



CIMAT

Centro de Investigación en Matemáticas A.C

Valores Extremos Aplicados a Datos Ambientales

T E S I S

que para obtener el grado de

Maestra en Ciencias con Especialidad en
Probabilidad y Estadística

Presenta:

Mónica Gladis Pérez Miranda

Director de Tesis:

Dr. Joaquín Ortega Sánchez

Guanajuato, Gto., México, Octubre del 2005.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) y al Centro de Investigación en Matemáticas (CIMAT), por financiar mis estudios de maestría.

A mi familia por su confianza.

A los miembros del jurado por darle una estructura adecuada a este trabajo. Al Dr. Joaquín Ortega por su fina dirección, paciencia y apoyo. A la Dra. Eloisa Díaz-Francés por enriquecer este trabajo, de igual modo al Ingeniero Roberto Mejía por darle un sentido práctico a los resultados generados.

A Armando, Claudia (Cd), Marina, Valentina, Claudia Nila, Esteban, Yanink, Armando y José Ramón Domínguez, por su invaluable apoyo.

C I M A T
BIBLIOTECA

I
020271

Contenido

Agradecimientos	I
Introducción	III
1. Marco teórico	1
1.1. Introducción	1
1.2. Dominio de atracción del máximo	4
1.2.1. El dominio de atracción de la distribución Fréchet	6
1.2.2. El dominio de atracción de la distribución Weibull	7
1.2.3. El dominio de atracción de la distribución Gumbel	9
1.3. La distribución generalizada de valores extremos	13
1.3.1. Inferencia para la DGVE	14
1.4. Excesos sobre un umbral (Series de duración parcial)	20
1.4.1. Inferencia para la DGP	24
1.5. Cluster de series	27
2. Aplicación	30
2.1. Ajuste de máximos anuales	30
2.1.1. Estación Cuitzillo	37
2.1.2. Estación Cointzio	39
2.1.3. Estación Morelia	43
2.2. Excesos sobre umbrales (Series de duración parcial)	46
2.2.1. Estación Cuitzillo	51
2.2.2. Estación Cointzio	56
2.2.3. Estación Morelia	61
2.3. Cluster de series	66
Conclusión	75
Bibliografía	76

Introducción

La teoría de valores extremos tiene una aplicación muy amplia en climatología. Aunque las distribuciones de valores extremos son las candidatas idóneas para describir la distribución de las observaciones más grandes en cierto período de tiempo, la distribución Generalizada de Pareto ha resultado ser muy útil en la reducción de varianza de los cuantiles estimados.

Este trabajo se basa en los datos de precipitación pluvial de 18 estaciones meteorológicas que tienen influencia sobre la cuenca hidrológica de la ciudad de Morelia, estado de Michoacán. Tomando como base un estudio del manejo de aguas pluviales en dicha zona, se realiza un estudio de valores extremos para la región y se estima el máximo nivel de precipitación que podría ocurrir en la región en los próximos 100 o 1000 años.

En el primer capítulo se presenta una introducción a la teoría de valores extremos, resultando de vital importancia el teorema de Gnedenko. La sección de caracterización del dominio de atracción del máximo es de gran utilidad cuando se conoce la distribución de los datos de origen. Al final de este capítulo se presenta la parte de inferencia estadística, esto es, como se obtienen los estimadores de los parámetros de interés, los intervalos de confianza, así como la estimación de los cuantiles.

En el segundo capítulo, se presenta un análisis de la estimación del parámetro de forma y del nivel de retorno haciendo uso de la verosimilitud perfil de la Distribución Generalizada de Valores Extremos y de la Distribución Generalizada de Pareto. Aunque el ajuste de las dos distribuciones se hizo para cada una de las estaciones, este análisis se presenta de manera detallada sólo para las estaciones Cointzio (16022), Morelia DGE (16081) y Cuitzillo (16028). Las dos primeras resultan ser, de acuerdo con los especialistas, las más representativas de la región, la tercera se incluyó para complementar el análisis.

Finalmente, se agrupan los datos de las 18 estaciones, usando técnicas de agrupación de series de tiempo, esto con el objetivo de tener un menor número de modelos que sean representativos de la región. Esta sección puede

servir como referencia en la futura realización estudios de regionalización, los cuales resultan de especial interés para los especialistas.

Para la realización de las gráficas y cálculos estadísticos se utilizó S-PLUS. Las funciones específicas se pueden obtener de <http://www.maths.bris.ac.uk/~masgc/isnev/summary.html>

Se recomienda de manera especial la lectura de Coles (2001) y Embrechts (1997) para mayor profundidad en la teoría de valores extremos así como sus aplicaciones.

Capítulo 1

Marco teórico

En esta sección presentamos algunos de los resultados más importantes en la Teoría de Valores Extremos. Es importante resaltar que al igual que el Teorema del Límite Central justifica a la distribución Normal como la única distribución límite para la suma de variables aleatorias independientes e idénticamente distribuidas (iid), el Teorema de Gnedenko nos proporciona tres únicas distribuciones límite para los máximos; la distribución Fréchet, Weibull y la Gumbel. Así mismo, la Distribución Generalizada de Pareto resulta ser una distribución muy útil en la modelación de un mayor número de observaciones extremas.

Presentamos el tema de Caracterización de Dominio de Atracción del Máximo el cual es aplicable cuando se conoce la distribución de las observaciones de origen, y aunque en este trabajo éste no es el caso consideramos importante incluir este tema por ser relevante dentro de la Teoría de Valores Extremos.

La última parte de este capítulo está enfocada en la estimación de los parámetros de interés.

1.1. Introducción

Sea X_1, X_2, \dots una sucesión iid de variables aleatorias con función de distribución común F . Sea

$$M_n = \max(X_1, \dots, X_n), \quad n \geq 2,$$

el máximo muestral. Luego la función de distribución de M_n está dada por

$$P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = F^n(x) \quad x \in \mathbf{R}, \quad n \in \mathbf{N}.$$

Denotemos por

$$x_F = \sup\{x \in \mathbf{R} : F(x) < 1\} \leq \infty,$$

el extremo derecho de F , entonces

$$M_n \xrightarrow[c.s.]{} x_F. \quad (1.1)$$

Donde $\xrightarrow[c.s.]{}x < x_F$, $F(x) < 1$ y

$$P(M_n \leq x) = F^n(x) \rightarrow 0, \quad n \rightarrow \infty.$$

Por lo tanto $M_n \xrightarrow[P]{} x_F$ y dado que la sucesión (M_n) es no decreciente en n , ésta converge *c.s.*

Encontrar expresiones analíticas para la función de distribución de M_n puede resultar difícil aún si F es completamente conocida. Al igual que en el caso de la distribución normal la cual es una aproximación útil de la distribución de $\sum X_i$, se trata de buscar una distribución límite que actúe como una aproximación de F^n . La relación 1.1 muestra que no existe una distribución límite no degenerada a menos que normalicemos M_n .

El rango completo de posibles distribuciones límite para M_n normalizado está dado por el siguiente teorema.

Teorema 1.1.1 (Gnedenko)

Sea (X_n) una sucesión iid de variables aleatorias. Si existen constantes de normalización $a_n > 0$, $b_n \in \mathbf{R}$ y alguna función de distribución no degenerada H tal que

$$a_n^{-1}(M_n - b_n) \xrightarrow[d]{} H, \quad (1.2)$$

entonces H pertenece al tipo de una de las siguientes funciones de distribución:

$$\begin{aligned} \text{Fréchet:} \quad \Phi_\alpha(x) &= \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} \quad \alpha > 0. \\ \text{Weibull:} \quad \Psi_\alpha(x) &= \begin{cases} \exp\{-(-x)^{-\alpha}\}, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0. \\ \text{Gumbel:} \quad \Lambda(x) &= \exp\{-\exp\{-x\}\}, \quad x \in \mathbf{R}. \end{aligned}$$

Nos referiremos a Φ_α , Ψ_α y Λ como las distribuciones de valores extremos (DVE) o de Fisher-Tippet.

La demostración de este teorema se puede encontrar en Resnick (1987). Las distribuciones de valores extremos cumplen con una propiedad muy importante que es la de que son las únicas distribuciones max-estables.

Definición 1.1.1 (Distribución max-estable)

Una función de distribución H es max-estable si satisface

$$H^n(a_n x + b_n) = H(x)$$

para constantes apropiadas $a_n > 0$ y $b_n \in \mathbf{R}$ y para toda $n \geq 2$.

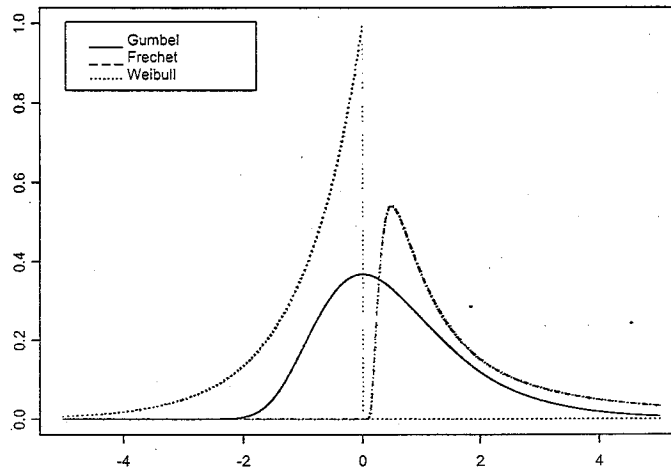


Figura 1.1: Densidades de las distribuciones de valores extremos. Para la distribución Fréchet y Weibull $\alpha = 1$

En particular, los tres casos en el Teorema 1.1.1 corresponden a:

- Fréchet: $M_n \stackrel{d}{=} n^{1/\alpha} X$
- Weibull: $M_n \stackrel{d}{=} n^{-1/\alpha} X$
- Gumbel: $M_n \stackrel{d}{=} X + \ln n$.

donde $\stackrel{d}{=}$ denota igualdad en distribución.

1.2. Dominio de atracción del máximo

Dada una distribución de valores extremos H , nuestro interés ahora es determinar que condiciones en la f.d. F implican que el máximo normalizado M_n converja débilmente a H .

Definición 1.2.1 (Dominio de atracción del máximo)

Decimos que la función de distribución F está en el dominio de atracción de la distribución de valor extremo H si existen constantes $a_n > 0$, $b_n \in \mathbf{R}$ tales que

$$P(M_n \leq a_n x + b_n) = F^n(a_n x + b_n) \rightarrow H(x), \quad x \in \mathbf{R}. \quad (1.3)$$

Notación: $F \in \text{DAM}(H)$.

El siguiente resultado será usado a lo largo de las siguientes secciones.

Proposición 1.2.1 (Caracterización de $\text{DAM}(H)$)

La f.d. F pertenece al dominio de atracción de la DVE H con constantes de normalización $a_n > 0$, $b_n \in \mathbf{R}$ si y sólo si

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n x + b_n) = -\log H(x), \quad x \in \mathbf{R}.$$

Donde $\bar{F} = 1 - F$. Cuando $H(x) = 0$ el límite se interpreta como ∞ .

Para caracterizar el dominio de atracción para la distribución Fréchet Φ_α y la distribución Weibull Ψ_α se requiere usar el concepto de las funciones de variación regular. El caso de la distribución Gumbel se trata de manera especial al final de esta sección.

Definición 1.2.2 (Funciones de variación regular)

Una función medible U es de variación regular en ∞ ($U \in \text{VR}_\alpha$) con índice α para algún $\alpha \in \mathbf{R}$, si para $x > 0$

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\alpha.$$

El siguiente conjunto de resultados examina las propiedades de las funciones de variación regular. Para más detalle sobre este tipo de funciones se puede consultar Embrechts, Klüppelberg y Mikosch (1997).

Teorema 1.2.1 (Karamata)

1. Si $\alpha \geq -1$ entonces $U \in VR_\alpha$ implica que $\int_0^x U(t)dt \in VR_{\alpha+1}$ y

$$\lim_{x \rightarrow \infty} \frac{xU(x)}{\int_0^x U(t)dt} = \alpha + 1. \quad (1.4)$$

Si $\alpha < -1$ (o si $\alpha = -1$ y $\int_x^\infty U(s)ds < \infty$) entonces $U \in VR_\alpha$ implica $\int_x^\infty U(t)dt$ es finita, $\int_x^\infty U(t)dt \in VR_{\alpha+1}$ y

$$\lim_{x \rightarrow \infty} \frac{xU(x)}{\int_x^\infty U(t)dt} = -\alpha - 1. \quad (1.5)$$

2. Si U satisface

$$\lim_{x \rightarrow \infty} \frac{xU(x)}{\int_0^x U(t)dt} = \lambda \in (0, \infty) \quad (1.6)$$

entonces $U \in VR_{\lambda-1}$. Si $\int_x^\infty U(t)dt < \infty$ y

$$\lim_{x \rightarrow \infty} \frac{xU(x)}{\int_x^\infty U(t)dt} = \lambda \in (0, \infty) \quad (1.7)$$

entonces $U \in VR_{-(\lambda+1)}$.

Teorema 1.2.2 (Representación de Karamata)

Una función $h \in VR_\alpha$ para algún $\alpha \in \mathbf{R}$ si y sólo si puede representarse como

$$h(x) = c(x) \exp \left\{ \int_1^x \frac{\epsilon(t)}{t} dt \right\} \quad (1.8)$$

para $x > 0$ donde $c: \mathbf{R}^+ \rightarrow \mathbf{R}^+$, $\epsilon: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ y

$$\lim_{x \rightarrow \infty} c(x) = c \in (0, \infty), \quad \lim_{t \rightarrow \infty} \epsilon(t) = \alpha. \quad (1.9)$$

El siguiente concepto define una relación de equivalencia entre dos funciones de distribución.

Definición 1.2.3 (Equivalencia asintótica)

Dos f.d. F y G son asintóticamente equivalentes si tienen el mismo extremo derecho, es decir, si $x_F = x_G$, y

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x)}{\overline{G}(x)} = c$$

para alguna constante $0 < c < \infty$.

Demostremos que los dominios de atracción de las DVE son cerrados respecto a esta relación, es decir, que si F y G son asintóticamente equivalentes entonces $F \in DAM(H)$ si y sólo si $G \in DAM(H)$. Más aún, para dos *f.d.* asintóticamente equivalentes es posible usar las mismas constantes de normalización.

El máximo muestral M_n es la versión empírica del cuantil $(1 - n^{-1})$ de la *f.d.* subyacente F . Los cuantiles corresponden a la inversa de una *f.d.*, la cual no siempre está bien definida.

Definición 1.2.4 (Función cuantil)

La inversa generalizada de la función de distribución F

$$Q(t) = F^{-}(t) = \inf\{x \in \mathbf{R} : F(x) \geq t\}, \quad 0 < t < 1,$$

es llamada la función cuantil de la función de distribución F . La cantidad $x_t = F^{-}(t)$ define el cuantil t de F .

1.2.1. El dominio de atracción de la distribución Fréchet

Teorema 1.2.3 (Dominio de atracción de Φ_α)

$F \in DAM(\Phi_\alpha)$, $\alpha > 0$, si y sólo si $\bar{F} \in VR_{-\alpha}$. En este caso

$$F^n(a_n x) \rightarrow \Phi_\alpha(x), \quad (1.10)$$

con

$$a_n = Q\left(1 - \frac{1}{n}\right) = (1/\bar{F})^{-}(n).$$

Note que este resultado implica que sólo distribuciones con $x_F = \infty$ pueden estar en $DAM(\Phi_\alpha)$. Esta clase de distribuciones tienen colas muy pesadas en el sentido de que $E(X^+)^\delta = \infty$ para $\delta > \alpha$. La demostración de este teorema se puede consultar en Embrechts *et al* (1997)

Von Mises encontró condiciones más sencillas para determinar si la densidad de una distribución pertenece a algún máximo dominio de atracción.

Corolario 1.2.1 (Condición de Von Mises)

Sea F absolutamente continua con densidad f que satisface para algún $\alpha > 0$

$$\lim_{x \rightarrow \infty} \frac{x f(x)}{\bar{F}(x)} = \alpha > 0, \quad (1.11)$$

entonces $F \in DAM(\Phi_\alpha)$.

El siguiente resultado nos adentra en la estructura del $DAM(\Phi_\alpha)$.

Proposición 1.2.2 (Propiedad de clausura del $DAM(\Phi_\alpha)$)

Sean F y G f.d. y supongamos que $F \in DAM(\Phi_\alpha)$ con constantes de normalización $a_n > 0$, es decir,

$$\lim_{n \rightarrow \infty} F^n(a_n x) = \Phi_\alpha(x), \quad x > 0. \quad (1.12)$$

Entonces

$$\lim_{n \rightarrow \infty} G^n(a_n x) = \Phi_\alpha(ax), \quad x > 0,$$

para algún $a > 0$ si y sólo si F y G son asintóticamente equivalentes con

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x)}{\bar{G}(x)} = a^\alpha.$$

La demostración se puede consultar Embrechts *et al* (1997). El teorema de representación de Karamata para funciones de variación regular implica que toda f.d. en $DAM(\Phi_\alpha)$, es asintóticamente equivalente a una f.d. absolutamente continua que satisface la condición de Von Mises. Luego podemos resumir de la siguiente manera: El dominio de atracción de Φ_α consiste de las f.d. que satisfacen la condición de Von Mises y las f.d. que son asintóticamente equivalentes a ellas.

1.2.2. El dominio de atracción de la distribución Weibull

Teorema 1.2.4 (Dominio de atracción de Ψ_α)

$F \in DAM(\Psi_\alpha)$, $\alpha > 0$, si y sólo si $x_F < \infty$ y $\bar{F}(x_F - x^{-1}) \in VR_\alpha$ cuando $x \rightarrow \infty$.

Si $F \in DAM(\Psi_\alpha)$, entonces

$$F^n(x_F + (x_F - a_n)x) \rightarrow \Psi_\alpha(x), \quad x < 0. \quad (1.13)$$

donde $a_n = Q(1 - \frac{1}{n}) = \left(\frac{1}{1-F}\right)^{\leftarrow}(n)$.

Así $DAM(\Psi_\alpha)$ consiste de distribuciones F con soporte acotado por la derecha. La prueba se puede consultar en Embrechts *et al* (1997).

Para este caso también es posible dar una condición más sencilla cuando la distribución tiene densidad.

Corolario 1.2.2 (Condición de Von Mises)

Sea F absolutamente continua con densidad f positiva en algún intervalo finito (z, x_F) . Si

$$\lim_{x \uparrow x_F} \frac{(x_F - x)f(x)}{\overline{F}(x)} = \alpha > 0, \quad (1.14)$$

entonces $F \in \text{DAM}(\Psi_\alpha)$.

Usando

$$F_*(x) = \begin{cases} 0, & x \leq 0, \\ F(x_F - x^{-1}), & x > 0, \end{cases} \quad (1.15)$$

podemos reformular la proposición 1.2.1 de la siguiente manera,

Proposición 1.2.3 (Propiedad de clausura de $\text{DAM}(\Psi_\alpha)$)

Sean F y G f.d. con $x_F = x_G < \infty$ y supongamos que $F \in \text{DAM}(\Psi_\alpha)$ con constantes de normalización $a_n > 0$, es decir,

$$\lim_{n \rightarrow \infty} F^n(a_n x + x_F) = \Psi_\alpha(x), \quad x < 0.$$

Entonces

$$\lim_{n \rightarrow \infty} G^n(a_n x + x_F) = \Psi_\alpha(ax), \quad x < 0,$$

para algún $a > 0$ si y sólo si F y G son asintóticamente equivalentes con

$$\lim_{x \rightarrow x_F} \frac{\overline{F}(x)}{\overline{G}(x)} = a^{-\alpha}.$$

Usando de nuevo el teorema de representación de Karamata para funciones de variación regular podemos ver que las funciones que pertenecen al dominio de atracción de Ψ_α son asintóticamente equivalentes a una f.d. absolutamente continua que satisface la condición de Von Mises dada en 1.14. Podemos resumir que el dominio de atracción de Ψ_α consiste de las f.d. que satisfacen la condición de Von Mises y las f.d. que son asintóticamente equivalentes a ellas.

1.2.3. El dominio de atracción de la distribución Gumbel

Funciones de Von Mises

Las funciones de Von Mises ayudan en la caracterización del dominio de atracción de la distribución Gumbel.

Definición 1.2.5 (Función de Von Mises)

Sea F una f.d. con extremo derecho $x_F \leq \infty$. Supongamos que existe $z < x_F$ tal que F tiene representación

$$\bar{F}(x) = c \exp \left\{ - \int_z^x \frac{1}{a(t)} dt \right\}, \quad z < x < x_F, \quad (1.16)$$

donde c es una constante positiva, $a(\cdot)$ es una función positiva y absolutamente continua (con respecto a la medida de Lebesgue) con densidad a' y $\lim_{x \uparrow x_F} a'(x) = 0$. Entonces decimos que F es una función de Von Mises y que $a(\cdot)$ es la función auxiliar de F .

Observación 1 Es interesante comparar la representación 1.16 con la representación de Karamata 1.8 para una función de variación regular. Si ponemos $a(x) = x/\epsilon(x)$ de modo que $\epsilon(x) \rightarrow \alpha \in [0, \infty)$ cuando $x \rightarrow \infty$, 1.16 se convierte en una función de variación regular con índice $-\alpha$. Veremos más adelante que la función auxiliar de una función de Von Mises con $x_F = \infty$ satisface $a(x)/x \rightarrow 0$. Esto quiere decir que $\bar{F}(x)$ decrece a 0 mucho más rápido que cualquier potencia.

Definición 1.2.6 (Funciones de variación rápida)

Una función medible y positiva h definida en $(0, \infty)$ es de variación rápida con índice $-\infty$ ($h \in \mathcal{R}_{-\infty}$) si

$$\lim_{z \rightarrow \infty} \frac{h(tx)}{h(x)} = \begin{cases} 0 & \text{si } t > 1, \\ \infty & \text{si } 0 < t < 1. \end{cases}$$

Las principales propiedades de estas funciones se resumen en el siguiente teorema.

Teorema 1.2.5 (Propiedades de las funciones de variación rápida)

1. Sea $h \in \mathcal{R}_{-\infty}$ una función no creciente, entonces para algún $z > 0$ y todo $\alpha \in \mathbf{R}$

$$\int_z^\infty t^\alpha h(t) dt < \infty,$$

y

$$\lim_{z \rightarrow \infty} \frac{x^{\alpha+1} h(x)}{\int_x^\infty t^\alpha h(t) dt} = \infty. \quad (1.17)$$

Si para algún $\alpha \in \mathbf{R}$, $\int_1^\infty t^\alpha h(t) dt < \infty$ y 1.17 se cumple, entonces $h \in \mathcal{R}_{-\infty}$.

2. Si $h \in \mathcal{R}_\infty$, entonces existen funciones c y δ tales que $c(x) \rightarrow c_0 \in (0, \infty)$, $\delta(x) \rightarrow -\infty$ cuando $x \rightarrow \infty$ y para algún $z > 0$,

$$h(x) = c(x) \exp \left\{ \int_z^x \frac{\delta(u)}{u} \right\}, \quad x \geq z. \quad (1.18)$$

El recíproco también es cierto.

Teorema 1.2.6 (Propiedades de las funciones de Von Mises)

Toda función de Von Mises es absolutamente continua en (z, x_F) con densidad positiva f . Podemos escoger la función auxiliar como $a(x) = \frac{\bar{F}(x)}{f(x)}$. Además tenemos las siguientes propiedades.

1. Si $x_F = \infty$, entonces $\bar{F} \in \mathcal{R}_{-\infty}$ y

$$\lim_{z \rightarrow \infty} \frac{x f(x)}{\bar{F}(x)} = \infty. \quad (1.19)$$

2. Si $x_F < \infty$, entonces $\bar{F}(x_F - x^{-1}) \in \mathcal{R}_{-\infty}$ y

$$\lim_{z \rightarrow x_F} \frac{(x_F - x) f(x)}{\bar{F}(x)} = \infty. \quad (1.20)$$

La prueba se puede consultar Embrechts *et al* (1997). Ahora podemos demostrar que las funciones Von Mises pertenecen al $\text{DAM}(\Lambda)$. Más aún, la forma específica de \bar{F} se sigue de calcular las constantes de normalización a_n de la función auxiliar.

Proposición 1.2.4 (Funciones de Von Mises y $\text{DAM}(\Lambda)$)

Si F es una función de Von Mises, entonces $F \in \text{DAM}(\Lambda)$. Las constantes de normalización pueden escogerse de acuerdo a las siguientes fórmulas

$$b_n = F^{\leftarrow}(1 - n^{-1}) \quad a_n = a(b_n), \quad (1.21)$$

donde a es la función auxiliar de F .

Caracterización del dominio de atracción de Λ

Las funciones Von Mises no caracterizan completamente el dominio de atracción de Λ . Sin embargo, una pequeña modificación en 1.16 nos lleva a una completa caracterización de $DAM(\Lambda)$.

Teorema 1.2.7 (Caracterización I de $DAM(\Lambda)$)

La f.d. F con extremo derecho $x_F \leq \infty$ pertenece a $DAM(\Lambda)$ si y sólo si existe algún $z < x_F$ tal que F tiene representación

$$\bar{F}(x) = c(x) \exp \left\{ - \int_z^x \frac{g(t)}{a(t)} dt \right\}, \quad z < x < x_F, \quad (1.22)$$

donde c y g son funciones medibles que satisfacen $c(x) \rightarrow c > 0$, $g(x) \rightarrow 1$ cuando $x \uparrow x_F$, y $a(x)$ es una función positiva y absolutamente continua con respecto a la medida de Lebesgue con densidad $a'(x)$ que satisface $\lim_{x \uparrow x_F} a'(x) = 0$.

Para una F con esta representación podemos escoger

$$b_n = Q(1 - n^{-1}) \quad a_n = a(b_n)$$

como constantes de normalización.

Una posible selección de a es

$$a(x) = \int_x^{x_F} \frac{\bar{F}(t)}{\bar{F}(x)} dt, \quad x < x_F. \quad (1.23)$$

Observación 2 Si X es una v.a. con f.d. F , la función a definida por 1.23 es la función media de excesos o vida remanente media:

$$a(x) = E(X - x | X > x), \quad x < x_F.$$

Una caracterización alternativa para el dominio de atracción de la distribución de Gumbel está dada por el siguiente teorema.

Teorema 1.2.8 (Caracterización II de $DAM(\Lambda)$)

La f.d. F pertenece al $DAM(\Lambda)$ si y sólo si existe alguna función positiva \tilde{a} tal que

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x + t\tilde{a}(x))}{\bar{F}(x)} = e^{-t}, \quad t \in \mathbf{R}. \quad (1.24)$$

Una posible selección es $\tilde{a} = a$ dada por la ecuación 1.23.

Al igual que en los casos anteriores, el dominio de atracción de la distribución Gumbel también es cerrado respecto a equivalencia asintótica, como indica el siguiente teorema.

Teorema 1.2.9 (Propiedad de clausura de $DAM(\Lambda)$ bajo equivalencia asintótica)

Sean F y G f.d. con el mismo extremo derecho $x_F = x_G$ y suponga que $F \in DAM(\Lambda)$ con constantes de normalización a_n y $b_n \in \mathbf{R}$; es decir,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \Lambda(x), \quad x \in \mathbf{R}. \quad (1.25)$$

Entonces.

$$\lim_{n \rightarrow \infty} G^n(a_n x + d_n) = \Lambda(x + b), \quad x \in \mathbf{R},$$

si y sólo si F y G son equivalentes asintóticamente, con

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x)}{\bar{G}(x)} = e^b$$

La prueba se puede consultar en Embrechts *et al* (1997).

Finalmente presentamos en el Cuadro 1.1 los dominios de atracción del máximo de algunas distribuciones conocidas (Castillo, 1988).

Distribución	Distribución del máximo
Normal	Gumbel
Exponencial	Gumbel
Log-Normal	Gumbel
Gama	Gumbel
Gumbel	Gumbel
Rayleigh	Gumbel
Uniforme	Weibull
Weibull	Weibull
Cauchy	Fréchet
Pareto	Fréchet
Fréchet	Fréchet

Cuadro 1.1: Dominio de atracción de algunas distribuciones conocidas

1.3. La distribución generalizada de valores extremos

En aplicaciones, las tres familias muestran representaciones muy diferentes del comportamiento de los valores extremos. Con un cambio conveniente de localización y escala, las tres DVE pueden ser representadas por una sola expresión que facilita el análisis, para esto se introduce un parámetro ξ tal que:

$$\begin{aligned} \xi = \alpha^{-1} > 0 & \text{ corresponde a la distribución Fréchet } \Phi_\alpha, \\ \xi = 0 & \text{ corresponde a la distribución Gumbel } \Lambda, \\ \xi = -\alpha^{-1} < 0 & \text{ corresponde a la distribución Weibull } \Psi_\alpha. \end{aligned}$$

Definición 1.3.1 (Representación von Mises-Jenkinson de las DVE: la distribución generalizada de valores extremos (DGVE)).

Defina la función de distribución H_ξ por:

$$H_\xi(x) = \begin{cases} \exp\{-(1 + \xi x)^{-1/\xi}\} & \xi \neq 0, \\ \exp\{-\exp\{-x\}\} & \xi = 0, \end{cases}$$

donde $1 + \xi x > 0$. El soporte de H_ξ corresponde a

$$\begin{aligned} x &> -\xi^{-1} & \text{para } \xi > 0, \\ x &< -\xi^{-1} & \text{para } \xi < 0, \\ x &\in \mathbf{R} & \text{para } \xi = 0. \end{aligned}$$

H_ξ es llamada la distribución generalizada de valores extremos estándar (DGVE). Se puede introducir la familia de localización-escala $H_{\xi;\mu,\sigma}$ reemplazando el argumento x por $(x - \mu)/\sigma$ para $\mu \in \mathbf{R}$, $\sigma > 0$. También nos referiremos a $H_{\xi;\mu,\sigma}$ como DGVE.

El siguiente teorema es uno de los resultados básicos en teoría de valores extremos y constituye la base para numerosas técnicas estadísticas. Primero tengamos presente la definición de la función cuantil F^{\leftarrow} de una *f.d.*; definamos

$$U(t) = F^{\leftarrow}(1 - t^{-1}), \quad t > 0.$$

Teorema 1.3.1 (Caracterización de $DAM(H_\xi)$)

Para $\xi \in \mathbf{R}$ las siguientes afirmaciones son equivalentes:

1. $F \in DAM(H_\xi)$.

2. Existe una función medible positiva $a(\cdot)$ tal que para $1 + \xi x > 0$,

$$\lim_{u \uparrow x_F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0, \\ e^{-x} & \text{si } \xi = 0. \end{cases} \quad (1.26)$$

3. Para $x, y > 0, y \neq 1$,

$$\lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1} & \text{si } \xi \neq 0, \\ \frac{\ln x}{\ln y} & \text{si } \xi = 0. \end{cases} \quad (1.27)$$

Observación 3 La condición 1.26 tiene una interpretación probabilística interesante. En efecto, sea X una variable aleatoria con f.d. $F \in \text{DAM}(H_\xi)$, entonces 1.26 puede ser reescrito como

$$\lim_{u \uparrow x_F} P\left(\frac{X - u}{a(u)} > x \mid X > u\right) = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0, \\ e^{-x} & \text{si } \xi = 0. \end{cases} \quad (1.28)$$

de aquí que 1.28 proporciona una aproximación a la distribución de los excesos sobre el umbral u . El factor de escala apropiado es $a(u)$.

1.3.1. Inferencia para la DGVE

La DGVE, H_θ , $\theta = (\xi, \mu, \psi) \in \mathbf{R} \times \mathbf{R} \times \mathbf{R}^+$, provee un modelo para la distribución de máximos normalizados de variables aleatorias *iid*. Al igual que si se tratara de alguna distribución de valores extremos, su aplicación consiste de separar los datos en bloques de igual longitud y ajustar la DGVE a los máximos en cada bloque. Generalmente el período de separación es de un año. El período de un año se escoge para compensar las estacionalidades dentro de ese año. De aquí que los datos originales pueden aparecer como

$$\begin{aligned} \mathbf{X}^{(1)} &= (X_1^{(1)}, \dots, X_s^{(1)}) \\ \mathbf{X}^{(2)} &= (X_1^{(2)}, \dots, X_s^{(2)}) \\ &\vdots \\ \mathbf{X}^{(n)} &= (X_1^{(n)}, \dots, X_s^{(n)}) \end{aligned}$$

donde se asume que los vectores $(\mathbf{X}^{(i)})$ son *iid*, pero dentro de cada vector $\mathbf{X}^{(i)}$ los componentes pueden ser dependientes. La muestra *iid* sobre la cual se realiza la inferencia consiste entonces de los valores

$$X_i = \max(X_1^i, \dots, X_s^i), \quad i = 1, \dots, n. \quad (1.29)$$

Por razones históricas y dado que s frecuentemente corresponde a un período de un año, la inferencia estadística para H_θ basada en los datos (1.29) se refiere como *ajuste de máximos anuales*.

Estimación de Máxima Verosimilitud

Supongamos que H_θ tiene densidad h_θ . Entonces la función de verosimilitud basada en los datos $\mathbf{X} = (X_1, \dots, X_n)$ está dada por

$$L(\theta; \mathbf{X}) = \prod_{i=1}^n h_\theta(X_i) I_{\{1+\xi(X_i-\mu)/\psi > 0\}}$$

Denotemos por $l(\theta; \mathbf{X}) = \ln L(\theta; \mathbf{X})$ a la función log-verosimilitud. El *estimador máximo verosímil (EMV)* de θ es igual a

$$\hat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta} l(\theta; \mathbf{X}),$$

es decir, $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ que maximiza $l(\theta; \mathbf{X})$ sobre un espacio parametral apropiado Θ . La log-verosimilitud para los parámetros de la DGVE cuando $\xi \neq 0$ es

$$\begin{aligned} l((\xi; \mu, \psi), \mathbf{X}) &= -n \ln \psi - (1 + 1/\xi) \sum_{i=1}^n \ln \left\{ 1 + \xi \left(\frac{x_i - \mu}{\psi} \right) \right\} \\ &\quad - \sum_{i=1}^n \left\{ 1 + \xi \left(\frac{x_i - \mu}{\psi} \right) \right\}; \end{aligned} \quad (1.30)$$

En el caso de $H_{0;\mu,\psi}$ la log-verosimilitud está dada por

$$l((0; \mu, \psi); \mathbf{X}) = -n \ln \psi - \sum_{i=1}^n \exp \left\{ - \left(\frac{X_i - \mu}{\psi} \right) \right\} - \sum_{i=1}^n \left(\frac{X_i - \mu}{\psi} \right). \quad (1.31)$$

Diferenciando la última función con respecto a μ y ψ tenemos las ecuaciones para el caso Gumbel:

$$\begin{aligned} 0 &= n - \sum_{i=1}^n \exp \left\{ - \frac{X_i - \mu}{\psi} \right\}, \\ 0 &= n + \sum_{i=1}^n \frac{X_i - \mu}{\psi} \left(\exp \left\{ - \frac{X_i - \mu}{\psi} \right\} - 1 \right) \end{aligned}$$

Claramente no existe una solución explícita para esas ecuaciones. La situación para H_ξ cuando $\xi \neq 0$ es aún más complicada, así que se requiere de métodos numéricos para obtener los EMV.

Una dificultad potencial con el uso de los métodos de verosimilitud para la DGVE tiene que ver con las condiciones de regularidad que son requeridas para que las propiedades asintóticas usuales asociadas con el estimador máximo verosímil sean válidas. Tales condiciones no son satisfechas por el

modelo de valores extremos generalizados debido a que los extremos de la DGVE son funciones de los valores parametrales: $\mu - \sigma/\xi$ es un extremo superior de la distribución cuando $\xi < 0$, y un extremo inferior cuando $\xi > 0$. Esta violación de las condiciones de regularidad usuales significa que la verosimilitud asintótica estándar resulta no ser automáticamente aplicable. En este caso se tienen los siguientes resultados:

- Cuando $\xi > -0,5$, los estimadores de máxima verosimilitud son regulares, en el sentido de que tienen las propiedades asintóticas usuales;
- Cuando $-1 < \xi < -0,5$, los estimadores de máxima verosimilitud se pueden obtener, pero no tienen las propiedades asintóticas estándar.
- Cuando $\xi < -1$, no pueden obtenerse los estimadores de máxima verosimilitud.

Sujeta a estas limitaciones, la distribución aproximada de $(\hat{\xi}, \hat{\mu}, \hat{\psi})$ es normal multivariada con media (ξ, μ, ψ) y matriz de varianzas-covarianzas igual a la inversa de la matriz de información observada evaluada en el estimador máximo verosímil. Los intervalos de confianza y otras formas de inferencia se obtienen inmediatamente de la normalidad aproximada del estimador.

Estimación del nivel de retorno

Como vimos antes, dado cualquier $p \in (0, 1)$, el cuantil x_p se define como $x_p = H_\theta^{-1}(1 - p)$. Un estimador natural para x_p , basado en X_1, \dots, X_n , está dado por

$$\hat{x}_p = H_{\hat{\theta}}^{-1}(1 - p).$$

Por la definición de H_θ tenemos que

$$\hat{x}_p = \begin{cases} \hat{\mu} - \frac{\hat{\psi}}{\hat{\xi}} \left(1 - (-\ln(1 - p))^{-\hat{\xi}}\right), & \text{para } \xi \neq 0, \\ \hat{\mu} - \hat{\psi} \ln\{-\ln(1 - p)\}, & \text{para } \xi = 0, \end{cases} \quad (1.32)$$

donde $H(x_p) = 1 - p$. En terminología común, x_p es el nivel de retorno asociado con el período de retorno $1/p$, esto es, se espera que el nivel x_p sea excedido en promedio una vez en $1/p$ años. De manera más precisa, x_p es excedido por el máximo anual en cualquier año particular con probabilidad p . Por ejemplo, el cuantil de probabilidad 0,95, $x_{0,95}$, se espera sea excedido una vez cada 20 años.

Dado que los cuantiles posibilitan que los modelos de probabilidad sean expresados en la escala de los datos, la relación entre el modelo de valores

extremos generalizado y sus parámetros se interpreta más fácilmente en términos de los cuantiles de la expresión (1.32). En particular, definiendo $y_p = -\ln(1-p)$,

$$\hat{x}_p = \begin{cases} \hat{\mu} - \frac{\hat{\psi}}{\hat{\xi}}(1 - y_p^{-\hat{\xi}}), & \text{para } \hat{\xi} \neq 0, \\ \hat{\mu} - \hat{\psi} \ln y_p, & \text{para } \hat{\xi} = 0, \end{cases}$$

se sigue que si se grafica x_p contra y_p en escala logarítmica la gráfica es lineal en el caso $\xi = 0$. Si $\xi < 0$ la gráfica es convexa con límite asintótico cuando $p \rightarrow 0$ en $\mu - \psi/\xi$; si $\xi > 0$ la gráfica es cóncava y no tiene cota finita. Ésta es una gráfica del nivel de retorno. Debido a la simplicidad de interpretación y a que la escala *comprime* la cola de la distribución se enfatiza el efecto de la extrapolación, los gráficos del nivel de retorno son particularmente convenientes tanto para presentación del modelo como para validación. Véase por ejemplo Figura 2.7, Capítulo 2.

Por otro lado tenemos que, por el método delta (Serfling, 1980),

$$\text{Var}(\hat{x}_p) \approx \nabla x_p^T V \nabla x_p, \quad (1.33)$$

donde V es la matriz de varianzas-covarianzas de $(\hat{\xi}, \hat{\mu}, \hat{\psi})$ y

$$\begin{aligned} \nabla x_p^T &= \left[\frac{\partial x_p}{\partial \xi}, \frac{\partial x_p}{\partial \mu}, \frac{\partial x_p}{\partial \psi} \right] \\ &= [\sigma \xi^{-2}(1 - y_p^{-\xi}) - \sigma \xi^{-1} y_p^{-\xi} \ln y_p, 1, -\xi^{-1}(1 - y_p^{-\xi})] \end{aligned}$$

evaluada en $(\hat{\xi}, \hat{\mu}, \hat{\psi})$.

Es usual que períodos de retorno largos, correspondan a valores pequeños de p , que son los de mayor interés. Si $\hat{\xi} < 0$ es posible hacer inferencia sobre el extremo derecho de la distribución, el cual es el *período de retorno infinito*, correspondiente a x_p con $p = 0$. El estimador máximo verosímil es $\hat{z}_0 = \hat{\mu} - \hat{\sigma}/\hat{\xi}$, y (1.33) es igual a

$$\nabla x_0^T = [\sigma \hat{\xi}^{-2}, 1, -\hat{\xi}^{-1}],$$

evaluada en $(\hat{\xi}, \hat{\mu}, \hat{\psi})$. Cuando $\hat{\xi} \geq 0$ el estimador máximo verosímil del extremo derecho es infinito.

Se requiere tener un cuidado especial cuando se interpreten los niveles de retorno, especialmente para niveles que corresponden a períodos de retorno grandes. Esto es porque la aproximación normal a la distribución del estimador máximo verosímil puede ser pobre. Mejores aproximaciones generalmente se obtienen haciendo uso de la función de verosimilitud perfil.

Verosimilitud Perfil

Un método alternativo, y usualmente más preciso en la estimación de parámetros de interés por separado de los restantes, llamados parámetros incidentales o de estorbo, está basado en la verosimilitud perfil. La log-verosimilitud para θ puede escribirse formalmente como $l(\theta_i, \theta_{-i})$, donde θ_{-i} denota todos los componentes de θ excluyendo θ_i , el parámetro de interés que podría ser por ejemplo el parámetro de forma ξ o un cuantil x_p . La log-verosimilitud perfil para el parámetro de interés θ_i se define como

$$l_p(\theta_i) = \max_{\theta_{-i}} l(\theta_i, \theta_{-i}).$$

Esto es, para cada valor de θ_i , la log-verosimilitud perfil es la log-verosimilitud maximizada con respecto a todos los otros componentes de θ . En otras palabras, $l_p(\theta_i)$ es el perfil de la superficie de la log-verosimilitud $l(\theta)$ vista desde el eje de las θ_i .

Esta definición se generaliza a la situación donde θ puede partitionarse en dos componentes, $(\theta^{(1)}, \theta^{(2)})$, de la cual $\theta^{(1)}$ es el vector de dimensión k de interés y $\theta^{(2)}$ corresponde a los restantes $(d - k)$ componentes. La log-verosimilitud perfil para $\theta^{(1)}$ se define ahora como

$$l_p(\theta^{(1)}) = \max_{\theta^{(2)} | \theta^{(1)}} l(\theta^{(1)}, \theta^{(2)}).$$

El siguiente resultado, conduce a un procedimiento para aproximar inferencias de los parámetros de interés, $\theta^{(1)}$.

Teorema 1.3.2 *Sea $\hat{\theta}_0$ el estimador máximo verosímil del parámetro d -dimensional $\theta_0 = (\theta^{(1)}, \theta^{(2)})$ bajo el modelo estadístico $h(\mathbf{X}; \theta_0)$, donde $\theta^{(1)}$ es un subconjunto de dimensión k de θ_0 . Entonces, bajo condiciones adecuadas de regularidad*

$$D_p(\theta^{(1)}) = 2\{l(\hat{\theta}_0) - l_p(\theta^{(1)})\} \xrightarrow{d} \chi_k^2.$$

El Teorema 1.3.2 se usa de tal manera que para un parámetro unidimensional θ_i , $C_\alpha = \{\theta_i : D_p(\theta_i) \leq c_\alpha\}$ es un intervalo de confianza de $(1 - \alpha)$, donde c_α es el cuantil $(1 - \alpha)$ de la distribución χ_1^2 . Para más detalles consultar Cox y Barndoff-Nielsen, página 89 (1989)

En el caso que nos interesa, la evaluación numérica de la verosimilitud perfil de cualquiera de los parámetros individuales de la DGVE se obtiene de manera inmediata. Por ejemplo, para obtener la verosimilitud perfil para ξ , fijamos $\xi = \xi_0$ en una rejilla fina de valores ξ_0 en un intervalo que incluya

al cero, por ejemplo $(-0,3,0,3)$, y maximizamos la log-verosimilitud (1.30) con respecto a los parámetros restantes, μ y ψ . Los valores maximizados correspondientes de la log-verosimilitud constituyen la verosimilitud perfil para ξ , de la cual el Teorema 1.3.2 nos conduce a la obtención de intervalos de confianza aproximados.

Esta metodología se puede aplicar cuando se requiere hacer inferencia sobre alguna combinación de parámetros. Cuando queremos obtener intervalos de confianza para un nivel de retorno especificado x_p , se requiere de una reparametrización del modelo generalizado de valores extremos, así que x_p es uno de los modelos parametrales, del cual se obtiene la log-verosimilitud perfil por maximización con respecto a los parámetros restantes en la manera usual.

Nótese que la aplicación del Teorema 1.3.2 puede producir intervalos de confianza asimétricos con respecto al estimador de máxima verosimilitud, lo cual refleja adecuadamente la posible asimetría de la superficie global de verosimilitud y de una estimación mucho más precisa del parámetro de interés. Esta es una gran ventaja frente al método tradicionalmente usado en teoría de valores extremos, el método delta (Castillo, 1980) que siempre arroja intervalos simétricos alrededor del estimado máximo verosímil y puede subestimar y excluir valores grandes y plausibles para el parámetro de interés.

1.4. Excesos sobre un umbral (Series de duración parcial)

La modelación de máximos anuales puede resultar en un desperdicio de información si disponemos de otros datos extremos en lugar de tomar sólo, por ejemplo, los máximos anuales. Por esta razón a continuación se expone otro enfoque para analizar valores extremos.

Antes de continuar, primero definamos la función de distribución de excesos.

Definición 1.4.1 (Función de distribución de excesos, función media de excesos)

Sea X una v.a. con f.d. F y extremo derecho x_F . Para $u < x_F$,

$$F_u(x) = P(X - u \leq x | X > u), \quad x \leq 0, \quad (1.34)$$

es la f.d. de excesos de la v.a. X sobre el umbral u . La función

$$e(u) = E(X - u | X > u)$$

es la función media de excesos de X .

En diferentes campos, F_u se conoce también como *f.d. de tiempo de vida residual* (confiabilidad) o como *f.d. de excesos de pérdida* (seguros).

Si F fuera conocida, la distribución de los excesos sobre un umbral en 1.34 también sería conocida. Dado que en aplicaciones prácticas este no es el caso, se requiere de aproximaciones distribucionales de los excesos, al igual que el caso de la DVEG la cual se usa como una aproximación a la distribución de los máximos. El siguiente teorema nos da la caracterización asintótica para modelar los excesos sobre un umbral.

Teorema 1.4.1 (La distribución generalizada de Pareto (DGP))

Sea X_1, \dots, X_n va's iid con f.d. $F \in \text{DAM}(H_\xi)$ para algún $\xi \in \mathbf{R}$. Escojamos un umbral u y denotemos por

$$N_u = \text{card}\{i : i = 1, \dots, n, X_i > u\}$$

el número de excedencias de u por X_1, \dots, X_n . Denotemos los correspondientes excesos por Y_1, \dots, Y_{N_u} . La f.d. de excesos está dada por

$$G_\xi(y) = \begin{cases} 1 - (1 + \xi y)^{-1/\xi} & \text{si } \xi \neq 0, \\ 1 - e^{-y} & \text{si } \xi = 0, \end{cases} \quad y \in D(\xi, \beta), \quad (1.35)$$

donde

$$D(\xi, \beta) = \begin{cases} [0, \infty) & \text{si } \xi \geq 0, \\ [0, -1/\xi] & \text{si } \xi < 0. \end{cases}$$

La relación 1.35 es llamada la distribución generalizada de Pareto (DGP). La familia de localización-escala $G_{\xi, \nu, \beta}$ se obtiene reemplazando el argumento y por $(y - \nu)/\beta$ con $\nu \in \mathbf{R}$, $\beta > 0$. También nos referiremos a $G_{\xi, \nu, \beta}$ como DGP.

Al igual que en el caso H_0 , G_0 se puede interpretar como el límite de G_ξ cuando $\xi \rightarrow 0$. Denotemos

$$G_{\xi, \beta}(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\beta}\right)^{-1/\xi} & \text{si } \xi \neq 0, \\ 1 - \exp^{-y/\beta} & \text{si } \xi = 0 \end{cases}, \quad y \in D(\xi, \beta), \quad (1.36)$$

donde

$$D(\xi, \beta) = \begin{cases} [0, \infty) & \text{si } \xi \geq 0, \\ [0, -\beta/\xi] & \text{si } \xi < 0. \end{cases}$$

Además

$$\beta = \psi + \xi(u - \mu). \quad (1.37)$$

Lo anterior implica que, si los máximos tienen DVEG, entonces los excesos sobre un umbral tienen una distribución aproximada dentro de la familia Generalizada de Pareto. Además los parámetros de la DGP están determinados de manera única por aquéllos asociados a la DVEG de los máximos. En particular, el parámetro ξ en 1.37 es igual al correspondiente en la DVEG.

Justificación del modelo de Pareto Generalizado

Sea X_1, \dots, X_n va's iid con f.d. F , para n suficientemente grande,

$$F^n(x) \approx \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\psi} \right) \right]^{-1/\xi} \right\}$$

para $\mu, \psi > 0$ y ξ . De aquí,

$$n \log F(x) \approx - \left[1 + \xi \left(\frac{x - \mu}{\psi} \right) \right]^{-1/\xi}. \quad (1.38)$$

Pero para valores de x grandes, una expansión de Taylor implica que

$$\log F(x) \approx -\{1 - F(x)\}.$$

Substituyendo en (1.38), se sigue que

$$1 - F(u) \approx \frac{1}{n} \left[1 + \xi \left(\frac{u - \mu}{\psi} \right) \right]^{-1/\xi}$$

para u grande. De manera similar, para $y > 0$,

$$1 - F(u + y) \approx \frac{1}{n} \left[1 + \xi \left(\frac{u + y - \mu}{\psi} \right) \right]^{-1/\xi}. \quad (1.39)$$

De aquí que

$$\begin{aligned} P\{X > u + y | X > u\} &\approx \frac{n^{-1} [1 + \xi(u + y - \mu)/\psi]^{-1/\xi}}{n^{-1} [1 + \xi(u - \mu)/\psi]^{-1/\xi}} \\ &= \left[1 + \frac{\xi(u + y - \mu)/\psi}{1 + \xi(u - \mu)/\psi} \right]^{-1/\xi} \\ &= \left[1 + \frac{\xi y}{\beta} \right], \end{aligned} \quad (1.40)$$

donde

$$\beta = \psi + \xi(u - \mu)$$

(1.40) es la cola de la DGP.

Selección del umbral

Los eventos extremos se identifican definiendo un umbral alto u , para el cual las excedencias son $\{X_i : X_i > u\}$. Como antes, definamos los excesos por $\{Y_j = X_j - u : X_j > u\}$. Por el teorema 1.4.1, los Y_j tienen una distribución que puede ser aproximada por la DGP. La inferencia consiste en ajustar la DGP a los excesos observados, seguida de la verificación del modelo y extrapolación.

El problema de seleccionar un umbral tiene que ver con encontrar un balance entre el sesgo y la varianza en la estimación de los parámetros de interés. En este caso, elegir un umbral demasiado bajo violaría la base asintótica del modelo, conduciendo al sesgo; la elección de un umbral demasiado alto conduciría a una varianza muy grande pues se incluye un menor número de observaciones (Coles, 2001). En la práctica, es usual adoptar un umbral tan bajo como sea posible sujeto a que se cumpla con la aproximación a la DGP. Para este propósito disponemos de dos métodos: uno es una técnica exploratoria llevada a cabo previo al ajuste del modelo; la segunda es una evaluación de la estabilidad de los parámetros estimados, basada en el ajuste

del modelo a través de diferentes umbrales.

El primer método está basado en la linealidad de la función media de excesos $e(u)$ de la DGP. Para una *v.a.* X con *f.d.* $G_{\xi, \beta}$,

$$e(u) = E(X - u | X > u) = \frac{\beta + \xi u}{1 - \xi}, \quad u \in D(\xi, \beta), \xi < 1,$$

de aquí $e(u)$ es función lineal de u . Por otro lado, la función media de excesos de una muestra dada X_1, \dots, X_n está definida como

$$e_n(u) = \frac{1}{N_u} \sum_{i \in \Delta_n(u)} (X_i - u), \quad u > 0,$$

donde como antes $N_u = \text{card}\{i : i = 1, \dots, n, X_i > u\} = \text{card}\Delta_n(u)$. Este resultado sugiere una manera gráfica de elegir u : se escoge $u > 0$ tal que $e_n(x)$ sea aproximadamente lineal para $x \geq u$. Una gráfica de la función media de excesos empírica consistiría de los puntos

$$\left\{ \left(u, \frac{1}{N_u} \sum_{i \in \Delta_n(u)} (X_i - u) \right) \right\}.$$

Se pueden agregar intervalos de confianza a la gráfica basándonos en la normalidad aproximada de las medias muestrales.

El segundo método consiste en ajustar la DGP a diferentes umbrales y mirar la estabilidad de los parámetros estimados. De acuerdo con el Teorema 1.4.1, si la DGP es un modelo adecuado para los excesos de un umbral u_0 , entonces los excesos de un umbral más alto u también debería seguir una DGP. Los parámetros de las distribuciones son idénticas. Sin embargo, denotemos por β_u el valor del parámetro de escala de la DGP para un umbral $u > u_0$, y de (1.37) se sigue que

$$\beta_u = \beta_{u_0} + \xi(u - u_0), \quad (1.41)$$

de manera que el parámetro de escala cambia con u , a menos que $\xi = 0$. Para resolver esto podemos reparametrizar el parámetro de escala como

$$\beta^* = \beta_u - \xi u,$$

el cual es constante con respecto a u . De aquí que β^* y ξ deberían ser constantes sobre u_0 , si u_0 es un umbral válido para los excesos con una DGP.

Por lo tanto, debemos graficar $\hat{\beta}^*$ y $\hat{\xi}$ contra u , junto con sus intervalos de confianza, y seleccionar u_0 como el valor más pequeño de u para el cual

los estimadores permanecen constantes. Los intervalos de confianza para $\hat{\xi}$ se obtienen directamente de la matriz de varianzas-covarianzas V . Los intervalos de confianza para $\hat{\beta}^*$ requieren el método delta, usando

$$\text{Var}(\beta^*) \approx \nabla \beta^{*T} V \nabla \beta^*,$$

donde

$$\nabla \beta^{*T} = \left[\frac{\partial \beta^*}{\partial \beta_u}, \frac{\partial \beta^*}{\partial \xi} \right] = [1, -u].$$

1.4.1. Inferencia para la DGP

Estimación por Máxima Verosimilitud

Habiendo determinado el umbral, los parámetros de la DGP pueden ser estimados por máxima verosimilitud. Supongamos que F es DGP con parámetros ξ y β , así que la densidad f está dada por

$$f(y) = \frac{\xi}{\beta} \left(1 + \xi \frac{y}{\beta}\right)^{-1/\xi-1}, \quad y \in D(\xi, \beta).$$

La función de log-verosimilitud es igual a

$$l((\xi, \beta); Y) = -n \ln \beta - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \ln \left(1 + \frac{\xi}{\beta} Y_i\right). \quad (1.42)$$

Note que los argumentos de la función tienen que satisfacer la restricción $Y_i \in D(\xi, \beta)$, de otro modo, $l((\xi, \beta); Y) = -\infty$. En el caso $\xi = 0$ la log-verosimilitud es

$$l(\beta; Y) = -n \ln \beta - \frac{1}{\beta} \sum_{i=1}^n Y_i.$$

Para resolver las ecuaciones de log-verosimilitud requerimos nuevamente de métodos numéricos. Estos métodos trabajan bien si $\xi > -1/2$, y en este caso se puede demostrar que

$$n^{1/2} \left(\hat{\xi}_n - \xi, \frac{\hat{\beta}_n}{\beta} - 1 \right) \xrightarrow{d} N(0, M^{-1}), \quad n \rightarrow \infty,$$

donde

$$M^{-1} = (1 + \xi) \begin{pmatrix} 1 + \xi & 1 \\ 1 & 2 \end{pmatrix}.$$

Las propiedades usuales de los EMV como consistencia y eficiencia asintótica se sostienen (Coles, 2001).

Niveles de retorno

Como antes, resulta más conveniente interpretar los modelos de valores extremos en términos de los cuantiles o niveles de retorno, en lugar de los parámetros individuales. Supongamos que la DGP con parámetros ξ y β es un modelo adecuado para los excesos de un umbral u , esto es,

$$P\{X > x | X > u\} = \left[1 + \xi \left(\frac{x-u}{\beta}\right)\right]^{-1/\xi}$$

de aquí que

$$P(X > x) = \zeta_u \left[1 + \xi \left(\frac{x-u}{\beta}\right)\right]^{-1/\xi} \quad (1.43)$$

donde $\zeta = P\{X > u\}$. El nivel x_m que es excedido en promedio una vez en m observaciones es la solución de

$$P(X > x_m) = \frac{1}{m},$$

de aquí que

$$\zeta_u \left[1 + \xi \left(\frac{x_m - u}{\beta}\right)\right]^{-1/\xi} = \frac{1}{m}. \quad (1.44)$$

Luego tendríamos,

$$x_m = u + \frac{\beta}{\xi} [(m\zeta_u)^\xi - 1], \quad (1.45)$$

m tiene que ser suficientemente grande para asegurar que $x_m > u$. En el caso $\xi = 0$,

$$x_m = u + \beta \ln(m\zeta_u). \quad (1.46)$$

Por construcción, x_m es el **nivel de retorno basado en m observaciones**. Si graficamos x_m de (1.45) y de (1.46) contra m en una escala logarítmica se producen las mismas características cualitativas como en las gráficas de nivel de retorno basadas en la DVEG: linealidad si $\xi = 0$; concavidad si $\xi > 0$; convexidad si $\xi < 0$.

La estimación de los niveles de retorno requieren la sustitución de los parámetros por sus valores estimados. Para β y ξ se tienen los EMV, para ζ_u , la probabilidad de que una observación individual exceda u , se tiene como estimador

$$\hat{\zeta}_u = \frac{N_u}{n},$$

la proporción muestral de puntos que exceden u . Dado que el número de excedencias de u sigue una distribución Binomial(n, ζ_u), $\hat{\zeta}_u$ es también el

EMV de ζ_u .

Por otro lado, un estimador del parámetro de localización está dado por:

$$\hat{v} = \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{N_u}{n} \right)^{\hat{\xi}} - 1 \right).$$

Los errores estándar o intervalos de confianza se pueden obtener por el método delta. La incertidumbre en el estimador de ζ_u se puede derivar de las propiedades de la distribución binomial, $Var(\hat{\zeta}_u) \approx \hat{\zeta}_u(1 - \hat{\zeta}_u)/n$, así que la matriz de varianzas-covarianzas de $(\hat{\zeta}_u, \hat{\beta}, \hat{\xi})$ es aproximadamente

$$V = \begin{bmatrix} \hat{\zeta}_u(1 - \hat{\zeta}_u)/n & 0 & 0 \\ 0 & v_{1,1} & v_{1,2} \\ 0 & v_{2,1} & v_{2,2} \end{bmatrix}$$

donde $v_{i,j}$ denota el (i, j) término de la matriz de varianzas-covarianzas de $\hat{\beta}$ y $\hat{\xi}$. De aquí por el método delta,

$$Var(\hat{x}_m) \approx \nabla x_m^T V \nabla x_m \quad (1.47)$$

donde

$$\begin{aligned} \nabla x_m^T &= \left[\frac{\partial x_m}{\partial \zeta_u}, \frac{\partial x_m}{\partial \beta}, \frac{\partial x_m}{\partial \xi} \right] \\ &= [\beta m^\xi \zeta_u^{\xi-1}, \xi^{-1} \{ (m\zeta_u)^\xi - 1 \}, \\ &\quad -\beta \xi^{-2} \{ (m\zeta_u)^\xi - 1 \} + \beta \xi^{-1} (m\zeta_u)^\xi \ln(m\zeta_u)], \end{aligned}$$

evaluada en $(\hat{\zeta}_u, \hat{\beta}, \hat{\xi})$.

Al igual que en el caso de la GEV, consideramos que se obtiene una mayor precisión en la estimación de los parámetros y niveles de retorno si consideramos la verosimilitud perfil. El caso β y ξ se obtiene de manera directa. Para el caso de los niveles de retorno se requiere de una reparametrización. De (1.45) y de (1.46) tenemos

$$\beta = \begin{cases} \frac{(x_m - u)^\xi}{(m\zeta_u)^\xi - 1}, & \text{si } \xi \neq 0; \\ \frac{x_m - u}{\ln(m\zeta_u)}, & \text{si } \xi = 0. \end{cases}$$

Fijando x_m , la sustitución en (1.42) nos lleva a una verosimilitud de un parámetro que puede ser maximizada con respecto a ξ . Como una función de x_m , esta es la log-verosimilitud perfil para el nivel de retorno basado en m observaciones.

1.5. Cluster de series

En esta sección nos proponemos reducir el número de estaciones agrupándolas de acuerdo a la intensidad de precipitación con el propósito de detectar unos pocos modelos representativos de este gran número de series. Debido a que los datos con los que estamos trabajando aparecen en un orden específico la aplicación directa de cualquier técnica de agrupación de datos afectaría los resultados de manera sustancial. De este modo, consideramos que la metodología más apropiada a usar es la formación de cluster en series de tiempo, la cual toma en cuenta precisamente dicha estructura de dependencia. Esta metodología consiste simplemente de ajustar un modelo *ARIMA* para cada una de las series, calcular los valores de π , que son los valores que caracterizan a la serie de tiempo; y posteriormente calcular la matriz de disimilitudes a la cual se puede aplicar alguna técnica de agrupamiento, en este caso la de cluster.

Antecedentes

La idea detrás de la metodología a aplicar es la de encontrar una medida de disimilitud entre series de tiempo, la cual tomará en cuenta la estructura de dependencias dentro de dichas series. En esta sección esbozaremos de manera general el procedimiento para agrupar series de tiempo, para mayor detalle sobre éste y otros métodos se puede consultar Domínguez (2000). Primeramente definimos algunos operadores que son de gran utilidad para definir los modelos.

El *operador de retraso* B se define como

$$B^m z_t = z_{t-m}.$$

También tenemos el *operador diferencia de retraso* ∇ , el cual se define por

$$\nabla^d z_t = (1 - B)^d z_t.$$

Finalmente tenemos el *operador diferencia estacional de retraso* ∇_s , dado por

$$\nabla_s^D z_t = (1 - B^s)^D z_t.$$

De este modo, sea a_t ruido blanco gaussiano y Z_t un proceso estocástico con media cero tal que

$$Z_t \sim \text{ARIMA}(p, d, q)(P, D, Q)_s, \quad (1.48)$$

dicho modelo se puede escribir de la siguiente manera

$$\varphi(B)Z_t = \theta(B)a_t \quad (1.49)$$

donde

$$\varphi(B) = \phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D$$

$$\theta(B) = \theta_q(B)\Theta_Q(B^s).$$

con

$$\phi_p(B) = 1 - \phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$$

$$\theta_q(B) = 1 - \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_P B^{Ps}.$$

Note que $\phi_p(\cdot)$ es una función, mientras que ϕ_p es un coeficiente. De igual modo es para el resto. Asumiendo que el modelo (1.49) es invertible, esto es, que las raíces del polinomio $\theta(B)$ estén fuera del círculo unitario, entonces $Z_t \in \mathcal{L}$, donde \mathcal{L} es la clase de modelos invertibles *ARIMA*.

Luego se sabe que si $Z_t \in \mathcal{L}$,

$$\begin{aligned} Z_t &= \bar{Z}_{t-1} + a_t \\ \bar{Z}_{t-1} &= \sum_{j=1}^{\infty} \pi_j Z_{t-j} = \{1 - \pi(B)\}Z_t \end{aligned}$$

donde \bar{Z}_{t-1} es independiente de a_t . Entonces

$$Z_t = \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \dots + a_t \quad (1.50)$$

donde el operador $AR(\infty)$ se define por $\pi(B) = \varphi(B)/\theta(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$.

Por lo tanto, dados valores iniciales y órdenes conocidos, cualquier proceso $Z_t \in \mathcal{L}$ es completamente caracterizado por la sucesión $\pi = (\pi_1, \pi_2, \dots)$, la cual especifica la distribución del proceso Z_t cuando a_t es un proceso gaussiano de ruido blanco. La sucesión π contiene toda la información útil acerca de la estructura del proceso estocástico ya que cualquier otra información necesaria para especificar Z_t es justamente a_t . Entonces una medida de la diversidad estructural entre $X_t \in \mathcal{L}$ y $Y_t \in \mathcal{L}$ se obtiene comparando sus

respectivas sucesiones de π .

Una métrica sobre \mathcal{L} puede definirse por la distancia

$$d(X, Y) = \left\{ \sum_{j=1}^{\infty} (\pi_{j,x} - \pi_{j,y})^2 \right\}^{1/2} \quad (1.51)$$

ya que $\sum_{j=1}^{\infty} \pi_j$, $\sum_{j=1}^{\infty} |\pi_j|$ y $\sum_{j=1}^{\infty} \pi_j^2$, son cantidades bien definidas, $d(X, Y)$ siempre existe para cualquier proceso en \mathcal{L} y satisface las propiedades clásicas de distancias, a saber, no negativa, simétrica y la desigualdad del triángulo. La definición de una distancia entre modelos de series de tiempo, permite en forma inmediata la aplicación de algoritmos de agrupamiento.

Capítulo 2

Aplicación

2.1. Ajuste de máximos anuales

Tomando como base el estudio del manejo de las aguas pluviales en la Zona Metropolitana de Morelia, Estado de Michoacán (Gutierrez, 2004), estamos interesados en realizar un estudio de valores extremos para la región y estimar el máximo nivel de precipitación que podría ocurrir en la región dentro de los próximos 100 o 1000 años.

De acuerdo con los especialistas, de 24 estaciones climatológicas que tienen influencia en la cuenca, sólo 18 son representativas para realizar el análisis, esto debido a la cantidad y calidad de información disponible en cada una de ellas. Por esta razón sólo consideramos dichas estaciones para realizar el estudio. El mapa de localización de las estaciones se muestra en la Figura 2.1 y la información a considerar se muestra en el Cuadro 2.1.

Los datos usados para el análisis fueron proporcionados por el Sistema Meteorológico Nacional a través de su página en internet y por el Instituto Mexicano de Tecnología del Agua (IMTA) a través de un CD con la base de datos ERIC II.

Para realizar el estudio de valores extremos procedemos de la siguiente manera. Primeramente hacemos un análisis basado en los máximos anuales en cada estación ajustando la DGVE. Una vez que se identifica el dominio de atracción del máximo a través del parámetro de forma de la DGVE se estiman los cuantiles de interés usando la DVE correspondiente.

En la Figura 2.2 presentamos la estimación del parámetro de forma ξ usando la log-verosimilitud perfil, de todas las estaciones meteorológicas. Se grafican todas las log-verosimilitudes entre 0 y 1 en el eje y para poder compararlas. A una altura de 0.1496 se pueden ubicar los intervalos de verosimilitud

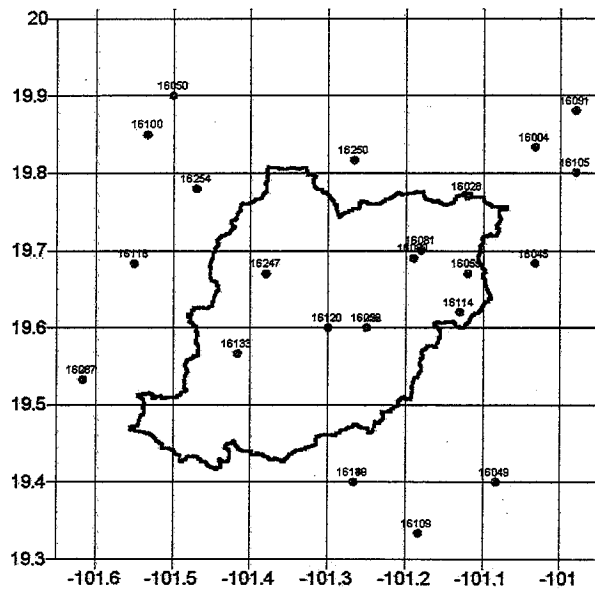


Figura 2.1: Ubicación de las estaciones climatológicas

Clave	Estación	Lluvia	Registro
16022	Cointzio	787.38	1940-2002
16028	Cuitzillo Gde.	609.21	1969-2001
16045	El Temascal, Charo	1312.61	1965-2003
16049	Etucuario, Madero	1058.79	1944-1988
16050	Huaniqueo	826.57	1948-2002
16055	Jesús de Monte	828.82	1940-2002
16081	Morelia DGE	786.15	1947-1986
16087	Patzcuaro, Patzcuaro	884.90	1969-1999
16091	Planta de Bombeo	623.73	1966-2003
16100	Puente San Isidro	801.26	1947-1993
16105	Quirio	671.56	1963-2003
16109	San Diego Curucupatzeo	1046.97	1940-2003
16114	San Miguel del Monte	937.35	1963-2003
16118	Santa F Quiroga	788.25	1963-2000
16120	Santiago Undameo	805.91	1953-1999
16133	Turicato, Turicato	829.58	1969-2002
16139	Villa Madero, Madero	1249.48	1943-1984
16140	Villa Madero, Madero	1195.76	1961-2000

Cuadro 2.1: Relación de estaciones climatológicas y período de registro. La columna lluvia se refiere a la lluvia media anual (mm)

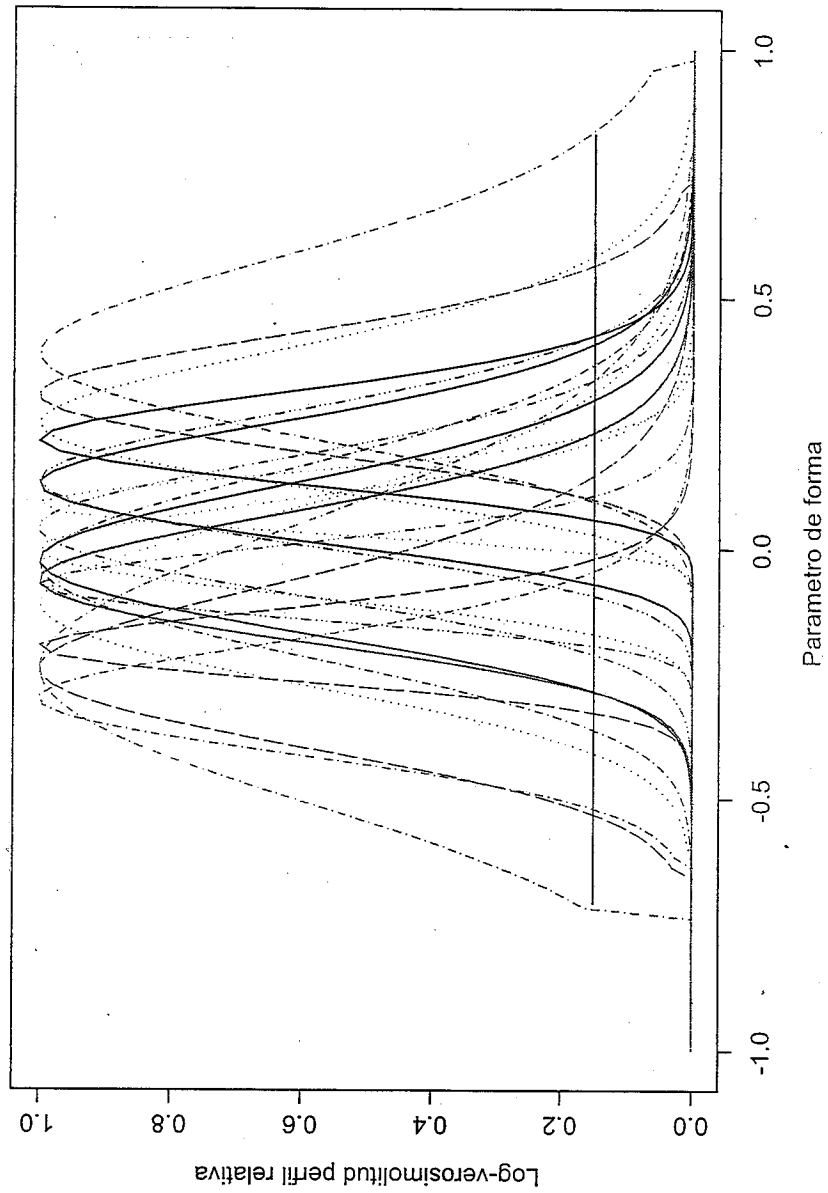


Figura 2.2: Análisis de máximos anuales de todas las estaciones

aproximados del 95 %.

Los parámetros estimados obtenidos por maximización de la log-verosimilitud de la DGVE se muestran en el Cuadro 2.2. Entre los paréntesis se muestran los errores estándar obtenidos bajo el supuesto de normalidad.

Clave	Estación	ψ	$\hat{\mu}$	$\hat{\xi}$
16022	Cointzio	37.0 (1.2)	8.1 (0.9)	0.214 (0.093)
16028	Cuitzillo Gde.	34.8 (2.1)	10.0 (1.5)	-0.110 (0.172)
16045	El Temascal, Charo	55.2 (2.8)	15.7 (2.0)	-0.300 (0.114)
16049	Etucuario, Madero	46.2 (1.8)	10.4 (1.5)	0.304 (0.117)
16050	Huaniqueo	40.9 (1.4)	8.8 (1.1)	0.133 (0.134)
16055	Jesús de Monte	27.4 (1.0)	6.5 (0.7)	0.126 (0.120)
16081	Morelia DGE	38.2 (1.7)	9.6 (1.2)	-0.004 (0.096)
16087	Patzcuaro, Patzcuaro	42.2 (2.9)	11.3 (2.3)	-0.239 (0.279)
16091	Planta de Bombeo	38.0 (1.8)	10.3 (1.2)	-0.197 (0.072)
16100	Puente San Isidro	41.2 (1.4)	7.8 (1.0)	0.040 (0.137)
16105	Quirio	35.5 (1.4)	7.7 (1.1)	-0.025 (0.148)
16109	San Diego Curucupatzeo	52.0 (2.6)	18.5 (1.9)	0.085 (0.059)
16114	San Miguel del Monte	39.6 (3.5)	16.5 (2.7)	-0.033 (0.187)
16118	Santa F Quiroga	30.7 (3.1)	15.9 (2.3)	-0.239 (0.165)
16120	Santiago Undameo	39.8 (1.4)	8.4 (0.9)	-0.073 (0.074)
16133	Turicato, Turicato	51.1 (2.5)	13.2 (1.8)	-0.064 (0.128)
16139	Villa Madero, Madero	54.8 (2.8)	15.4 (2.3)	0.246 (0.140)
16140	Villa Madero, Madero	53.1 (3.3)	15.7 (3.0)	0.387 (0.182)

Cuadro 2.2: Estimadores de máxima verosimilitud de ξ , μ y ψ y errores estándar

Sin embargo, dado que se obtiene mayor precisión usando la log-verosimilitud perfil, en el Cuadro 2.5 se muestran los intervalos obtenidos por este método para cada una de las estaciones. Además presentamos la plausibilidad de $\xi = 0$ en el modelo. Si resultara que dicho valor es mayor a 0,5, entonces preferimos el modelo Gumbel sobre las otras distribuciones para estimar los cuantiles. El modelo Gumbel es un modelo sencillo pues sólo tiene dos parámetros y las estimaciones derivadas de su ajuste son más confiables.

De acuerdo con este criterio las estaciones que se modelaron con el modelo Gumbel son: 16028, 16050, 16055, 16081, 16087, 16100, 16105, 16114, 16120 y 16133. Las estaciones que se modelaron con el modelo Fréchet son: 16022, 16049, 16109, 16139 y 16140. Finalmente las estaciones que se modelaron con el modelo Weibull son: 16045, 16091 y 16118.

Para simplificar esta parte del análisis, sólo presentaremos los resultados

de las estaciones climatológicas Cuitzillo Grande (16028), Presa Cointzio (16022) y Morelia DGE (16081). Son de especial interés estas estaciones pues resultan ser, de acuerdo con los especialistas, las estaciones más representativas de la cuenca. En la Figura 2.3 se presentan los máximos anuales para cada una de estas estaciones. En algunas estaciones ciertos años presentaron niveles de precipitación muy bajos y afectaban de manera visible el ajuste de la DGVE, en tales caso se realiza el ajuste incluyendo y excluyendo dichas observaciones para comparar y decidir cual proporciona una mejor estimación de los cuantiles.

Clave	Estación	Intervalo de verosimilitud	$R(\xi = 0)$
16022	Cointzio	(0,060,0,425)	0.010
16028	Cuitzillo Gde.	(-.410,0.310)	0.830
16045	El Temascal, Charo	(-.520,-0.050)	0.091
16049	Etucuario, Madero	(0.100,0.560)	0.012
16050	Huaniqueo	(-.090,0.420)	0.570
16055	Jesús de Monte	(-.070,0.400)	0.510
16081	Morelia DGE	(-.170,0.210)	1.000
16087	Patzcuaro, Patzcuaro	(-.720,0.330)	0.710
16091	Planta de Bombeo	(-.323,-.012)	0.100
16100	Puente San Isidro	(-.200,0.327)	0.960
16105	Quirio	(-.275,0.290)	0.985
16109	San Diego Curucupatzeo	(0.000,0.230)	0.230
16114	San Miguel del Monte	(-.360,0.365)	0.985
16118	Santa F Quiroga	(-.530,0.145)	0.440
16120	Santiago Undameo	(-.180,0.100)	0.670
16133	Turicato, Turicato	(-.280,0.232)	0.890
16139	Villa Madero, Madero	(0.025,0.580)	0.070
16140	Villa Madero, Madero	(0.100,0.830)	0.020

Cuadro 2.3: Intervalo de verosimilitud de 0.1496, el cual equivale a un intervalo aproximado del 95 % para el parámetro de forma ξ y verosimilitud relativa en $\xi = 0$.

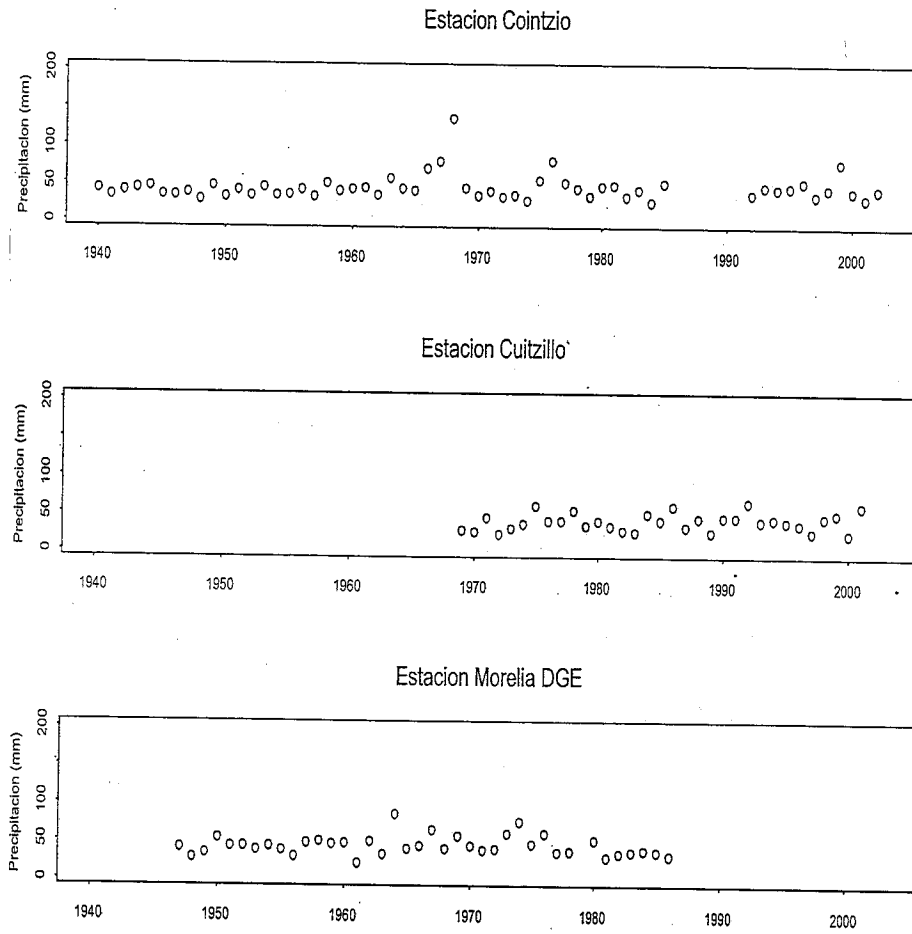


Figura 2.3: Máximos anuales de precipitación

2.1.1. Estación Cuitzillo

La estación Cuitzillo el Grande se ubica hacia el noreste de la cuenca de Morelia y cuenta con un registro de 33 años. El ajuste de la DGVE a los máximos anuales nos muestra los siguientes resultados. El intervalo de verosimilitud aproximado del 95 % para el parámetro de forma está dado por $[-0.41, 0.31]$. La distribución Weibull sería la distribución natural a usar en este caso como modelo para los máximos pues obtuvimos un valor de $\xi = -0.110$. Sin embargo, el valor de $\xi = 0$ tiene una plausibilidad muy alta de 0.83 como lo muestra la Figura 2.4, así que el modelo Gumbel puede ser una buena elección para ajustar a los datos. Una forma de ver cual es el mejor modelo es obteniendo la razón de verosimilitud entre los dos modelos, en este caso, si dicha razón es muy diferente de 1 entonces un modelo se ve más favorecido que el otro, así tenemos,

$$\frac{L_{Gumbel}(\hat{\theta}|\mathbf{X})}{L_{DGVE}(\hat{\theta}|\mathbf{X})} = 0,8355.$$

Dicho valor no es muy diferente de 1, así que no hay preferencia por uno de los modelos, sin embargo, en este caso es más recomendable usar la distribución Gumbel para estimar los cuantiles pues como comentamos antes se tiene mayor precisión por ser un modelo más sencillo.

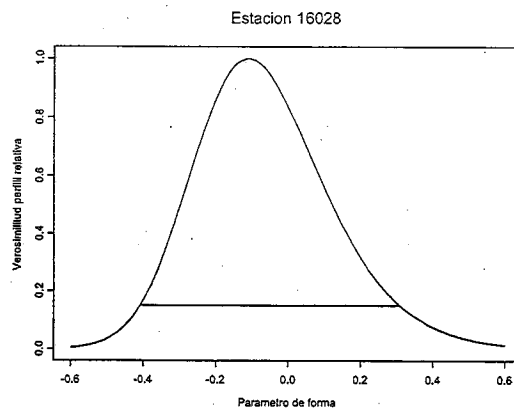


Figura 2.4: Verosimilitud perfil para ξ . A una altura de 0.1496 se muestra con una línea el intervalo de verosimilitud.

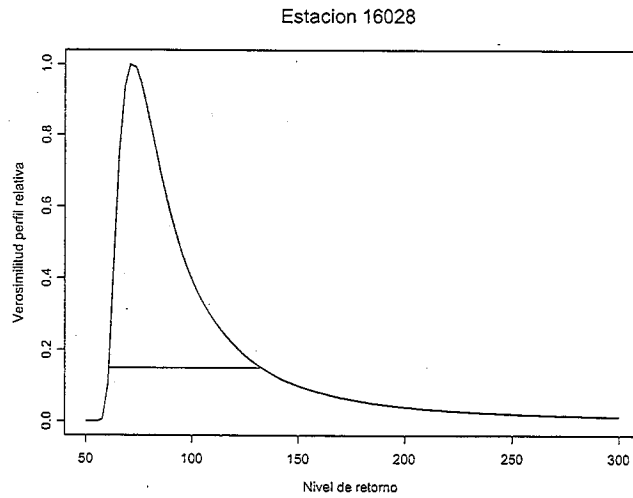


Figura 2.5: Verosimilitud perfil para el nivel de retorno de 100 años en la estación Cuitzillo (16028)

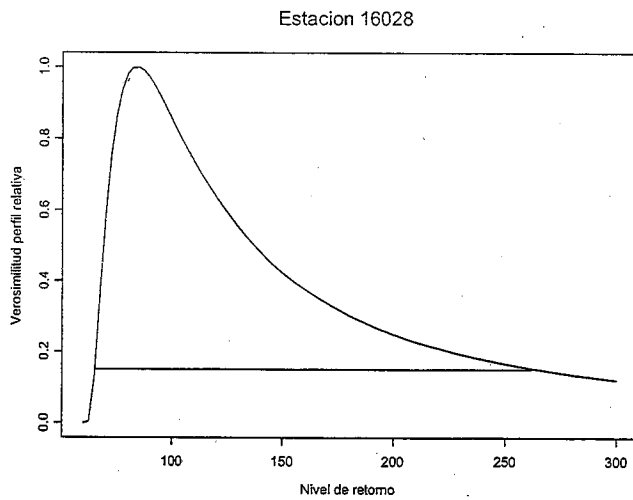


Figura 2.6: Verosimilitud perfil para el nivel de retorno de 1000 años en la estación Cuitzillo (16028)

Las Figuras 2.5 y 2.6 muestran la log-verosimilitud perfil para los niveles de retorno 100 y 1000 años, respectivamente. Los valores de $\hat{x}_{0,01}$ y $\hat{x}_{0,001}$ son 71.5 y 83.8 con intervalos de confianza aproximados del 95 % [60,132] y [65,262], respectivamente.

Las gráficas de diagnóstico para evaluar la precisión del modelo Gumbel se muestran en la Figura 2.7. Las gráficas de probabilidad y de cuantiles confirman la validez del modelo ajustado. El histograma no aporta mucha información en cuanto al ajuste del modelo.

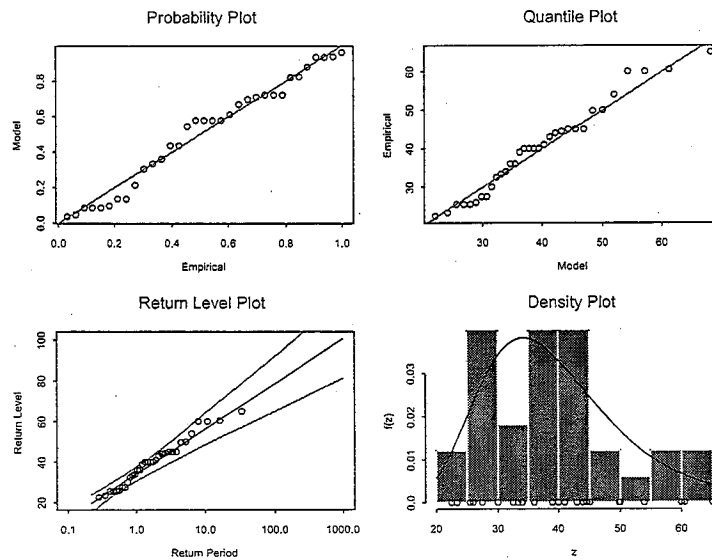


Figura 2.7: Gráficas de diagnóstico para el ajuste del modelo Gumbel de la Estación Cuitzillo (16028)

2.1.2. Estación Cointzio

La Estación Cointzio se ubica en la zona centro de la cuenca de Morelia y cuenta con un período de registro de 58 años. Existe una precipitación máxima registrada en la estación de 135.80 mm en el año de 1968, en contraste en 1986 se tiene un nivel de precipitación máximo de 3.7 mm. Dicho valor es demasiado bajo si lo comparamos con el resto de la serie, y no podríamos decir mucho acerca de su confiabilidad pues por ser tan bajo podría tratarse de un outlier, por tal razón decidimos eliminarlo del análisis pues

su inclusión nos llevaría a ajustar un modelo que no es del todo adecuado como veremos enseguida.

Exclusión de observaciones

En el Cuadro 2.4 se muestran los parámetros estimados considerando ambos casos. Evidentemente se trata de dos modelos diferentes, de incluir la observación de 1986 tendríamos que los datos se ajustan a un modelo Gumbel ($\xi \approx 0$), en caso excluirla tendríamos un ajuste del modelo Fréchet

Parámetros	Incluyendo 1986	No incluyendo 1986
ξ	-0.028	0.214
μ	36.652	37.019
ψ	13.294	8.106

Cuadro 2.4: Estimadores de máxima verosimilitud de la DGVE

($\xi > 0$). Este último modelo nos proporciona estimaciones más confiables sobre los cuantiles de interés pues está captando el comportamiento más cercano posible de los valores extremos, como se puede apreciar en las gráficas de diagnóstico de los dos modelos. De este modo, elegimos trabajar sin la observación de 1986.

El intervalo de confianza aproximado del 95 % obtenido por verosimilitud perfil nos da para $\xi = 0,214$, $[0,06, 0,425]$. La Figura 2.10 muestra la log-verosimilitud perfil para ξ .

La Figura 2.11 muestra la log-verosimilitud perfil para los niveles de retorno 100 y 1000 años, con valores de $\hat{x}_{0,01} = 100,5$ y $\hat{x}_{0,001} = 166$, con intervalos de confianza aproximados del 95 % $[78,165]$ y $[105,410]$, respectivamente.

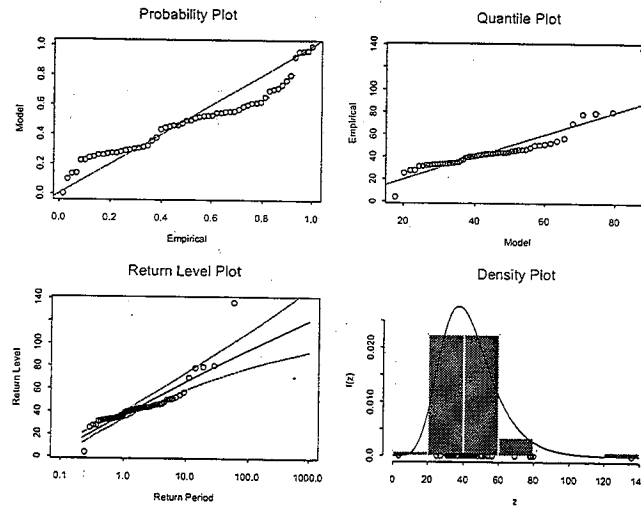


Figura 2.8: Gráficas de diagnóstico para el ajuste de la DGVE de la Estación Cointzio (16022), con 1986

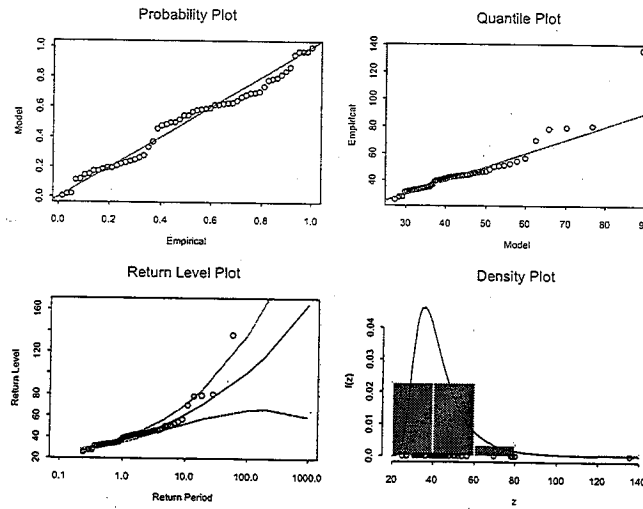


Figura 2.9: Gráficas de diagnóstico para el ajuste de la DGVE de la Estación Cointzio (16022), sin 1986

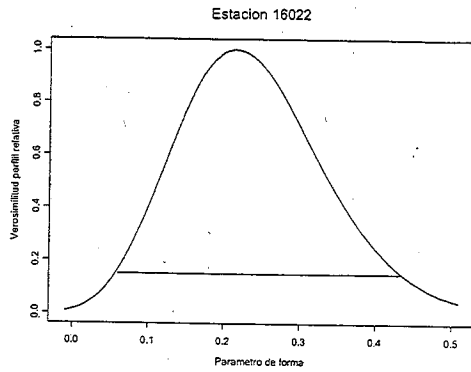


Figura 2.10: Verosimilitud perfil para ξ en la Estación Cointzio (16022)

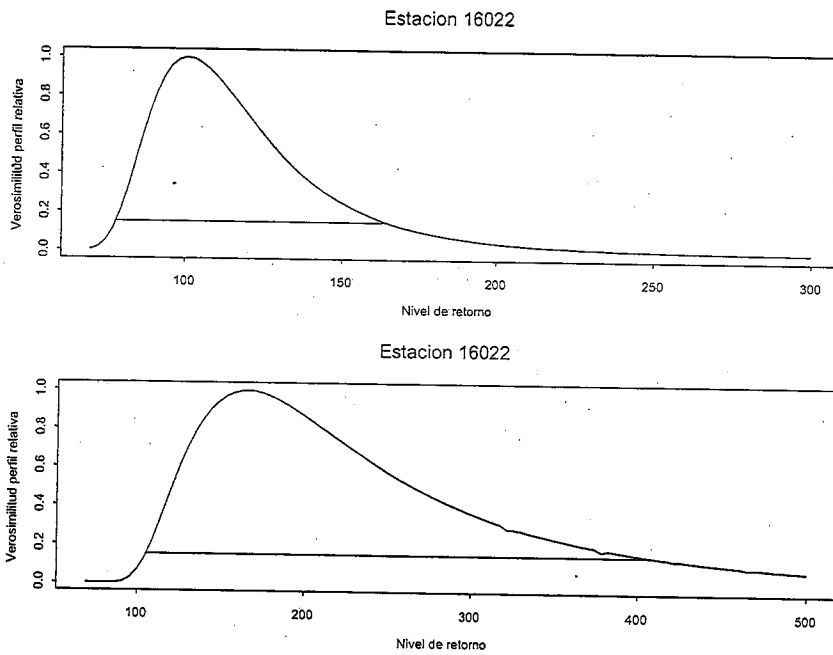


Figura 2.11: Verosimilitud perfil para el nivel de retorno de 100 y 1000 años en la Estación Cointzio (16022)

2.1.3. Estación Morelia

La Estación Morelia ubicada en la zona noreste de la cuenca y muy cercana a las estaciones arriba analizadas, cuenta con un período de registro de 40 años. Los resultados obtenidos de ajustar la DGVE a los máximos anuales se muestran a continuación. Tenemos un intervalo de verosimilitud aproximado del 95 % de $[-.170, 0.21]$ para $\xi = -0,004$. La Figura 2.12 muestra la log-verosimilitud perfil para ξ la cual es bastante simétrica alrededor del 0, de aquí que el modelo Gumbel es el modelo a usar para la estimación de los cuantiles.

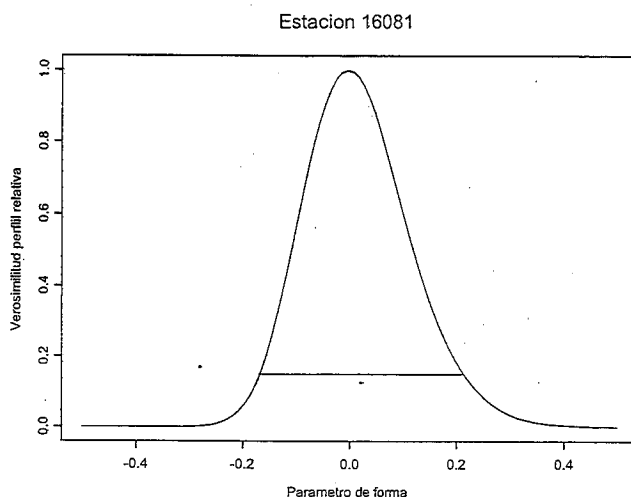


Figura 2.12: Verosimilitud perfil para ξ en la Estación Morelia (16081)

Las Figuras 2.13 y 2.14 muestran la log-verosimilitud perfil para los niveles de retorno 100 y 1000 años, respectivamente. Los valores de $\hat{x}_{0,01}$ y $\hat{x}_{0,001}$ son 82 y 103.54, con intervalos de confianza aproximados del 95 % $[70,120]$ y $[81,199]$, respectivamente. Las gráficas de diagnóstico para evaluar la precisión del modelo Gumbel se muestran en la Figura 2.15. Se tiene un buen ajuste del modelo Gumbel como lo muestran las gráficas de probabilidad y cuantiles.

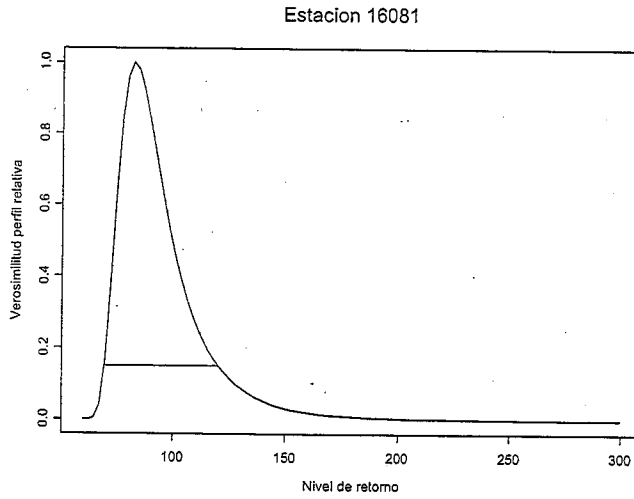


Figura 2.13: Verosimilitud perfil para el nivel de retorno de 100 años en la Estación Morelia (16081)

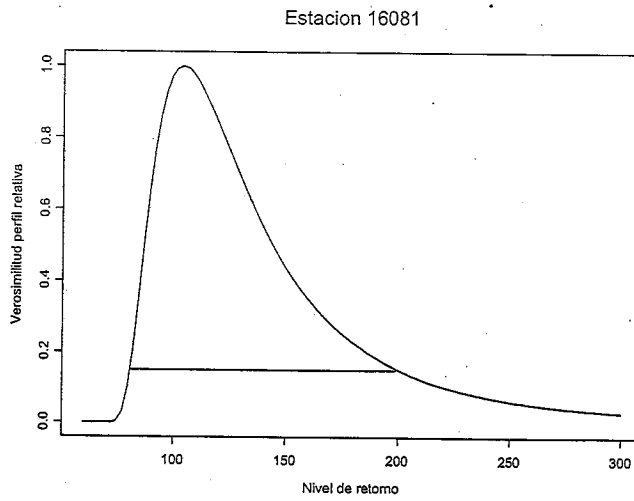


Figura 2.14: Verosimilitud perfil para el nivel de retorno de 1000 años en la Estación Morelia (16081)

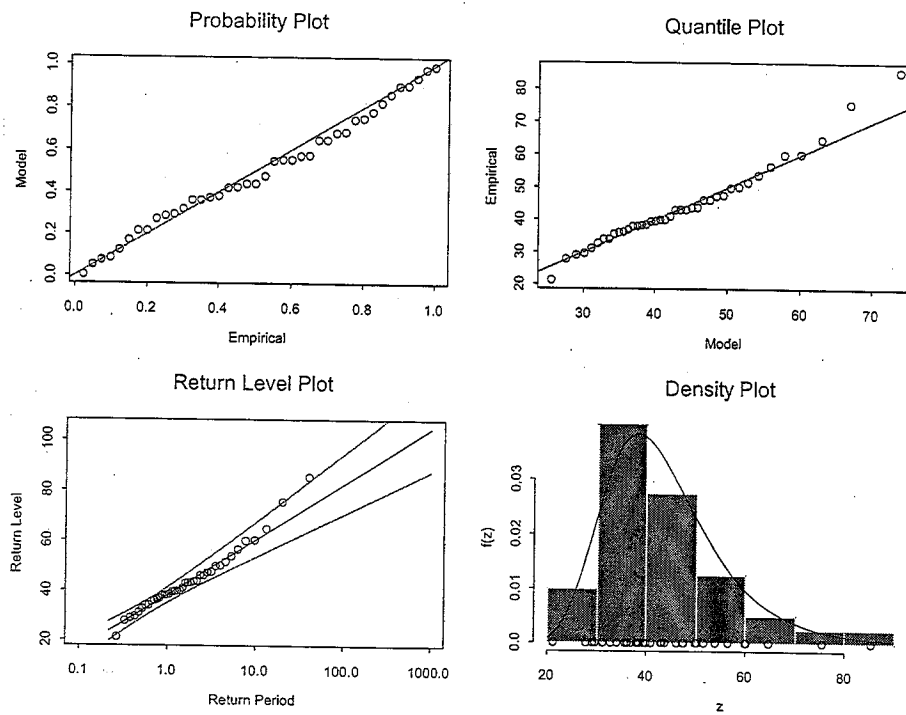


Figura 2.15: Gráficas de diagnóstico para el ajuste del modelo Gumbel de la Estación Morelia (16081)

2.2. Excesos sobre umbrales (Series de duración parcial)

La Figura 2.16 muestra la distribución temporal porcentual de la precipitación en la cuenca hidrológica de la ciudad de Morelia. Se observa que más del 90% de la precipitación se presenta durante los meses de mayo, junio, julio, agosto, septiembre y octubre. Debido a esto decidimos tomar sólo dichos meses en cada una de las estaciones para realizar el análisis de excesos sobre umbrales, además de que con este método se controla el efecto de estacionalidad dentro de cada año.

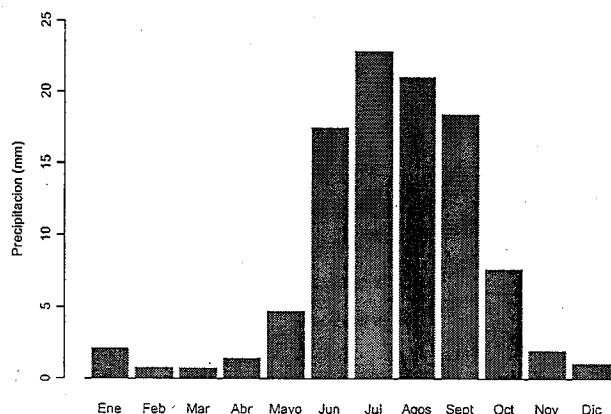


Figura 2.16: Distribución porcentual de la lluvia mensual en la cuenca hidrológica de la ciudad de Morelia calculada con base en las 18 estaciones

Para cada estación se hizo un ajuste de la DGP a los excesos sobre cierto umbral, el cual varió de estación a estación. Los parámetros estimados obtenidos por maximización de la log-verosimilitud de la DGP se muestran en el Cuadro 2.6. Se muestran también los errores estándar bajo el supuesto de normalidad. Dado que los intervalos de verosimilitud son más confiables que los obtenidos por el método delta, en el Cuadro 2.6 se muestran los intervalos de verosimilitud para el parámetro de forma, además se presenta la plausibilidad de $\xi = 0$ y donde $R(\xi = 0) > 0,5$ se modelan los datos una DGP donde ξ es igual a cero. En la Figura 2.17 presentamos la verosimili-

Clave	Estación	$\hat{\sigma}$	$\hat{\xi}$	Umbral	Tasa
16022	Cointzio	7.2(0.9)	0.198(0.093)	30.40	1.24
16028	CuitzilloGde.	10.7(2.0)	-0.169(0.131)	30.40	1.03
16045	ElTemascal,Charo	14.3(1.2)	-0.113(0.059)	33.15	3.59
16049	Etucuario,Madero	7.9(0.7)	0.166(0.070)	29.40	3.17
16050	Huaniqueo	9.8(1.4)	-0.002(0.101)	35.50	1.18
16055	JesúsdeMonte	5.1(0.8)	0.086(0.113)	30.00	0.87
16081	MoreliaDGE	9.0(0.9)	0.011(0.071)	24.50	2.46
16087	Patzcuaro,Patzcuaro	14.4(2.7)	-0.232(0.145)	29.00	1.52
16091	PlantadeBombeo	10.3(1.0)	-0.127(0.060)	25.00	2.35
16100	PuenteSanIsidro	10.3(1.0)	-0.091(0.069)	27.00	2.32
16105	Quirio	9.1(1.1)	-0.085(0.087)	25.00	2.02
16109	SanDiegoCurucupatzeo	13.3(0.8)	0.079(0.036)	30.00	4.33
16114	SanMigueldeMonte	15.4(1.5)	-0.103(0.068)	25.00	3.41
16118	SantaFQuiroga	13.1(1.9)	-0.247(0.094)	30.00	1.27
16120	SantiagoUndameo	10.1(0.9)	-0.091(0.052)	25.00	2.70
16133	Turicato,Turicato	12.2(1.2)	-0.061(0.065)	30.00	3.32
16139	VillaMadero,Madero	11.8(1.1)	0.127(0.067)	35.00	2.88
16140	VillaMadero,Madero	12.8(1.2)	0.113(0.061)	30.00	3.68

Cuadro 2.5: Estimadores de máxima verosimilitud de ξ , σ , entre paréntesis errores estándar, umbral y tasa de excedencia

Clave	Estación	Intervalo de verosimilitud	$R(\xi = 0)$
16022	Cointzio	(0,050, 0,415)	0.020
16028	Cuitzillo Gde.	(-.390,0.150)	0.525
16045	El Temascal, Charo	(-.210,0.020)	0.240
16049	Etucuaró, Madero	(0.050,0.330)	0.008
16050	Huaniqueo	(-.160,0.023)	1.000
16055	Jesús de Monte	(-.080,0.370)	0.700
16081	Morelia DGE	(-.103,0.175)	0.980
16087	Patzcuaro, Patzcuaro	(-.480,0.100)	0.360
16091	Planta de Bombeo	(-.220,0.020)	0.240
16100	Puente San Isidro	(-.200,0.060)	0.490
16105	Quirio	(-.230,0.120)	0.670
16109	San Diego Curucupatzco	(0.020,0.160)	0.040
16114	San Miguel del Monte	(-.210,0.050)	0.410
16118	Santa F Quiroga	(-.400,-0.020)	0.120
16120	Santiago Undameo	(-.170,0.030)	0.340
16133	Turicato, Turicato	(-.170,0.080)	0.685
16139	Villa Madero, Madero	(0.015,0.280)	0.050
16140	Villa Madero, Madero	(0.020,0.250)	0.080

Cuadro 2.6: Intervalo de verosimilitud de 0.1496, el cual equivale a un intervalo aproximado del 95% para el parámetro de forma ξ y verosimilitud relativa en $\xi = 0$.

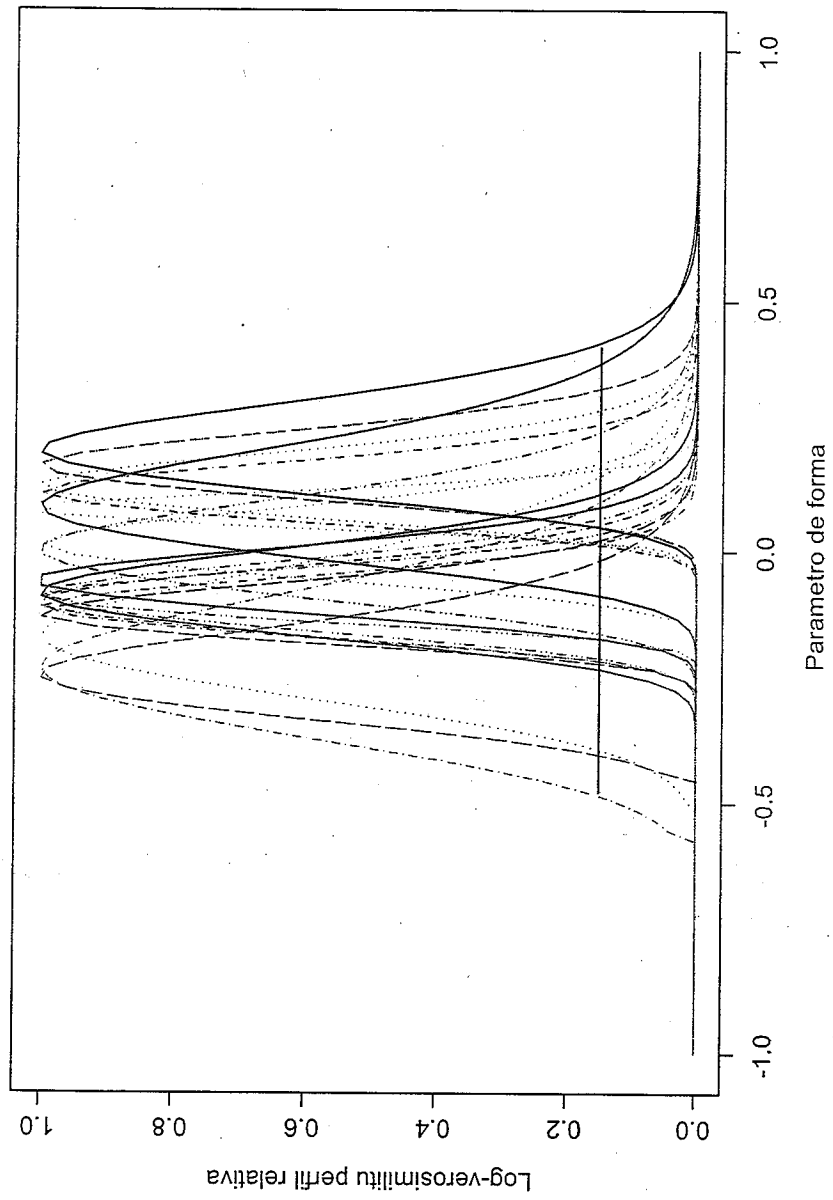


Figura 2.17: Verosimilitud perfil del parámetro de forma

tud perfil para el parámetro de forma de todas la estaciones climatológicas. En la Figura 2.18 se muestran los estimadores de ξ para cada una de las estaciones obtenidos en los dos ajustes (máximos anuales y excesos sobre umbrales). Recordemos que el parámetro de forma teóricamente debería ser el mismo en los dos ajustes.

Al igual que antes, sólo mostramos de manera detallada el ajuste de los excesos sobre umbrales para las estaciones más representativas de la cuenca de la ciudad de Morelia; Cuitzillo, Coitzio y Morelia DGE.

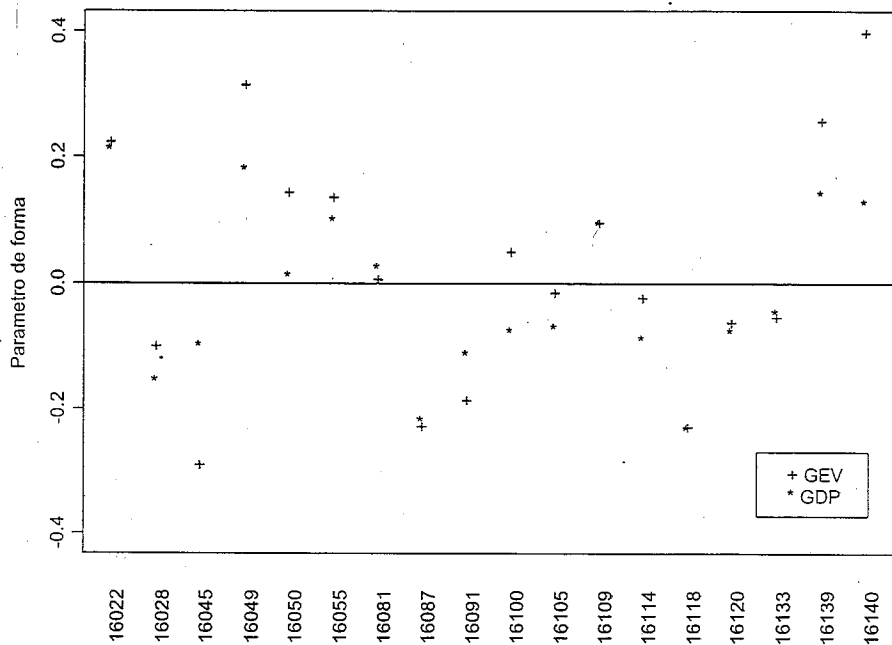


Figura 2.18: Comparación del parámetro de forma

2.2.1. Estación Cuitzillo

El primer paso consiste en determinar las excedencias a partir de las cuales se hará la inferencia. En la Figura 2.19 tenemos un ejemplo que muestra las excedencias que resultarían al fijar cierto umbral. Para saber cual es el umbral más adecuado usamos la función media de excesos y la gráfica de estabilidad para elegirlo.

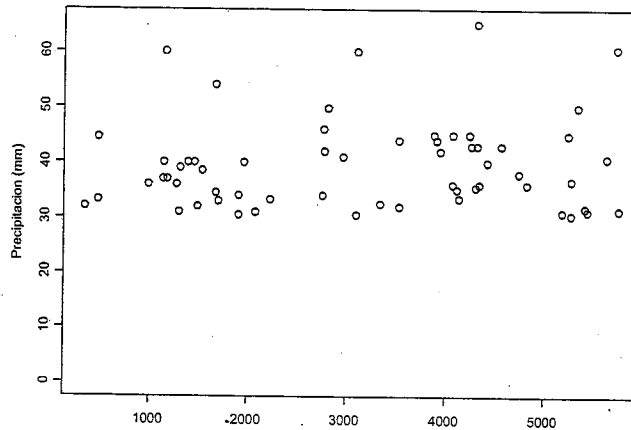


Figura 2.19: Precipitación diaria en exceso del umbral 30.4 en los meses de Mayo a Octubre 1969-2001, Estación Cuitzillo (16028)

La gráfica 2.20 muestra la función media de excesos empírica con intervalos de confianza aproximados del 95%. La interpretación en este caso no resulta sencilla, note que la gráfica es decreciente, sin embargo, no se puede determinar a simple vista cual sería el umbral a partir de cual la gráfica presenta un comportamiento "lineal", cualquier valor mayor a 20 podría ser la elección.

Las gráficas de $\hat{\beta}^*$ y $\hat{\xi}$ a diferentes umbrales u se muestran en la Figura 2.21. El intervalo en el cual se examina la estabilidad de los parámetros se elige de manera arbitraria. Para elegir el umbral adecuado buscamos un punto a partir del cual la gráfica se comporte estable y que además nos permita incluir un mayor número de observaciones por año sin que el porcentaje de excedencias sea demasiado grande (digamos menos de un 5%). Tampoco es

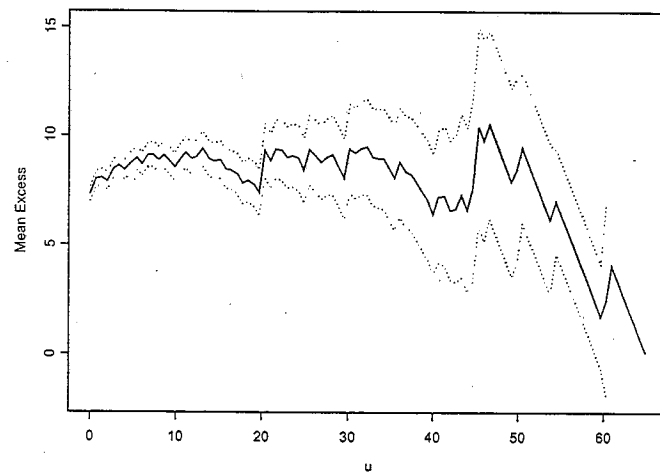


Figura 2.20: Función media de excesos de la Estación Cuitzillo (16028)

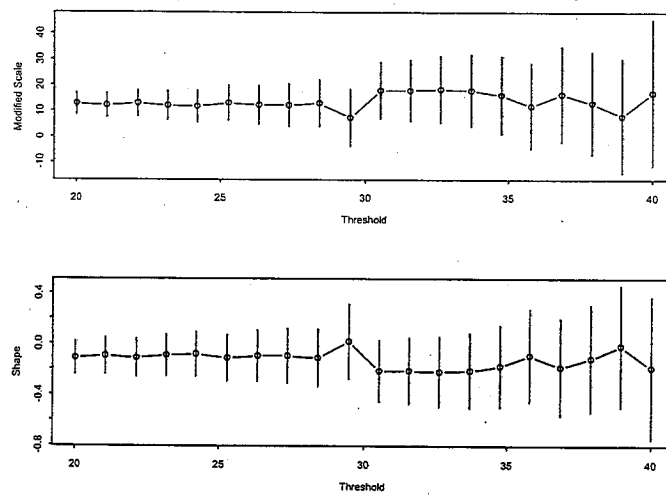


Figura 2.21: Parámetros estimados a diferentes umbrales para datos de precipitación, Estación Cuitzillo (16028)

deseable un umbral demasiado alto pues aumenta la varianza al tener muy pocas excedencias. Así, en este caso el umbral que cumple con dichas condiciones es $u = 30,4$.

Los estimadores de máxima verosimilitud usando este umbral son.

$$(\hat{\xi}, \hat{\beta}) = (-0,169, 10,7)$$

El parámetro de forma ξ es poco menor que el obtenido por el ajuste de máximos anuales como se muestra en la Figura 2.22 y el intervalo de verosimilitud aproximado de confianza del 95 % de $[-0.39, 0.15]$.

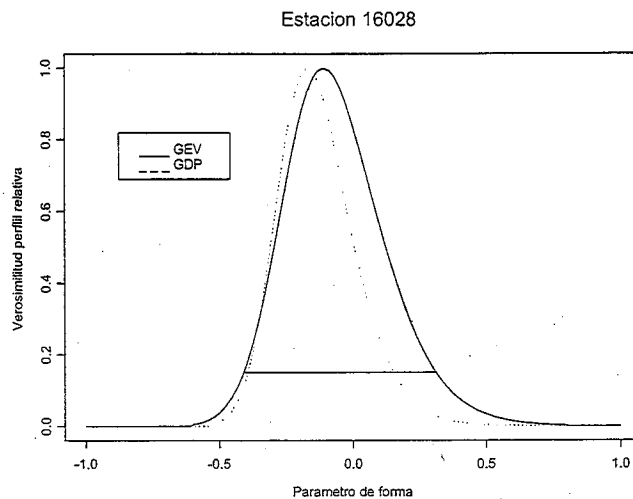


Figura 2.22: Verosimilitud perfil para ξ basada el modelo de excesos

En las Figuras 2.23 y 2.24 se muestran las verosimilitudes perfiles para los niveles de retorno a 100 y 1000 años, respectivamente. Vemos que las gráficas son más simétricas que las obtenidas en la sección anterior y los intervalos son más angostos. Los valores de $\hat{x}_{0,01}$ y $\hat{x}_{0,001}$ son 68 y 76, con intervalos de confianza del 95 % $[59.8, 99]$ y $[65, 148]$, respectivamente.

En la Figura 2.25 presentamos las gráficas de diagnóstico del modelo. Las excedencias se ajustan bien a la DGP al igual los máximos anuales se ajustan a la DGVE. En todo caso preferimos hacer uso de la DGP para hacer inferencia en los cuantiles, pues nos proporciona intervalos de confianza más precisos.

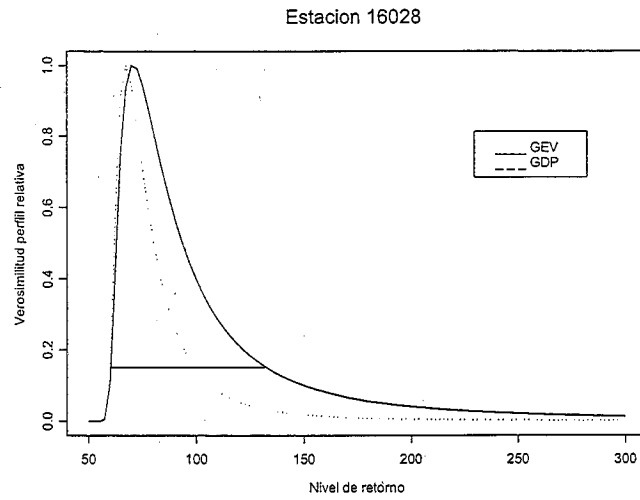


Figura 2.23: Verosimilitud perfil para el nivel de retorno en 100 años

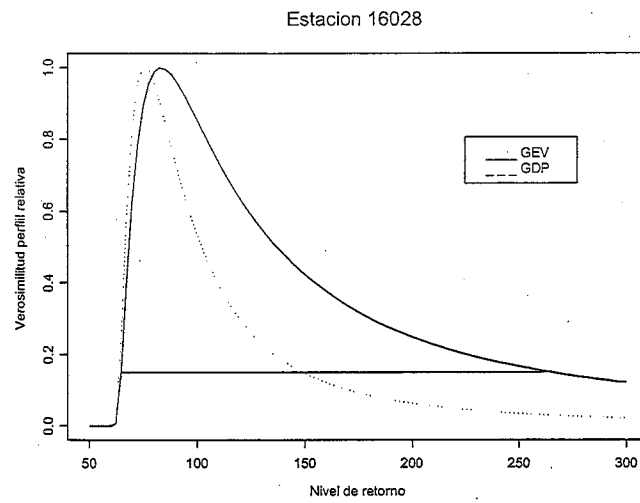


Figura 2.24: Verosimilitud perfil para el nivel de retorno en 1000 años

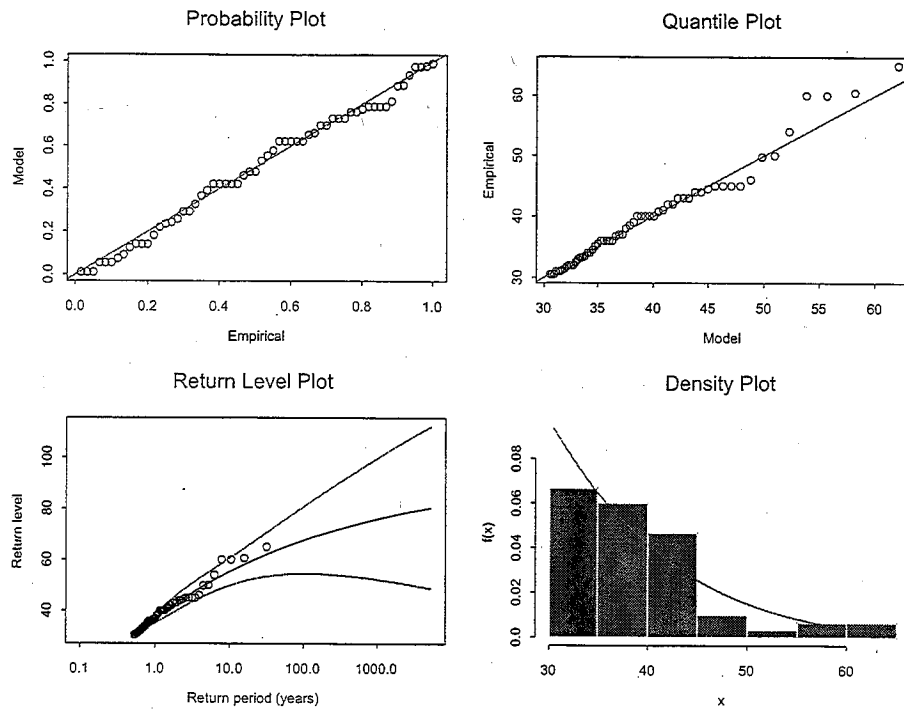


Figura 2.25: Gráficas de diagnóstico

Finalmente debemos mencionar que se obtuvieron 60 excedencias en el umbral $u = 30,4$ de un total de 5796 observaciones, por lo tanto el estimador de máxima verosimilitud de la probabilidad de excedencia es $\hat{\zeta}_u = 0,010$ con varianza aproximada $Var(\hat{\zeta}_u) = 1,7 \times 10^{-6}$.

2.2.2. Estación Cointzio

En la Figura 2.26 mostramos las excedencias obtenidas de fijar un umbral en $u = 30,4$. Para elegir este umbral se examinaron nuevamente la función media de excesos y la de estabilidad de los parámetros. El comportamiento

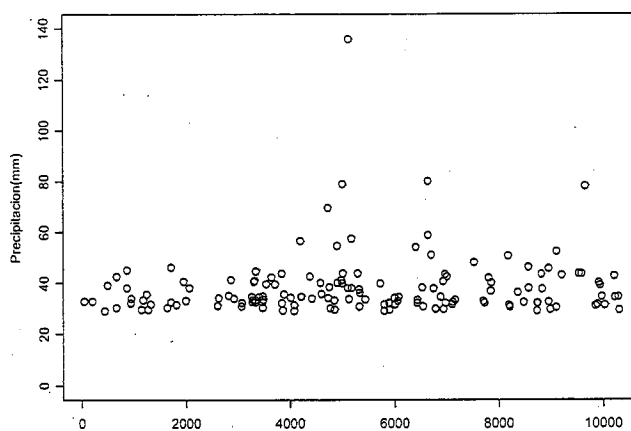


Figura 2.26: Precipitación diaria en exceso del umbral 30.4 en los meses de Mayo a Octubre 1947-1986, Estación Cointzio (16022)

de la función media de excesos (Figura 2.27) es muy irregular en umbrales altos, se observa mayor estabilidad entre $u = 20$ y $u = 60$. Las gráficas de $\hat{\beta}^*$ y $\hat{\xi}$ a diferentes umbrales u se muestran en la Figura 2.28. A partir del umbral $u = 30,4$ parece haber estabilidad en los parámetros (excepto en umbrales altos) y la tasa de excedencia es aceptable (1,2%). Notemos como la gráfica de ξ toma sólo valores en el dominio positivo lo cual nos indica que los extremos siguen el comportamiento de una distribución Fréchet.

Los estimadores de máxima verosimilitud son este caso.

$$(\hat{\xi}, \hat{\beta}) = (0,197, 7,224).$$

La estimación obtenida por log-verosimilitud perfil nos da un intervalo aproximado de confianza del 95% para ξ , $[0.05, 0.415]$. La Figura 2.29 muestra la verosimilitud perfil para ξ .

Los valores de $\hat{x}_{0,01}$ y $\hat{x}_{0,001}$ son 101 y 163, con intervalos de confianza del

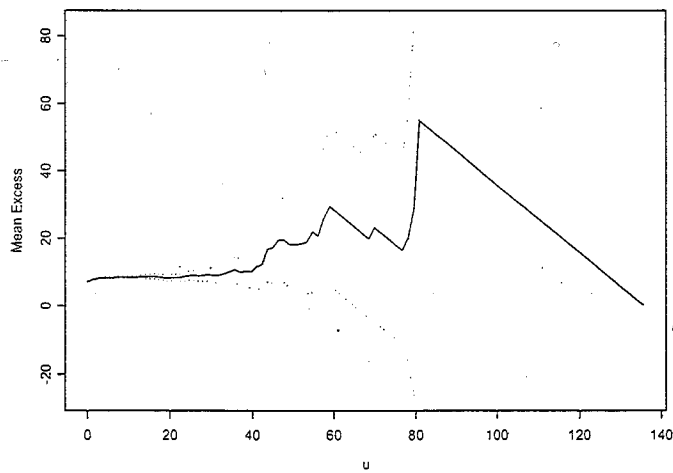


Figura 2.27: Función media de excesos de la Estación Cointzio (16022)

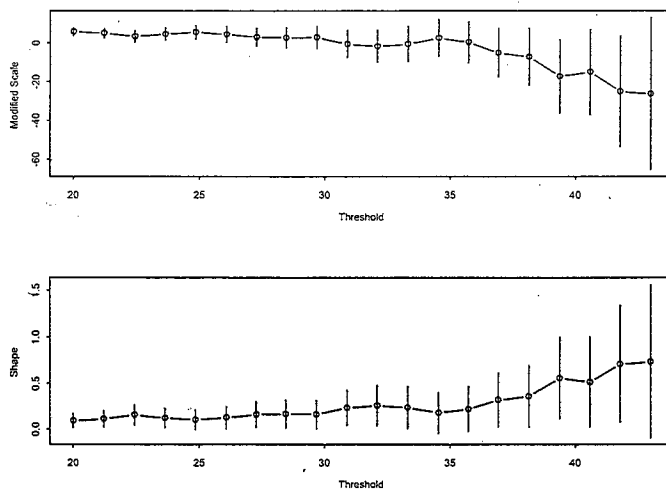


Figura 2.28: Parámetros estimados a diferentes umbrales para datos de Precipitación, Estación Cointzio (16022)

95 % [79,165] y [106,400], respectivamente. Estos valores son muy parecidos a los obtenidos en el ajuste de máximos, así que no hubo una mejoría notable en cuanto a la reducción de la varianza como consecuencia de una mayor inclusión de observaciones. En las Figuras 2.30 y 2.31 se muestran las verosimilitudes perfiles para los niveles de retorno a 100 y 1000 años, respectivamente. En la Figura 2.32 presentamos las gráficas de diagnóstico

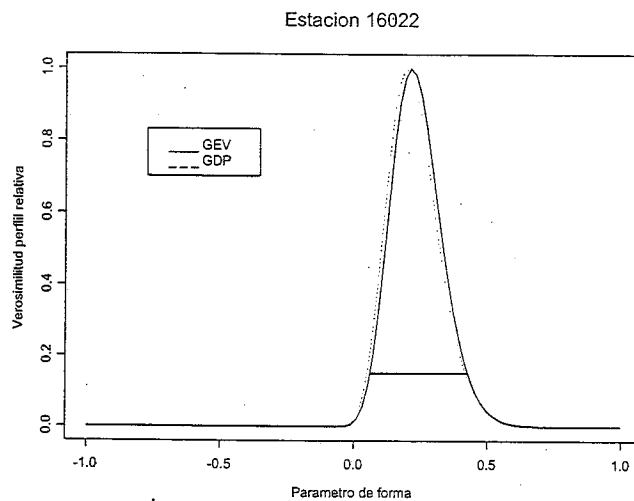


Figura 2.29: Verosimilitud perfil para ξ basada el modelo de excesos

del modelo.

Finalmente, se obtuvieron 129 excedencias en el umbral $u = 30,4$ de un total de 10329 observaciones, por lo tanto el estimador de máxima verosimilitud de la probabilidad de excedencia es $\hat{\zeta}_u = 0,012$ con varianza aproximada $Var(\hat{\zeta}_u) = 1,19 \times 10^{-6}$.

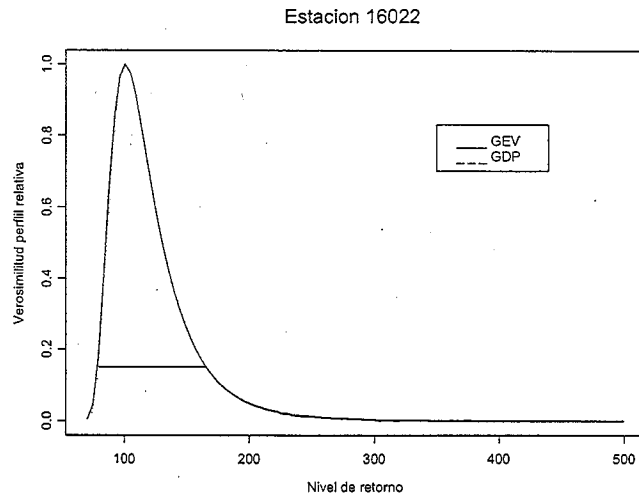


Figura 2.30: Verosimilitud perfil para el nivel de retorno en 100 años

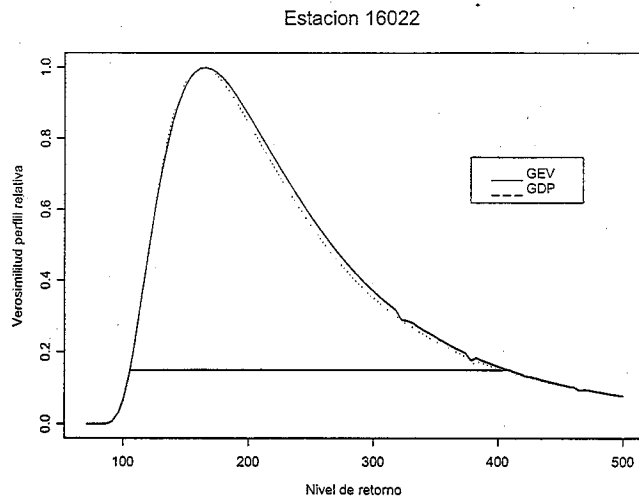


Figura 2.31: Verosimilitud perfil para el nivel de retorno en 1000 años

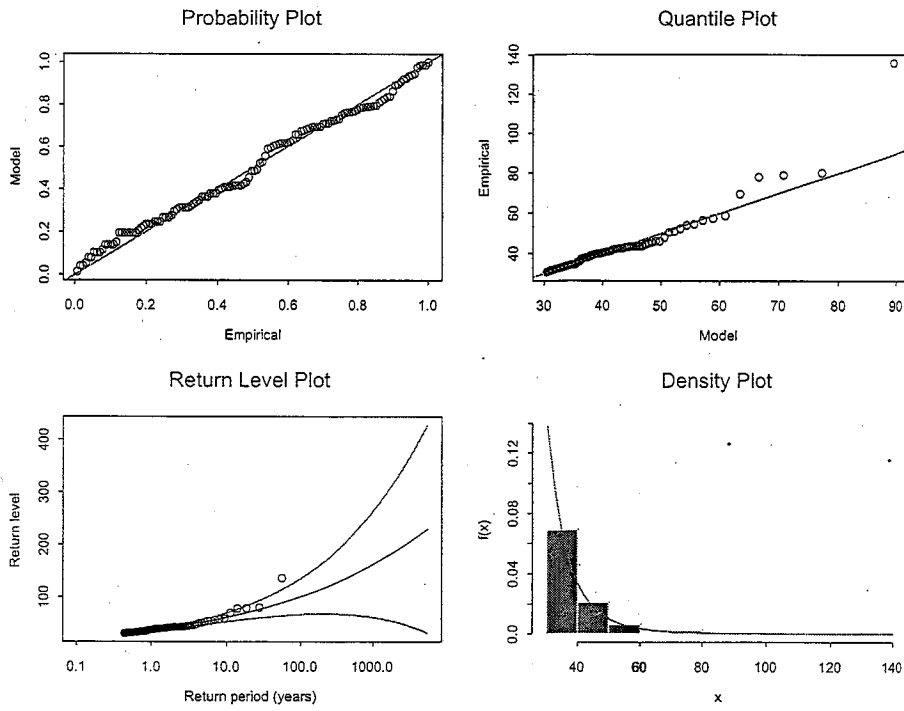


Figura 2.32: Gráficas de diagnóstico

2.2.3. Estación Morelia

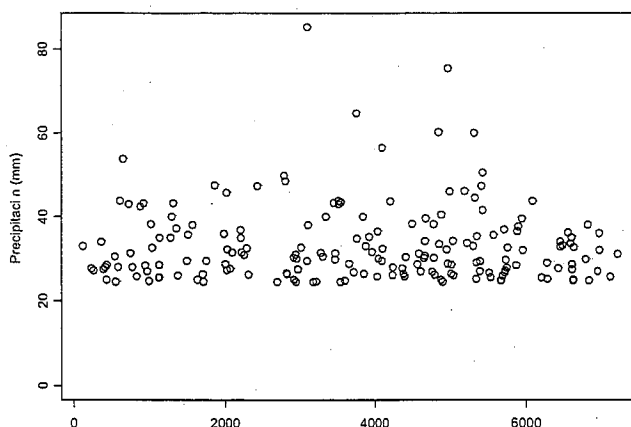


Figura 2.33: Precipitación diaria en exceso del umbral 24.5 en los meses de Mayo a Octubre 1947-1986, Estación Morelia (16081)

En la Figura 2.33 mostramos las excedencias de precipitación diaria de la Estación Morelia. La gráfica 2.34 muestra la función media de excesos empírica con intervalos de confianza aproximados del 95%. A partir de umbral $u = 20$ la gráfica sigue una tendencia horizontal, lo que sugiere que el modelo que podría ajustar a nuestros datos es el Gumbel.

Las gráficas de $\hat{\beta}^*$ y $\hat{\xi}$ a diferentes umbrales u se muestran en la Figura 2.35. Vemos que ésta gráfica tiene un comportamiento bastante regular, el umbral que cumple con las especificaciones planteadas anteriormente es $u = 24,5$. Además aunque los intervalos de confianza son muy amplios para umbrales altos, la tendencia alrededor del 0 es muy clara. Los estimadores de máxima verosimilitud son este caso

$$(\hat{\xi}, \hat{\beta}) = (0,010, 8,950)$$

La estimación obtenida por log-verosimilitud perfil nos da un intervalo de confianza aproximado del 95%, $[-0.103, 0.175]$. La Figura 2.36 muestra la verosimilitud perfil para ξ .

En las Figuras 2.37 y 2.38 se muestran las verosimilitudes perfiles para los

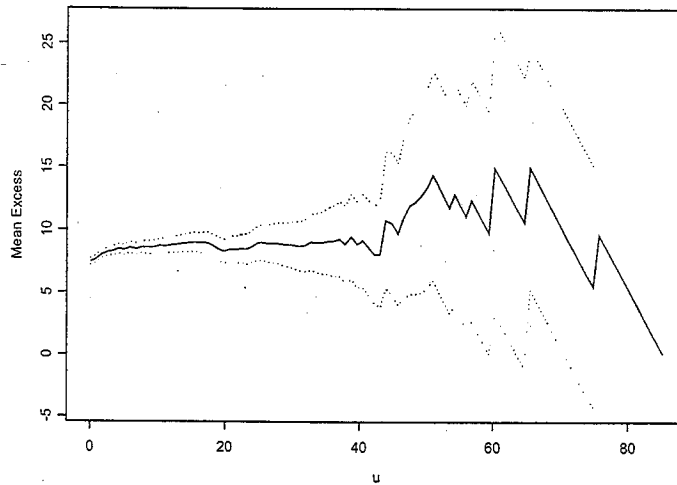


Figura 2.34: Función media de excesos de la Estación Morelia (16081)

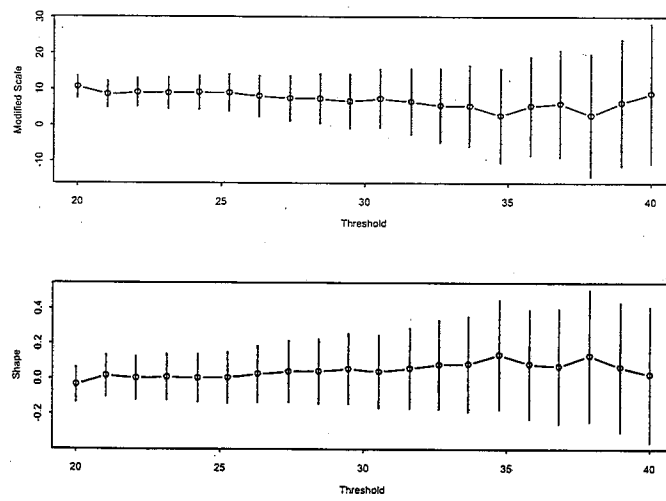


Figura 2.35: Parámetros estimados a diferentes umbrales para datos de Precipitación, Estación Morelia (16081)

niveles de retorno a 100 y 1000 años, respectivamente. Los valores de $\hat{x}_{0,01}$ y $\hat{x}_{0,001}$ son 81.5 y 103.5, con intervalos de confianza del 95% [68.5,113] y [81,177], respectivamente. En la Figura 2.39 presentamos las gráficas de

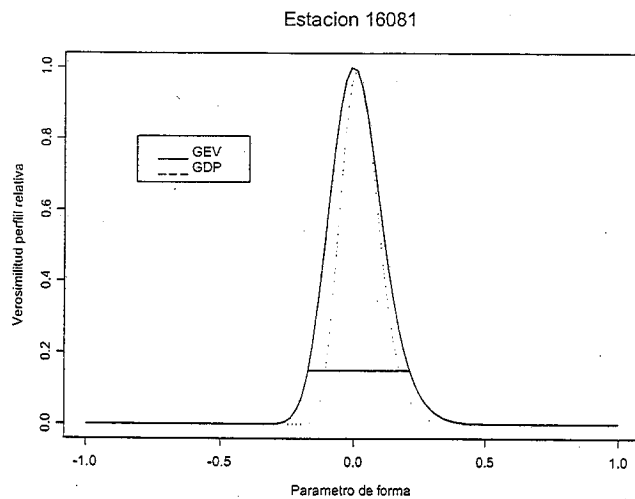


Figura 2.36: Verosimilitud perfil para ξ basada el modelo de excesos

diagnóstico del modelo y vemos que se tiene un buen ajuste.

Finalmente se obtuvieron 179 excedencias en el umbral $u = 24,5$ de un total de 7259 observaciones, por lo tanto el estimador de máxima verosimilitud de la probabilidad de excedencia es $\hat{\zeta}_u = 0,024$ con varianza aproximada $Var(\hat{\zeta}_u) = 3,3 \times 10^{-6}$.

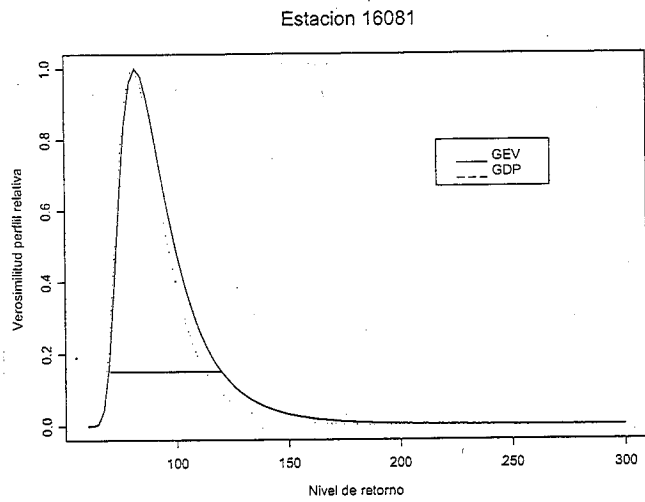


Figura 2.37: Verosimilitud perfil para el nivel de retorno en 100 años

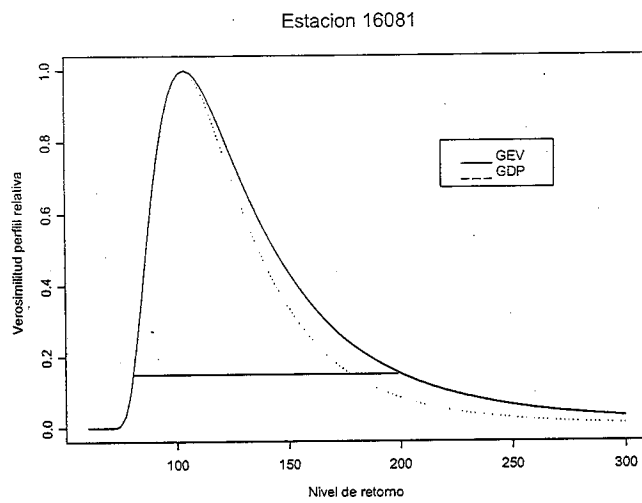


Figura 2.38: Verosimilitud perfil para el nivel de retorno en 1000 años

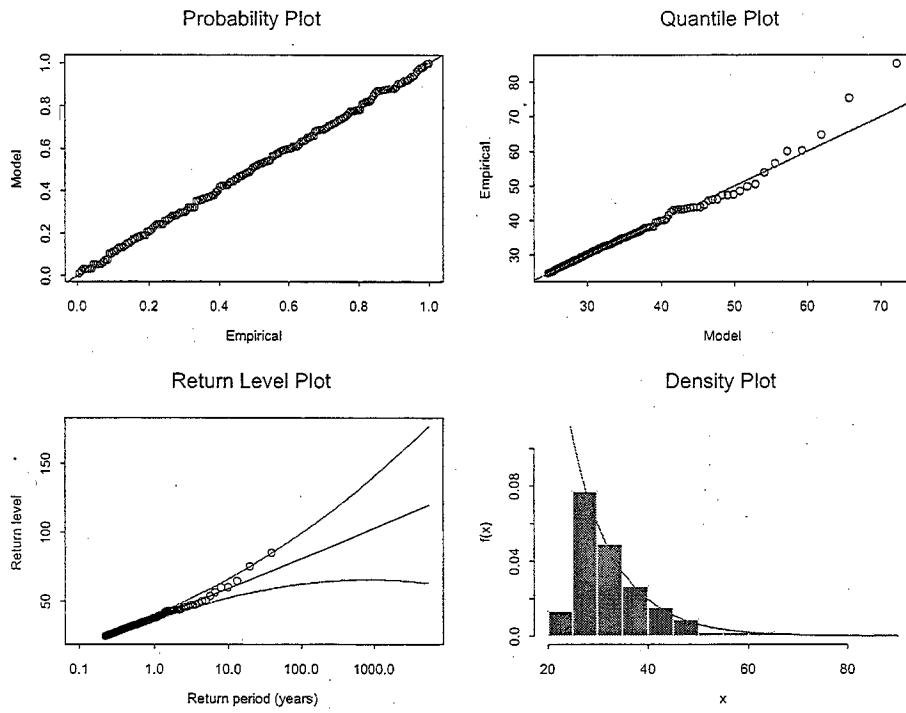


Figura 2.39: Gráficas de diagnóstico

2.3. Cluster de series

Como vimos en las secciones anteriores, dentro de las estaciones climatológicas el comportamiento a largo plazo es muy distinto en cada una de ellas. Este comportamiento está determinado por el modelo que se ajusta a los valores extremos, de manera particular por el parámetro de forma obtenido. Partiendo de este hecho, cuando se agrupan las estaciones esperamos que dicha clasificación esté influenciada, además del contexto geográfico, por ese comportamiento a largo plazo.

En la Figura 2.40 se muestra el dendograma obtenido. Este dendograma se obtuvo usando una técnica de formación de cluster jerárquica y por el método de la liga completa.¹

Podemos ver que las estaciones que de manera individual registraron un valor de ξ mayor a cero tienen a estar unidas (16022, 16140, etc.), de igual modo pasa con ciertas estaciones que tuvieron un valor de ξ menor a cero. A partir de aquí consideramos que con tres grupos podemos captar la información necesaria para hacer inferencia. En el Cuadro 2.7 se muestra la manera en que se agrupan las estaciones de acuerdo a este criterio.

Ahora en cada uno de los grupos formados podemos combinar la informa-

Grupo 1	Grupo 2	Grupo 3
16022	16028	16045
16049	16091	16055
16050	16105	16114
16081	16133	
16087		
16100		
16109		
16118		
16120		
16139		
16140		

Cuadro 2.7: Grupos

ción y ver si de este modo obtenemos mejores estimadores de los cuantiles. Puesto que en la sección anterior vimos que obteníamos mejores estimaciones usando excesos sobre umbrales, en esta sección nos enfocaremos en esa metodología.

¹El programa utilizado para esta aplicación se puede obtener en Domínguez(2000).

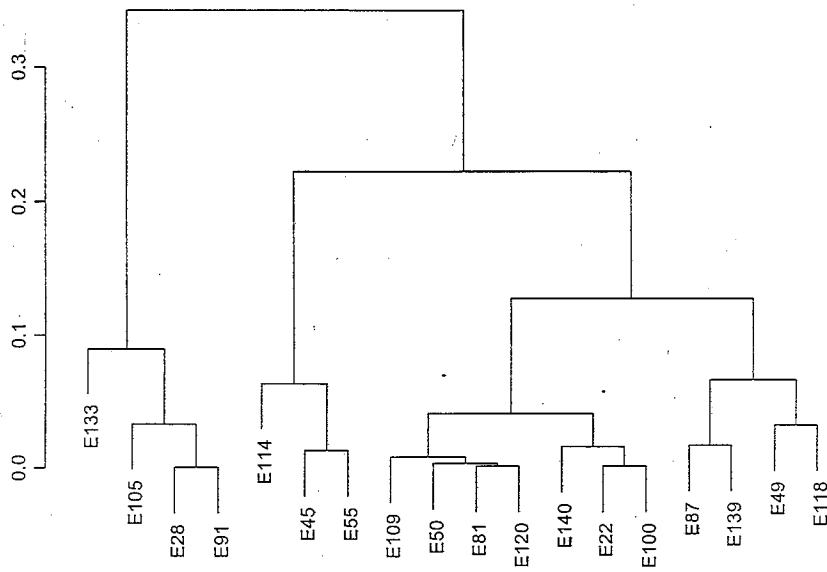


Figura 2.40: Dendrograma de las estaciones meteorológicas de la cuenca de Morelia

Grupo 1

En este grupo encontramos a cinco de las seis estaciones que, de acuerdo a la estimación de la DGP tuvieron $\xi > 0$. El ajuste de los excesos está basado en las observaciones diarias de precipitación de mayo a octubre de cada año de todas las estaciones reunidas en el grupo.

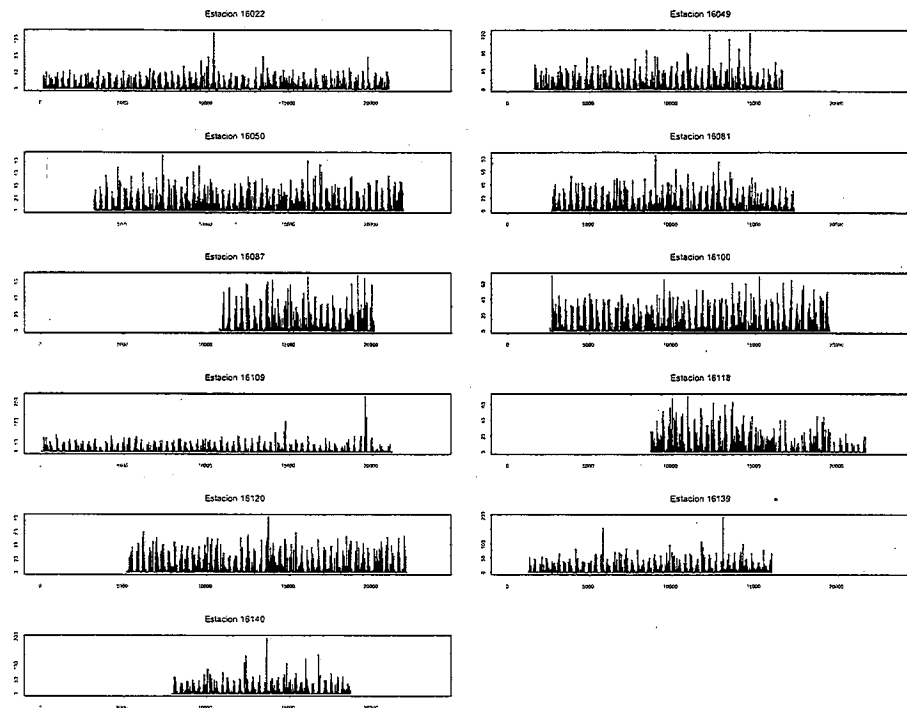


Figura 2.41: Grupo 1

La estimación obtenida por log-verosimilitud perfil nos da $\xi = 0,176$ con intervalos aproximados de confianza del 95 % de $[0.113, 0.25]$. La Figura 2.42 muestra la verosimilitud perfil para ξ .

Las estimaciones que nos interesan son las de los cuantiles $\hat{x}_{0,01}$ y $\hat{x}_{0,001}$. Comparemos con los niveles de retorno obtenidos para la estación Cointzio y la estación Morelia (DGE). En la gráfica 2.43 podemos ver que los intervalos se reducen de manera significativa, las gráficas obtenidas para el Grupo 1 son muy simétricas, por lo cual esperamos que las estimaciones sean de

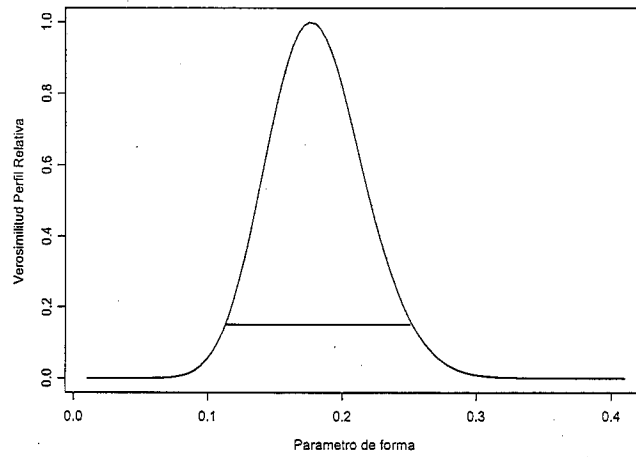


Figura 2.42: Verosimilitud perfil para ξ basada el modelo de excesos

Estimación	Grupo 1	Cointzio	Morelia DGE
$\hat{x}_{0,01}$	125.5	101	81.5
IC	[126,170]	[79,165]	[68.5,113]
$\hat{x}_{0,001}$	197	163	103.5
IC	[182,300]	[106,400]	[81,177]

Cuadro 2.8: Comparación

cierto modo más confiables. El nivel de retorno aumenta considerablemente, y este valor se puede comparar con los más recientes datos extremos de precipitación y corroborar la validez del modelo ajustado.

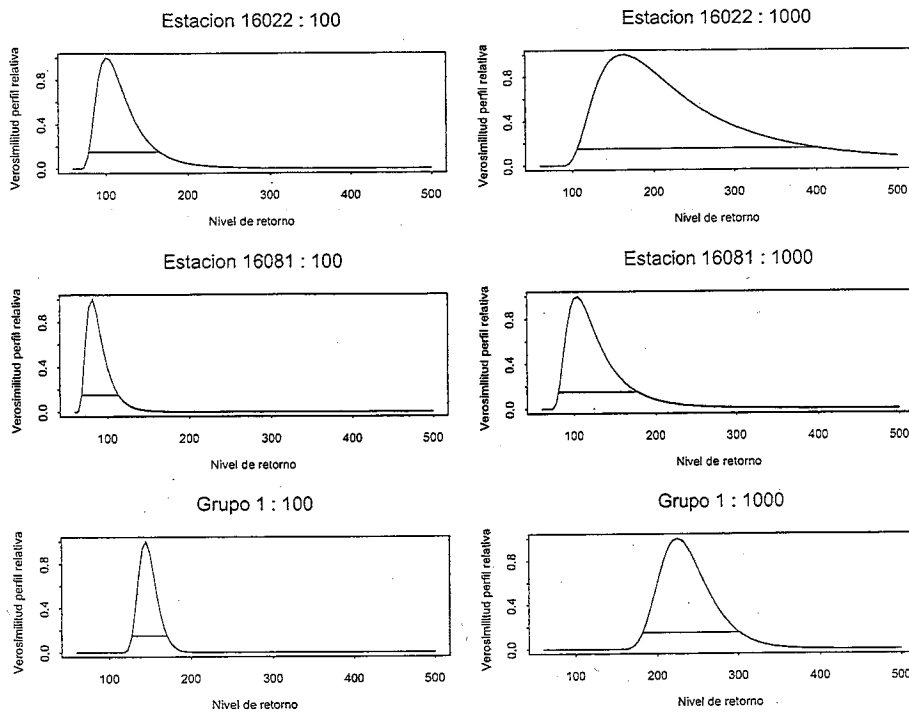


Figura 2.43: Verosimilitud perfil para el nivel de retorno

Grupo 2

Al igual que combinamos la información en el Grupo 1 ahora lo hacemos para el Grupo 2. Este grupo se caracteriza por tener estaciones cuyos datos se

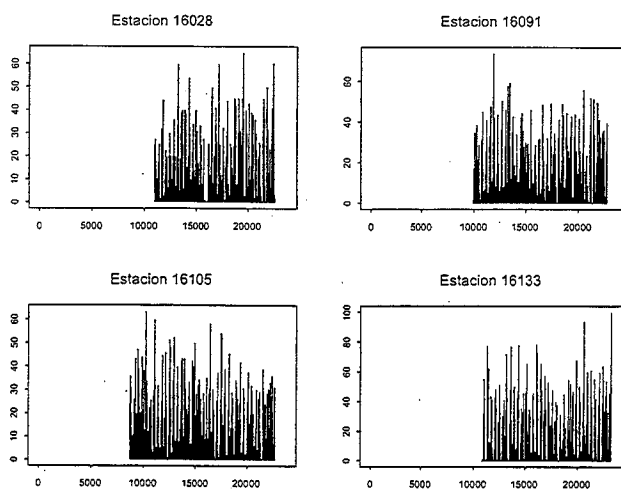


Figura 2.44: Grupo 2

ajustaron a una distribución Weibull. Aunque la estimación de los cuantiles en cada una de dichas estaciones fue aceptable (Ver Anexo), al agruparlas se tiene una gran mejoría. Primeramente presentamos la estimación del parámetro de forma de la DGP. La estimación obtenida es $\hat{\xi} = -0,062$ con un intervalo de confianza aproximado del 95 % $[-.135, 0.032]$. Aunque este valor es muy cercano a cero se mantiene ligeramente menor, lo cual es deseable ya que los datos que conforman el grupo presentan un $\xi < 0$.

Estimación	Grupo 2	Cuitzillo
$\hat{x}_{0,01}$	87	68
IC	[79.5,101]	[59.8,99]
$\hat{x}_{0,001}$	102.5	76
IC	[89,130]	[65,148]

Cuadro 2.9: Comparación

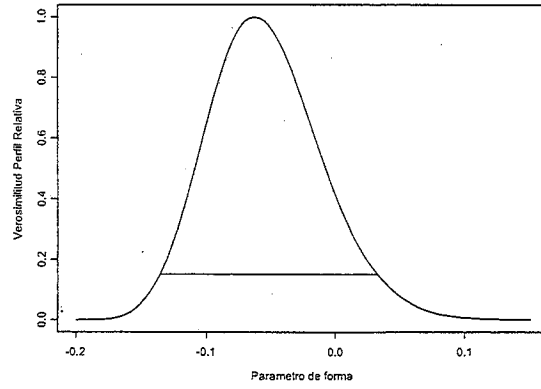


Figura 2.45: Verosimilitud perfil para ξ basada el modelo de excesos

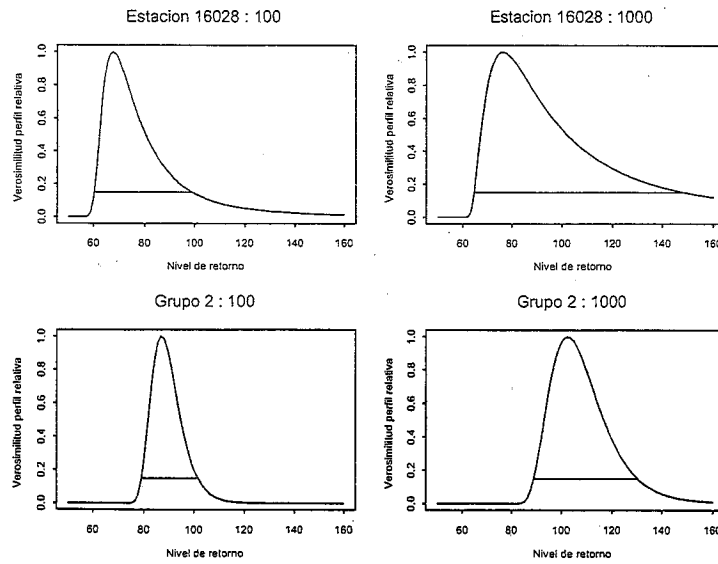


Figura 2.46: Verosimilitud perfil para el nivel de retorno

Grupo 3

En este grupo las excedencias registradas en la estación Jesús del Monte (16055) se ajustaron a una DGP con $\xi > 0$ mientras que Temazcal (16045) y San Miguel de Monte (16114) presentaron un valor $\xi < 0$. Estas estaciones tienen una relación cercana geográficamente.

La estimación obtenida del parámetro de forma fue $\hat{\xi} = -0,015$ con intervalo de confianza aproximado del 95 % $[-0,098, 0,08]$. En la Figura 2.48 podemos ver que al mezclar la información el parámetro de forma es muy cercano a

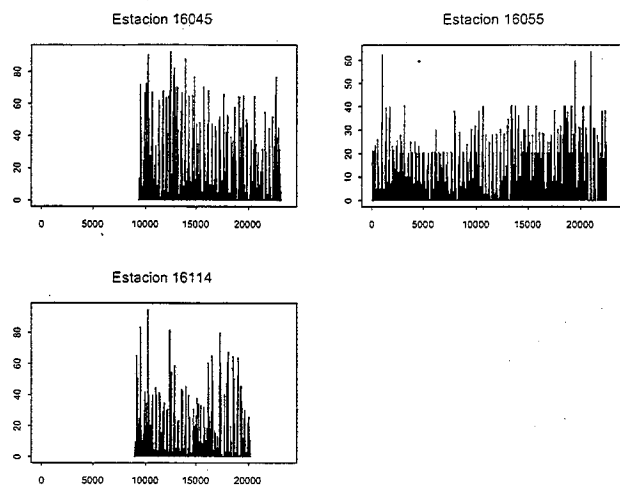


Figura 2.47: Grupo 3

cero.

En la Figura de la estimación del nivel de retorno comparamos con la estación 16055 que fue la que presentó más problemas al hacer las estimaciones de los cuantiles.

Estimación	Grupo 3	Jesús del Monte
$\hat{x}_{0,01}$	105.5	62
IC	[93,128]	[54,95]
$\hat{x}_{0,001}$	130	83
IC	[107,175]	[63,197]

Cuadro 2.10: Comparación

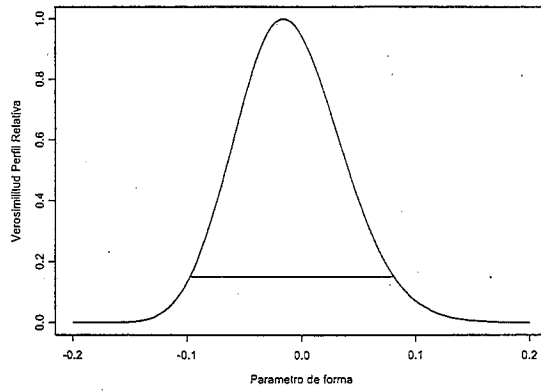


Figura 2.48: Verosimilitud perfil para ξ basada el modelo de excesos

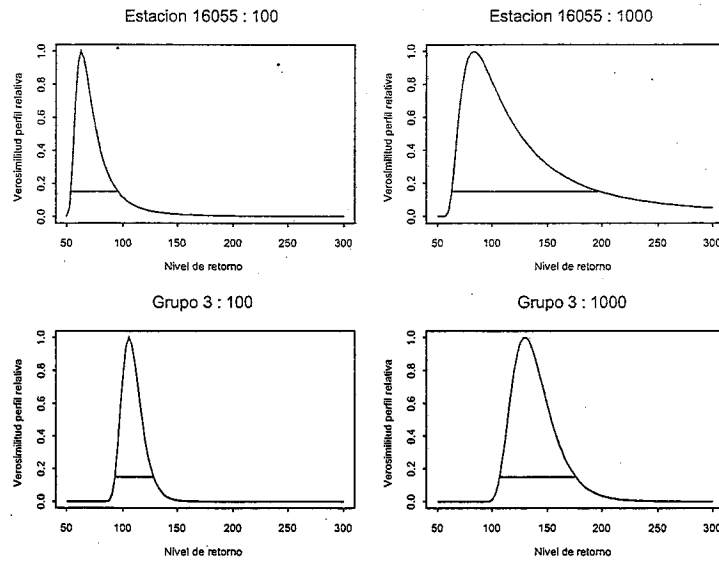


Figura 2.49: Verosimilitud perfil para el nivel de retorno

Conclusión

Como era de esperarse hubo mucha coherencia en la estimación del parámetro de forma obtenida por ajuste de la DGVE y la DGP en cada una de las 18 estaciones. Sin embargo las estimaciones de los cuantiles obtenidas a partir de la DGVE fueron menos precisas y en algunos casos se presentaron gráficas de la verosimilitud relativa con límite derecho infinito e incluso algunas mostraron irregularidades para niveles de retorno muy grandes. En contraste, con el ajuste de la DGP se tuvieron mejores estimadores de los cuantiles al mostrar gráficas de verosimilitud más simétricas y con intervalos de confianza aproximados del 95 % más angostos.

Las estimaciones del parámetro de forma obtenidas por ajuste de la DGVE y la DGP son cercanas al cero, sin embargo, enfocarnos sólo en este modelo para tratar de estimar los cuantiles en algunas estaciones nos puede llevar a tener resultados no del todo correcto.

La gráfica de nivel de retorno es una herramienta muy eficaz en la validación del modelo pues muestra tanto las estimaciones empíricas de los cuantiles como los intervalos de confianza generados por el modelo ajustado. En el ajuste de la DGP la gráfica de la función media de excesos no fue de mucha utilidad para la elección del umbral pues en todos los casos se mostró muy irregular y difícil de interpretar, en su lugar se usó la gráfica de estabilidad de los parámetros a diferentes umbrales por ser más específica visualmente en la elección de umbral.

Con el agrupamiento de los datos se obtuvieron estimaciones más precisas de los cuantiles incluso que las obtenidas por el ajuste de excesos sobre umbrales. Las estaciones que de manera individual registraron un valor de ξ mayor a cero tienen a estar unidas, de igual modo pasa con ciertas estaciones que tuvieron un valor de ξ menor a cero y las que registraron un valor de ξ igual a cero. Al parecer el factor geográfico no influyó en la manera en como se agruparon las estaciones. Finalmente, aunque no se pudo profundizar en esta parte del trabajo, puede servir como una herramienta en el desarrollo futuro de estudios de regionalización.

Bibliografía

- Aparicio, F. J. (2004) *Fundamentos de hidrología de superficie*, Limusa, México.
- Barndoff-Nielsen y Cox, D. R. (1989) "Asymptotic techniques for use in statistics". Chapman and Hall, London.
- Billingsley, P. (1995) *Probability and Measure*, 3a. ed. Wiley, New York.
- Buishand, T.A. (1989) "Statistics of extremes in climatology", *Statistica Neerlandica*, 43, 1-29.
- Castillo, E. (1988) *Extreme Value Theory in Engineering*. Academic Press, San Diego.
- Coles, S. (2001) *An introduction to Statistical Modelling of Extreme Values*. Springer, New York.
- Coles, S. G y Pericchi, L. (2003) "Anticipating catastrophes through extreme value modelling". *Applied Statistics*, 52,405-416.
- Coles, S. G. y Tawn, J. A. (1996) "A Bayesian analysis of extreme rainfall data". *Appl. Statist.*, 45, 463-478.
- Davison, A. C. y Smith, R. L. (1990) "Models for exceedances over high thresholds (with discussion)". *J. R. Statis. Soc. B*, 52, 393-442.
- Domínguez Molina J. R. (2001) "Formación de Conglomerados y valores atípicos en Series de Tiempo". *Tesis no publicada*. Maestría en Estadística. Programa conjunto CIMAT-Universidad de Guanajuato.
- Embrechts, P., Klüppelberg, C. y Mikosh, T. (1997) *Modelling Extremal Events*, Springer, Berlin Heidelberg.

Galambos, J. (1995). *The Asymptotic Theory of Extreme Orders Statistics*. Krieger, Florida. Segunda Edición.

Gutierrez, Alfonso. (2004) "Estudio del manejo de aguas pluviales en la zona metropolitana de Morelia (ZMM), Estado de Michoacán". Instituto Mexicano de Tecnología del Agua, pp. 29.

Reiss, R. D. y Thomas, M. (2001) *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhauser Verlag AG, Segunda Edición.

Resnick, S. I. (1986) *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.

Serfling, Robert. J. (1980) "Aproximation theorems of mathematical statistics". John Wiley, New York.

Smith, R. L. (1986) "Extreme value theory based on the r largest annual events". *Journal of Hydrology*, 86, 27-43.

Smith, R. L. (1990) "Extreme value analysis of environmental time series: an application to trend detection in ground level ozone (with discussion)". *Statist. Sci.*, 4, 367-393.