	800

Año hidrológico	Máximo diario (m ³ /s)	Máximo instantáneo (m ³ /s)
1959	205	302
1960	100	114
1961	89	143
1962	170	207
1963	126	128
1964	107	
1965	217	362
1966	123	161
1967	101	155
1968	153	161
1969	326	461
1970	118	133
1971	504	556
1972	273	377
1973	101	163
1974	68	70
1975	267	347
1976	179	319
1977	402	524
1978	86	87
1979	189	226
1980	71	87
1981	40	46
1982		12000
1983	29	31
1984	270	347

Año hidrológico	Máximo diario (m ³ /s)	Máximo instantáneo (m ³ /s)
1985	137	511
1986	295	413
1987	795	5200
1988	135	207

Tabla 6.2. Avenidas históricas del río Júcar a su paso por Carcagente (Centro de Estudios Hidrográficos, 1983).

Año	Caudal Punta (m ³ /s)
1632	Superior a 6.200
1778	6.200
1805	8.400
1814	6.400
1864	13.000
1923	4.800

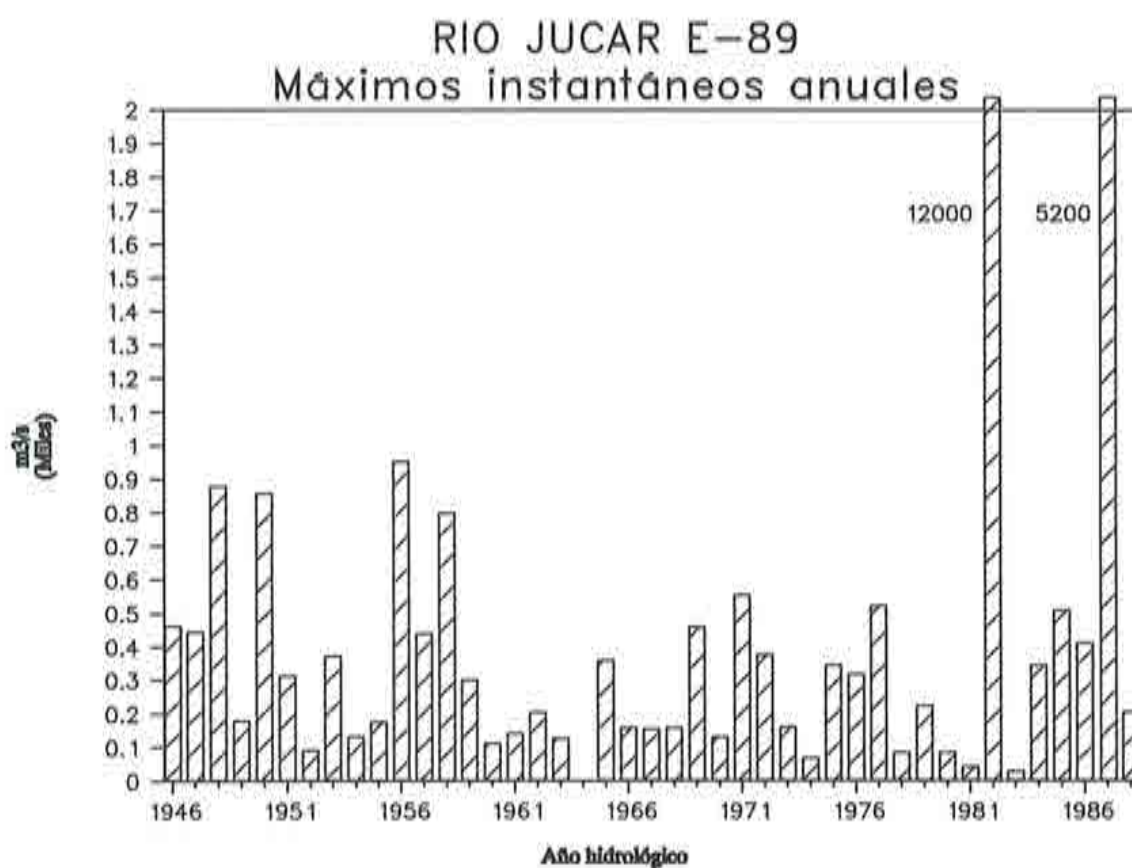


Figura 6.1. Caudales máximos instantáneos anuales del río Júcar. Estación de aforos 89.

6.2.2.- Río Turia

El punto de la cuenca del río Turia escogido para el análisis es su paso por la ciudad de Valencia. La estación de aforos utilizada es la E-25 en la presa de Manises, inmediatamente aguas arriba de la ciudad. La variable estadística utilizada, análogamente al río Júcar, es el máximo instantáneo anual. De esta estación se disponen datos sistemáticos de avenidas (máximos instantáneos anuales) de 1945-46 a 1988-89, que se encuentran recogidos en la Tabla (6.3) y Figura (6.2), siendo un total de 41 años no continuos de datos. La cuenca vertiente a esta estación es de 6.300 km² aproximadamente, aunque análogamente a lo que ocurre en el río Júcar, la cuenca productora de las avenidas del río es mucho menor.

La información histórica que se ha utilizado básicamente proviene de Almela (1957), Comisión Nacional de Protección Civil (1983), Carmona (1986 y 1990) y Las Provincias (1988).

La primera noticia que se tiene de una inundación de la ciudad de Valencia, es del año 1088 en la época musulmana (Huici, 1969). Según Carmona (1986 y 1990) desde 1321 hasta 1977 se han producido en Valencia 22 desbordamientos, 11 avenidas y 15 noticias de avenidas menos importantes, es decir, un total de 48 sucesos. Sin embargo, la constitución en 1589 de la "Fábrica Nova del Riu" y el comienzo de la construcción de los pretilos del antiguo cauce, suponen la modificación del mecanismo de desbordamiento del río a su paso por la ciudad.

Afortunadamente, los pretilos del río suponen un elemento censor de las avenidas históricas muy fácil de detectar en las crónicas, como es el del desbordamiento por encima de los mismos. Cuando esto se producía, el primer punto en inundarse eran las poblaciones de Campanar y Marchalenes (hoy en día barrios de la ciudad) por encontrarse en una depresión aguas abajo de un estrechamiento del río (coincidente con el actual puente del 9 de Octubre). Esto explica en parte el menor desarrollo que han tenido a lo largo de la historia los barrios de la margen izquierda del Turia.

El caudal correspondiente a este límite de percepción se ha considerado igual al caudal máximo de la avenida de 1949, que fue de 2300 m³/s. Esta avenida llenó por completo el cauce del río provocando desbordamientos localizados (Carmona, 1986).

Como se observa en la Tabla (6.4), de 1590 a 1944 (inicio de la serie sistemática) se produjeron de 3 a 7 inundaciones de los poblados de Campanar y Marchalenes, dependiendo de la fuente. El número más probable es 7: las avenidas de 1651, 1676, 1689, 1731, 1776, 1870 y 1879. Sin embargo, la construcción de los pretilos del río se realizó a lo largo del siglo XVII, por lo que probablemente no se puede adoptar durante este siglo el mismo límite de

percepción que en siglos posteriores. A esta hipótesis de información histórica se le va a denominar de período largo. Una alternativa es considerar el período de 1710 (intermedio entre 1689 y 1731) a 1944, en el que el umbral de percepción fue superado 4 veces. A esta hipótesis se la ha denominado de período corto. En el análisis posterior se utilizan ambas hipótesis, seleccionando aquella que de lugar a resultados más homogéneos a nivel regional.

Tabla 6.3. Datos de aforos de la estación 25 del río Turia en la presa de Manises.

Año hidrológico	Máximo diario (m ³ /s)	Máximo instantáneo (m ³ /s)
1944-45	86	
1945	54	139
1946	45	90
1947	38	63
1948	310	2300
1949	21	26
1950	108	260
1951	51	117
1952	32	76
1953	73	514
1954	20	84
1955	30	90
1956	104	
1957	2676	3700
1958		
1959		
1960	59	88
1961	133	155
1962	45	199

Año hidrológico	Máximo diario (m ³ /s)	Máximo instantáneo (m ³ /s)
1963	52	60
1964	37	58
1965	42	79
1966	27	150
1967	75	918
1968	34	90
1969	74	133
1970	21	25
1971	50	150
1972	38	136
1973	29	35
1974	21	43
1975	19	34
1976	23	40
1977	52	238
1978	25	37
1979	37	49
1980	22	32
1981	17	42
1982	17	34
1983	25	117
1984	33	64
1985	18	48
1986	19	48
1987	39	42
1988	72	144

Tabla 6.4. Inundaciones de la ciudad de Valencia entre 1589 y 1944 según diversas fuentes (S = inundación en la ciudad, ? = solo referencia, * = no consta).

AÑOS	ALMELA	CARMONA	LAS PROVINCIAS	PROTECCION CIVIL
1651	S	S	S	S
1672	?	*	*	S
1676	S	S	*	S
1689	S	S	*	?
1731	S	S	S	S
1770	*	*	*	?
1776	S	?	?	?
1783	?	?	?	*
1870	S	?	*	S
1897	S	S	S	S

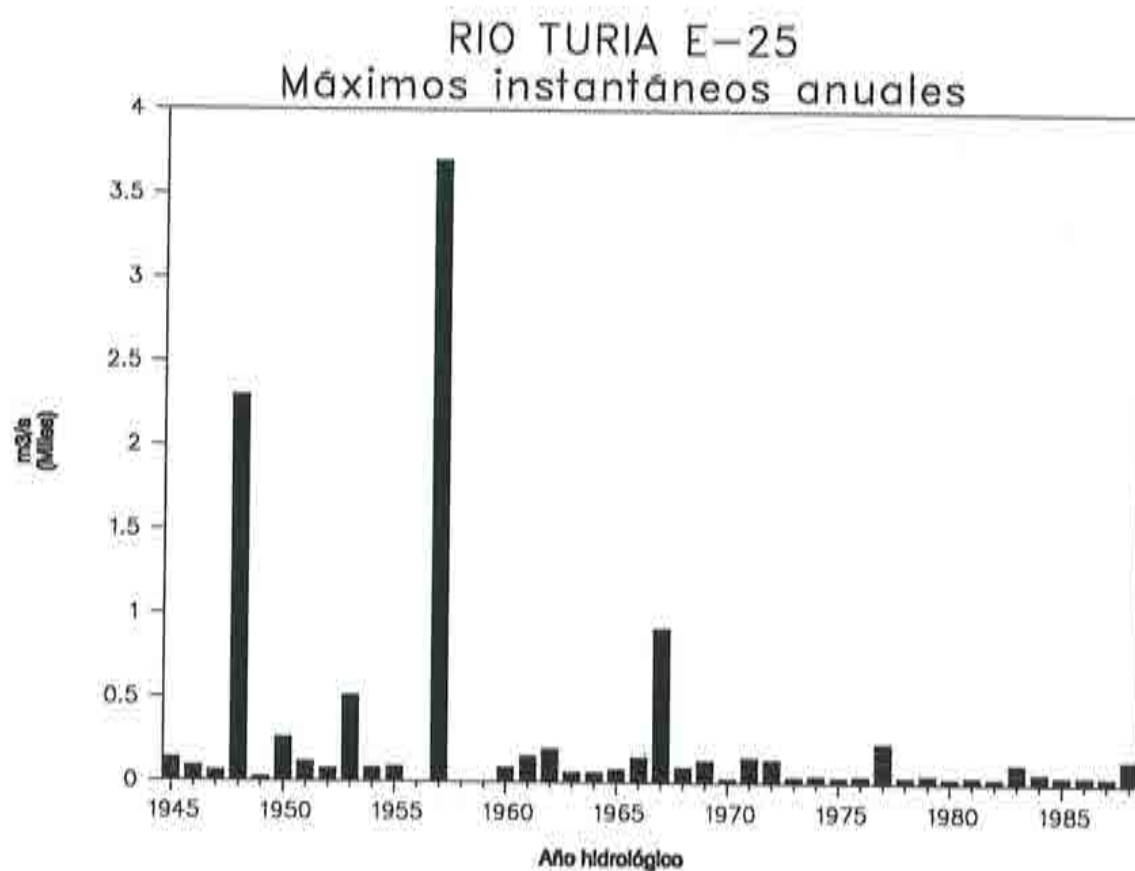


Figura 6.2. Caudales máximos instantáneos anuales del río Turia. Estación de aforos 25.

6.2.3.- Río Segura

La estación de aforos utilizada es la 28 situada en Orihuela. La superficie de la cuenca que drena en la estación es de 13606 km². Se trata de la única estación que funciona actualmente en la cuenca baja del río (exceptuando la estación de aforos de Guardamar).

Sin embargo, la serie sistemática de máximos anuales instantáneos es de solo 15 datos y no continuos como se refleja en la Tabla (6.5) y Figura (6.3). Por ello, en este caso, se ha utilizado como variable estadística los caudales máximos diarios (Figura 6.4).

Para pasar el dato de información histórica instantáneo a diario, se ha realizado una regresión lineal con los 12 datos comunes. El resultado es:

$$Q_t = 1.05033Q_d \quad (6.1)$$

El coeficiente de la regresión, r^2 , es de 0.9795 y como se observa en la Figura (6.5) el ajuste es bastante bueno.

Las avenidas en Orihuela pueden ser debidas tanto a crecidas del río Guadalentín, río que desemboca en el Segura aguas abajo de Murcia, como a la cuenca baja del río Segura. Sin embargo, las crecidas del Guadalentín se veían atenuadas por el hecho de que las aguas del río se desparramaban por la huerta antes de ser recogidas por un azarbe y conducidas al Segura aguas arriba de Orihuela. Este mecanismo se modifica en 1825 con la construcción del Canal del Reguerón, con lo que el régimen de avenidas se acentúa. De hecho, el Congreso contra Inundaciones de Levante celebrado en 1885 (Bautista, 1989), el informe de los ingenieros García Hernández y Gaztelo Maritorea de 1886 (Bautista, 1989), y el Ayuntamiento de Orihuela en sesión del 2 de Mayo de 1889 (Canales, 1990), acusan al Guadalentín de las inundaciones que se producen en Orihuela.

Couchoud (1965) realizó una recopilación de todas las avenidas sufridas en las huertas de Murcia y Orihuela desde 1545 hasta 1879. De su lectura se deduce que las tres mayores fueron las riadas de San Lucas en 1545, San Calixto en 1651 y Santa Teresa en 1879.

El impacto en la comunidad mundial provocado por los sucesos de la riada de Santa Teresa de 1879, en pleno siglo de Las Luces, hace que ésta se tome como referencia en avenidas posteriores y que se preste mayor atención a las mismas.

Por todo ello, se adopta como período histórico desde 1825 (construcción del Canal del Reguerón) hasta el comienzo de la serie sistemática en 1911, ambos inclusive.

Couchoud (1965) en ese período referencia dos grandes avenidas y lo que es más importante, informa acerca de su magnitud:

- Riada de Santa Brígida en 1834, con 938 m³/s de punta en Murcia (893 m³/s de caudal medio diario).
- Riada de Santa Teresa en 1879, con 1890 m³/s de punta en Murcia (1799 m³/s de caudal medio diario).

Canales (1990) analizando el período 1875-1925 afirma que las avenidas de San Florencio en 1884 y San Valentín en 1895 produjeron más daños en la huerta de Orihuela que la de Santa Teresa. La de San Andrés de 1916 también produjo importantes destrozos pero debido a su larga duración. Su caudal medio fue de 305 m³/s (tabla 6.5) por lo que no se puede considerar como una avenida importante, y en cualquier caso se encuentra dentro de la serie sistemática.

Por tanto, la información histórica utilizada es del tipo binomial censurada. El umbral de percepción histórico adoptado es de 800 m³/s, que es un caudal medio diario ligeramente inferior al de la avenida sistemática de 1946 (menor de las avenidas extraordinarias). Por tanto, en el período histórico desde 1825 hasta 1911 (87 años) hubo 4 avenidas importantes, superiores a la de Santa Brígida de 1834, que a su vez es superior al límite anteriormente declarado.

A efectos de la obtención de la curva de densidad probabilidad acumulada (f.d.a.) empírica (ver Anejo A para una descripción de la misma), el orden de las 7 avenidas superiores al límite de percepción cuyo valor se conoce (históricas y sistemáticas) es el siguiente:

- 1ª. La avenida histórica de Sta. Teresa de 1879 con 1799 m³/s.
- 4ª. La avenida sistemática de 1948 con 950 m³/s.
- 5ª. La avenida sistemática de 1945 con 928 m³/s.
- 6ª. La avenida histórica de Sta. Brígida de 1834 con 893 m³/s.
- 7ª. La avenida sistemática de 1946 con 845 m³/s.

Tabla 6.5. Datos de aforos de la estación 28 del río Segura en Orihuela.

Año hidrológico	Máximo diario (m ³ /s)	Máximo instantáneo (m ³ /s)
1912-13	56	
1913		
1914		
1915	204	
1916	305	
1917	92	
1918	170	
1919	208	
1920	208	
1921	83	
1922	180	
1923		
1924	259	
1925	116	
1926	266	
1927	158	
1928	240	
1929	138	
1930	56	
1931	121	
1932	109	
1933	23	
1934	60	
1935	70	
1936	130	

Año hidrológico	Máximo diario (m ³ /s)	Máximo instantáneo (m ³ /s)
1937	29	
1938	108	
1939	295	
1940	80	
1941	152	
1942	248	337
1943	434	472
1944	391	408
1945	928	950
1946	845	
1947	108	
1948	950	971
1949	59	140
1950	252	
1951	219	
1952	131	
1953	214	
1954	161	
1955	28	
1956	10	
1957		
1958		
1959		
1960		
1961		
1962		

Año hidrológico	Máximo diario (m ³ /s)	Máximo instantáneo (m ³ /s)
1963		
1964	169	274
1965	243	
1966	185	198
1967	46	
1968	46	
1969	86	
1970	42	
1971	19	21
1972	75	75
1973	6	
1974	60	99
1975	37	
1976	37	
1977	35	
1978	32	
1979	50	
1980	24	
1981	17	
1982	31	31
1983	20	

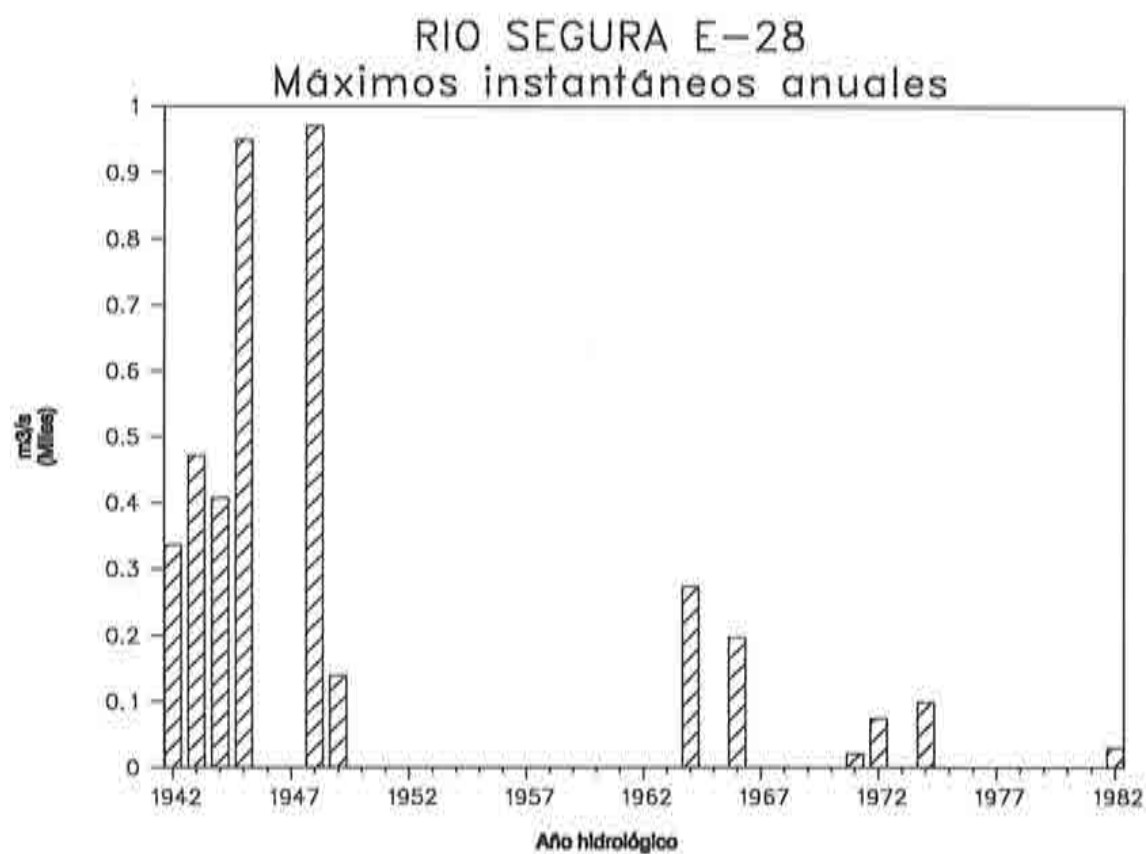


Figura 6.3. Caudales máximos instantáneos anuales del río Segura. Estación de aforos 28.

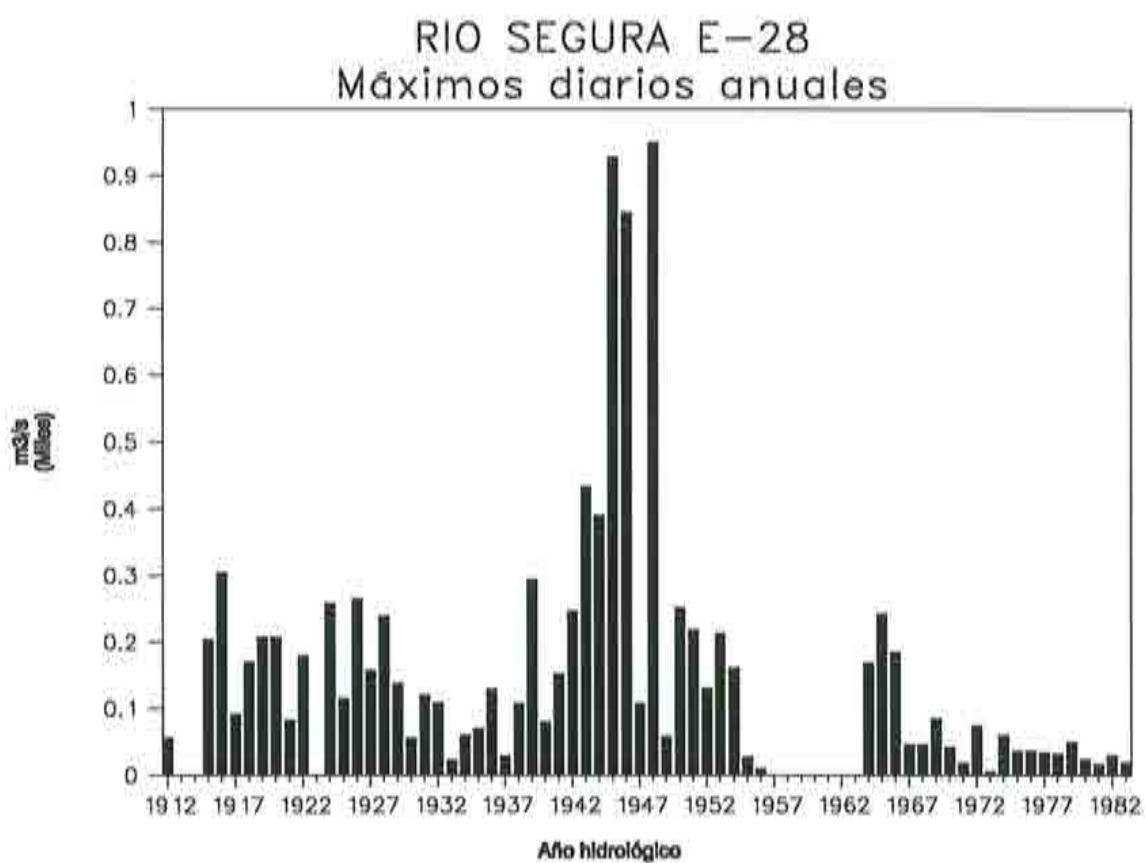


Figura 6.4. Caudales máximos diarios anuales del río Segura, Estación de aforos 28.

RIO SEGURA E-28

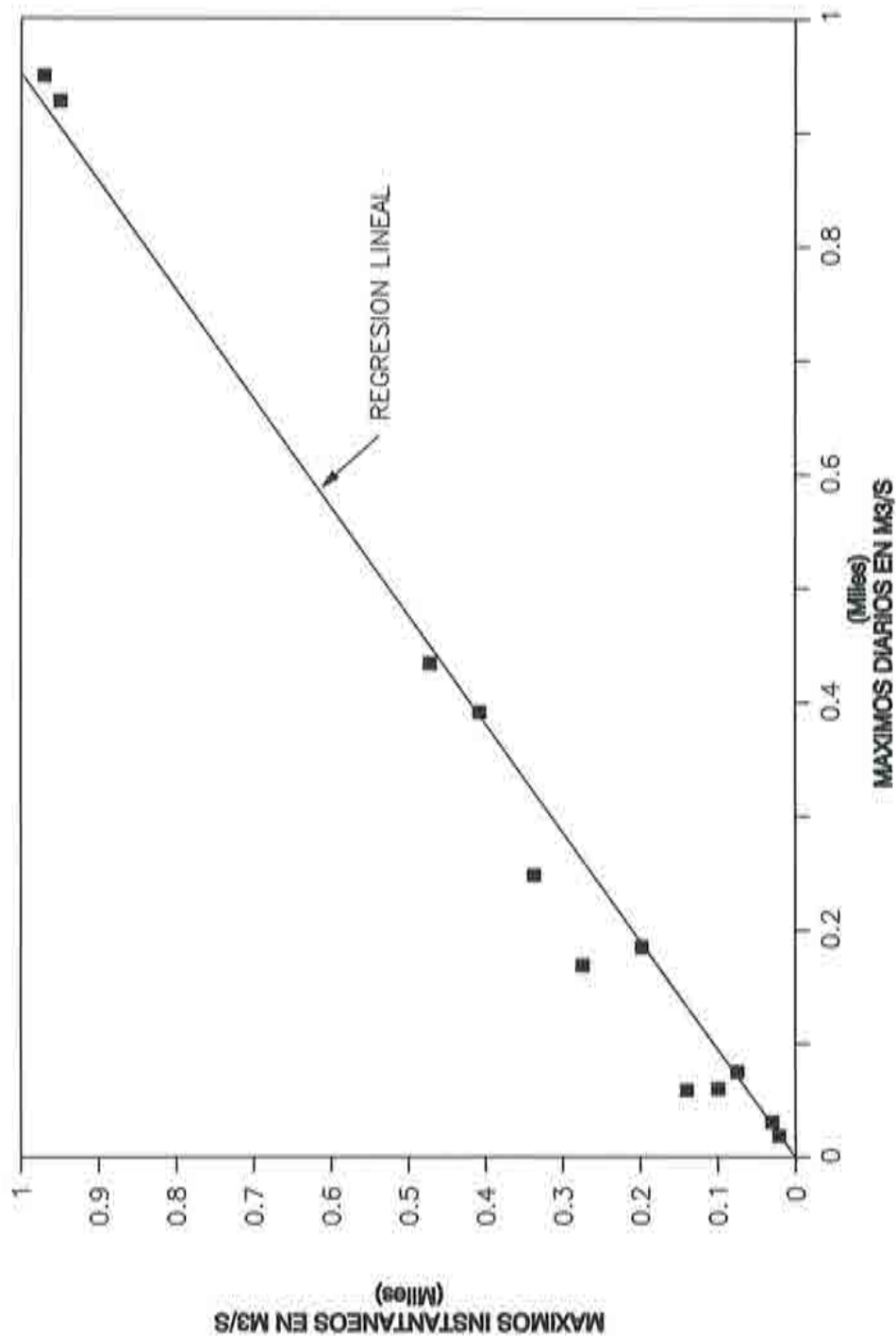


Figura 6.5. Regresión lineal entre los caudales máximos instantáneos y diarios anuales del río Segura en la estación de aforos 28.

6.3.- ANALISIS LOCAL DEL RIO JUCAR

En primer lugar se ha ajustado una función de distribución de tipo Gumbel a los 42 datos sistemáticos por el Método de Máxima Verosimilitud. Los parámetros y los cuantiles estimados se encuentran en la Tabla (6.6). La representación en escala Gumbel de este ajuste se encuentra en la Figura (6.6), comparado con la f.d.a. empírica utilizando información histórica (es decir, con toda la información disponible).

Como se observa en la misma, el ajuste de una distribución de Gumbel a los datos sistemáticos del río Júcar es desastroso. Claramente, los máximos instantáneos anuales derivan de dos poblaciones distintas que solo pueden ser representadas por una distribución TCEV. El ajuste de una TCEV a los datos sistemáticos se resume en la Tabla (6.6) y se representa en la Figura (6.7). En la serie sistemática del río Júcar solo existen dos avenidas extraordinarias. Por lo tanto, la incertidumbre en la estimación de una rama derecha de la distribución es alta. A pesar de ello, la comparación con la información histórica no utilizada indica que el ajuste es aceptable.

La utilización de la información histórica con ajuste de una distribución TCEV queda reflejado en la Tabla (6.6) y Figura (6.8). La incertidumbre en la estimación de los cuantiles correspondientes a las avenidas extraordinarias (período de retorno superior a 25 años aproximadamente) se reduce en gran medida, al disponer de mayor información sobre la rama derecha de la distribución.

Tabla 6.6. Parámetros y cuantiles estimados para el río Júcar con diversas hipótesis de análisis local.

Parámetros por M.V.	Gumbel sistemática	TCEV sistemática	TCEV histórica
$\lambda_1=$	1,846	3.757	3.775
$\theta_1=$	0.002079	0.006205	0.006276
$\lambda_2=$	0	0.06172	0.07223
$\theta_2=$	0	0.0001454	0.0002147
Período de retorno	Cuantiles en m ³ /s		
10	1377.3	697.38	711.67
50	2171.7	7678.9	5934.5
100	2507.5	12480.	9186.7
500	3283.6	23575.	16702.
1000	3617.3	28344.	19933.

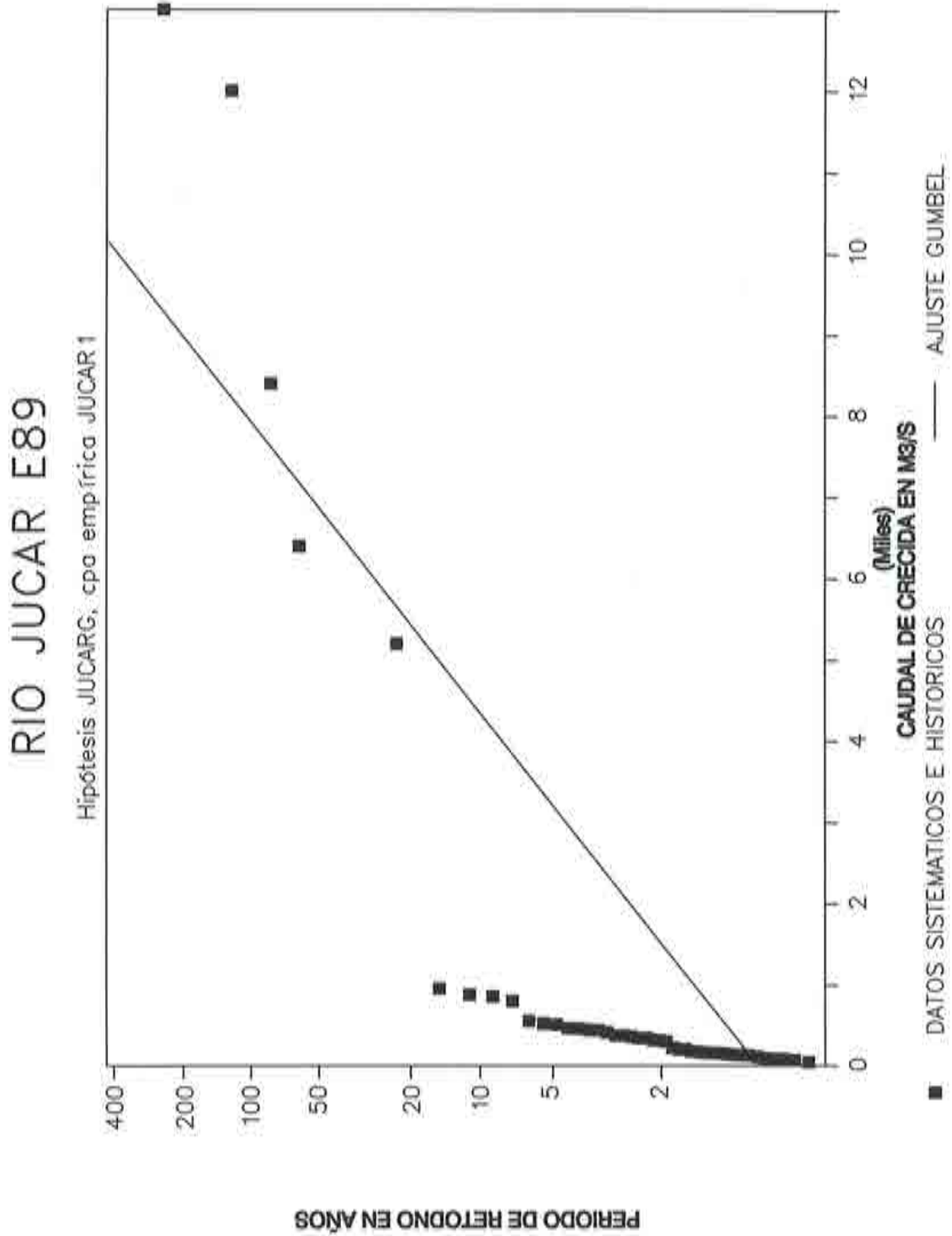


Figura 6.6. Ajuste de una distribución Gumbel con los datos sistemáticos exclusivamente en el río Júcar. F.d.a. empírica considerando la información histórica.

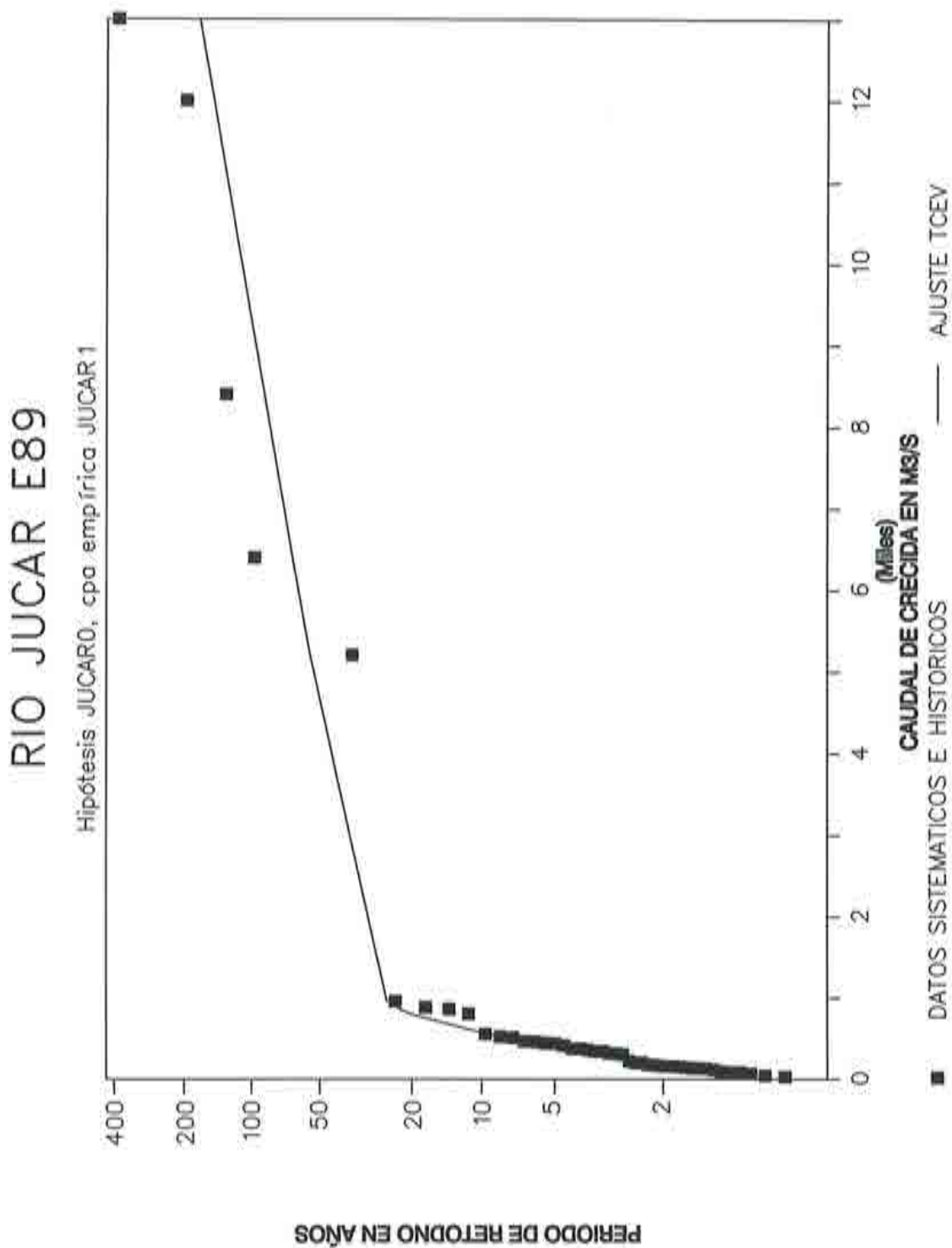


Figura 6.7. Ajuste de una distribución TCEV con los datos sistemáticos exclusivamente en el río Júcar. F.d.a. empírica considerando la información histórica.

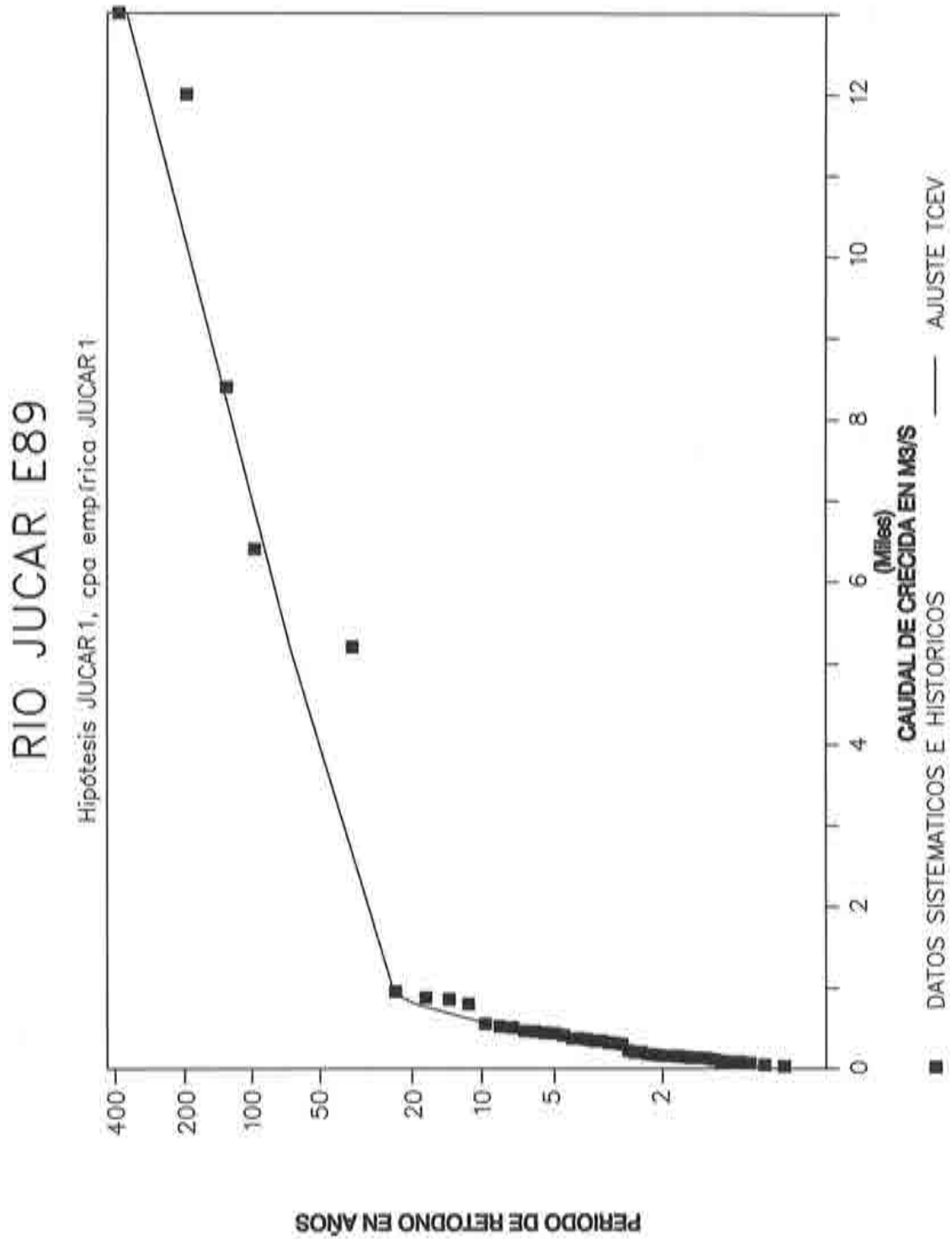


Figura 6.8. Ajuste de una distribución TCEV con los datos sistemáticos e históricos al río Júcar. F.d.a. considerando la información histórica.

6.4.- ANALISIS LOCAL DEL RIO TURIA

Análogamente al río Júcar, el río Turia presenta dos poblaciones de crecidas de características estadísticas diferenciadas, como queda reflejado en la Figura (6.9). El ajuste de una distribución de Gumbel (Tabla 6.7) es totalmente erróneo.

La utilización de una distribución TCEV considerando solamente la serie sistemática de aforos (Tabla 6.7 y Figura 6.10) tiene en cuenta este fenómeno; pero, dado que solo se tienen 4 datos sistemáticos en la rama derecha de la función, la incertidumbre en el ajuste de esta rama es alta. Además, como se observa en la Figura (6.10), este ajuste da cuantiles superiores a la f.d.a. empírica con información histórica de período corto. La razón fundamental es que, precisamente en la serie sistemática, se encuentra la máxima avenida de todo el período considerado.

En principio, la inclusión de la información histórica de la hipótesis de período largo (Tabla 6.7 y Figura 6.10) podría mejorar esta situación. Sin embargo, según se observa en la Figura (6.11), esto no es así. El ajuste no es compatible con la información histórica, por lo que esta última debe tener alguna incongruencia. La incongruencia probablemente se encuentra en el hecho de considerar el umbral de percepción constante durante todo el período histórico desde el inicio de la construcción de los pretilos del río.

Si se considera la hipótesis con un período histórico menor, en el que ya sí se puede garantizar la constancia del umbral de percepción, como se observa en la Figura (6.12) el ajuste es aceptable. Por tanto, en el análisis regional esta hipótesis de información histórica será la utilizada. Los resultados se encuentran en la Tabla (6.7).

Tabla 6.7. Parámetros y cuantiles estimados para el río Turia con diversas hipótesis de análisis local.

Parámetros por M.V.	Gumbel sistemática	TCEV sistemática	TCEV histórica larga	TCEV histórica corta
$\lambda_1 =$	1.710	5.117	5.124	5.132
$\theta_1 =$	0.005364	0.02633	0.02644	0.02654
$\lambda_2 =$	0	0.1393	0.1335	0.1320
$\theta_2 =$	0	0.0007194	0.0008073	0.0008416
Período de retorno	Cuantiles en m ³ /s			
10	519.55	390.37	310.23	292.82
50	827.45	2684.0	2339.4	2230.5
100	957.61	3654.6	3204.3	3060.0
500	1258.4	5897.5	5202.9	4977.1
1000	1387.7	6861.8	6062.1	5801.2

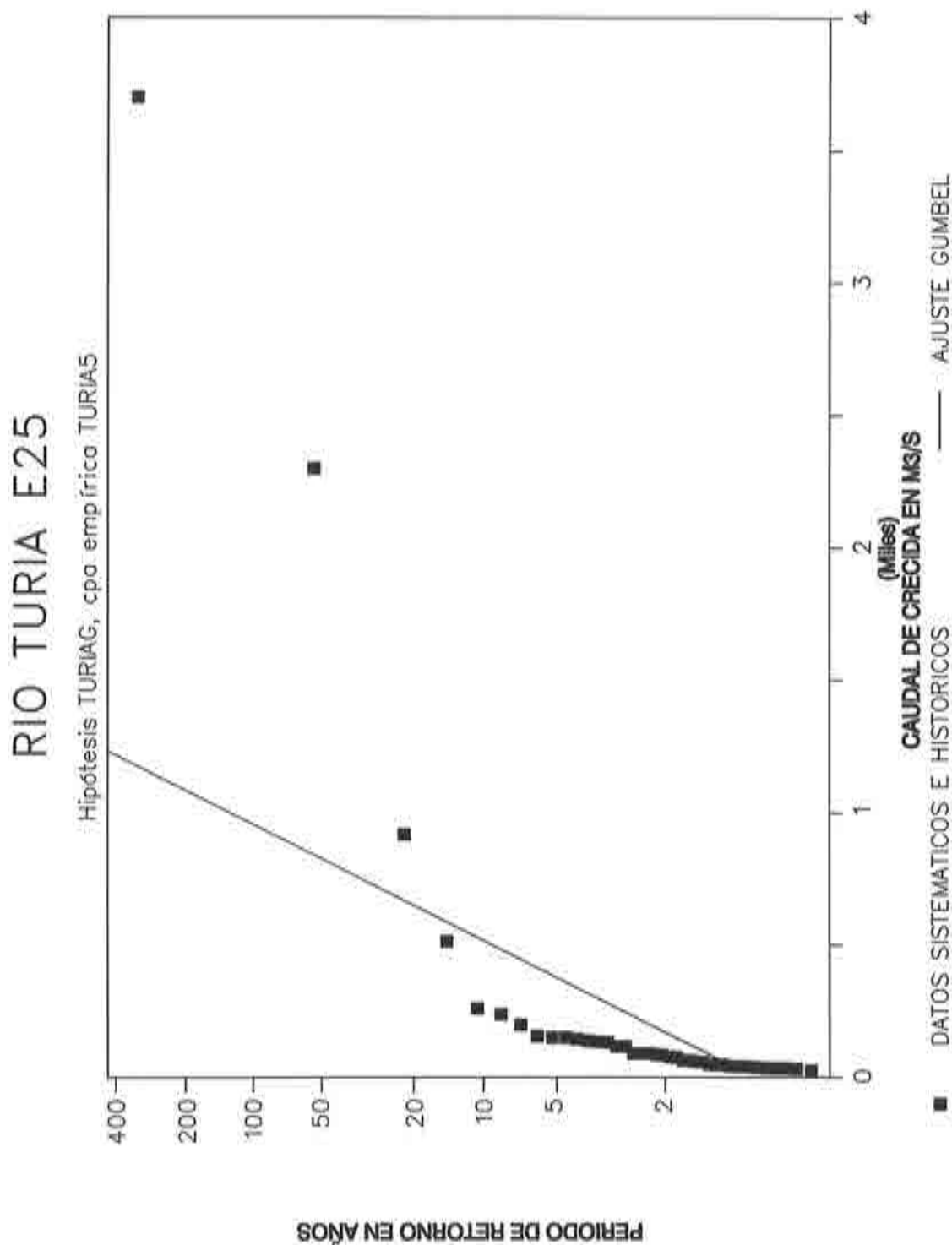


Figura 6.9 Ajuste de una distribución Gumbel a los datos sistemáticos del río Turia. F.d.a. empírica con la información histórica de la hipótesis corta.

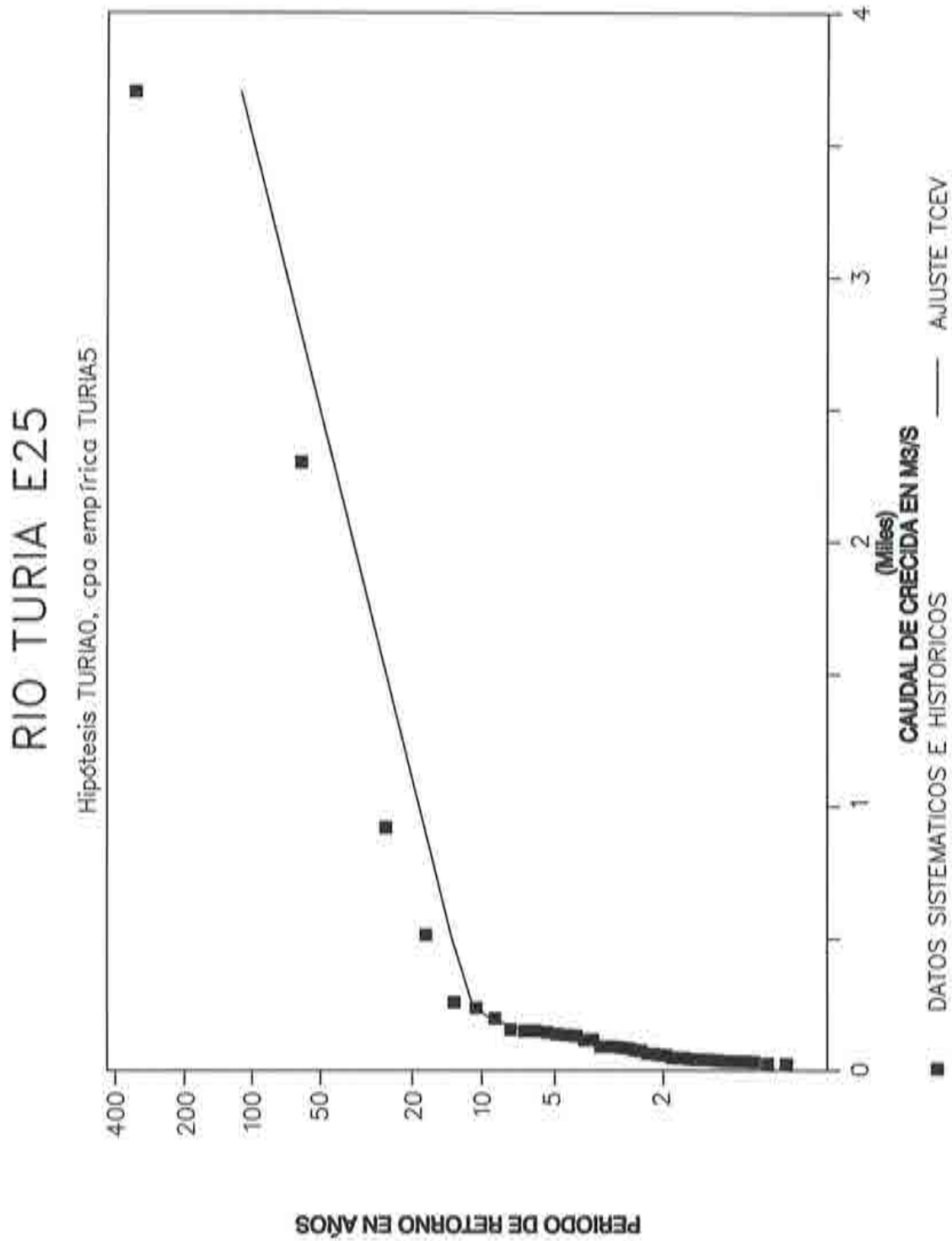


Figura 6.10 Ajuste de una distribución TCEV a los datos sistemáticos del río Turia. F.d.a. empírica con la información histórica de la hipótesis corta.

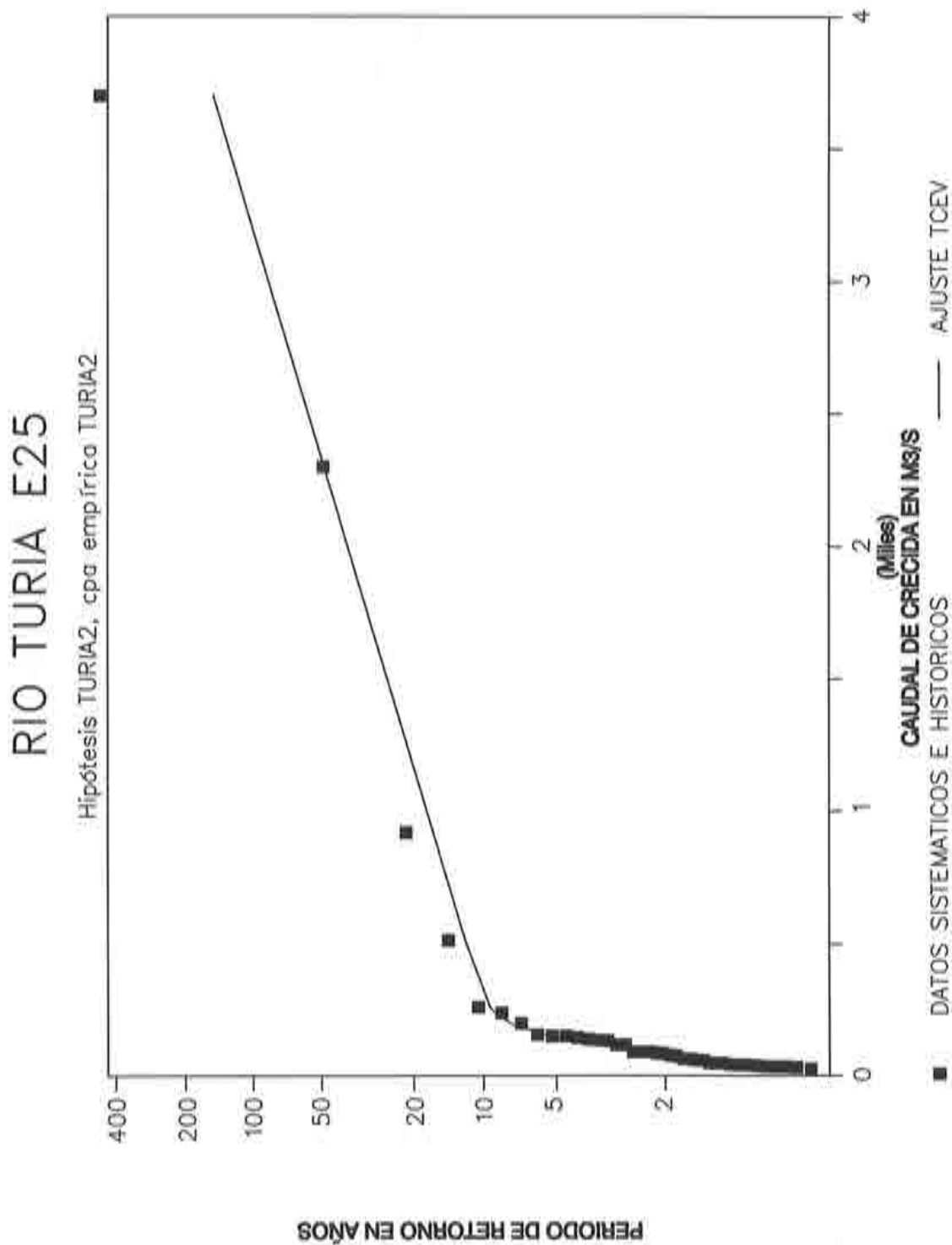


Figura 6.11 Ajuste de una TCEV con los datos sistemáticos e históricos (hipótesis larga) del río Turia. F.d.a. incorporando la información histórica con la hipótesis larga.

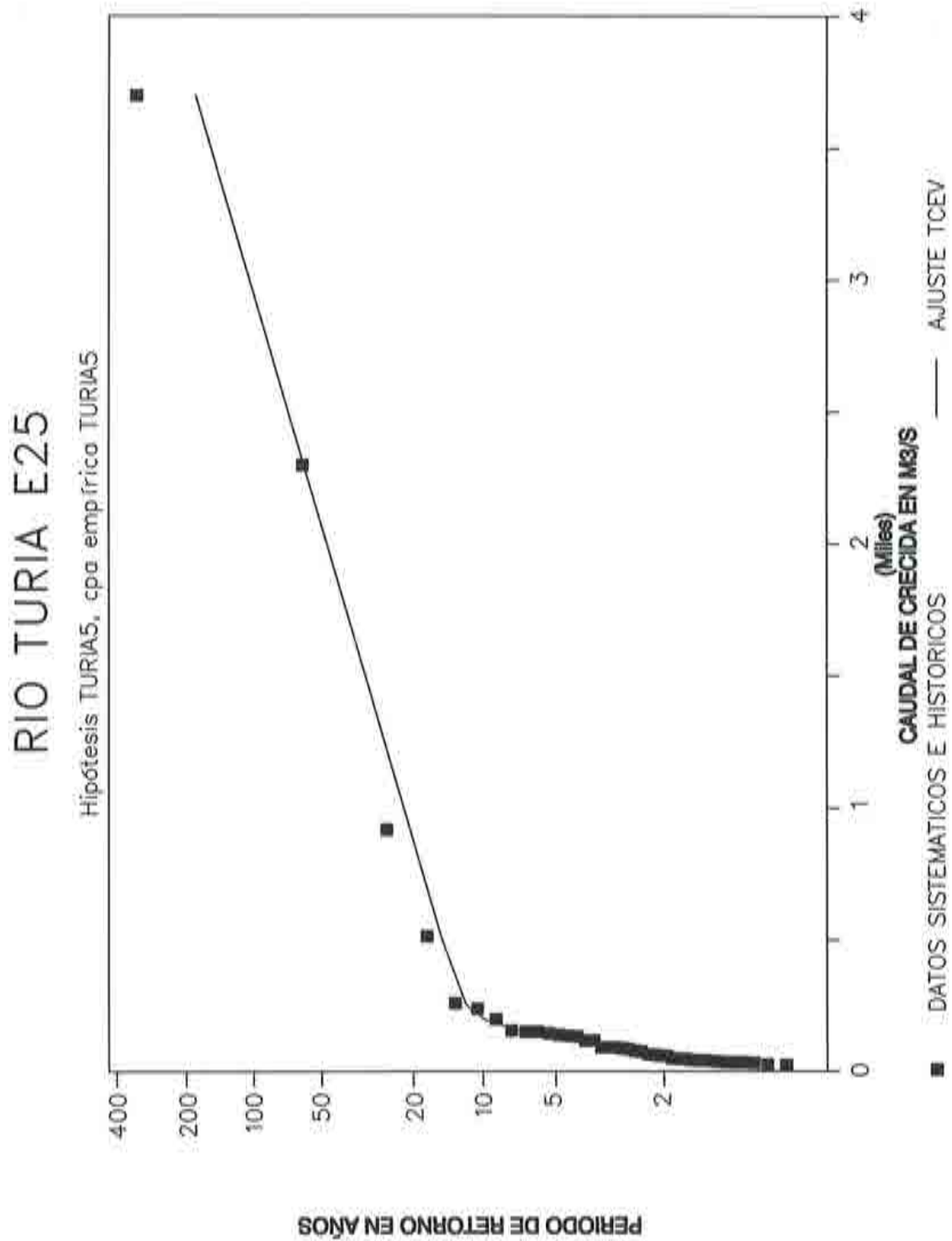


Figura 6.12 Ajuste de una distribución TCEV con los datos sistemáticos e históricos (hipótesis corta) del río Turia, F.d.a. incorporando la información histórica según la hipótesis corta.

6.5.- ANALISIS CONJUNTO DE LOS RIOS JUCAR Y TURIA

Como se describe en el Apartado 6.5, ha sido imposible analizar conjuntamente los ríos Júcar, Turia y Segura por incompatibilidad de homogeneidad regional de este último.

En primer lugar se ha realizado la estandarización de las series sistemáticas mediante el ajuste de una función Gumbel a las avenidas ordinarias. Los parámetros de la estandarización se encuentran en la Tabla (6.8).

Tabla 6.8. Estandarización de los ríos Júcar y Turia.

	Júcar E-89	Turia E-25
Longitud serie	42	41
"outliers"	2	4
$\lambda=$	3.725	4.874
$\theta=$	0.006086	0.02477

Del análisis individual de los ríos Júcar y Turia se deduce que la información histórica es necesaria para una buena estimación de la rama derecha de la distribución. Además, esa información histórica del río Turia es la correspondiente a la hipótesis corta.

El ajuste de una función TCEV a los datos regionales y su posterior desestandarización se resume en la Tabla (6.9) y en las Figuras (6.13) y (6.14).

Tabla 6.9. Parámetros y cuantiles estimados para los ríos Júcar y Turia con la hipótesis de análisis regional.

Parámetros por M.V.	TCEV Júcar	TCEV Turia
$\lambda_1=$	3.799	5.028
$\theta_1=$	0.006350	0.02584
$\lambda_2=$	0.09460	0.09544
$\theta_2=$	0.0002003	0.0008152
Período de retorno	Cuantiles en m ³ /s	
10	793.24	205.77
50	7707.6	1904.7
100	11193.	2761.2
500	19248.	4740.4
1000	22711.	5591.3

Para el río Júcar, se considera que el ajuste regional es bueno, aunque presenta ciertas diferencias con el ajuste local:

1.- El punto de inflexión de la curva TCEV desciende de 14 a 11 años de período de retorno, lo que origina un incremento del valor de los cuantiles importante de alrededor del 25% para períodos de retorno medios (50 a 100 años).

2.- Un incremento de la pendiente de la rama derecha que compensa en parte el efecto anterior para períodos de retorno altos, dejando el incremento alrededor del 15%.

En cuanto al río Turia, el análisis regional mejora el ajuste de la función TCEV respecto al análisis local, como se observa comparando las Figuras (6.12) y (6.14). Las

diferencias son:

- 1.- El punto de inflexión de la curva TCEV asciende de 9 a 11 años de período de retorno. Ello da lugar a un descenso muy importante de los cuantiles de 10 años de período de retorno por cambiar de rama de la curva TCEV. Para períodos de retorno medios se produce una disminución en torno al 10%.
- 2.- Aumento ligero de la pendiente de la rama derecha, que para largos períodos de retorno da lugar a disminuciones de los cuantiles del 4%.
- 3.- En el análisis regional, la avenida de 1957 tiene un período de retorno de 214 años, mientras que con el análisis local es de solo 171 años. El resultado regional está más de acuerdo con el hecho de considerar la avenida de 1957 como la mayor sufrida por la ciudad de Valencia en todo el período considerado.

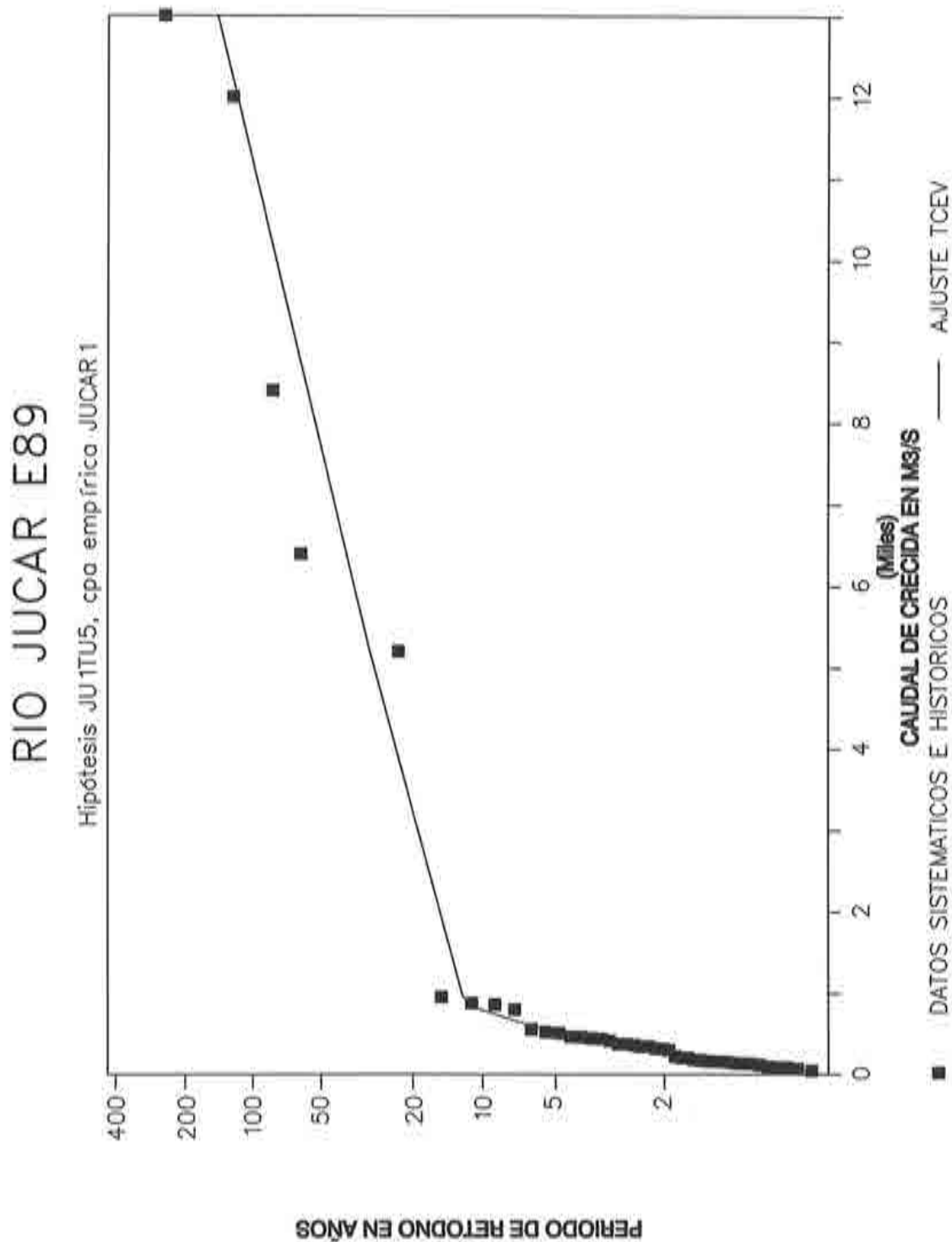


Figura 6.13 Ajuste de una distribución TCEV al río Júcar en el análisis regional, F.d.a. empírica considerando la información histórica local.

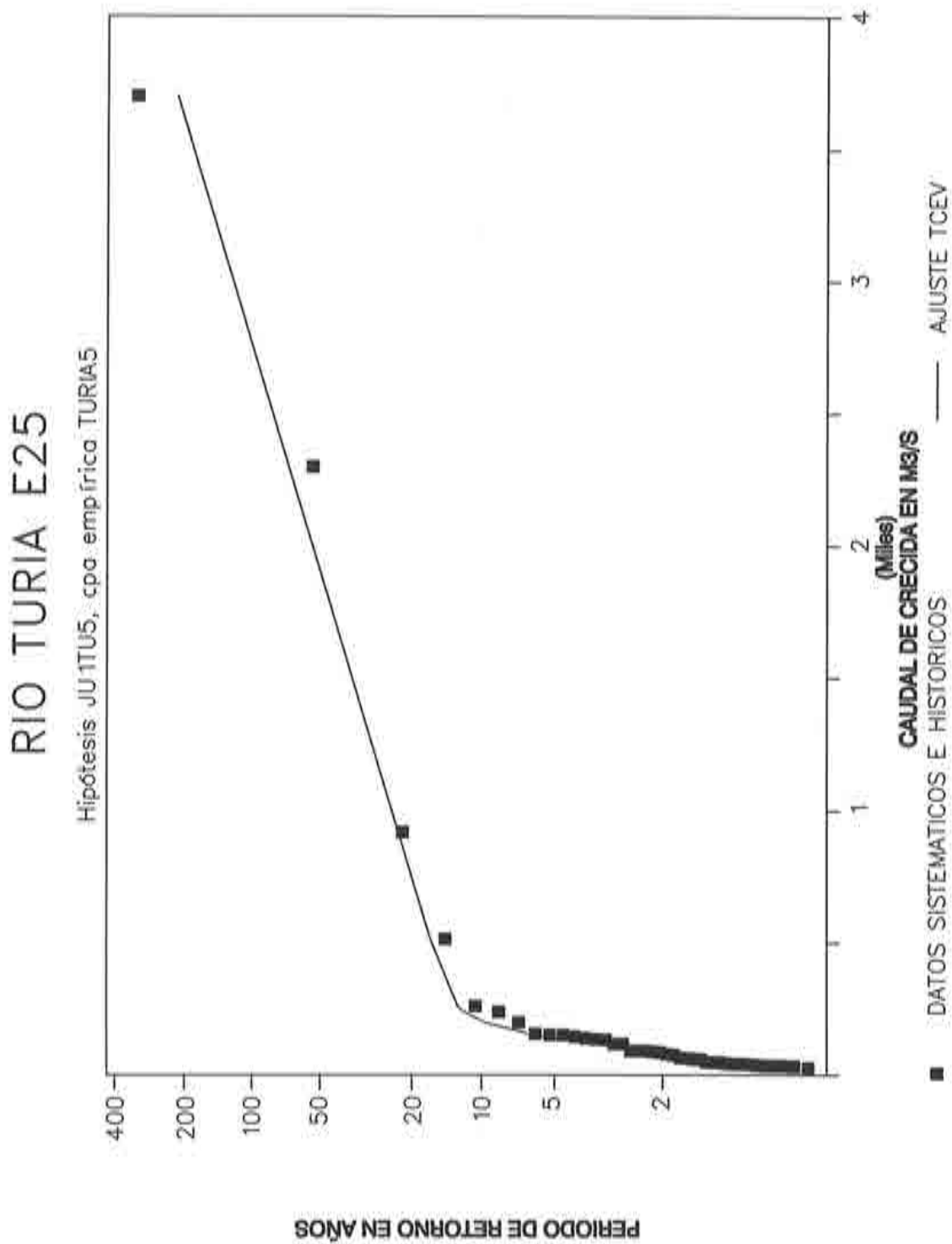


Figura 6.14 Ajuste de una distribución TCEV al río Turia en el análisis regional. F.d.a. empírica considerando la información histórica local según la hipótesis corta.

6.6.- ANALISIS LOCAL DEL RIO SEGURA

Dado que no existía prácticamente datos de máximos instantáneos anuales, el análisis del río Segura se ha realizado utilizando la serie sistemática de máximos diarios. El ajuste de una función Gumbel a estos datos (Tabla 6.10) es desastroso, según se observa en la Figura (6.15).

El ajuste de una TCEV a los datos sistemáticos (Tabla 6.10) no es bueno, incluso si se compara con la f.d.a. empírica de los datos sistemáticos (Figura 6.16). El error cometido en la rama derecha de la función de distribución es debido a que solo se tienen tres datos sistemáticos muy próximos entre ellos.

Se hace obligatorio el uso de mayor cantidad de información estadística. Si se utiliza la información histórica, el ajuste de la rama derecha mejora en gran medida, según se observa en la Figura (6.17). Los resultados se acompañan en la Tabla (6.10).

Sin embargo, el intento de análisis regional de los tres ríos resultó infructuoso (Tabla 6.10 y Figura 6.18). Los cuantiles estimados utilizando exclusivamente las series sistemáticas de los tres ríos son muy inferiores a los obtenidos con la información histórica local del Segura. Esto podría deberse a un error en la información histórica del río Segura, pero dejaría a las avenidas de los años 40 con $900 \text{ m}^3/\text{s}$ de magnitud, con un período de retorno de solo 18 años, cosa que es excesivamente frecuente. La justificación que se considera como más realista es el diferente comportamiento de la cuenca del Segura, frente a las cuencas del Júcar o Turia.

Las avenidas extraordinarias del río Segura son menores, en términos de la serie estandarizada, y más frecuentes que las del Júcar y Turia, lo que se refleja en los valores de los parámetros λ_2 (frecuencia) y θ_2 (magnitud). La explicación física sería la existencia de una cuenca más impermeable, que daría lugar a un umbral de escorrentía mucho menor.

Tabla 6.10. Parámetros y cuantiles estimados para el río Segura con diversas hipótesis de análisis local y regional.

Parámetros por M.V.	Gumbel sistemática	TCEV sistemática	TCEV histórica	TCEV regional sistemática
$\lambda_1=$	2.660	4.827	3.258	3.234
$\theta_1=$	0.01017	0.05608	0.01687	0.01460
$\lambda_2=$	0	1.386	0.2540	0.09036
$\theta_2=$	0	0.006830	0.002492	0.0005314
Período de retorno	Cuantiles en m ³ /s			
10	317.58	377.25	375.64	322.40
50	480.04	619.07	1015.8	2818.9
100	548.7	721.31	1295.9	4132.9
500	707.43	957.55	1943.4	7169.1
1000	775.66	1059.1	2221.7	8474.4

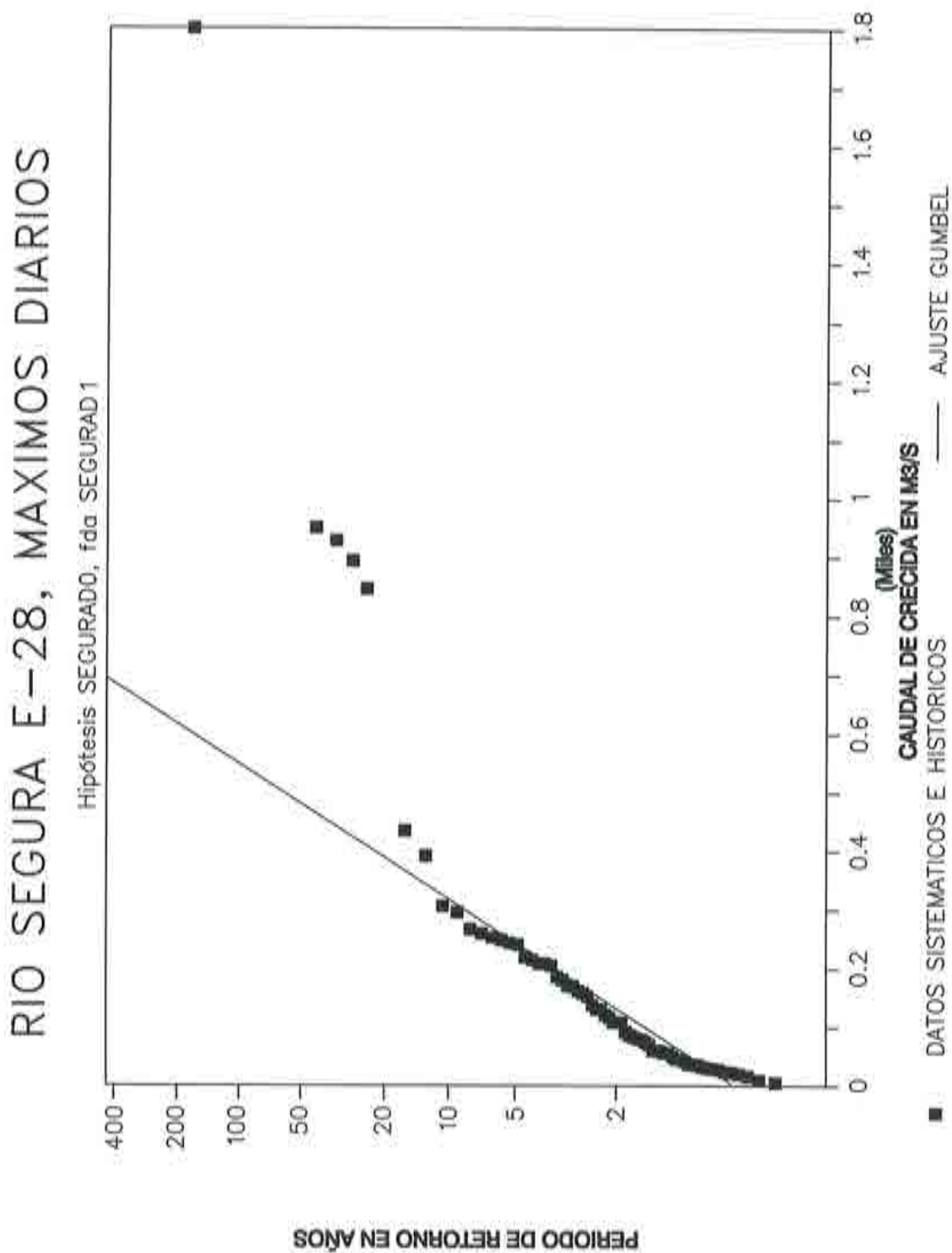


Figura 6.15 Ajuste de una distribución Gumbel al río Segura con la serie sistemática exclusivamente. F.d.a. empírica considerando la información histórica.

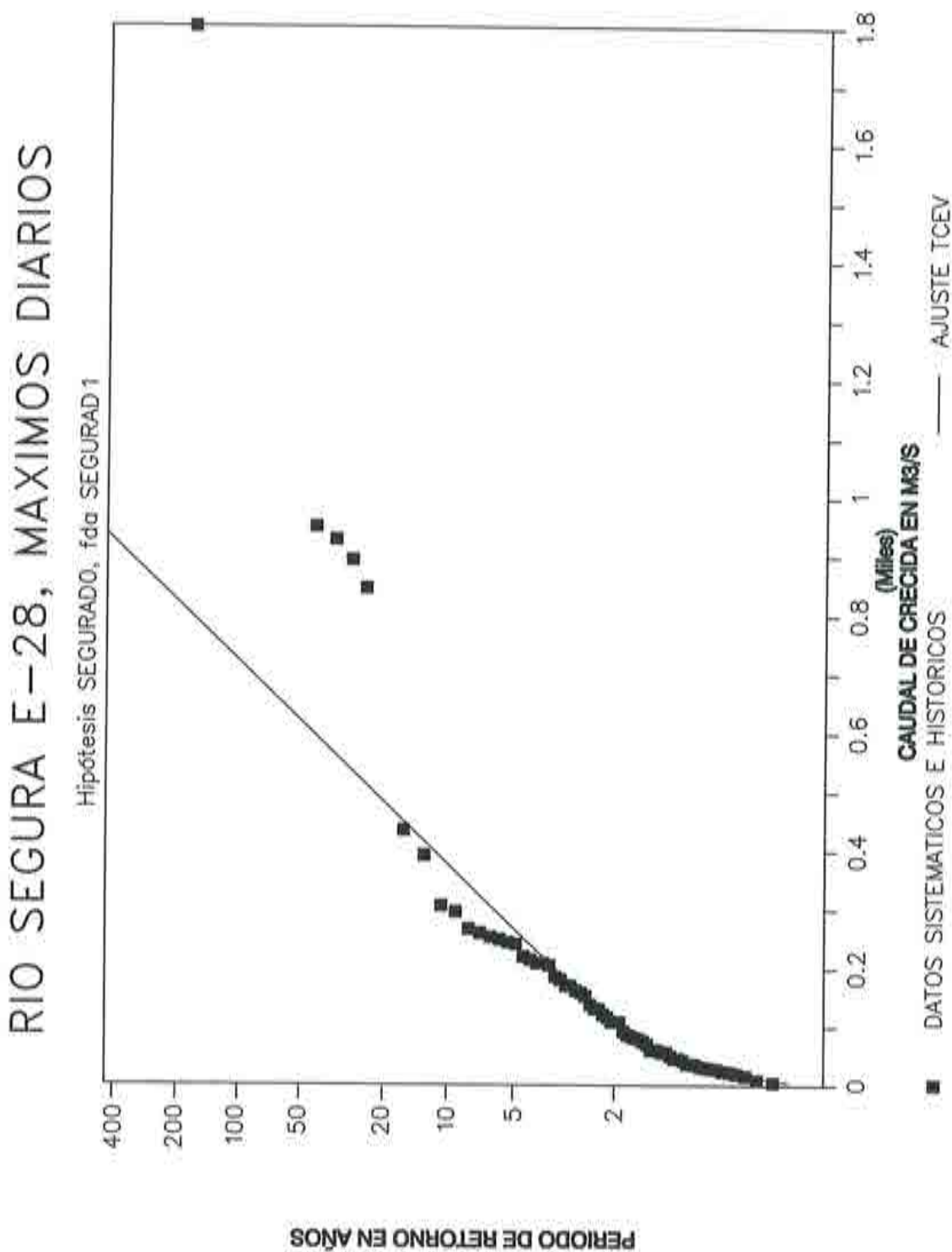


Figura 6.16 Ajuste de una distribución TCEV al río Segura con la serie sistemática exclusivamente. F.d.a. empírica considerando la información histórica.

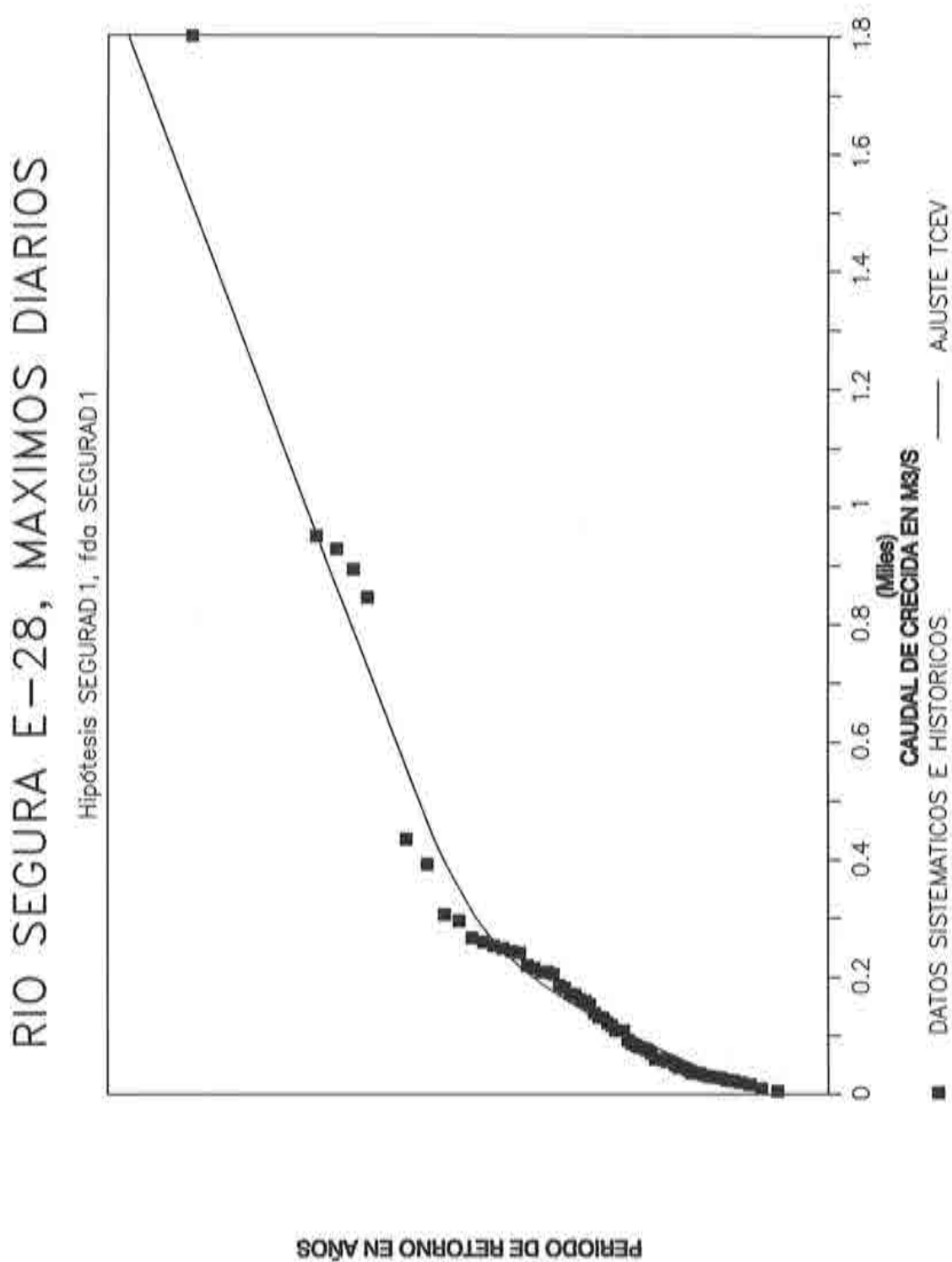


Figura 6.17 Ajuste de una distribución TCEV al río Segura con datos sistemáticos e históricos. F.d.a. empírica considerando la información histórica.

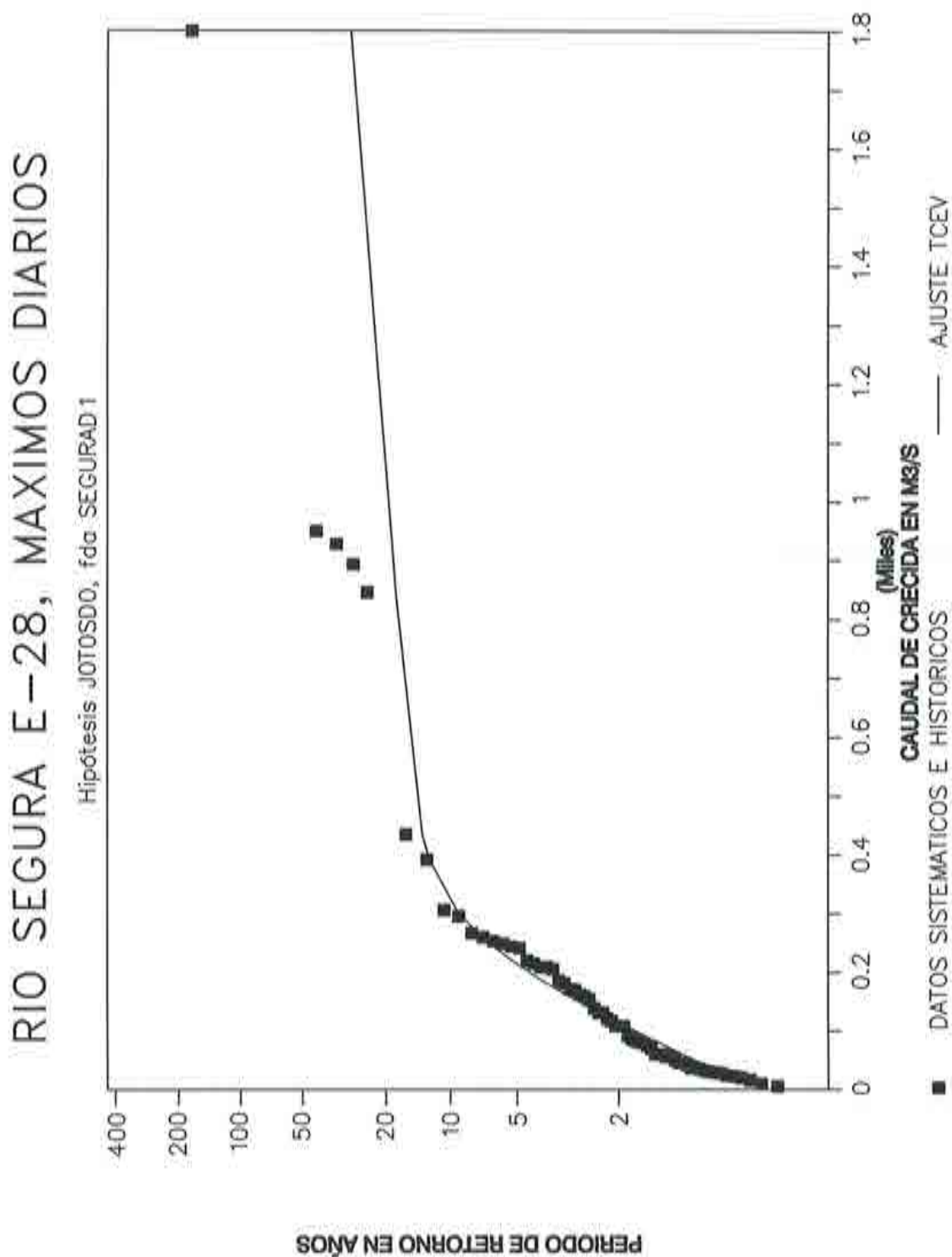


Figura 6.18 Ajuste de una distribución TCEV al río Segura dentro de un análisis regional de los datos sistemáticos exclusivamente. F.d.a. empírica considerando la información histórica local.

6.7.- SOLUCION ADOPTADA

Los parámetros estimados de las distribuciones TCEV adoptados en los tres ríos, se adjuntan en la Tabla (6.11).

Tabla 6.11. Parámetros de la distribución TCEV adoptados en cada río.

Parámetros por M.V.	JUCAR	TURIA	SEGURA
$\lambda_1=$	3.799	5.028	3.258
$\theta_1=$	0.006350	0.02584	0.01687
$\lambda_2=$	0.09460	0.09544	0.2540
$\theta_2=$	0.0002003	0.0008152	0.002492

Los parámetros del río Segura son los obtenidos mediante un análisis local con la inclusión de la información histórica, mientras que los del Júcar y Turia se han obtenido mediante un análisis regional incluyendo también la información histórica disponible. Del río Turia se ha utilizado el periodo histórico corto, por no poderse garantizar la constancia del umbral de percepción y dar un ajuste relativamente peor.

Tan importante como el conocer el valor de los cuantiles para cada periodo de retorno, es tener una idea de la precisión de la estimación. La mayor o menor amplitud del intervalo de confianza es una buena medida de esta precisión. Si se asume que el estimador del cuantil X_T sigue una distribución normal, los límites del intervalo de confianza del 100(1- α)% vienen dados por:

$$L_{1-\alpha}(X_T) = X_T - \sqrt{\text{Var}(X_T)} u_{1-\alpha/2} \quad (2)$$

$$U_{1-\alpha}(X_T) = X_T + \sqrt{\text{Var}(X_T)} u_{1-\alpha/2} \quad (3)$$

donde:

$L(\cdot)$ = Límite inferior del intervalo

$U(\cdot)$ = Límite superior del intervalo

$u_{1-\alpha/2}$ = Cuantil de probabilidad de no excedencia 1- $\alpha/2$ de la distribución normal

estandarizada.

Se puede aproximar la varianza del estimador por su varianza asintótica. Si la estimación es local, se podrá hacer uso de las diferentes ecuaciones de los CRLBs vistas en el Capítulo 5 para obtener la varianza asintótica, en función del tipo de información histórica utilizada.

Si la estimación es regional, habrá que tener en cuenta la estandarización previa de las series según se desarrolla en el último apartado del Capítulo 5. Los cuantiles correspondientes a diversos periodos de retorno, así como los intervalos de confianza del 95% obtenidos con las ecuaciones (6.2) y (6.3) se presentan en la Tabla (6.12).

Tabla 6.12. Cuantiles e intervalos de confianza del 95%.

Período de retorno	Júcar E-89		Turia E-25		Segura E-28 (d)	
	Cuantil (m ³ /s)	Intervalo (m ³ /s)	Cuantil (m ³ /s)	Intervalo (m ³ /s)	Cuantil (m ³ /s)	Intervalo (m ³ /s)
5	522,11	±139,20	139,14	±34,97	221,97	±62,82
10	793,23	±337,93	205,77	±83,82	375,63	±195,46
15	1587,4	±2724,4	400,81	±669,93	525,61	±253,35
20	3055,9	±2795,0	761,70	±688,57	642,38	±274,64
40	6580,9	±3313,6	1627,8	±821,47	925,22	±380,73
50	7707,6	±3668,9	1904,7	±910,57	1015,7	±427,90
75	9748,8	±4446,1	2406,2	±1104,5	1179,8	±521,83
100	11193,	±5063,9	2761,2	±1258,2	1295,9	±592,54
500	19248,	±8935,7	4740,4	±2219,3	1943,4	±1014,2
1000	22711,	±10692,	5591,3	±2654,9	2221,8	±1201,8
10000	34208,	±16643,	8416,2	±4130,1	3145,8	±1832,8

Gráficamente se reflejan estos resultados en las Figuras (6.19), (6.20) y (6.21). El ensanchamiento brusco del límite inferior en el "codo" de la TCEV puede ser debido a:

- 1.- La no normalidad del estimador del cuantil, que podría presentar un cierto sesgo positivo.
- 2.- Para los ríos Júcar y Turia, la hipótesis de independencia entre el estimador del cuantil regional y de los parámetros de estandarización se hace tanto más incorrecta cuanto menor sea el periodo de retorno.

Por último, habría que resaltar el incremento de fiabilidad en los estimadores a medida que se incorpora información histórica o regional en el análisis. Esto se refleja en la Tabla (6.13) para el cuantil de 1000 años de periodo de retorno.

Tabla 6.13. Evolución del intervalo de confianza del 95% del cuantil X_{1000} de una TCEV, para diversos grados de información.

	Análisis local	An. local con info. histórica	An. regional con info. histórica
Júcar E-89	± 24665	± 12711	± 10692
Turia E-25	± 6133.9	± 3251.4	± 2654.9
Segura E-28 (d)	± 1503.5	± 1201.8	-

Se puede observar en dicha tabla que el aumento de fiabilidad en la estimación es espectacular con la incorporación de la información histórica, siendo algo menor al realizar el análisis regional. Con ello se hace prácticamente obligatorio, si se encuentra disponible como es este caso, el empleo de la información histórica.

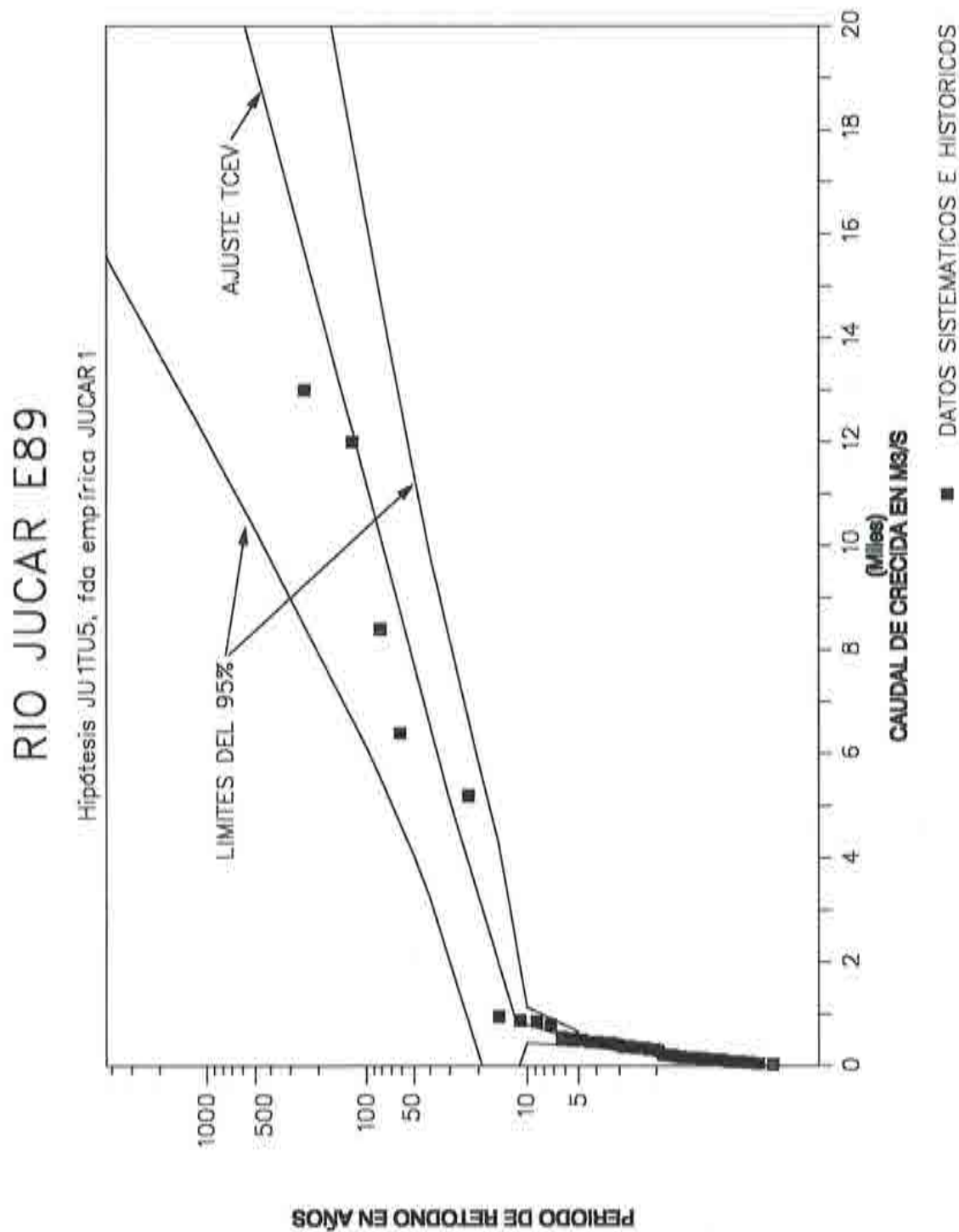


Figura 6.19 Ajuste adoptado y limites de confianza del 95 % para el río Júcar.

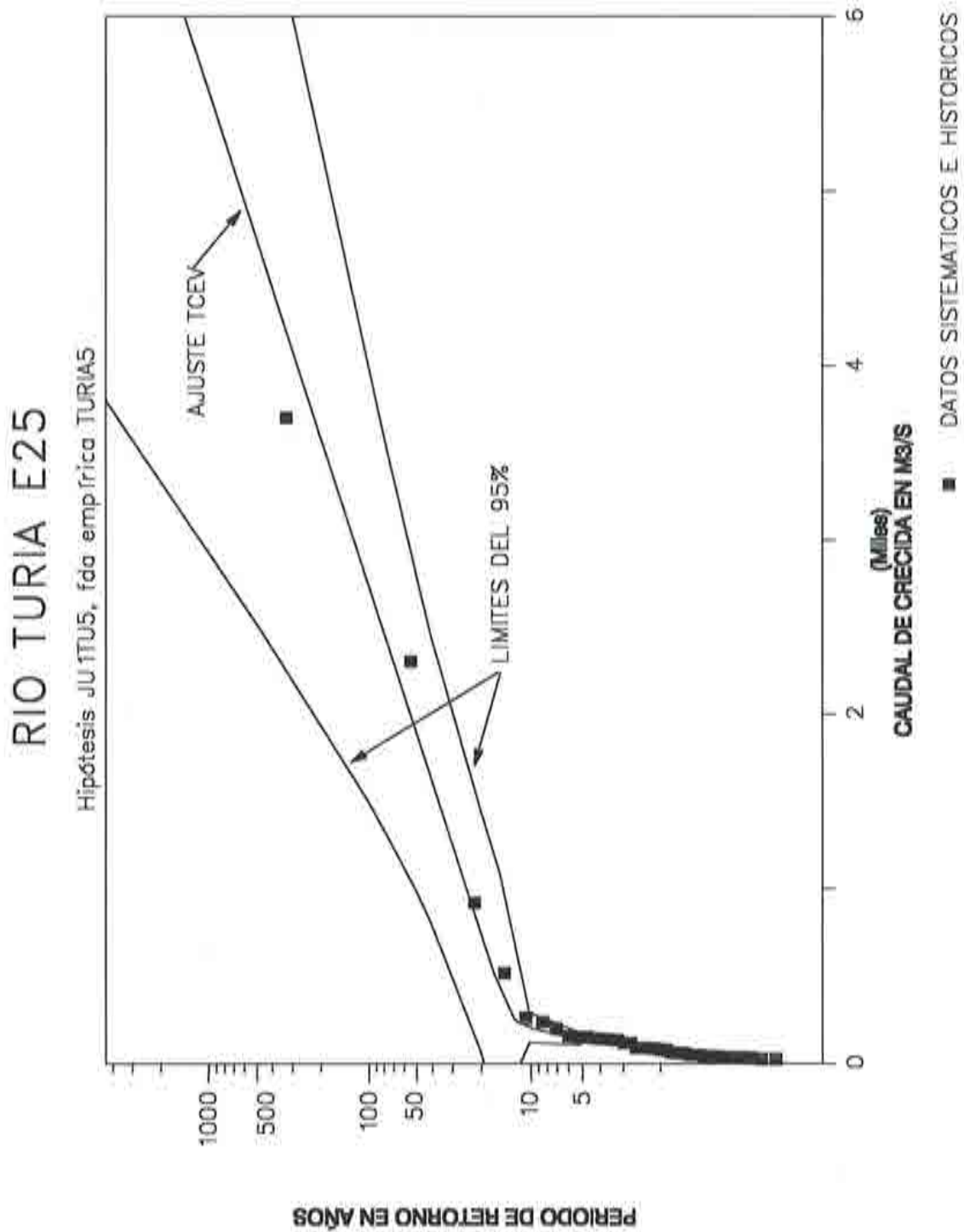


Figura 6.20 Ajuste adoptado y limites de confianza del 95 % para el rio Turia.

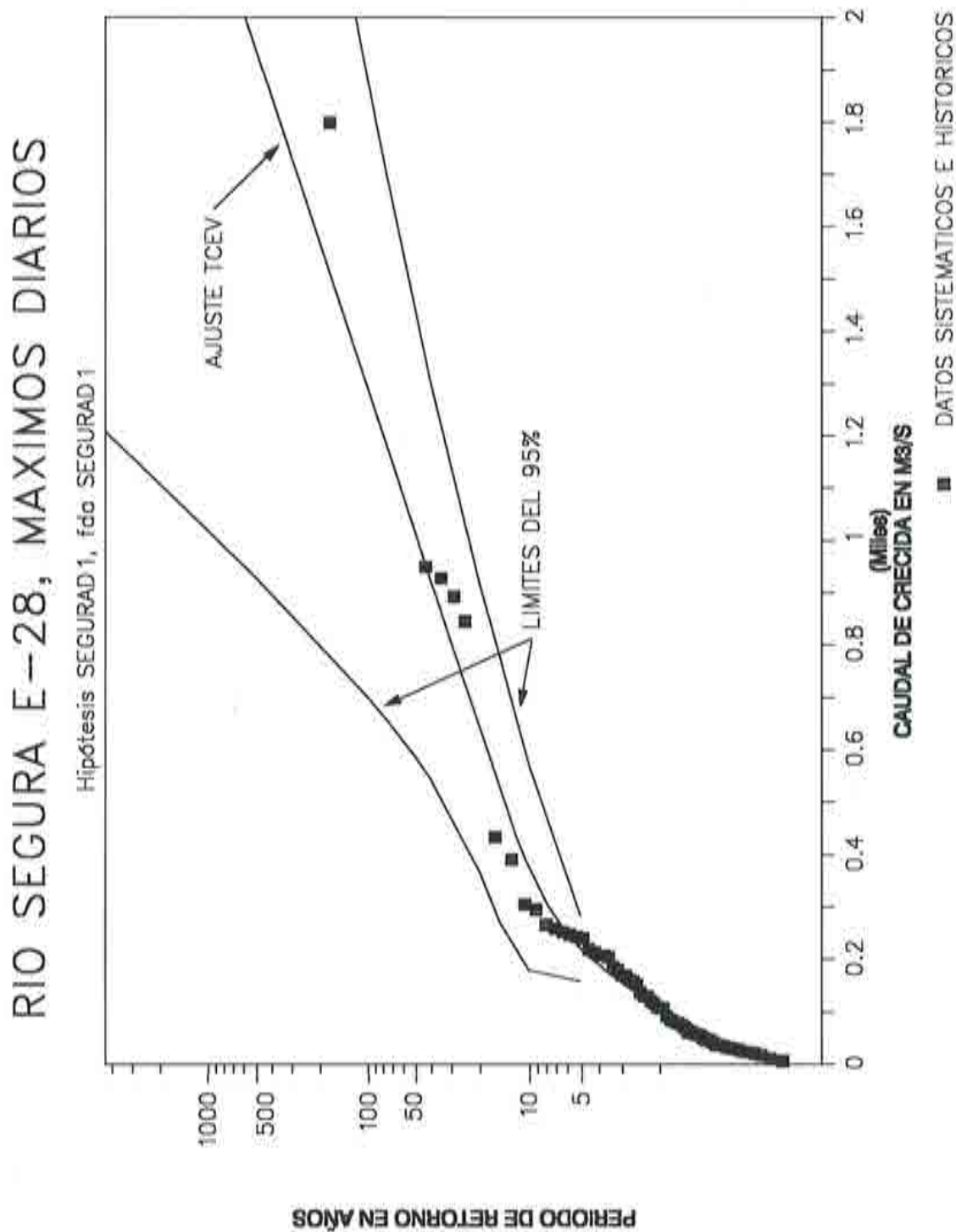


Figura 6.21 Ajuste adoptado y limites del 95 % para el rio Segura.

7.- SUMARIO

Los objetivos del presente trabajo se centran en las posibilidades de mejoras en la metodología clásica del análisis de la frecuencia de las avenidas de los ríos. Las principales aportaciones originales realizadas en este trabajo de investigación son:

- i) El desarrollo de modelos estadísticos para la familia de distribuciones "Extreme Value" (EV) y la distribución "Two Component Extreme Value" (TCEV) capaces de incorporar diferentes tipos de información histórica. Para el caso de la distribución TCEV además es posible enmarcar dicha incorporación dentro de un análisis regional del problema. Estos modelos son generalizables a cualquier otra distribución.
- ii) Haber explorado las ventajas de la utilización de la información histórica adicionalmente al uso de la serie sistemática de datos de aforos, obteniendo propiedades muy interesantes que pueden orientar el uso de dicha información.
- iii) La aplicación a los ríos Júcar, Turia y Segura.

El análisis de la frecuencia de las avenidas consiste básicamente en obtener los cuantiles correspondientes a diversos periodos de retorno. La estimación de estos cuantiles es altamente variable debido a la presencia de errores, utilización de periodos de retorno mucho mayores que la longitud de la serie sistemática, variabilidad muestral de las poblaciones de avenidas anuales y la posible existencia de múltiples mecanismos de generación de las crecidas de los ríos.

Por tanto, la mejora de la fiabilidad de los estimadores de los cuantiles se convierte en fundamental. Un aumento de la fiabilidad puede conseguirse incrementando la información disponible mediante un análisis regional y/o la incorporación de información histórica. El análisis regional consiste en considerar, después de realizar algún tipo de transformación estadística, que los datos de diferentes estaciones de aforos provienen de una misma superpoblación. La información histórica es el conocimiento que se tiene acerca de avenidas anteriores al registro sistemático de datos. Este conocimiento puede provenir de "la memoria humana" (libros, periódicos, marcas, etc...) o de estudios geomorfológicos sobre las paleoavenidas. Desde un punto de vista estadístico no existen diferencias entre estos dos posibles orígenes.

De la revisión del estado del arte y, sobre todo, de la experiencia adquirida, se puede clasificar la información histórica en tres tipos:

- i) Información histórica binomial censurada. Durante el periodo histórico se conoce el número de avenidas que han superado un determinado umbral de percepción.

ii) Información histórica censurada. Además de conocer el número, se conoce el valor de las avenidas que superaron el mencionado umbral.

iii) Información histórica de máxima avenida. Lo único que se conoce es el valor de la mayor crecida ocurrida durante el período histórico.

La información histórica a utilizar puede encontrarse de alguna de estas formas, o ser transformada ficticiamente a alguna de estas formas. Información ficticia puede ser debida al hecho de emplear un umbral de percepción mayor del real, o despreñar parte de la información disponible (por ejemplo, disponer de información censurada y utilizarla como de máxima avenida).

Los modelos estadísticos utilizados se basan en el Método de Estimación por Máxima Verosimilitud, que para series de datos de longitud media y larga se muestra como el método más robusto, especialmente frente al más conocido Método de los Momentos.

La investigación comienza con la familia EV, a la que pertenecen las distribuciones Gumbel, LogGumbel y Weibull, como un paso previo a la distribución TCEV. Esta última, de reciente creación y mucho más completa que las anteriores, era el objetivo final por ser capaz de reproducir los efectos "pata de perro" y "separación", presentes en los ríos mediterráneos.

La utilización de la información histórica se traduce en una disminución del error cuadrático medio del estimador del cuantil. Este error cuadrático medio es necesario si se quieren establecer los límites de confianza de los cuantiles estimados. En esta investigación se ha aproximado de dos formas: analíticamente a partir de la varianza asintótica o Límite Inferior de Cramer-Rao, y mediante simulaciones de Montecarlo, obteniendo el error cuadrático medio muestral.

El análisis de las ventajas de la utilización de la información histórica se ha realizado a través de la denominada ganancia estadística. La ganancia estadística representa la mejora, en términos del error cuadrático medio, al utilizar algún tipo de información histórica frente al uso exclusivo de la serie sistemática. Si se parte de la varianza asintótica se obtiene la ganancia estadística asintótica. Si se utiliza el error cuadrático medio muestral se llega a una ganancia estadística muestral. Aunque realmente esta última se trata de una variable aleatoria, se ha considerado como más aproximada a la verdadera ganancia estadística. El problema estriba en que su obtención requiere un consumo importante de tiempo de ordenador, con lo que es difícil obtener sus propiedades. Estas propiedades se consideran fundamentales para el conocimiento de la influencia del uso de la información histórica en el aumento de

fiabilidad de los estimadores de los cuantiles.

Afortunadamente se ha demostrado que las expresiones asintóticas son una buena aproximación de las expresiones muestrales y se pueden generalizar las propiedades de las ganancias estadísticas asintóticas a las verdaderas ganancias estadísticas. El conocimiento de la varianza asintótica permite también el establecer límites de confianza de los estimadores.

La mayor parte de estas propiedades son comunes en todos los casos y se pueden resumir en:

- i) Con las hipótesis consideradas (reflejadas en el Apartado 3.2) siempre se incrementa la fiabilidad de los estimadores de los cuantiles utilizando cualquier tipo de información histórica adicional. Si se utiliza información histórica binomial censurada o censurada, la ganancia estadística supera el 50 %, para las situaciones habituales en Hidrología.
- ii) Para la familia de distribuciones EV, las ganancias estadísticas son independientes de los parámetros de la distribución correspondiente, es decir, independientes de las características estadísticas de la población. Por otra parte, en la distribución TCEV depende de los dos parámetros estandarizados en vez de los cuatro parámetros de la misma (esta dependencia es muy ligera para el caso de información censurada).
- iii) El principal factor en la variación de la ganancia estadística con información binomial censurada o censurada es el umbral de percepción histórico. Para el caso de la información histórica de máxima avenida el factor más influyente es la longitud del período histórico.
- iv) En la mayor parte de los casos que pueden presentarse en la realidad, la información acerca de la máxima avenida es la más pobre de las tres, por lo que solo debe usarse cuando sea el único tipo de información disponible.
- v) Siempre la ganancia estadística con información censurada es la mayor. Aunque para la familia de distribuciones EV salvo que exista un límite de percepción extraordinariamente bajo, apenas existen diferencias entre las ganancias estadísticas utilizando información censurada y binomial censurada. La diferencia es tanto menor cuanto mayor sea el umbral de percepción. Por tanto, teniendo en cuenta que los errores pueden ser mayores con el primer tipo de información, se recomienda el uso de la información como binomial censurada.

vi) Un efecto similar, aunque menor, ocurre con la distribución TCEV. En este caso solo con escasez de información histórica (umbral de percepción alto) puede ser recomendable el uso de la información como binomial censurada en vez de censurada.

vii) Dado el comportamiento asintótico de las ganancias estadísticas cuando crece la longitud del período histórico, no interesa su prolongación mediante la realización de estudios sobre paleoavenidas si existe información histórica binomial censurada o censurada.

viii) Dentro de un análisis regional con la distribución TCEV, la información histórica de una de las estaciones de aforos se transmite al resto de estaciones, especialmente para períodos de retorno medios y altos.

Como una demostración de la posible aplicación de todo lo desarrollado, se analizaron las avenidas de los ríos Júcar, Turia y Segura. En estos ríos resultó desastroso el ajuste clásico de una distribución Gumbel, por dos razones:

i) La no reproducción del efecto "pata de perro", observable en la comparación gráfica entre las funciones de densidad de probabilidad acumulada ajustada y empírica. Este fenómeno es debido al origen dual de las avenidas del Levante español, que da lugar a una población de avenidas ordinarias y otra de extraordinarias.

ii) La existencia de coeficientes de sesgo (en los datos sistemáticos utilizados en los ríos Júcar, Turia y Segura 5.26, 4.39 y 2.82 respectivamente) muy alejados del valor de 1.14 correspondiente a la distribución Gumbel.

En el otro extremo, el empleo de la distribución TCEV se puede considerar como un éxito.

Los ríos Júcar y Turia pudieron ser estudiados conjuntamente, a partir de las series de caudales máximos instantáneos anuales. Los máximos diarios, a pesar de ser más abundantes, no pudieron utilizarse por no adecuarse correctamente a ninguna de las distribuciones probadas. Del río Júcar se disponía de un umbral ficticio de percepción de período de retorno 37 años aproximadamente. Al tratarse de un umbral bajo, se empleó la información censurada disponible. En el río Turia el umbral de percepción era más alto (69 años de período de retorno) con lo que la información utilizada fue binomial censurada.

El río Segura no resultó homogéneo con los anteriores, por lo que hubo de ser

analizado individualmente. Además, se empleó como variable estadística el caudal máximo diario, al ser la serie de máximos instantáneos anuales excesivamente corta. Aunque el límite de percepción es de 29 años, los valores de las avenidas se obtienen en la ciudad de Murcia, aguas arriba de la posición de la estación de aforos de Orihuela, por lo que la información histórica utilizada en este río fue del tipo binomial censurada.

Los resultados son espectaculares. Para el análisis regional con información histórica de los ríos Júcar y Turia la disminución de los intervalos de confianza del estimador del cuantil de 1000 años de período de retorno es del 57%, frente al análisis local con la serie sistemática exclusivamente. Para el estudio individual de río Segura la disminución al incorporar la información histórica es del 20%. Se hace por tanto obligatorio el uso de la información histórica disponible en la estimación de los cuantiles.

Ninguna investigación en su conclusión puede haber cubierto todos los aspectos del problema tratado. Futuras líneas de trabajo que no han sido cubiertas podrían ser:

- i) Analizar la posible generalización de las propiedades encontradas sobre la información histórica a otras distribuciones. Puede ocurrir que, por ejemplo, la independencia de las ganancias estadísticas respecto de los parámetros de las distribuciones de la familia EV sea una propiedad común a todas las distribuciones de dos parámetros.
- ii) No se ha considerado la existencia de errores tanto en los datos de la serie sistemática como en la información histórica. Habría que investigar cual es la estructura de estos errores y en que medida influyen en la utilización de la información histórica.
- iii) En análisis regional, la existencia de fuertes correlaciones entre las estaciones de aforos puede empobrecer los resultados, si bien esto no se ha estudiado hasta la actualidad para la distribución TCEV y menos considerando la incorporación en todas o en parte de las estaciones de algún tipo de información histórica.
- iv) La incorporación de un límite superior representado por la avenida máxima probable (PMF) debe de incrementar la fiabilidad de los estimadores de alto período de retorno. Sin embargo, el problema radica en realizar esta incorporación de información adicional con funciones de distribución como la TCEV que no se encuentran limitadas por la derecha.

8.- REFERENCIAS

- Acreman, M.C., C.D. Sinclair, 1986. "Classification of Drainage Basins According to their Physical Characteristics; an Application for Flood Frequency in Scotland". *Journal of Hydrology*, v. 84, p. 365-380.
- Adamowski, K. and W. Feluch, 1990. "Non-parametric Flood Frequency Analysis with Historical Information". *ASCE Jour. of Hydr. Eng.*, 116(e), 1035-1047.
- Almela y Vives, F., 1957. "Las Riadas del Turfa (1321-1949)". Publicaciones del Archivo Municipal Excmo. Ayuntamiento de Valencia.
- Arnell, N., M. Beran, J.R. Hosking, 1986. "Unbiased Plotting Positions for the General Extreme Value Distribution". *Journal of Hydrology*, v. 86, p. 59-69.
- Arnell, N.W., S. Gabriele, 1988. "The Performance of the Two-Component Extreme Value Distribution in Regional Flood Frequency Analysis". *Water Res. Res.*, v. 24, p. 879-887.
- Baker, V., G. Pickup, H.A. Polach, 1983. "Desert Paleofloods in Central Australia". *Nature*, v. 301, p. 502-504.
- Bautista Martín, J., 1989. "Las Avenidas en la Cuenca del Segura y los Planes de Defensa". Reunión Científica Internacional sobre Avenidas Fluviales e Inundaciones en la Cuenca del Mediterráneo, Universidad de Alicante.
- Benson, M.A., 1950. "Use of Historical Data in Flood Frequency Analysis", *Eos, Trans. AGU*, v. 31(3), p. 419-424.
- Boes, D.C., J. Heo, J.D. Salas, 1989. "Regional Flood Quantile Estimation for a Weibull Model". *Water Res. Res.*, v. 25, p. 979-990.
- Bratley, P., B.L. Fox, L.E. Schrage, 1987. "A Guide to Simulation". Springer-Verlag, 397 p.
- Calvo, F., 1989. "Grandes Avenidas e Inundaciones Históricas". Reunión Científica Internacional sobre Avenidas Fluviales e Inundaciones en la Cuenca del Mediterráneo, Universidad de Alicante.
- Canales Martínez, G., 1989. "Las Inundaciones en la Vega Baja del Segura (1875-1925)". Reunión Científica Internacional sobre Avenidas Fluviales e Inundaciones en la Cuenca del

Mediterráneo, Universidad de Alicante.

- Carmona González, P., 1986. "El Cuaternario de los Mantos Aluviales del Bajo Turia". Tesis Doctoral de la Facultad de Geografía e Historia de Valencia.
- Carmona González, P., 1990. "La Formació de la Plana Al.luvial de València". Institució Valenciana d'Estudis i Investigació.
- Centro de Estudios Hidrográficos, 1983. "Actualización de los Estudios de Desagüe de la Autopista Tarragona-Alicante, Tomo IV: Cruce con el río Júcar". MOPU.
- Cohn, T.A., J.R. Stedinger, 1987. "Use of Historical Information in a Maximum Likelihood Framework". *Journal of Hydrol.*, v. 96, p. 215-233.
- Comisaría de Aguas de la C.H.J., 1983. "Estudio de Máximas Avenidas y de Sequías de la Cuenca Hidrográfica del Júcar". Confederación Hidrográfica del Júcar.
- Comisión Nacional de Protección Civil, 1983. "Estudio de Inundaciones Históricas. Mapa de Riesgos Potenciales de la Cuenca del Júcar".
- Condie, R., 1986. "Flood Samples from a Three-Parameter Lognormal Population with Historic Information: the Asymptotic Standard Error of Estimate of the T-year Flood". *Jour. of Hydrol.*, v. 85, p. 139-150.
- Condie, R., K.H. Lee, 1982. "Flood Frequency Analysis with Historic Information". *Jour. of Hydrol.*, v. 58, p. 47-61.
- Costa, J.E., 1978. "Holocene Stratigraphy in Flood Frequency Analysis". *Water Resour. Res.*, v. 14(4), p. 626-632.
- Couchoud Sebastián, R., 1965. "Hidrología Histórica del Segura". CEDEX, MOPU.
- Cunnane, C., J.E. Nash, 1971. "Bayesian Estimation of Frequency of Hydrologic Events". *Proc. Symposium de Varsovia en "Matematical Models in Hydrology"*.
- Cunnane, C., 1978. "Unbiased Plotting Positions. A Review". *J. of Hyd.*, v. 37, p. 205-222.
- Dalrymple, T., 1960. "Flood Frequency Analysis". U.S. Geological Survey Water Supply Paper 1543-A.

- De Coursey, D.G., 1973. "Objective Regionalization of Peak Flow Rates". Proc. of the Second Inter. Symp. in Hydrology, sep. 1972 Fort Collins (USA).

- Fiorentino, M., P. Versace, F. Rossi, 1985. "Regional Flood Frequency Estimation Using the Two-Component Extreme Value Distribution". Hydrological Sciences Journal, v. 30, p. 51-64.

- Francés, F., J.D. Salas, D.C. Boes, 1994. "Flood Frequency Analysis with Systematic and Historical or Paleoflood Data Based on the Two-Parameter General Extreme Models". Water Res. Res., 30 (6), 1653-1664.

- Gerard, R., E.W. Karpuk, 1979. "Probability Analysis of Historical Flood Data". J. Hydraul. Div. Am. Soc. Civ. Eng., v. 105(HY9), p. 1153-1165.

- Gómez Ortega, J., F. Lizárraga, E. de Churruca, 1865. "Estudio de las Inundaciones del Júcar" Facsimil Editado por la C.H.J. en 1988.

- Greenwood, J.A., J.M. Landwehr, N.C. Matalas, J.R. Wallis, 1979. "Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form". W.R.R., v. 15, p. 1049-1054.

- Guo, S. L. and C. Cunnane, 1991. "Evaluation of the usefulness of historical and paleological floods in quantile estimation". Jour. of Hydrol., 129, 245-262.

- Hirsh, R. M., J.R. Stedinger, 1987. "Plotting Positions for Historical Floods and their Precision". Water Res. Res., v.23 (4), p. 715-727.

- Hosking, J.R., 1984. "Testing Whether the Shape Parameter is Zero in the Generalized Extreme-Value Distribution". Biometrika, v.71, n.2, p. 367-374.

- Hosking, J.R.M., J.R. Wallis, 1986a. "Palaeoflood Hydrology and Flood Frequency Analysis". Water Resour. Res., v. 22(4), p. 543-550.

- Hosking, J.R.M., J.R. Wallis, 1986b. "The Value of Historical Data in Flood Frequency Analysis". Water Resour. Res., v. 22(11), p. 1606-1612.

- Houghton, J.C., 1978. "Birth of a Parent: the Wakeby Distribution for Modeling Flood Flows". W.R.R., v. 14, p. 1105-1109.

-
- Huici, A., 1969. "La Historia musulmana de Valencia y su Región". Ed. Anubar, Valencia.
 - Hupp, C.R., 1982. "Extension of Flood Record for Passage Creek, Virginia, Utilizing Tree Ring Analysis". *Bull. Ecol. Soc. Am.*, v. 63(2), p. 78.
 - Kendall, M.G., and A. Stuart, 1967. "The advanced theory of statistics", v. II, Hafner.
 - Kochel, R.C., V.R. Baker, 1982. "Paleoflood Hydrology". *Science*, v. 215(4531), p. 353-361.
 - Kottegoda, N.T., 1984. "Investigation of Outliers in Annual Maximum Flow Series". *Journal of Hydrology*, v. 72, p. 105-137.
 - Kuczera, G., 1982a. "Combining Site-Specific and Regional Information: an Empirical Bayes Approach". *W.R.R.*, v. 18, p. 306-314.
 - Kuczera, G., 1982b. "Robust Flood Frequency Models". *W.R.R.*, v. 18, p. 315-324.
 - Landwehr, J.M., N.C. Matalas, J.R. Wallis, 1979. "Probability Weighted Moments Compared with some Traditional Techniques in Estimating Gumbel Parameters and Quantiles". *W.R.R.*, v. 15, p. 1055-1064.
 - Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1980. "Quantile Estimation with more or less Flood-like Distributions". *W.R.R.*, v. 16, p. 547-555.
 - Lane, W. E., 1987. "Paleohydrology Data and Flood Frequency Estimation". *Proceedings of the International Symposium on Flood Frequency and Risk Analysis*, Louisiana State University, Baton Rouge, Louisiana, (May 1986), edited by V. P. Shing, pp 287-298.
 - Las Provincias, 1988. "Historia Viva de Valencia".
 - Leese, M. N., 1973. "Use of Censored Data in the Estimation of Gumbel Distribution Parameters for Annual Maximum Flood Series". *Water Res. Res.*, v. 9, p. 1534-1542.
 - Lettenmaier, D.P., Potter, K.W., 1985. "Testing Flood Frequency Estimation Methods Using a Regional Generation Model". *W.R.R.*, v. 21, p. 1903-1914.
 - Matalas, N.C., J.R. Slack, J.R. Wallis, 1975. "Regional Skew in Search of a Parent". *W.R.R.*, v. 11, p. 815-826.

- Ministerio de OOPP y Urbanismo, 1967. "Instrucción para el Proyecto, Construcción y Explotación de Grandes Presas". Madrid.
- Mood, A.M., F.A. Graybill, D.C. Boes, 1974. "Introduction to the Theory of Statistics". McGraw-Hill, 564 p.
- Muñoz Bravo, J., 1989. "Enseñanzas de las Avenidas Históricas en la Cuenca del Segura". Reunión Científica Internacional sobre Avenidas Fluviales e Inundaciones en la Cuenca del Mediterráneo, Universidad de Alicante.
- Nash, J.E., Shaw, B.L., 1966. "Flood Frequency as a Function of Catchment Characteristics", Proc. Symp. de Londres en "River Flood Hydrology".
- Natural Environment Reseach Council, 1975. "Flood Studies Report". Hydrologic Studies, v. 1, Londres.
- Phien, H. N. and 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. and K. Adamowski, 1993. "Asymptotic variance of flood quantile in log Pearson type III distribution with historical information". Jour. of Hydrol., 143, 481-503.
- Press, W.H., B.P. Flannery, S.A. Tevkolsky, W.T. Vetterlig, 1989. "Numerical Recipies". Cambridge University Press, 702 p.
- Potter, W.D., 1958. "Upper and Lower Frequency Curves for Peak Rates of Runnof". EOS, Trans. AGU, v. 39, p. 100-105.
- Raynal, J., 1985. "Bivariate Extreme Value Distributions Applied for Flood Frequency Analysis". Ph. D. en CSU, Fort Collins (USA).
- Rossi, F., M. Fiorentino, P. Versace, 1984. "Two-Component Extreme Value Distribution for Flood Frequency Analysis". W.R.R., v. 20, p. 847-856.
- Shen, H.W., M. Bryson, I. Ochoa, 1980. "Effect of Tail Behavoir Assumptions on Flood Predictions". W.R.R., v. 16, p. 361-364.
- Stedinger, J.R., 1983. "Estimating a Regional Flood Frequency Distribution". W.R.R., v. 19,

p. 503-510.

- Stedinger, J.R. T.A. Cohn, 1986. "Flood Frequency Analysis with Historical and Paleoflood Information". *Water Resour. Res.*, v. 22(5), p. 785-793.

- Stedinger, J.R. T.A. Cohn, 1987. "Historical Flood Frequency Data: its Value and use". *Proceedings of the International Symposium on Flood Frequency and Risk Analysis*, Louisiana State University, Baton Rouge, Louisiana, (May 1986), edited by V. P. Shing, pp 273-286.

- Stedinger, J. R., V.R. Baker, 1987. "Surface Water Hydrology: Historical and Paleoflood Information". *Reviews of Geophysics*, v. 25 (2), p. 119-124.

- Tasker, G. D., W.O. Thomas, 1978. "Flood-Frequency Analyses with Prerecord Information". *Journ. of the Hydraul. Div. ASCE*, v. 104 (2), p. 249-259.

- Tasker, G.D., 1982. "Comparing Methods of Hydrologic Regionalization". *Water Resources Bulletin*, v. 18, p. 965-970.

- U.S. Water Resources Council, 1967. "A uniform Technique for Determining Flood Flow Frequencies". *Bulletin 15*, Hydrology Committee, Washington D.C.

- U.S. Water Resources Council, 1976. "Guidelines for Determining Flood Frequency". *Bulletin 17A*, Hydrology Committee, Washington D.C.

- U.S. Water Resources Council, 1982. "Guidelines for Determining Flood Frequency". *Bulletin 17B*, Hydrology Committee, Washington D.C.

- Wall, D. J., D. F. Kibler, M. E. Hastings, 1987. "Use of Historic Flood Information in Estimating Flood Peaks on Ungaged Watersheds". *Water Res. Bulletin*, v. 23 (5), p. 919-926.

- Wall, D. J., D. F. Kibler, D. W. Newton, J. C. Herrin, 1987. "Flood Peak Estimates from Limited At-Site Historic Data". *Jour. of Hydraul. Eng. ASCE*, v. 113 (9).

- Wallis, J.R., 1980. "Risk and Uncertainties in the Evaluation of Flood Events for the Design of Hydrologic Structures". *Seminar on Extreme Hydrologic Events*, Fondazione Politecnica del Mediterraneo, Erice (Italia).

- Waylen, P., M. Woo, 1982. "Prediction of Annual Floods Generated by Mixed Processes". *W.R.R.*, v. 18, p. 1283-1286.

- White, E.L., 1975. "Factor Analysis of Drainage Basin Properties: Classification of Flood Behaviour in Terms of Basin Geomorphology". W.R. Bulletin, v. 11, p. 676-687.
- Wiltshire, S.E., 1986. "Identification of Homogeneous Regions for Flood Frequency Analysis". Jour. of Hydr., v. 84, p. 287-302.
- Zhang, Y., 1982. "Plotting Positions of Annual Flood Extremes Considering Extraordinary Values". W.R.R., v. 18, p. 859-864.

**ANEJO A.- OBTENCION DE LA FUNCION DE DENSIDAD DE
PROBABILIDAD ACUMULADA EMPIRICA**

La función de densidad de probabilidad acumulada (f.d.a.) empírica tiene dos utilidades principales en Hidrología:

- i) En Estadística no paramétrica para obtener los cuantiles de un cierto periodo de retorno sin ajustar ninguna distribución.
- ii) Servir como criterio de bondad de ajuste gráfico de las diferentes funciones de distribución utilizadas.
- iii) Detección de "outliers".

La f.d.a. empírica se obtiene asignando una probabilidad de no excedencia a la variable estadística ordenada.

Cunnane (1978), y posteriormente Arnell et al. (1982), obtiene un estimador de la f.d.a. empírica, con las propiedades de ser insesgado y de tener el mínimo error cuadrático medio. La expresión general es:

$$F(X'_i) = \frac{i - \alpha}{N + 1 - 2\alpha} \quad (\text{A.1})$$

donde:

$F(\cdot)$ = probabilidad de no excedencia

X'_i = serie en orden ascendente

i = número de orden

N = longitud de la serie

α = coeficiente que depende de la población de la variable estadística

Cunnane (1978) da algunos ejemplos del coeficiente α :

$\alpha = 0$ para la distribución uniforme (fórmula de Weibull).

$\alpha = 0.44$ para distribución Gumbel o EV1 (fórmula de Gringorten).

$\alpha = 3/8$ para distribución Normal (fórmula de Blom).

$\alpha = 0.4$ mejor compromiso si se desconoce la población.

En este trabajo, cuando se utilizan solo datos de la serie sistemática se emplea este último coeficiente.

Para incorporar la información histórica en la f.d.a. empírica se ha utilizado una de las expresiones analizadas por Hirsh y Stedinger (1987), que es:

$$F(X_j) = 1 - \frac{j}{k'+1} \frac{k'}{N+M} \quad j = 1, \dots, k' \quad (\text{A.2})$$

$$F(X_j) = 1 - \frac{k'}{N+M} - \frac{N+M-k'}{N+M} \frac{j-k'}{N-e+1} \quad j = k'+1, \dots, N+k \quad (\text{A.3})$$

donde:

X_j^* = serie en orden descendente

j = número de orden

M = período histórico

k = avenidas superiores al límite de percepción durante el período histórico

k' = avenidas superiores al límite de percepción

e = avenidas superiores al límite de percepción durante el período sistemático = $k' - k$

Esta expresión, que denominan "E-fórmula", tiene mayor precisión que los estimadores clásicos tipo fórmula de Hazen o fórmula de Weibull.

ANEJO B.- RESULTADOS PARA LA DISTRIBUCION WEIBULL

Como un anejo al Capítulo 4 se va dejar constancia de los resultados finales obtenidos para las otras dos distribuciones de la familia EV. En primer lugar se hace con la distribución Weibull, cuya función de densidad de probabilidad acumulada es:

$$F_X(x) = 1 - e^{-(\theta x)^\gamma} \quad x \geq 0 \quad (\text{B.1})$$

donde:

θ = parámetro de escala

γ = parámetro de forma

Derivando la ecuación (B.1) se obtiene la función de densidad de probabilidad de la distribución Weibull:

$$f_X(x) = \theta \gamma (\theta x)^{\gamma-1} e^{-(\theta x)^\gamma} \quad x \geq 0 \quad (\text{B.2})$$

La media de la distribución viene dada por:

$$\mu = E[X] = \frac{1}{\theta} \Gamma\left(1 + \frac{1}{\gamma}\right) \quad (\text{B.3})$$

La varianza de la distribución Weibull resulta ser:

$$\sigma^2 = E[(X-\mu)^2] = \frac{1}{\theta^2} \left[\Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma^2\left(1 + \frac{1}{\gamma}\right) \right] \quad (\text{B.4})$$

Y el coeficiente de sesgo es:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\Gamma\left(1 + \frac{3}{\gamma}\right) - 3\Gamma\left(1 + \frac{1}{\gamma}\right)\Gamma\left(1 + \frac{2}{\gamma}\right) + 2\Gamma^3\left(1 + \frac{1}{\gamma}\right)}{\left[\Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma^2\left(1 + \frac{1}{\gamma}\right) \right]^{3/2}} \quad (\text{B.5})$$

Que es la misma expresión que para la distribución LogGumbel. Si el parámetro de escala tuviera las mismas dimensiones que en el caso LogGumbel, todos los momentos centrales tendrían la misma expresión.

Los logaritmos de la función de verosimilitud para los 4 niveles de información considerados son:

$$LL_{SY}(\Theta) = NL\theta + NL\gamma + (\gamma - 1) \sum_{i=1}^N L(\theta x_i) - \sum_{i=1}^N (\theta x_i)^\gamma \quad (\text{B.6})$$

$$LL_{BC}(\Theta) = NL\theta + NL\gamma + (\gamma - 1) \sum_{i=1}^N L(\theta x_i) - \sum_{i=1}^N (\theta x_i)^\gamma + (M - k)L[1 - e^{-(\theta Q_\mu)^\gamma}] - k(\theta Q_\mu)^\gamma \quad (\text{B.7})$$

$$LL_{CF}(\Theta) = (M - k)Lq + (N + k)\gamma L\theta + (N + k)L\gamma + (\gamma - 1) \sum_{i=1}^N Lx_i - \sum_{i=1}^N (\theta x_i)^\gamma + \\ + (\gamma - 1) \sum_{j=1}^k Ly_j - \sum_{j=1}^k (\theta y_j)^\gamma \quad (\text{B.8})$$

$$LL_{MF}(\Theta) = (N + 1)\gamma L\theta + (N + 1)L\gamma + (\gamma - 1) \sum_{i=1}^N Lx_i - \sum_{i=1}^N (\theta x_i)^\gamma + (\gamma - 1)Lw - (\theta w)^\gamma + \\ + (M - 1)L[1 - e^{-(\theta w)^\gamma}] \quad (\text{B.9})$$

Y las tres ganancias estadística asintóticas resultan ser:

$$SG_1 = 1 - R_1 \left[1 + \frac{rD^2L^2 \left(\frac{D}{K} \right) \frac{p}{q}}{a - 2bLK + L^2K} \right] \quad (\text{B.10})$$

donde:

$$\frac{1}{R_1} = 1 + \frac{rD^2(L^2D - 2bLD + a) \frac{p}{q}}{a - b^2} \quad (\text{B.11})$$

muy similar a la ganancia estadística de la distribución LogGumbel, pero donde ahora:

$$D = L H$$

$$K = L T$$

$$SG_2 = 1 - R_2 \left[1 + r \frac{a' + p + pDLD \left(\frac{D}{q} - 1 \right) - 2 \left[b' + pDLD \left(\frac{D}{q} - 1 \right) \right] LK + p \left(1 - \frac{D^2}{q} \right) L^2K}{a - 2bLK + L^2K} \right] \quad (\text{B.12})$$

donde:

$$\frac{1}{R_2} = 1 + \frac{r^2}{a-b^2} \left\{ b'^2 + p \left(1 - \frac{D^2}{q} \right) \left[a' + p + pDL D \left(\frac{D}{q} - 1 \right) + pD^2 L^2 D \right] \right\} + \frac{r}{a-b^2} \left[a' - 2bb' + p + ap \left(1 - \frac{D^2}{q} \right) + (1-2b)pDL D \left(\frac{D}{q} - 1 \right) \right] \quad (\text{B.13})$$

$$SG_3 = 1 - R_3 \left\{ N + \frac{1 + M(I_4 + (M+1)I_5 - [I_6 + (M-1)(I_8 - I_7)]LK + [I_1 + (M-1)(I_3 - I_2)]L^2K)}{a - 2bLK + L^2K} \right\} \quad (\text{B.14})$$

donde:

$$\frac{1}{R_3} = \frac{[N + MI_1 + M(M-1)(I_3 - I_2)][Na + 1 + MI_4 + M(M-1)I_5] - [Nb + MI_6 + M(M-1)(I_8 - I_7)]^2}{N(a-b^2)} \quad (\text{B.15})$$

con las funciones de la longitud del período histórico M:

$$I_1 = \sum_{i=0}^{M-1} (-1)^i \binom{M-1}{i} \frac{1}{(i+1)^2} \quad (\text{B.16})$$

$$I_2 = \sum_{i=0}^{M-2} (-1)^i \binom{M-2}{i} \frac{1}{(i+2)^2} \quad (\text{B.17})$$

$$I_3 = \sum_{i=0}^{M-3} (-1)^i \binom{M-3}{i} \frac{1}{(i+2)^3} \quad (\text{B.18})$$

$$I_4 = \sum_{i=0}^{M-1} (-1)^i \binom{M-1}{i} \frac{L^2(1+i) - 2L(1+i)\Gamma'(2) + \Gamma''(2)}{(i+1)^2} \quad (\text{B.19})$$

$$I_5 = \sum_{i=0}^{M-3} (-1)^i \binom{M-3}{i} \frac{2L^2(2+i) - 2L(2+i)\Gamma'(3) + \Gamma''(3)}{(i+2)^3} \quad (\text{B.20})$$

$$I_6 = \sum_{i=0}^{M-1} (-1)^i \binom{M-1}{i} \frac{\Gamma'(2) - L(i+1)}{(i+2)^2} \quad (\text{B.21})$$

$$I_7 = \sum_{i=0}^{M-2} (-1)^i \binom{M-2}{i} \frac{\Gamma'(2) - L(i+2)}{(i+2)^2} \quad (\text{B.22})$$

$$I_8 = \sum_{i=0}^{M-3} (-1)^i \binom{M-3}{i} \frac{\Gamma'(3) - 2L(i+2)}{(i+2)^3} \quad (\text{B.23})$$

ANEJO C.- RESULTADOS PARA LA DISTRIBUCION GUMBEL

Como última representante de la familia EV, se resumen en este Anejo los resultados finales para la distribución Gumbel, cuya función de densidad de probabilidad acumulada es:

$$F_X(x) = \exp(-\lambda e^{-\theta x}) \quad (\text{C.1})$$

donde:

θ = parámetro de escala

λ = parámetro de forma

Derivando la ecuación (C.1) se obtiene la función de densidad de probabilidad de la distribución Gumbel:

$$f_X(x) = \lambda \theta e^{-\theta x} \exp(-\lambda e^{-\theta x}) \quad (\text{C.2})$$

La media de la distribución viene dada por:

$$\mu = E[X] = \frac{1}{\theta} (L\lambda + C) \quad (\text{C.3})$$

donde C es la constante de Euler igual a 0.5772.

La varianza de la distribución Gumbel resulta ser:

$$\sigma^2 = E[(X-\mu)^2] = \frac{\pi^2}{6\theta^2} \quad (\text{C.4})$$

Y el coeficiente de sesgo es:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{-\psi''(1)}{\left(\frac{\pi^2}{6}\right)^{3/2}} \quad (\text{C.5})$$

donde $\Psi(\cdot)$ es la función psi de Euler, cuya derivada segunda es:

$$\psi''(x) = -2 \sum_{i=0}^{\infty} \frac{1}{(x+i)^3} \quad (\text{C.6})$$

resultando un coeficiente de sesgo constante e igual a 1.29857.

Los logaritmos de la función de verosimilitud para los 4 niveles de información considerados son:

$$LL_{SY}(\Theta) = NL\theta + NL\lambda - \lambda \sum_{i=1}^N e^{-\theta x_i} - \theta \sum_{i=1}^N x_i \quad (C.7)$$

$$LL_{BC}(\Theta) = -(M-k)\lambda e^{-\theta Q_n} + kLp + NL\theta + NL\lambda - \lambda \sum_{i=1}^N e^{-\theta x_i} - \theta \sum_{i=1}^N x_i \quad (C.8)$$

$$LL_{CE}(\Theta) = -(M-k)e^{-\theta Q_n} + (N+k)L\theta + (N+k)L\lambda - \lambda \sum_{i=1}^N e^{-\theta x_i} - \theta \sum_{i=1}^N x_i - \lambda \sum_{j=1}^k e^{-\theta y_j} - \theta \sum_{j=1}^k y_j \quad (C.9)$$

$$LL_{MF}(\Theta) = (N+1)L\theta + (N+1)L\lambda + LM - \lambda \sum_{i=1}^N e^{-\theta x_i} - \theta \sum_{i=1}^N x_i - M\lambda e^{-\theta w} - \theta w \quad (C.10)$$

Y las tres ganancias estadística asintóticas resultan ser:

$$SG_1 = 1 - R_1 \left[1 + \frac{rD^2 L^2 \left(\frac{D}{K} \right) \frac{q}{p}}{a - 2bLK + L^2 K} \right] \quad (C.11)$$

donde:

$$\frac{1}{R_1} = 1 + \frac{rD^2(L^2D - 2bLD + a) \frac{q}{p}}{a - b^2} \quad (C.12)$$

muy similar a la ganancia estadística de la distribución LogGumbel, pero donde ahora:

$$D = L [-L (1-1/H)]$$

$$K = -L (1-1/T)$$

$$SG_2 = 1 - R_2 \left[1 + r \frac{a' + p + qDL^2D - 2b'LK + pL^2K}{a - 2bLK + L^2K} \right] \quad (C.13)$$

donde:

$$\frac{1}{R_2} = 1 + \frac{r}{a-b^2} [a' + p + qDL^2D - 2bb' + ap] + \frac{r^2}{a-b^2} (p^2 + pqDL^2D + pa' - b'^2) \quad (\text{C.14})$$

$$SG_3 = 1 - R_3 \left[N+1 + \frac{LM(LM-2b+2LK)}{a-2bLK+L^2K} \right] \quad (\text{C.15})$$

donde:

$$\frac{1}{R_3} = \frac{(N+1)^2}{N} + \frac{L^2M}{a-b^2} \quad (\text{C.16})$$