



# Análisis de Valores Extremos

## Modelización Espacial

Autora: Adriana Medialdea Villanueva  
Dirigido por: José Miguel Angulo Ibáñez

---

Máster en Estadística Aplicada  
Facultad de Ciencias. Universidad de Granada.

12 de Septiembre de 2016



# Índice general

<b>Resumen</b>	<b>5</b>
<b>1. Teoría de valores extremos univariante</b>	<b>9</b>
1.1. Distribución de valores extremos generalizada . . . . .	9
1.2. Modelos de máximos por bloques . . . . .	11
1.2.1. Estimación de los parámetros de la distribución GEV . . .	12
1.2.2. Niveles de retorno . . . . .	12
1.3. Modelos de excedencias de umbrales . . . . .	13
1.3.1. Introducción . . . . .	13
1.3.2. Distribución de Pareto generalizada . . . . .	13
1.3.3. Estimación de los parámetros de la distribución de Pareto generalizada . . . . .	14
1.3.4. Selección del umbral . . . . .	17
1.3.5. Niveles de retorno . . . . .	19
1.4. Modelización de extremos en secuencias de datos dependientes .	20
<b>2. Teoría de valores extremos multivariante</b>	<b>23</b>
2.1. Introducción . . . . .	23
2.2. Cópulas . . . . .	23
2.3. Distribución de valores extremos multivariante . . . . .	24
2.3.1. Caso bivariante . . . . .	24
2.3.2. Extensión para n variables . . . . .	26
2.3.3. Transformaciones . . . . .	26
2.4. Convergencia del vector de máximos . . . . .	27
2.5. Funciones de dependencia . . . . .	28
2.6. Modelización y estimación de parámetros . . . . .	30
2.6.1. Caso bivariante . . . . .	30

2.6.2.	Extensión para $n$ variables. Estimación mediante cópulas	31
2.7.	Niveles de retorno	31
2.7.1.	Caso bivalente	31
2.7.2.	Extensión para $n$ variables. Medida de intensidad	31
2.8.	Modelos de excedencias de umbrales	32
2.8.1.	Caso bivalente	32
2.8.2.	Extensión a $n$ variables	34
<b>3.</b>	<b>Extremos espaciales</b>	<b>37</b>
3.1.	Introducción	37
3.2.	Procesos max-estables espaciales	37
3.2.1.	Representación espectral de los procesos max-estables	38
3.2.2.	Modelos max-estables	39
3.3.	Medidas de dependencia para extremos espaciales	40
3.3.1.	Función del coeficiente extremal	40
3.3.2.	F-madograma	41
3.3.3.	$\lambda$ -madograma	41
3.4.	Simulación de procesos espaciales max-estables	42
3.4.1.	Simulaciones no condicionadas	42
3.4.2.	Simulaciones condicionadas	42
3.5.	Ajuste de un proceso max-estable Fréchet a los datos	43
3.5.1.	Mínimos cuadrados	43
3.5.2.	Máxima verosimilitud	44
3.6.	Selección del modelo	44
3.6.1.	Criterio de información de Takeuchi	44
3.6.2.	Estadístico de la tasa de verosimilitud	45
<b>4.</b>	<b>Aplicación con R. Extremos espaciales</b>	<b>47</b>
4.1.	Ajuste de la distribución de valores extremos generalizada a los datos	48
4.2.	Transformación de los datos a la distribución Fréchet.	52
4.3.	Ajuste del proceso max-estable a los datos transformados.	53
4.3.1.	Modelo 1	53
4.3.2.	Modelo 2	54
4.3.3.	Diagnos y elección del modelo	55
4.3.4.	Predicciones.	59

<i>ÍNDICE GENERAL</i>	5
<b>5. Conclusión</b>	<b>61</b>



# Resumen

La Teoría de Valores Extremos es una disciplina estadística cuyos avances fundamentales son relativamente recientes. Está enfocada al análisis del comportamiento estocástico de los valores extremos de un proceso.

Con amplias aplicaciones en hidrología, investigación medioambiental y meteorología, seguros, aplicaciones financieras y geología entre otras, tiene como objetivo cuantificar el comportamiento de una serie de valores de un proceso, siendo estos valores más extremos que los que usualmente se observan en él.

Esta teoría se desarrolla a partir de los resultados de Fréchet (1927), Fisher y Tippett (1928) y von Mises (1936), que sientan las bases de la teoría probabilística de valores extremos unidimensional; sin embargo, su desarrollo teórico no tendrá lugar hasta principios de los años 70 con la tesis doctoral de de Haan (1970) y los resultados de inferencia estadística desarrollados por Pickands (1975); estos resultados supusieron las primeras contribuciones a la Teoría de Valores Extremos Multivariante y motivaron el desarrollo de modelos alternativos basados en excedencias de umbrales.

En la actualidad, los avances en esta teoría se está centrando principalmente en el desarrollo de modelos y métodos para valores extremos de fenómenos espaciales y otras estructuras más complejas.

En el presente trabajo se realiza una revisión de los principales enfoques existentes en el análisis de valores extremos; prestando especial atención a la dimensión espacial, junto con una aplicación práctica a datos geoestadísticos.

El primer capítulo está dedicado a la teoría univariante, donde se presenta el punto de partida de esta disciplina a partir de la condición de valores extremos; además de los modelos clásicos de máximos por bloques y excedencias de umbrales junto con las distintas distribuciones a las que dan lugar cada uno de ellos. Finalmente, se introduce la modelización para el caso de datos dependientes.

El tercer capítulo está enfocado a las distribuciones de valores extremos multivariantes y su aproximación para máximos por bloques y excedencias de umbrales, prestando especial atención a su estructura de dependencia a través de cópulas.

En el cuarto capítulo se trata la modelización de la dependencia tras la inclusión de una dimensión espacial, para lo que serán fundamentales los procesos max-estables y las distribuciones finito-dimensionales que se derivarán a través de ellos.

Finalmente, en el capítulo cinco se realiza una aplicación práctica con el software R y la herramienta SpatialExtremes.



# Capítulo 1

## Teoría de valores extremos univariante

### 1.1. Distribución de valores extremos generalizada

El modelo para el que se desarrolla la teoría de valores extremos está enfocado a describir el comportamiento estadístico de

$$M_n = \text{máx}\{X_1, \dots, X_n\},$$

donde  $X_1, \dots, X_n$  es una secuencia de variables aleatorias independientes con distribución común  $F$  y  $M_n$  representa el máximo del proceso sobre  $n$  unidades de tiempos de observación.

La distribución de  $M_n$  podría obtenerse de manera exacta a partir de la distribución de las  $n$  variables, teniendo en cuenta las propiedades de independencia:

$$\begin{aligned} F_{M_n}(z) &= Pr[M_n \leq z] = Pr[X_1 \leq z, \dots, X_n \leq z] \\ &= Pr[X_1 \leq z] \times \dots \times Pr[X_n \leq z] = [F(z)]^n. \end{aligned} \quad (1.1)$$

y derivando, se obtendría su función de densidad:

$$f_{M_n}(z) = n (F_{M_n}(z))^{n-1} f(z). \quad (1.2)$$

La función de distribución anteriormente calculada converge a cero cuando  $n \rightarrow \infty$  para  $z > z^*$  y a uno para  $z \leq z^*$ , con  $z^* = \sup\{z : F(z) < 1\}$ . Por lo que, para obtener una distribución límite no degenerada, será necesario llevar a cabo una normalización.

Esta solución, basada en el Teorema Central del Límite, consiste en la búsqueda de secuencias de constantes  $\{b_n; n \geq 1\}$  y  $\{a_n; n \geq 1\}$  tales que la distribución de

$$M_n^* = \frac{M_n - b_n}{a_n} \quad (1.3)$$

converge a una distribución no degenerada cuando  $n \rightarrow \infty$ , es decir,

$$\lim_{n \rightarrow \infty} F^n(a_n z + b_n) = G(z). \quad (1.4)$$

El rango completo de distribuciones límite que podrá seguir  $M_n^*$  vendrá dado por el Teorema de Valores Extremos:

**Teorema 1.1.1** *Si existen sucesiones de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tales que*

$$P\left(\frac{M_n - b_n}{a_n} \geq z\right) \rightarrow G(z), \quad \text{cuando } n \rightarrow \infty,$$

*siendo  $G$  una función de distribución no degenerada, entonces  $G$  debe pertenecer a una de las siguientes familias:*

- I. **Gumbel:**  $G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty$
- II. **Fréchet:**  $G(z) = \begin{cases} 0, & z \leq b \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b \end{cases}$
- III. **Weibull:**  $G(z) = \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)\right]^\alpha\right\}, & z < b \\ 1, & z \geq b \end{cases}$

*con parámetros  $a > 0$ ,  $b$  y, en el caso de las familias ii y iii,  $\alpha > 0$ .*

Las anteriores distribuciones se conocen como **distribuciones de valores extremos** y serán las únicas a las que pueda converger la variable  $M_n^*$ , independientemente de como se distribuya  $F$ .

Las familias de distribuciones Fréchet, Gumbel y Weibull se pueden combinar en una única familia con función de distribución

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad (1.5)$$

definida en  $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$ , donde los parámetros de localización, escala y forma satisfacen, respectivamente,  $-\infty < \mu < \infty$ ,  $\sigma > 0$  y  $-\infty < \xi < \infty$ . Esta familia de distribuciones se conoce como **familia de distribuciones de valores extremos generalizada (GEV)** (o familia de distribuciones de Von Mises).

La especificación de  $\xi$  determinará el comportamiento de la cola de la distribución, de forma que según el valor que tome este parámetro se tendrá una de las siguientes distribuciones:

- **Gumbel** si  $\xi = 0$
- **Fréchet** si  $\xi > 0$
- **Weibull** si  $\xi < 0$

De esta forma, el teorema 1.1.1 se replanteará como sigue:

**Teorema 1.1.2** *Si existen sucesiones de constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tales que*

$$Pr\{(M_n - b_n) / a_n \leq z\} \rightarrow G(z), \quad \text{cuando } n \rightarrow \infty, \quad (1.6)$$

*para una distribución  $G$  no degenerada, entonces  $G$  pertenece a la familia de distribuciones GEV,*

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}. \quad (1.7)$$

El concepto de max-estabilidad, que se introduce a continuación, está estrechamente relacionado con el teorema anterior.

**Definición 1.1.1** *Una distribución  $G$  se denomina max-estable si, para todo  $n = 2, 3, \dots$ , existen constantes  $\alpha_n > 0$  y  $\beta_n$  tales que*

$$G^n(\alpha_n z + \beta_n) = G(z). \quad (1.8)$$

Es decir, la propiedad de max-estabilidad la satisfarán aquellas distribuciones en las que la operación de tomar máximos muestrales conduzca a una distribución idéntica, aunque con distintos parámetros de localización y escala.

Esta propiedad se relaciona con el Teorema de Valores Extremos a partir del siguiente resultado:

**Teorema 1.1.3** *Una distribución es max-estable si y solo si es una distribución de valores extremos generalizada.*

## 1.2. Modelos de máximos por bloques

La familia de distribuciones GEV será útil para modelizar la distribución de los **máximos por bloques**.

El procedimiento consiste en agrupar los datos en bloques de igual tamaño y, a continuación, ajustar la distribución GEV al conjunto de los máximos correspondientes a cada uno de los bloques.

El principal problema que presenta este método reside en la elección del tamaño de los bloques, para la cual habrá que encontrar un equilibrio entre el sesgo y la varianza.

La elección de bloques muy pequeños conducirá a una pobre aproximación del modelo, con lo que se aumentará el sesgo al estimar y extrapolar. Por el contrario, la elección de bloques demasiado grandes aumentará la varianza de las estimaciones.

Por cuestiones prácticas, en secuencias de datos temporales mensuales, se suelen tomar bloques de longitud anual, de esta manera los máximos se distribuyen de manera similar en cada uno de los bloques.

### 1.2.1. Estimación de los parámetros de la distribución GEV

Sean  $Z_1, \dots, Z_k$  v.a.i.i.d. con distribución GEV. La función de log-verosimilitud para  $\xi \neq 0$  es

$$\log L(\mu, \sigma, \xi) = -k \log \sigma - (1 + 1/\xi) \sum_{i=1}^k \log \left[ 1 + \xi \left( \frac{z_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^k \left[ 1 + \xi \left( \frac{z_i - \mu}{\sigma} \right) \right]^{-1/\xi},$$

con

$$1 + \xi \left( \frac{z_i - \mu}{\sigma} \right) > 0, \quad \text{para } i = 1, \dots, k, \quad (1.9)$$

mientras que para el caso  $\xi = 0$ ,

$$\log L(\mu, \sigma) = -k \log \sigma - \sum_{i=1}^k \left( \frac{z_i - \mu}{\sigma} \right) - \sum_{i=1}^k \exp \left\{ - \left( \frac{z_i - \mu}{\sigma} \right) \right\}. \quad (1.10)$$

La maximización de estas ecuaciones no tiene solución analítica; sin embargo, para un conjunto dado de datos, la maximización se obtiene de manera directa usando algoritmos numéricos de optimización.

### 1.2.2. Niveles de retorno

Se denomina niveles de retorno a los cuantiles de la distribución de valores extremos generalizada; se notarán como  $z_p$  y se obtendrán como la inversa de la función de distribución GEV:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} \left[ 1 - \{-\log(1-p)\}^{-\xi} \right], & \text{para } \xi \neq 0 \\ \mu - \sigma \log \{-\log(1-p)\}, & \text{para } \xi = 0, \end{cases} \quad (1.11)$$

de forma que  $z_p$  es el nivel de retorno asociado al periodo de retorno  $1/p$ , es decir, se espera que el nivel  $z_p$  sea excedido en promedio una vez cada  $1/p$  unidades de tiempo. Equivalentemente,  $z_p$  será excedido en una unidad de tiempo con probabilidad  $p$ .

La inferencia de los niveles de retorno, para una probabilidad  $p$  dada, se llevará a cabo por sustitución directa de las estimaciones de los parámetros del modelo.

## 1.3. Modelos de excedencias de umbrales

### 1.3.1. Introducción

Modelizar únicamente máximos de bloques es una aproximación poco eficiente en el análisis de valores extremos si algunos de los bloques contienen eventos más extremos que el resto. En este caso, será adecuado el uso de modelos de excedencias de umbrales.

Sea  $X_1, X_2, \dots$  una secuencia de variables aleatorias independientes e idénticamente distribuidas con función de distribución marginal  $F$ . Se considerará un evento extremo aquel que sobrepase el valor de un umbral  $u$  y se notarán las excedencias de este umbral como  $Y = X - u$ .

De esta forma, el comportamiento de las excedencias de  $X$  sobre el umbral  $u$  vendrá dado por la probabilidad condicionada:

$$Pr\{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0. \quad (1.12)$$

Si se conociera la distribución  $F$ , la distribución de las excedencias de umbral sería igualmente conocida; sin embargo, esto no sucede en la práctica, por lo que será necesario el uso de la distribución GEV como aproximación.

### 1.3.2. Distribución de Pareto generalizada

La distribución de Pareto generalizada se utilizará como aproximación para la distribución límite de las excedencias de umbrales.

**Teorema 1.3.1** *Sea  $X_1, X_2, \dots$  una secuencia de variables aleatorias independientes con distribución común  $F$ , y sea*

$$M_n = \text{máx}\{X_1, \dots, X_n\}. \quad (1.13)$$

*Notaremos por  $X$  a un término arbitrario  $X_i$  de esta secuencia y se supondrá que  $F$  satisface el teorema 1.1.2, es decir,*

$$Pr\{M_n \leq z\} \approx G(z), \quad \text{cuando } n \rightarrow \infty, \quad (1.14)$$

donde

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (1.15)$$

para  $\mu, \sigma > 0$  y  $\xi$ . Entonces, para un umbral  $u$  suficientemente grande, la función de distribución de  $(X - u)$ , condicionada a  $X > u$ , tendrá la forma

$$H(y) = \begin{cases} 1 - \left( 1 + \frac{\xi y}{\tilde{\sigma}} \right)^{-1/\xi}, & y > 0 \quad \text{si } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), & y > 0 \quad \text{si } \xi = 0, \end{cases} \quad (1.16)$$

estando definida en  $\{y : y > 0 \text{ y } \left( 1 + \frac{\xi y}{\tilde{\sigma}} \right)^{\frac{1}{\xi}} > 0\}$ , donde

$$\tilde{\sigma} = \sigma + \xi(u - \mu). \quad (1.17)$$

Esta familia se denomina **familia de distribuciones de Pareto generalizada**, y viene caracterizada por los parámetros de escala  $\tilde{\sigma}$  y de forma  $-\infty < \xi < +\infty$ .

Según el teorema anterior, si los máximos por bloques siguen una distribución  $G$ , entonces la distribución de las excedencias de umbral se encuentra dentro de la familia de distribuciones de Pareto generalizada.

El comportamiento de la distribución de Pareto generalizada está determinado por el parámetro  $\xi$ :

- Si  $\xi < 0$ , la distribución de los excesos tiene como límite superior  $u - \tilde{\sigma}/\xi$ .
- Si  $\xi \geq 0$ , la distribución no tiene límite superior. En concreto, para el caso  $\xi > 0$  se tiene la distribución de Pareto ordinaria.

En el caso en el que  $\xi = 0$ , la distribución se corresponde con una exponencial de parámetro  $1/\tilde{\sigma}$ .

A partir de este punto, por simplicidad en la notación, el parámetro de escala de la distribución de Pareto Generalizada,  $\tilde{\sigma}$ , pasará a notarse como  $\sigma$ .

### 1.3.3. Estimación de los parámetros de la distribución de Pareto generalizada

Dado un valor del umbral  $u$  y el número de datos,  $k$ , de la muestra original  $X_1, \dots, X_n$  que sobrepasan el umbral, la estimación de los parámetros  $\sigma$  y  $\xi$  se puede llevar a cabo mediante diferentes métodos.

#### Método de máxima verosimilitud

El logaritmo de la función de verosimilitud para una muestra  $y_1, \dots, y_k$  de variables aleatorias independientes e idénticamente distribuidas con distribución de Pareto generalizada viene dado por

$$\log L(\sigma, \xi) = -k \log \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^k \log \left(1 + \frac{\xi y_i}{\sigma}\right), \quad \text{si } \xi \neq 0 \quad (1.18)$$

$$\log L(\sigma, \xi) = -k \log \sigma - \frac{1}{\sigma} \sum_{i=1}^k y_i, \quad \text{si } \xi = 0. \quad (1.19)$$

Para maximizar esta función; en el caso en que  $\xi \neq 0$ , será necesario llevar a cabo la siguiente reparametrización:

$$(\sigma, \xi) \rightarrow (\tau, \xi) \quad \text{con } \tau = \frac{\sigma}{\xi},$$

con lo que se tiene

$$\log L(\tau, \xi) = -k \log \xi + k \log \tau - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^k \log(1 + \tau y_i).$$

Los estimadores máximo-verosímiles  $\hat{\tau}_{k,n}^{ML}$  y  $\hat{\xi}_{k,n}^{ML}$  se obtienen a partir de la ecuación

$$\frac{1}{\hat{\tau}_{k,n}^{ML}} - \left(\frac{1}{\hat{\xi}_{k,n}^{ML}} + 1\right) \frac{1}{k} \sum_{i=1}^k \frac{y_i}{1 + \hat{\tau}_{k,n}^{ML} y_i} = 0,$$

donde

$$\hat{\xi}_{k,n}^{ML} = \frac{1}{k} \sum_{i=1}^k \log(1 + \hat{\tau}_{k,n}^{ML} y_i),$$

Para la obtención de las estimaciones de los parámetros será necesario utilizar técnicas numéricas, debido a que la maximización de la función de log-verosimilitud no se puede llevar a cabo de forma analítica.

### Método de los momentos ponderados

Los momentos ponderados de una variable aleatoria  $Y$  con función de distribución  $F$  se definen como

$$M_{p,r,s} = E \{Y^p [F(Y)]^r [1 - F(Y)]^s\},$$

para  $p, r$  y  $s$  números reales. Para el caso de la distribución de Pareto generalizada se considerará  $p = 1, r = 0$  y  $s = 0, 1, 2, \dots$ , obteniéndose

$$M_{1,0,s} = \frac{\sigma}{(s+1)(s+1+\xi)}, \quad \text{para } \xi < 1, \quad (1.20)$$

siendo el equivalente muestral

$$\hat{M}_{1,0,s} = \frac{1}{k} \sum_{j=1}^k \left(1 - \frac{j}{k+1}\right)^s y_{j,k}.$$

Los estimadores de los parámetros se obtienen resolviendo (1.20) para  $s = 0$  y  $s = 1$ :

$$\begin{aligned} \hat{\xi}_{PWM} &= 2 - \frac{\hat{M}_{1,0,0}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}}, \\ \hat{\sigma} &= \frac{2\hat{M}_{1,0,0}\hat{M}_{1,0,1}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}}. \end{aligned}$$

La aplicación de este método presenta algunos problemas. En el caso en que  $\xi \geq 1$ , los momentos no existen; y para el caso en el que  $\xi < 0$ , las estimaciones pueden ser inconsistentes con los datos observados.

### Método del percentil básico

Este método es válido en el caso en que  $\xi \neq 0$ ; en el caso contrario, el parámetro  $\sigma$  se podrá estimar de forma eficiente mediante el método de máxima verosimilitud.

Se considerarán dos estadísticos de orden,  $Y_{i,k}$  y  $Y_{j,k}$ , correspondientes a la muestra  $Y_1, \dots, Y_k$ . La función de distribución acumulativa de una distribución de Pareto generalizada, evaluada en estos estadísticos para sus correspondientes valores percentuales, da como resultado un sistema de ecuaciones con dos incógnitas:

$$1 - (1 + \hat{\tau}_{i,j} y_{i,k})^{-\frac{1}{\hat{\xi}_{i,j}}} = p_{i,n}, \quad (1.21)$$

$$1 - (1 + \hat{\tau}_{i,j} y_{j,k})^{-\frac{1}{\hat{\xi}_{i,j}}} = p_{j,n}, \quad (1.22)$$

donde  $\tau = \xi/\sigma$  y  $p_{i,n} = \frac{i}{n+1}$ .

Resolviendo este sistema, se obtiene

$$\hat{\xi}_{i,j} = \frac{\log(1 + \hat{\tau}_{i,j} y_{i,k})}{-\log(1 - p_{i,n})},$$

$$\hat{\sigma}_{i,j} = \frac{\hat{\xi}_{i,j}}{\hat{\tau}_{i,j}}.$$



Los estimadores finales se obtienen calculando  $\hat{\xi}_{i,j}$  y  $\hat{\sigma}_{i,j}$  para todos los pares de estadísticos de orden  $Y_{i,k} < Y_{j,k}$ :

$$\hat{\xi}_{EPM} = \text{mediana} \left\{ \hat{\xi}_{i,j}; \quad i < j \right\},$$

$$\hat{\sigma}_{EPM} = \text{mediana} \left\{ \hat{\sigma}_{i,j}; \quad i < j \right\}.$$

### 1.3.4. Selección del umbral

Para identificar eventos extremos se definirá un umbral  $u$ ; los eventos que superen este umbral,  $\{x_i : x_i > u\}$ , se etiquetarán como  $x_{(1)}, \dots, x_{(k)}$ , y se definirán las excedencias de umbrales  $y_j = x_{(j)} - u$ , para  $j = 1, \dots, k$ .

Teniendo en cuenta el teorema (1.3.1), los valores  $y_j$  se considerarán realizaciones independientes de una variable aleatoria cuya distribución se podrá aproximar mediante la familia de Pareto generalizada.

Para la elección del umbral habrá que tener en cuenta que una elección de un umbral muy bajo aumentaría el sesgo del modelo, mientras que la elección de un umbral demasiado alto generaría un número de excedencias muy pequeño, con lo que aumentaría su varianza.

Se dispone de dos métodos para la elección del umbral:

**Método 1:** Este método se basa en la esperanza de la distribución de Pareto generalizada, de forma que si  $Y$  sigue esta distribución con parámetros  $\sigma$  y  $\xi$ ,

$$E(Y) = \frac{\sigma}{1 - \xi}, \quad \text{con } \xi > 1. \quad (1.23)$$

Si suponemos que la distribución de Pareto generalizada es válida para modelizar las excedencias de un cierto umbral al que notaremos  $u_0$ , para un término arbitrario  $X$  de la serie se tendrá:

$$E(X - u_0 | X > u_0) = \frac{\sigma_{u_0}}{1 - \xi}, \quad \text{con } \xi > 1,$$

donde  $\sigma_{u_0}$  será el parámetro de escala de la distribución de las excedencias del umbral  $u_0$ . Es decir, si la distribución de Pareto es válida para las excedencias del umbral  $u_0$ , será igualmente válida para cualquier umbral  $u > u_0$  mediante una sustitución del parámetro de escala por  $\sigma_u$ . De esta forma, y teniendo en cuenta (1.17), se tiene

$$E(X - u | X > u) = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} + \xi u}{1 - \xi}. \quad (1.24)$$

Se espera que los valores proporcionados por esta expresión varíen linealmente al variar  $u$  en los niveles en los que el ajuste de la distribución de Pareto generalizada es apropiada, lo que se traduce en el siguiente procedimiento.

Se representarán los puntos

$$\left\{ \left( u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right) : u < x_{max} \right\},$$

donde  $x_{(1)}, \dots, x_{(n_u)}$  son las observaciones que exceden  $u$ , y  $x_{max}$  es el máximo  $X_i$ . Este gráfico se denomina gráfico de vida media residual.

Se tomará como umbral el valor máximo para el que el gráfico muestre una tendencia lineal.

**Método 2** Consiste en ajustar la distribución de Pareto generalizada en un rango de umbrales y elegir el umbral adecuado según la estabilidad de los parámetros estimados.

Según la caracterización de la familia de distribuciones de Pareto generalizada, si una distribución de Pareto generalizada es útil para modelizar las excedencias de un umbral  $u_0$ , las excedencias de un umbral  $u$  más alto que el anterior también seguirán una distribución de Pareto generalizada, siendo sus parámetros de forma idénticos, mientras que sus parámetros de escala satisfacen la siguiente relación:

$$\sigma_u = \sigma_{u_0} + \xi(u - u_0) \quad \text{si } \xi \neq 0. \quad (1.25)$$

En el caso en que  $\xi = 0$  se realizará la siguiente reparametrización:

$$\sigma^* = \sigma_u - \xi u.$$

Se representarán  $\hat{\sigma}^*$  y  $\hat{\xi}$  frente a  $u$ . Las estimaciones de ambos parámetros deberían ser constantes por encima de  $u_0$  si  $u_0$  es un umbral válido, para excedencias que siguen la distribución de Pareto generalizada.

Los intervalos de confianza para  $\hat{\xi}$  se obtendrán a través de la matriz de varianzas-covarianzas  $V$ , mientras que los intervalos de confianza para  $\hat{\sigma}^*$  requieren el uso del siguiente resultado:

**Teorema 1.3.2 (Método Delta).** Sea  $\hat{\theta}$  el estimador máximo-verosímil de  $\theta$  con matriz de varianzas-covarianzas  $\Sigma_{\hat{\theta}}$ . Sea  $\psi = h(\theta)$  un vector de parámetros de dimensión  $r$  definido como función de  $\theta$ . Entonces,

$$\hat{\psi} \sim N_r(\psi; \nabla_{\theta}^T \psi \Sigma_{\hat{\theta}} \nabla_{\theta} \psi),$$

donde  $\nabla_{\theta} \psi$  es la matriz de orden  $k \times r$  de las derivadas parciales de  $\psi$  respecto  $\theta$ , dada por

$$\nabla_{\theta} \psi = \begin{bmatrix} \frac{\partial \psi_1}{\partial \theta_1} & \frac{\partial \psi_2}{\partial \theta_1} & \cdots & \frac{\partial \psi_r}{\partial \theta_1} \\ \frac{\partial \psi_1}{\partial \theta_2} & \frac{\partial \psi_2}{\partial \theta_2} & \cdots & \frac{\partial \psi_r}{\partial \theta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \psi_1}{\partial \theta_k} & \frac{\partial \psi_2}{\partial \theta_k} & \cdots & \frac{\partial \psi_r}{\partial \theta_k} \end{bmatrix}.$$

Por tanto, se tendría que

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

donde

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

### 1.3.5. Niveles de retorno

Se denomina niveles de retorno,  $x_k$ , a los cuantiles de la distribución de valores extremos, con  $1/k$  la probabilidad de que  $x_k$  sea superado al menos una vez en una unidad de tiempo.

Suponemos que las excedencias de un umbral  $u$  siguen una distribución de Pareto generalizada; es decir, para  $x > u$ ,

$$\text{Pr}\{X > x | X > u\} = \left[ 1 + \xi \left( \frac{x - u}{\sigma} \right) \right]^{-1/\xi}. \quad (1.26)$$

Por lo que, notando  $\zeta_u = \text{Pr}\{X > u\}$ , el nivel  $x_k$  que se excede en promedio cada  $k$  observaciones es la solución de la ecuación

$$\zeta_u \left[ 1 + \xi \left( \frac{x_k - u}{\sigma} \right) \right]^{-1/\xi} = \frac{1}{k}. \quad (1.27)$$

Despejando el término  $x_k$ , se tiene que el nivel de retorno para excedencias de umbrales viene dado por:

$$x_k = \begin{cases} u + \frac{\sigma}{\xi} \left[ (k\zeta_u)^\xi - 1 \right] & \text{para } \xi \neq 0, \\ u + \sigma \log(k\zeta_u) & \text{para } \xi = 0. \end{cases} \quad (1.28)$$

Para realizar la estimación de los niveles de retorno se sustituirán los parámetros por sus estimaciones máximo verosímiles.

Un concepto relacionado con los niveles de retorno es el tiempo de retorno, o los tiempos en los que se superaría un cierto umbral; estos tiempos se definen recursivamente como:

$$\tau_1 = \text{mín} \{k : X_k > u\},$$

$$\tau_r = \text{mín} \{k > \tau_{r-1} : X_k > u\}, \quad r > 1.$$

## 1.4. Modelización de extremos en secuencias de datos dependientes

Hasta ahora se ha tenido en cuenta que los datos se habían obtenido a partir de sucesiones de variables aleatorias independientes; sin embargo, en la mayoría de casos prácticos esta asunción no es realista. En estos casos, lo usual será considerar que la serie es estacionaria.

Una serie se considerará estacionaria si sus distribuciones de probabilidad se mantienen estables a lo largo del tiempo. En este caso, las variables serán mutuamente dependientes, pero sus propiedades estocásticas permanecerán constantes a lo largo del tiempo.

La estacionariedad se define de como:

**Teorema 1.4.1** *Sea  $X_1, X_2, \dots$  una serie de tiempo. Esta serie se dice estacionaria si para cada conjunto de índices  $1 \leq t_1 < t_2 < \dots < t_m$ , la distribución conjunta de  $(X_{t_1}, X_{t_2}, \dots, X_{t_m})$  coincide con la distribución conjunta de  $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_m+h})$ , para cualquier  $h$ .*

Coles (2001) da la siguiente definición:

**Definición 1.4.1** *Una serie estacionaria  $X_1, X_2, \dots$  satisface la condición de los  $D(u_n)$  si, para todo  $i_1 < \dots < i_p < j_1 < \dots < j_q$  con  $j_i - i_p > l$ , se tiene que*

$$\begin{aligned} & |P[X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, \dots, X_{j_q} \leq u_n] \\ & - P[X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n] P[X_{j_1} \leq u_n, \dots, X_{j_q} \leq u_n]| \leq \alpha(n, l), \end{aligned}$$

donde  $\alpha(n, l_n) \rightarrow 0$  para alguna sucesión  $l_n$  de forma que  $\frac{l_n}{n} \rightarrow 0$  cuando  $n \rightarrow \infty$ .

Para secuencias de variables independientes la diferencia en probabilidad anterior es igual a cero. En el caso en que las variables no son independientes, si están suficientemente alejadas entre sí la diferencia en probabilidad es lo suficientemente cercana a cero para no tener efecto en las leyes de límites para extremos, lo que se resume en el siguiente teorema:

**Teorema 1.4.2** *Sea  $X_1, X_2, \dots$  un proceso estacionario y sea  $M_n = \max\{X_1, \dots, X_n\}$ . Entonces, si  $\{a_n > 0\}$  y  $\{b_n\}$  son secuencias de constantes tales que*

$$Pr\{(M_n - b_n)/a_n \leq z\} \rightarrow G(z),$$

donde  $G$  es una función de distribución no degenerada y la condición  $D(u_n)$  se satisface con  $u_n = a_n z + b_n$  para cualquier número real  $z$ ,  $G$  pertenece a la familia de distribuciones de valores extremos generalizada.

Este resultado implica que dada una sucesión de variables suficientemente lejanas y dependientes entre sí (en el sentido especificado por la condición  $D(u_n)$ ),

los valores máximos de estas series estacionarias se distribuyen según la misma distribución límite que en el caso de sucesiones de variables independientes. Sin embargo, los parámetros de la distribución se verán afectados por la dependencia.

**Teorema 1.4.3** *Sea  $X_1, X_2, \dots$  un proceso estacionario y sea  $X_1^*, X_2^*, \dots$  una sucesión de variables independientes con la misma distribución marginal. Se define  $M_n = \max\{X_1, \dots, X_n\}$  y  $M_n^* = \max\{X_1^*, \dots, X_n^*\}$ . Bajo condiciones adecuadas de regularidad,*

$$\Pr\{(M_n^* - b_n)/a_n \leq z\} \rightarrow G_1(z)$$

*cuando  $n \leftarrow \infty$  para sucesiones  $\{a_n > 0\}$  y  $\{b_n\}$ , donde  $G_1$  es una función de distribución no degenerada, si y solo si*

$$\Pr\{(M_n - b_n)/a_n \leq z\} \rightarrow G_2(z),$$

*donde*

$$G_2(z) = G_1^\theta(z)$$

*para una constante  $\theta$  tal que  $0 < \theta \leq 1$ .*

De esta forma, si  $M_n^*$  sigue una distribución GEV de parámetros  $(\mu, \sigma, \xi)$ ,  $M_n$  seguirá una distribución GEV con parámetros  $(\mu^*, \sigma^*, \xi)$  donde

$$\begin{aligned}\mu^* &= \mu - \frac{\sigma}{\xi} (1 - \theta^{-\xi}), \\ \sigma^* &= \sigma \theta^\xi.\end{aligned}$$



## Capítulo 2

# Teoría de valores extremos multivariante

### 2.1. Introducción

En el estudio de extremos para dos o más procesos, si bien cada uno de ellos puede ser modelizado individualmente usando técnicas univariantes, sin embargo, también será interesante estudiar las relaciones que puedan existir entre ellos. En particular, puede ocurrir que alguna combinación de estos procesos sea de mayor interés que los procesos individuales.

Sean  $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,p})$ ,  $i \in \{1, \dots, n\}$ , una secuencia de vectores aleatorios independientes e idénticamente distribuidos con función de distribución conjunta  $F$  y marginales  $F_1, \dots, F_p$ . El vector de máximos, que se notará como  $M$ , para el caso multivariante se define como:

$$M_j = \max_{i \in \{1, \dots, n\}} \{X_{i,j}\}; \quad \text{para } j = 1, \dots, p,$$

$$M = \{M_1, \dots, M_p\}.$$

Este vector no se tiene que corresponder necesariamente con un vector observado en la serie original.

### 2.2. Cópulas

En esta sección se describe uno de los procedimientos más usados para representar la estructura de dependencia existente entre varias funciones de densidad univariantes.

A través las cópulas, se podrán analizar y estimar de forma separada las componentes univariantes y la estructura de dependencia multivariante. La unión de

las diferentes estimaciones conducirá a una estimación de la función de densidad conjunta.

Sea  $X = (X_1, \dots, X_p)$  un vector aleatorio con función de distribución  $F$ .

Supongamos que cada una de las variables  $X_j$  es continua con función de distribución marginal  $F_j$ ; entonces, aplicando una transformación de probabilidad,  $Y_j = F_j(X_j)$  tendrá distribución uniforme  $(0, 1)$ .

La función de distribución  $C$  del vector  $Y = (Y_1, \dots, Y_p)$  es la cópula de  $F$ . Es decir,

$$C(u) = P(Y_1 \leq u_1, \dots, Y_p \leq u_p) = F(F_1^{-1}(u_1), \dots, F_p^{-1}(u_p)),$$

donde  $F_j^{-1}$  es la función cuantil de  $F_j$ .

Se podrá volver a la distribución original desde la cópula mediante la transformación cuantil:

$$F(X) = C(F_1(x_1), \dots, F_p(x_p)).$$

## 2.3. Distribución de valores extremos multivariante

### 2.3.1. Caso bivalente

Coles (2001) propone una caracterización asintótica de la distribución de valores extremos, a partir de la función de distribución conjunta de  $M^* = (M_1/n, M_2/n)$ :

**Teorema 2.3.1** *Sea  $M^* = (M_1^*, M_2^*)$ , con  $(X_{i1}, X_{i2})$ ,  $i = 1, \dots, n$ , vectores independientes con distribución marginal Fréchet. Entonces, si para  $n \rightarrow \infty$*

$$Pr \{M_1^* \geq x_1, M_2^* \geq x_2\} \xrightarrow{d} G(x_1, x_2), \quad (2.1)$$

donde  $G$  es una función de distribución no degenerada, se tiene que

$$G(x_1, x_2) = \exp \{-V(x_1, x_2)\}, \quad x_1 > 0, x_2 > 0,$$

donde

$$V(x_1, x_2) = 2 \int_0^1 \max\left(\frac{\omega}{x_1}, \frac{1-\omega}{x_2}\right) dH(\omega) \quad (2.2)$$

y  $H$  es una función de distribución en  $[0, 1]$  que satisface la condición

$$\int_0^1 \omega dH(\omega) = 1/2.$$

A diferencia del caso univariante, esta distribución engloba a infinitas familias de distribuciones.



### Ejemplo 1

Si se distribuye la masa de la función  $H$  en los puntos  $\omega = 0$  y  $\omega = 1$ , se obtiene

$$V(x_1, x_2) = x_1^{-1} + x_2^{-1}$$

y la correspondiente distribución de valores extremos generalizada será

$$G(x_1, x_2) = \exp\{-\{x_1^{-1} + x_2^{-1}\}\}, \quad x_1 > 0, \quad x_2 > 0.$$

Esta función se puede factorizar en función de  $x_1$  y  $x_2$ , por lo que ambas variables serían independientes.

### Ejemplo 2

Si se concentra toda la masa de la función  $H$  en  $\omega = 0,5$ , se obtiene,

$$G(x_1, x_2) = \exp\{-\max(x_1^{-1}, x_2^{-1})\}, \quad x_1 > 0, \quad x_2 > 0,$$

que correspondería al caso de variables perfectamente dependientes,  $X_1 = X_2$ .

La clase completa de funciones de distribución de valores extremos bivariate se puede obtener mediante una generalización de las distribuciones marginales. Tomando

$$\begin{aligned} \tilde{x}_1 &= \left[ 1 + \xi_{x_1} \left( \frac{x_1 - \mu_{x_1}}{\sigma_{x_1}} \right) \right]^{1/\xi_{x_1}}, \\ \tilde{x}_2 &= \left[ 1 + \xi_{x_2} \left( \frac{x_2 - \mu_{x_2}}{\sigma_{x_2}} \right) \right]^{1/\xi_{x_2}}, \end{aligned}$$

se obtiene

$$G(x_1, x_2) = \exp\{-V(\tilde{x}_1, \tilde{x}_2)\},$$

donde la función  $V$  satisface (2.2) para cualquier  $H$ .

Las distribuciones marginales serán GEV con parámetros  $(\mu_{x_j}, \sigma_{x_j}, \xi_{x_j})$  para  $j = 1, 2$ .

La función  $V$  será homogénea de orden -1, en el sentido de que para cualquier constante  $a > 0$  se tiene

$$\begin{aligned} V(a^{-1}x_1, a^{-1}x_2) &= 2 \int_0^1 \max\left(\frac{\omega}{a^{-1}x_1}, \frac{1-\omega}{a^{-1}x_2}\right) \\ &= 2a \int_0^1 \max\left(\frac{\omega}{x_1}, \frac{1-\omega}{x_2}\right) = aV(x_1, x_2). \end{aligned}$$

### 2.3.2. Extensión para n variables

La siguiente caracterización del comportamiento límite del vector de máximos se va a desarrollar basándose en la pertenencia a un dominio de atracción.

Sea  $\mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^d$ . Si  $X_i = (X_{i,1}, \dots, X_{i,p})$ ,  $i = 1, \dots, n$ , son vectores aleatorios independientes e idénticamente distribuidos de dimensión  $p$  con función de distribución  $F$ , se puede considerar la existencia de secuencias  $(\mathbf{a}_n)_n$  y  $(\mathbf{b}_n)_n$  pertenecientes a  $\mathbb{R}^p$ , con  $a_{n,j} > 0$  y  $b_{n,j} \in \mathbb{R}$ ,  $\forall j = 1, \dots, p$ , y una función de distribución  $G$  con marginales no degeneradas tales que, cuando  $n \rightarrow \infty$ ,

$$Pr \left\{ \left( \max_{i=1, \dots, n} X_i - \mathbf{b}_n \right) / \mathbf{a}_n \leq x \right\} = F^n(\mathbf{a}_n \mathbf{x} + \mathbf{b}_n) \rightarrow G(\mathbf{x}),$$

donde  $G$  es la función de distribución multivariante de valores extremos y  $F$  pertenece al dominio de atracción de  $G$  para el máximo.

#### Ejemplo 1

Se considera la función de distribución normal multivariante  $F_{\mathcal{N}}$ , con marginales univariantes  $\mathcal{N}(0, 1)$  y correlaciones  $EX_i X_j < 1, \forall i, j = 1, \dots, p$ . El dominio de atracción será el producto de distribuciones marginales Gumbel:

$$F_{\mathcal{N}}^n(a_n \mathbf{x} + \mathbf{b}_n) \rightarrow G(x) = \prod_{j=1}^p \exp \{-e^{-x_j}\},$$

con constantes

$$a_n = (2 \log n)^{-1/2},$$

$$\mathbf{b}_n = b_n \mathbf{1},$$

donde

$$b_n = (2 \log n)^{1/2} - 1/2 (\log(\log n) + \log 2\pi) / (2 \log n)^{1/2}$$

$$\mathbf{1} = (1, \dots, 1).$$

### 2.3.3. Transformaciones

Algunas transformaciones que serán de utilidad son las siguientes:

**Transformación de un modelo log-Gumbel en un modelo Fréchet.** Una variable aleatoria Gumbel con parametros de localización y escala  $\log \sigma$  y  $1/\alpha$  se corresponde con una distribución Fréchet con parámetros de forma y escala  $\alpha$  y  $\sigma$ , respectivamente.

**Transformación de un modelo Weibull en un modelo Fréchet.** Una variable aleatoria Weibull con parametros de forma y escala  $-\alpha$  y  $1/\sigma$  se corres-

ponde con una distribución Fréchet con parámetros de forma y escala  $\alpha$  y  $\sigma$ , respectivamente.

**Transformación cuantil.** Si  $U$  es una variable aleatoria distribuida uniformemente en  $(0, 1)$ , entonces  $F^{-1}(U)$  tiene función de distribución  $F$ .

**Transformación de probabilidad.** De forma inversa a la transformación cuantil, si  $X$  es una variable aleatoria con función de distribución continua  $F$ , entonces  $F(X)$  tiene función de distribución uniforme  $(0,1)$ .

## 2.4. Convergencia del vector de máximos

La convergencia débil de una secuencia de vectores aleatorios implica la convergencia débil de cada uno de los componentes. Por lo tanto, de la misma manera que en el caso univariante, será razonable aplicar transformaciones a cada uno de los vectores marginales y considerar la secuencia de variables aleatorias

$$\left( \frac{M_n - b_{n,j}}{a_{n,j}} : j = 1, \dots, p \right),$$

para constantes de normalización  $a_{n,j} > 0$  y  $b_{n,j}$ . Cada uno de los componentes  $(M_n - b_{n,j}/a_{n,j})$ ,  $j = 1, \dots, p$ , converge débilmente cuando  $n \rightarrow \infty$  a una distribución max-estable univariante, si se cumplen las condiciones de convergencia.

Sin embargo, la convergencia de cada una de las  $p$  componentes es estrictamente más débil que la convergencia del vector conjunto de máximos. Se necesitará además establecer la estructura de dependencia de la distribución conjunta  $F$  de los vectores aleatorios  $X_i$ . Una manera de describir esta dependencia será a través de cópulas:

$$Pr [X_i \leq x] = F(x) = C_1(F_1(x_1), \dots, F_p(x_p)).$$

Suponiendo continuidad de las funciones marginales, la cópula  $C_1$  de la función de distribución  $F$  es única y puede obtenerse como la función de distribución conjunta de los vectores aleatorios  $(F_1(X_{i1}), \dots, F_p(X_{ip}))$ .

La cópula del vector de máximos y, por tanto, de cualquier vector creado mediante componentes transformadas, vendrá dada por

$$C_i(u) = \left\{ C_1 \left( u_1^{1/i}, \dots, u_p^{1/i} \right) \right\}^i.$$

Puesto que los vectores de máximos convergen en distribución a una distribución límite no degenerada, la secuencia de cópulas  $C_i$  también debe converger.

La cópulas obtenidas a través de los límites de  $C_i$  cuando  $i \rightarrow \infty$  se denominan cópulas de valores extremos, esto es, una cópula  $C$  es una cópula de valores extremos si existe una cópula  $C_1$  tal que, cuando  $i \rightarrow \infty$ ,

$$\lim_{n \rightarrow \infty} \left\{ C_1 \left( u_1^{1/i}, \dots, u_p^{1/i} \right) \right\}^i = C(u_1, \dots, u_p).$$

La cópula  $C_1$  forma parte del dominio de atracción de  $C$ , al pertenecer esta última a la clase de posibles cópulas límite.

Una cópula  $C$  es max-estable si, para todo  $u \in [0, 1]^d$  y  $k = 1, 2, \dots$ ,

$$C(u) = \left\{ C\left(u_1^{1/k}, \dots, u_d^{1/k}\right) \right\}^k.$$

Por tanto, la clase de cópulas de valores extremos coincide con la clase de cópulas max-estables.

En resumen, las distribuciones límite no degeneradas de vectores de máximos transformados tienen marginales de valores extremos y cópulas de valores extremos o max-estables. En concreto, si

$$Pr \left[ \bigcap_{j=1}^p \left\{ \frac{M_{n,j} - b_{n,j}}{a_{n,j}} \leq x_j \right\} \right] \xrightarrow{\omega} G(x_1, \dots, x_d), \quad n \rightarrow \infty,$$

entonces,

$$G(x_1, \dots, x_d) = C(G_1(x_1), \dots, G_p(x_p)).$$

con marginales de valores extremos  $G_1, \dots, G_p$  y cópula de valores extremos  $C$ .

## 2.5. Funciones de dependencia

Se partirá del límite que caracteriza a la cópula de valores extremos para, a continuación, tomar logaritmos y aplicar una expansión lineal. De esta forma, se llegará a la expresión equivalente

$$\begin{aligned} & \lim_{n \rightarrow \infty} n \left\{ 1 - C_1(1 - n^{-1}x_1, \dots, 1 - n^{-1}x_p) \right\} \\ & = -\log C(e^{-x_1}, \dots, e^{-x_p}) = \ell(x), \quad x \in [0, \infty)^d. \end{aligned}$$

El límite  $\ell$  se conoce como función de dependencia de cola estable de  $C$ .

Si se considera el vector aleatorio  $X = (X_1, \dots, X_p)$ ,

$$\begin{aligned} & 1 - C_1(1 - x_1/n, \dots, 1 - x_p/n) \\ & = Pr[F_1(X_1) > 1 - x_1/n \cup \dots \cup F_p(X_p) > 1 - x_p/n] = \ell(x). \end{aligned}$$

Esta probabilidad hace referencia al evento en el que al menos una de las  $p$  componentes excede un alto percentil de su distribución.

La cópula de cola,  $R$ , introducida por Schmidt and Stadtmüller (2006), plantea el caso opuesto donde las  $p$  componentes exceden simultáneamente un alto percentil,

$$\begin{aligned} \lim_{n \rightarrow \infty} n \Pr [F_1(X_1) > 1 - x_1/n \cap \dots \cap F_p(X_p) > 1 - x_p/n] \\ = R(x), \end{aligned}$$

para  $x \in [0, \infty)^p$ .

Ambas funciones,  $\ell(x)$  y  $R(x)$ , son homogéneas.

Se aplicará la restricción de estas funciones al 1-símplex

$$\Delta_{p-1} = \{(\omega_1, \dots, \omega_p) \in [0, 1]^p : \omega_1 + \dots + \omega_p = 1\}.$$

La restricción  $\Delta_{p-1}$  sobre  $\ell$  se denomina función de dependencia de Pickands. Por la propiedad de homogeneidad de  $\ell$ ,

$$\ell(x) = (x_1 + \dots + x_p) D(\omega_1, \dots, \omega_p), \quad \omega_j = \frac{x_j}{x_1 + \dots + x_p}.$$

La función  $\ell(x)$  hace referencia a la probabilidad de la unión de los eventos  $\{F_j(X_j) > 1 - x_j/n\}$ , siendo la probabilidad de cada uno de estos eventos por separado  $x_j/n$ , con  $x_j \leq n$ ; como consecuencia, las cotas elementales de  $\ell(x)$  serán

$$\begin{aligned} \max(x_1/n, \dots, x_p/n) &\leq \\ &\leq \Pr [F_1(X_1) > 1 - x_1/n \cup \dots \cup F_p(X_p) > 1 - x_p/n] \leq x_1/n + \dots + x_p/n. \end{aligned}$$

Multiplicando por  $n$  y haciéndola tender a infinito, se obtiene

$$\max(x_1, \dots, x_p) \leq \ell(x_1, \dots, x_p) \leq x_1 + \dots + x_p, \quad x \in [0, \infty)^p.$$

Según esta relación anterior, una cópula de valores extremos deberá satisfacer la relación

$$u_1 \cdots u_p \leq C(u_1, \dots, u_p) \leq \max(u_1, \dots, u_i).$$

Las cotas inferior y superior de los dos casos anteriores se corresponden, respectivamente, con los casos extremos de independencia y dependencia perfecta.

La cópula  $C$  puede ser dada en términos de la función de dependencia de cola a través de

$$C(u_1, \dots, u_p) = \exp\{-\ell(-\log u_1, \dots, -\log u_p)\}, \quad u \in (0, 1]^p.$$

En teoría de valores extremos, a menudo es conveniente estandarizar a otras distribuciones distintas a la uniforme  $(0, 1)$ . Las tres formas más comunes son las distribuciones de Fréchet, de Gumbel y exponencial inversa, que se representan respectivamente como

$$C\left(e^{-1/x_1}, \dots, e^{-1/x_p}\right) = \exp\{-\ell(1/x_1, \dots, 1/x_p)\}, \quad x \in (0, \infty)^p,$$

$$C\left(e^{-e^{-x_1}}, \dots, e^{-e^{-x_p}}\right) = \exp\{-\ell(e^{-x_1}, \dots, e^{-x_p})\}, \quad x \in \mathbb{R}^p,$$

$$(e^{x_1}, \dots, e^{x_p}) = \exp\{-\ell(-x_1, \dots, -x_p)\}, \quad x \in (-\infty, 0)^p,$$

## 2.6. Modelización y estimación de parámetros

### 2.6.1. Caso bivalente

A partir de las series de realizaciones independientes  $(x_{1,1}, x_{2,1}), \dots, (x_{1,n}, x_{2,n})$  se obtendrá la secuencia de bloques de máximos  $(z_{1,1}, z_{2,1}), \dots, (z_{1,m}, z_{2,m})$ .

Las series de bloques de máximos se considerarán de forma independiente y se modelizarán como en el caso univariante usando la distribución GEV, es decir, cada  $z_{i,j}$  será tratada como una realización independiente de una variable aleatoria  $Z_i$ , para cada  $i = 1, 2$ , donde

$$Z_i \sim GEV(\mu_i, \sigma_i, \xi_i).$$

A continuación se obtendrán las estimaciones de máxima verosimilitud de los parámetros según la metodología univariante, de forma que la variable transformada

$$\tilde{Z}_i = \left[1 + \hat{x}_i \left(\frac{Z_i - \hat{\mu}_i}{\hat{\sigma}_i}\right)\right]^{1/\hat{x}_i}$$

se distribuye aproximadamente según una distribución Fréchet estándar. Los pares  $(\tilde{z}_{1,j}, \tilde{z}_{2,j})$  se obtienen sustituyendo en la expresión anterior las observaciones  $(z_{1,j}, z_{2,j})$ .

La función de densidad de este modelo es

$$g(x, y) = \{V_x(x, y)V_y(x, y) - V_{xy}(x, y)\} \exp\{-V(x, y)\}, \quad x > 0, y > 0,$$

donde  $V_x$ ,  $V_y$  y  $V_{x,y}$  son las derivadas parciales y mixtas de  $V$ , respectivamente. Seleccionando un modelo con parámetro  $\theta$  para  $V$ , y a partir de su función de verosimilitud

$$L(\theta) = \prod_{i=1}^m g(\tilde{z}_{1,i}, \tilde{z}_{2,i}),$$

se obtendrán las estimaciones de máxima verosimilitud de  $\theta$ .

## 2.6.2. Extensión para n variables. Estimación mediante cópulas

Sean  $x_i = (x_{i,1}, \dots, x_{i,d})$ ,  $i = 1, \dots, k$ , realizaciones de una variable que sigue una distribución de valores extremos  $G$  con vectores de parámetros de localización, escala y forma  $\mu$ ,  $\sigma$ ,  $\xi$ , respectivamente.

A través de la cópula  $G$  se tiene la representación

$$G(x) = C \left( (G_{\mu_j, \sigma_j, \xi_j}(x_j))_{j \leq d} \right),$$

donde  $G_{\mu_j, \sigma_j, \xi_j}$  es la  $j$ -ésima distribución marginal de  $G$ .

Para estimar  $G$  se deberán construir estimadores de  $\mu$ ,  $\sigma$ ,  $\xi$  y  $C$ . Las estimaciones de los parámetros se calcularán de forma independiente para cada función de distribución marginal  $G_{\mu_j, \sigma_j, \xi_j}$  mediante máxima verosimilitud.

La estimación de  $C$  se basará en los vectores transformados:

$$z_i = \left( G_{\hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j}(x_{i,j}) \right)_{j \leq d'}$$

## 2.7. Niveles de retorno

### 2.7.1. Caso bivalente

Se denominará variable estructural a la operación

$$Z = \phi(M_{x_1}, M_{x_2}),$$

con función de distribución

$$Pr \{Z \leq z\} = \int_{A_z} g(x_1, x_2) dx_1 dx_2,$$

donde  $A_z = \{(x_1, x_2) : \phi(x_1, x_2) \leq z\}$ .

El nivel de retorno en  $N$  unidades de tiempo de la variable estructural  $Z$  es la solución de

$$G_Z(z) = 1 - 1/N,$$

donde  $G_Z$  es la función de distribución de  $Z$ .

### 2.7.2. Extensión para n variables. Medida de intensidad

Aplicando la teoría univariante se tiene que el tiempo de retorno de la observación  $X_j$  es  $Y_j = 1/\{1 - F_j(X_j)\}$ , con distribución de Pareto, ya que  $Pr[Y_j > y] = Pr[F_j(X_j) > 1 - 1/y] = 1/y$  para  $y \leq 1$ .

Suponiendo que la cópula  $C_1$  se encuentra en el dominio de atracción de una cópula de valores extremos con función de dependencia  $\ell$ , el vector aleatorio  $Y = (Y_1, \dots, y_p)$  satisface

$$\begin{aligned} t \{1 - C_1(1 - x_1/t, \dots, 1 - x_p/t)\} &= t \Pr \left[ \bigcup_{j=1}^p \{Y_j > t/x_j\} \right] \\ &= t \Pr [Y/t \in ([0, \infty]^p \setminus [0, 1/x])] \rightarrow \ell(x), \quad t \rightarrow \infty. \end{aligned}$$

En el espacio  $\mathbb{E}_d = [0, \infty]^p \setminus \{0\}$ , existe una medida  $\mu$ , que se conocerá como **medida de intensidad**, tal que

$$E \left[ \sum_{i=1}^n I(Y_i/n \in \cdot) \right] = n \Pr [Y/n \in \cdot] \xrightarrow{v} \mu(\cdot), \quad n \rightarrow \infty,$$

siendo  $\lim_{n \rightarrow \infty} n E[f(Y/n)] = \int_{\mathbb{E}_p} f(x) d\mu(x)$  para toda función  $f$  continua y acotada en  $\mathbb{E}_p$  y que se hace cero en su origen, de forma que, cuando  $n$  crece, el vector  $Y/n$  es desplazado cerca del origen, en las cercanías donde la función  $f$  es cero. Por tanto, la medida de intensidad solo afecta a la cola superior de la distribución de  $Y$ .

La medida  $\mu$  representa el número esperado de observaciones en el suconjunto sobre el que se aplique esta medida.

La función  $\ell$  de dependencia actúa como función de distribución de la medida de intensidad, de forma que

$$\ell(x) = \mu([0, \infty]^p \setminus [0, 1/x]).$$

## 2.8. Modelos de excedencias de umbrales

La aproximación mediante métodos de máximos por bloques presenta varios inconvenientes; en primer lugar, todos los datos excepto el máximo de cada bloque son descartados; además, el vector compuesto por estos máximos no se corresponde con observaciones reales.

Una aproximación más flexible y eficiente es considerar las excedencias sobre un umbral.

### 2.8.1. Caso bivalente

Sean  $(x_{1,1}, x_{1,2}), \dots, (x_{n,1}, x_{n,2})$  realizaciones independientes de un par de variables aleatorias  $(X_1, X_2)$  con función de distribución conjunta  $F$ .

Para dos umbrales  $u_{x_1}$  y  $u_{x_2}$  se obtendrán dos funciones de probabilidad de Pareto con sus correspondientes parámetros para cada una de las distribuciones marginales de  $F$ .

Mediante la siguiente transformación,



$$\tilde{X}_1 = - \left( \log \left\{ 1 - \zeta_{x_1} \left[ 1 + \frac{\xi_{x_1} (X_1 - u_{x_1})}{\sigma_{x_1}} \right]^{-1/\xi_{x_1}} \right\} \right)^{-1},$$

$$\tilde{X}_2 = - \left( \log \left\{ 1 - \zeta_{x_2} \left[ 1 + \frac{\xi_{x_2} (X_2 - u_{x_2})}{\sigma_{x_2}} \right]^{-1/\xi_{x_2}} \right\} \right)^{-1},$$

se obtiene el par  $(\tilde{X}_1, \tilde{X}_2)$  con función de distribución conjunta  $\tilde{F}$  y marginales Fréchet para  $X_1 > u_{x_1}$  y  $X_2 > u_{x_2}$ . Por tanto,

$$\begin{aligned} \tilde{F}(\tilde{x}_1, \tilde{x}_2) &= \left\{ \tilde{F}^n(\tilde{x}_1, \tilde{x}_2) \right\}^{1/n} \approx [\exp \{-V(\tilde{x}_1/n, \tilde{x}_2/n)\}]^{-1/n} \\ &= \exp \{-V(\tilde{x}_1, \tilde{x}_2)\}, \end{aligned}$$

por la propiedad de homogeneidad de  $V$ . Finalmente, dado que  $F(x_1, x_2) = \tilde{F}(\tilde{x}_1, \tilde{x}_2)$ , se tiene que

$$F(x_1, x_2) \approx G(x_1, x_2) = \exp \{-V(\tilde{x}_1, \tilde{x}_2)\}, \quad x_1 > u_{x_1}, \quad x_2 > u_{x_2}.$$

La inferencia en este caso es complicada, ya que dado un par bivalente, puede ocurrir que solo uno de sus componentes supere el umbral. Para solucionar este problema, se definen los siguientes espacios:

$$R_{0,0} = (-\infty, u_{x_1}) \times (-\infty, u_{x_2}),$$

$$R_{1,0} = [u_{x_1}, \infty) \times (-\infty, u_{x_2}),$$

$$R_{0,1} = (-\infty, u_{x_1}) \times [u_{x_2}, \infty),$$

$$R_{1,1} = [u_{x_1}, \infty) \times [u_{x_2}, \infty),$$

de forma que, por ejemplo, un punto  $(x_1, x_2) \in R_{0,1}$  si la componente  $x_1$  no supera el umbral  $u_{x_1}$  y la componente  $x_2$  supera el umbral  $u_{x_2}$ .

La inferencia del modelo aproximado podrá llevarse a cabo por máxima verosimilitud para puntos pertenecientes a la región  $R_{1,1}$ ; para puntos en otras regiones, al no poder aplicarse  $\tilde{F}$ , será necesario censurar la componente correspondiente en la función de verosimilitud.

Por ejemplo, si se supone que  $(x_1, x_2) \in R_{1,0}$ , entonces,  $x_1 > u_{x_1}$  pero  $x_2 < u_{x_2}$  y los datos solo proporcionarán información de la componente  $x_1$ ; de esta forma, la contribución a la verosimilitud de este punto será

$$Pr \{X_1 = x_1, X_2 \leq u_{x_2}\} = \left. \frac{\partial F}{\partial x} \right|_{(x_1, u_{x_2})}.$$

Así, para cada una de las regiones se usará la función de verosimilitud:

$$L(\theta; (x_{1,1}, x_{1,2}), \dots, (x_{n,1}, x_{n,2})) = \prod_{i=1}^n \psi(\theta; (x_{i,1}, x_{i,2})),$$

donde

$$\psi(\theta; (x_1, x_2)) = \begin{cases} \left. \frac{\partial^2 F}{\partial x_1 \partial x_2} \right|_{(x_1, x_2)} & (x_1, x_2) \in R_{1,1} \\ \left. \frac{\partial F}{\partial x_1} \right|_{(x_1, u_{x_2})} & (x_1, x_2) \in R_{1,0} \\ \left. \frac{\partial F}{\partial x_2} \right|_{(u_{x_1}, x_2)} & (x_1, x_2) \in R_{0,1} \\ F(u_{x_1}, u_{x_2}) & (x_1, x_2) \in R_{0,0} \end{cases}.$$

Mediante la maximización de la función de log-verosimilitud se obtendrán las estimaciones y los errores estándar de los parámetros de  $F$ . De la misma forma que en el caso de máximos por bloques, la inferencia puede simplificarse llevando a cabo estimaciones marginales para después aplicar las transformaciones a las variables. De este modo, al realizar la maximización, la función de verosimilitud únicamente dependerá de los parámetros de  $V$ .

Se puede utilizar otro método alternativo si existe una variable estructural  $Z = \phi(X_1, X_2)$ ; en este caso se aplicarán técnicas univariantes de excedencias de umbrales a cada una de las series  $z_i = \phi(x_{i,1}, x_{i,2})$ .

## 2.8.2. Extensión a $n$ variables

Sea  $X_i = (X_{i,1}, \dots, X_{i,p})$ , con  $i = 1, \dots, n$ , una muestra de observaciones de dimensión  $p$ , independientes e idénticamente distribuidas con función de distribución conjunta  $F$ , perteneciente al dominio de atracción de una distribución de valores extremos multivariante  $G$ .

Se han propuesto varios métodos de análisis de excedencias de umbrales, todos ellos basados en la aproximación de  $F$  en una región extrema definida usando las condiciones del dominio de atracción especificado en el siguiente teorema.

**Teorema 2.8.1** *Sea  $X$  un vector aleatorio de dimensión  $p$  con función de distribución  $F$  y marginales Fréchet, y sea  $\{X_i\}$ ,  $i = 1, \dots, n$ , una secuencia extraída de  $F$ . Sea también  $\mathbb{1}(\cdot)$  una función indicadora tal que  $\mathbb{1}(z)$  es 1 si  $z \in C$  y 0 en otro caso. La función de distribución  $F$  pertenece al dominio de atracción de una distribución  $GEV$  multivariada con función de distribución  $G$  con marginales Fréchet si y solo si*

$$(a) \quad \lim_{s \rightarrow \infty} \frac{\log F(sx)}{\log F(s1)} = \frac{\log G(x)}{\log G(1)}$$

$$(b) \quad \lim_{s \rightarrow \infty} \frac{1-F(sx)}{1-F(s1)} = \frac{\log G(x)}{\log G(1)}$$

$$(c) \lim_{s \rightarrow \infty} P(X \leq sx | X > s) = \frac{1}{-\log G(1)} \log \left( \frac{G(x)}{G(\min(x, 1))} \right)$$

(d) El proceso puntual  $N_n$ , definido como

$$N_n(\cdot) = \sum_{i=1}^n \mathbb{1}(X_i/n \in \cdot),$$

converge débilmente a un proceso de Poisson no homogéneo en  $E = [0, \infty]^d \setminus \{0\}$  con medida de intensidad  $\mu$  definida como

$$\mu \left\{ y \in E : \|y\| > r, \frac{y}{\|y\|} \in A \right\} = \frac{H(A)}{r}.$$

Smith et al (1997) utilizan la condición (b) para obtener la aproximación  $F(x_1, \dots, x_p) \approx 1 + \log G(x_1, \dots, x_p)$ , válida en una región donde  $f_j(x_j)$  es cercana a 1 para cada  $x_j$ ,  $j = 1, \dots, p$ . Sin embargo, este método presenta el inconveniente de que cuando  $F$  es asintóticamente independiente el sesgo de la aproximación es significativo.

Ledford y Tawn (1996) corrigen el sesgo de la aproximación anterior usando la condición (a) para obtener  $F(x_1, \dots, x_p) = G(x_1, \dots, x_p)$ , de forma que  $f_j(x_j)$  es cercana a 1 para cada  $x_j$ ,  $j = 1, \dots, p$ .

En ambos métodos se utiliza un modelo paramétrico para  $G$ , obteniéndose a través de las aproximaciones una parametrización de la cola de  $F$ . Los parámetros de los modelos se estimarán por máxima verosimilitud mediante un mecanismo de censura. En concreto, si  $u_j$  es un umbral tal que  $F_j(u_j)$  es cercana a 1, todas las observaciones  $X_{i,j} \leq u_j$  estarán censuradas a la derecha por  $u_j$ ; de esta forma, la contribución a la verosimilitud de una variable  $X_i$  es

$$l_i = \frac{\partial^q}{\partial x_{j_1} \cdots \partial x_{j_q}} F(x_1, \dots, x_p)$$

evaluada en  $(\max(X_{i,1}, u_1), \dots, \max(X_{i,p}, u_p))$ , donde  $j_l$  es el índice de las variables marginales para las que  $X_{i,j_l} > u_{j_l}$  y  $q$  es el número de variables que cumplen la condición. Normalmente, se consideran las  $F_j$  marginales Fréchet, siendo los  $u_j$  umbrales localizados en altos cuantiles de esta distribución.

Coles y Tawn (1991) proponen otro método de análisis de excedencias de umbrales partiendo de la condición (d). Para  $n$  grande, el proceso puntual  $N_n$  de esta condición se comporta como un proceso de Poisson no homogéneo en  $E$ . Si existen un número  $n_A \leq n$  de observaciones,  $X_{(1)}, \dots, X_{(n_A)}$ , pertenecientes a  $A = (0, u]^c$ , donde  $u$  es un vector de umbrales, la verosimilitud para la aproximación Poisson es

$$L = \exp \left\{ -\mu(\tilde{A}) \right\} \prod_{k=1}^{n_A} \mu(d\tilde{X}_{(k)}),$$

donde  $\tilde{A} = A/n$  y  $\tilde{X}_{(k)} = X_{(k)}/n$ . Los parámetros se estimarán por máxima verosimilitud.

Se puede desarrollar otro método alternativo de excedencias de umbrales a partir de la condición (c) aproximando la distribución condicional de  $X_i$  dado que  $X_i \in A$  mediante

$$W(x) = \frac{1}{-\log G(1, \alpha)} \log \left( \frac{G(x, \alpha)}{G(\min(x, 1), \alpha)} \right), \quad x > 0,$$

donde  $A = (0, u1]^c$  y  $u$  es un vector de umbrales.

Mediante una parametrización de  $G$  se obtiene un modelo paramétrico para  $W$  cuya estimación se realizara por máxima verosimilitud.

Hasta ahora se ha considerado que la distribución  $F$  tiene marginales Fréchet; sin embargo, en la práctica las distribuciones marginales  $F_j$  son desconocidas y deberán ser estimadas. Una solución habitual a este problema es la realización de una aproximación semiparamétrica, donde  $F_j(x_j)$  se estima a través de la distribución empírica para  $x_j \leq u_j$ , y una modelización mediante la distribución de Pareto generalizada para  $x_j > u_j$ :

$$F_j(x_j) = \begin{cases} 1 - \lambda_j \left( 1 + \xi_j \left[ \frac{x_j - u_j}{\sigma_j} \right] \right)_+^{-1/\xi_j}, & \text{si } x_j > u_j \\ (n+1)^{-1} \sum_{i=1}^n \mathbb{1}(X_{i,j} \leq x_j), & \text{si } x_j \leq u_j, \end{cases}$$

donde  $\lambda_j > 0$ ,  $\sigma_j > 0$  y  $\xi_j \in \mathbb{R}$  son parámetros desconocidos que deberán estimarse. Mediante las transformaciones  $-1/\log F_j(X_{i,j})$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, p$  se obtienen las observaciones estandarizadas a escala Fréchet y se pueden aplicar los procedimientos vistos en esta sección.

## Capítulo 3

# Extremos espaciales

### 3.1. Introducción

En algunos casos, principalmente en el estudio de fenómenos medioambientales, se tendrá una dimensión espacial y el objetivo será modelizar la dependencia espacial entre extremos a partir de observaciones localizadas en una malla. Para ello, se usarán procesos max-estables, es decir, procesos estocásticos en los que las distribuciones finito-dimensionales serán distribuciones de valores extremos.

Los extremos espaciales son de interés en el estudio de riesgos naturales como olas de calor, altas precipitaciones y nevadas, inundaciones, etc.

Como ejemplo de aplicaciones del estudio de extremos espaciales, en el contexto de las precipitaciones se tiene a Davidson, Padoan y Ribatet (2012), que realizaron una comparación entre modelos de variables latentes y modelos basados en procesos max-estables para un conjunto de precipitaciones estivales máximas recogidas entre 1962-2008 en 52 localizaciones en la región de Plateau (Suiza).

En cuanto a inundaciones costeras, Coles y Tawn (1990), consideraron los niveles máximos de la marea en la costa Británica, construyendo un modelo logístico bivariante para localizaciones cercanas.

### 3.2. Procesos max-estables espaciales

Para el caso de extremos espaciales son fundamentales los procesos max-estables, análogos infinito-dimensionales a los vectores aleatorios max-estables, definidos en la sección anterior.

En este caso, los extremos estarán definidos en el dominio espacial  $R \subset \mathbb{R}^d$ ,  $d \geq 1$ , y se supondrá en el caso teórico que  $R$  es un subconjunto compacto de  $\mathbb{R}^d$  y que todos los procesos estocásticos considerados siguen patrones de muestreo continuo, aunque estas últimas condiciones se podrán relajar en la práctica.

Sea  $Z(r)$  un proceso estocástico con marginales no degeneradas definido en

un conjunto  $R$ . Este proceso sera max-estable si todas sus distribuciones  $n$ -dimensionales satisfacen la propiedad de max-estabilidad, es decir:

**Definición 3.2.1** *Un proceso estocástico  $\{Z(r) : r \in R\}$  se dice max-estable si existen sucesiones de funciones continuas  $\{a_n(r) > 0 : r \in R, n \geq 1\}$  y  $\{b_n(r) \in \mathbb{R} : r \in R, n \geq 1\}$  tales que, para todo  $n \geq 1$ ,*

$$\left\{ \frac{\max_{i=1, \dots, n} Z_i(r) - b_n(r)}{a_n(r)} : r \in R \right\} \stackrel{d}{=} \{Z(r) : r \in R\},$$

donde  $\{Z_i(r) : r \in R, i \geq 1\}$  es una sucesión de réplicas independientes de  $\{Z(r) : r \in R\}$ .

La relevancia de los procesos max-estables en la modelización de extremos espaciales estará basada en argumentos asintóticos, según según de Haan (1984):

**Teorema 3.2.1** *Sea  $X_1(r), X_2(r), \dots$  una sucesión de réplicas independientes de un proceso estocástico  $\{X_n(r) : r \in R\}$ . Si existen funciones continuas  $c_n > 0$  y  $d_n \in \mathbb{R}$  tales que el proceso límite  $\{Z_n(r) : r \in R\}$ , definido como*

$$\frac{\max_{i=1, \dots, n} X_i(r) - d_n(r)}{c_n(r)} \rightarrow Z_n(r), \quad r \in R, n \rightarrow \infty,$$

es no degenerado, entonces, el proceso  $\{Z_n(r) : r \in R\}$  es un proceso max-estable.

Una descripción más precisa de los procesos max-estables se puede obtener a partir de su representación espectral.

### 3.2.1. Representación espectral de los procesos max-estables

El proceso aleatorio  $\{Z(r) : r \in R\}$  será un proceso max-estable simple si tiene marginales Fréchet.

**Teorema 3.2.2** *(de Haan, 1984; Penrose, 1992) Cualquier proceso max-estable simple no degenerado  $\{Z(r) : r \in R\}$  definido en un conjunto compacto  $R \subset \mathbb{R}^d$ ,  $d \geq 1$ , con patrones de muestreo continuos satisface*

$$Z(r) \stackrel{d}{=} \max_{i \geq 1} \zeta_i f_i(r), \quad r \in R,$$

donde  $\{(\zeta_i, f_i) : i \geq 1\}$  son las coordenadas de un proceso de Poisson en  $(0, \infty) \times \mathcal{C}$  con medida de intensidad  $\zeta^{-2} d\zeta \nu(df)$  para alguna medida local finita  $\nu$  definida en el espacio de las funciones continuas no negativas  $\mathcal{C}$  en  $R$ , tal que

$$\int f(r) \nu(df) = 1, \quad r \in R.$$

### 3.2.2. Modelos max-estables

En esta sección se van a introducir algunos modelos max-estables que serán relevantes en el estudio de extremos espaciales.

#### Modelo de Smith (1990)

Smith (1990) introdujo la siguiente representación paramétrica de un proceso max-estable continuo en el tiempo:

$$Z(r) = \max_{i \geq 1} X_i f(r, U_i),$$

donde  $\{X_i, U_i\}_{i \geq 1}$  es una realización de un proceso de Poisson definido en  $(0, \infty) \times R_U$  con medida de intensidad  $x^{-2} dx \nu(du)$  y  $f(\cdot)$  es una función de densidad no negativa en  $R \times R_U$ , con integral finita respecto a una medida positiva  $\nu$  en  $R_U$ .

El proceso  $Z(\cdot)$  será un proceso max-estable simple si imponemos la restricción  $\int_{R_U} f(r, u) \nu(du) = 1$ .

En este proceso, que Smith (1990) interpretó en términos de precipitaciones,  $X_i$  y  $U_i$  son, respectivamente, el tamaño y el tipo de tormenta  $i$ .

Smith (1990) definió el proceso de valores extremos gaussianos como

$$P(Z(r) \leq z_1, Z(r+h) \leq z_2) = \exp \left\{ -\frac{1}{z_1} \Phi \left( \frac{a(h)}{2} + \frac{1}{a(h)} \log \frac{z_2}{z_1} \right) - \frac{1}{z_2} \Phi \left( \frac{a(h)}{2} + \frac{1}{a(h)} \log \frac{z_1}{z_2} \right) \right\}, \quad h \in \mathbb{R}^d,$$

donde  $\Phi(\cdot)$  es la función de distribución normal estándar y  $a(h) = \sqrt{h' \Sigma^{-1} h}$ . Esta representación se usará en modelos medioambientales.

#### Modelo de Schlather (2002)

Schlather (2002) propuso una extensión del modelo de Smith (1990), sustituyendo la función determinística  $f(\cdot)$  por una forma aleatoria  $Y(\cdot)$ :

$$Z(r) = \max_{i \geq 1} X_i Y_i(r - U_i),$$

donde  $\{X_i, U_i\}_{i \geq 1}$  es una realización de un proceso de Poisson en  $(0, \infty) \times \mathbb{R}^d$  con medida de intensidad  $x^{-2} dx \times \mu^{-1} du$  e  $\{Y_i(\cdot)\}_{i \geq 1}$  son réplicas independientes de una función aleatoria no negativa  $Y(\cdot)$  definida en  $\mathbb{R}^d$ , tal que  $\mu = \mathbb{E} \left[ \int_{\mathbb{R}^d} Y(r) dr \right] \in (0, \infty)$ .

Schlather (2002) también introdujo una clase de procesos max-estables basada en procesos estocásticos estacionarios con media finita.

Sea  $\{X_i\}_{i \geq 1}$  una realización de un proceso de Poisson definido en  $(0, \infty)$  con medida de intensidad  $\mu^{-1}x^{-2}dx$  y sean  $\{W_i(\cdot)\}_{i \geq 1}$  replicas independientes de un proceso estocástico estacionario  $W(\cdot)$  definido en  $\mathbb{R}^2$ , con  $\mu = \mathbb{E}(\max\{0, W(0)\}) \in (0, \infty)$ .

El proceso estocástico  $Z(\cdot)$  definido por

$$Z(r) = \max_{i \geq 1} X_i W_i(r)$$

es un proceso max-estable simple estacionario, con distribución multivariante, para cualquier subconjunto  $\{r_1, \dots, r_m\}$  de  $R$ :

$$P(Z(r_1) \leq z_1, \dots, Z(r_m) \leq z_m) = \exp \left\{ -\mathbb{E} \left( \sup_{1 \leq i \leq m} \frac{W(r_i)}{z_i} \right) \right\}.$$

Schlather (2002) propuso tomar  $Y_i(\cdot)$  como un proceso estacionario gaussiano con función de correlación  $\rho$ . El proceso max-estable resultante se conoce como proceso de extremos gaussiano y su distribución bivariante viene dada por

$$P(Z(r) \leq z_1, Z(r+h) \leq z_2) = \exp \left\{ -\frac{1}{2} \left( \frac{1}{z_1} + \frac{1}{z_2} \right) \times \left( 1 + \sqrt{1 - 2(\rho(h) + 1) \frac{z_1 z_2}{(z_1 + z_2)^2}} \right) \right\}.$$

### 3.3. Medidas de dependencia para extremos espaciales

#### 3.3.1. Función del coeficiente extremal

El coeficiente extremal se utilizará para medir la dependencia entre los extremos de dos vectores aleatorios  $(X, Y)$  con distribución de extremos bivariante y distribución marginal común  $F(r)$ . Este coeficiente se define por la identidad

$$P(\max(X, Y) \leq x) = F^\theta(x),$$

donde el rango de  $\theta$  será  $[1, 2]$ , siendo  $\theta = 1$  dependencia perfecta y  $\theta = 2$  independencia entre los máximos.

Para un proceso max-estable  $Z(r)$ , la función del coeficiente extremal,  $\theta(\cdot)$ , viene dada por (Schlather y Tawn, 2003)

$$P(\max(Z(r), Z(r+h)) \leq z) = \exp\{-\theta(h)/z\}.$$

Algunas propiedades de  $\theta(\cdot)$  son:



1.  $2 - \theta(\cdot)$  es una función semi-definida positiva.
2. La función  $\theta(\cdot)$  no es diferenciable en 0, excepto en el caso  $\theta(h) = 1, \forall h$
3. Si  $Z(\cdot)$  es un proceso estocástico isotropico,  $\theta(\cdot)$  solo tiene un punto de discontinuidad en 0.

### 3.3.2. F-madograma

En geoestadística, la herramienta básica para medir la dependencia espacial es el semi-variograma, que se define como

$$\gamma(h) = \frac{1}{2} \mathbb{E} \left[ \{W(r) - W(r+h)\}^2 \right], \quad r, r+h \in R.$$

La herramienta análoga al semi-variograma para la teoría de extremos fue definida por Cooley et al. (2006) y se conoce como  $F$ -madograma:

$$\nu_F(h) = \frac{1}{2} \mathbb{E} [ |F\{Z(r+h)\} - F\{Z(r)\}| ], \quad r, r+h \in R1$$

donde  $F$  es la función de distribución acumulativa de  $Z(r)$ .

El rango de valores del  $F$ -madograma varía de 0 a 1/6, lo que se corresponde con dependencia completa e independencia respectivamente.

En la práctica se estimará de la siguiente forma:

$$\hat{\nu}_F(h) = \frac{1}{2n(n+1)} \sum_{i=1}^n |R_i(r) - R_i(r+h)|, \quad R_i(r) = \sum_{\ell=1}^n 1_{\{Z_\ell(r) \leq Z_i(r)\}},$$

donde  $Z_1, \dots, Z_n$  son réplicas independientes de un proceso max-estable  $Z$ .

### 3.3.3. $\lambda$ -madograma

El coeficiente extremal, así como el  $F$ -madograma, no representan de forma completa la dependencia espacial de un campo aleatorio.

Para resolver este problema, Naveau et al. (2009) introdujo el  $\lambda$ -madograma, que se define como

$$\nu_\lambda(r_1, r_2) = \frac{1}{2} \mathbb{E} [ |F^\lambda\{Z(r_1)\} - F^{1-\lambda}\{Z(r_2)\}| ], \quad \forall \lambda \in [0, 1].$$

De esta forma, mediante la variación de  $\lambda$  se tendrá  $P[Z(r_1) \leq z_1, Z(r_2) \leq z_2]$ , donde  $z_1 = \lambda z$  y  $z_2 = (1 - \lambda) z$ , y de esta forma se podrá explorar todo el espacio.

## 3.4. Simulación de procesos espaciales max-estables

Las simulaciones espaciales permiten la recuperación de información en cualquier localización.

Se llamarán simulaciones condicionadas aquellas que se ajustan a los datos observados; en otro caso, se denominarán no condicionadas. La simulación de las primeras será especialmente útil ya que se podrán combinar diferentes simulaciones condicionadas para obtener estimadores de cualquier medida de interés, como cuantiles, probabilidades de excedencia de umbrales o niveles de retorno.

### 3.4.1. Simulaciones no condicionadas

Los procesos max-estables se han definido basándose en infinitas réplicas de un proceso estocástico; sin embargo, en la práctica solo se podrá obtener un número finito de realizaciones de un proceso. A pesar de esto existen resultados teóricos y procedimientos ad-hoc mediante los que será posible obtener simulaciones exactas o aproximadas de algunos modelos max-estables.

En Schlather (2002) se dan algunas condiciones para la obtención de simulaciones exactas para la clase de procesos max-estables

$$Z(r) = \max_{i \geq 1} X_i W_i(r).$$

Suponemos que el proceso estocástico  $W(\cdot)$  está uniformemente acotado por una constante positiva  $C < \infty$ . Para  $T_0 = 0$  y  $k = 1, 2, \dots$ , el algoritmo de simulación será el siguiente:

1. Se genera  $E_k \sim \mathcal{E}$  y se hace  $T_k = T_{k-1} + E_k$ ,  $X_k = T_k^{-1}$ ,
2. se genera  $W_k(\cdot) \sim W(\cdot)$ ,
3. si  $CX_k > \max_{1 \leq i \leq k} X_i W_i(r)$ , volver a 1; en otro caso  $Z(r) = \max_{1 \leq i \leq k} X_i W_i(r)$ ,

donde  $\mathcal{E}$  es una distribución exponencial con media unitaria,  $(T_i)_{i \geq 1}$  es un proceso de Poisson definido en  $(0, \infty)$  con medida de intensidad  $dt$  y  $(X_i)_{i \geq 1}$  es un proceso de Poisson definido en  $(0, \infty)$  con medida de intensidad  $x^{-2}dx$ .

Se obtendrá una buena aproximación al elegir un  $C$  tal que  $P(W(r) > C)$  sea lo suficientemente pequeña.

Para el proceso  $Z(r) = \max_{i \geq 1} X_i Y_i(r - U_i)$  se puede aplicar un algoritmo análogo, ya que  $Y(\cdot)$  es un proceso uniformemente acotado.

### 3.4.2. Simulaciones condicionadas

La simulación condicionada, a diferencia de la no condicionada, no puede obtenerse de forma exacta.

Wang y Stoev (2011) dan una solución aproximada a este problema. Supongamos que se observa un proceso max-estable  $Z(\cdot)$  en las ubicaciones  $r_1, \dots, r_n$ . La distribución de un vector max-estable  $(Z(r_1), \dots, Z(r_n))'$  se podrá aproximar mediante el vector multivariante  $(\tilde{Z}(r_1), \dots, \tilde{Z}(r_n))'$ , donde  $\tilde{Z}(r_i) = \max_{k=1, \dots, p} \phi_k(r_i) Y_k$ , para  $i = 1, \dots, n$ , siendo  $\phi_k(\cdot)$  funciones determinísticas no negativas e  $Y_k$  variables aleatorias independientes con distribución  $\alpha$ -Fréchet, es decir,  $P(Y_k \leq y) = \exp(-\sigma_k^\alpha y^{-\alpha})$ , para  $\alpha, \sigma_k, y > 0$ .

El algoritmo de Wang y Stoev (2011) genera muestras a partir de la probabilidad condicionada de  $(Y_1, \dots, Y_p)'$  dadas  $(\hat{Z}(r_1), \dots, \hat{Z}(r_n))' = (Z(r_1), \dots, Z(r_n))'$  y predice  $Z(r_*)$  en cualquier ubicación arbitraria  $r_*$  a partir de las muestras  $(Y_i^{(sim)}, \dots, Y_p^{(sim)})'$ , usando la siguiente aproximación:

$$\hat{Z}^{(sim)}(r_*) = \max_{k=1, \dots, p} \phi_k(r_*) Y_k^{(sim)}.$$

### 3.5. Ajuste de un proceso max-estable Fréchet a los datos

En esta sección se van a presentar dos aproximaciones diferentes para realizar el ajuste de procesos max-estables a los datos. La primera estará basada en mínimos cuadrados y para la segunda se usará el estimador compuesto máximo-verosímil propuesto por Lindsay (1988).

#### 3.5.1. Mínimos cuadrados

El procedimiento consiste en minimizar la función

$$C(\psi) = \sum_{i < j} \left( \frac{\theta_{i,j} - \tilde{\theta}_{i,j}}{s(\tilde{\theta}_{i,j})} \right)^2,$$

donde  $\psi$  es el vector de parámetros del proceso max-estable,  $\theta_{i,j}$  es el coeficiente extremal predicho para el modelo y  $\tilde{\theta}_{i,j}$  es un estimador semi-paramétrico definido como

$$\tilde{\theta}_{i,j} = \frac{n}{\sum_{k=1}^n \min \{ Z_k(r_i)^{-1}, Z_k(r_j)^{-1} \}}$$

con desviación estándar  $s(\tilde{\theta}_{i,j})$ .

### 3.5.2. Máxima verosimilitud

Este método presenta el inconveniente de que la verosimilitud del modelo no se conoce de forma analítica para dimensiones mayores o iguales que tres. Sin embargo, dado que la densidad bivalente sí es conocida, Ribatet (2009) sugiere el uso de funciones de verosimilitud dos a dos. La expresión de la log-verosimilitud dos a dos viene dada por

$$\ell_p(\mathbf{z}; \psi) = \sum_{i < j} \sum_{k=1}^{n_{i,j}} \log f\left(z_k^{(i)}, z_k^{(j)}; \psi\right),$$

donde  $\mathbf{z}$  es el conjunto total de observaciones de la región,  $n_{i,j}$  es el número de observaciones entre las localizaciones  $i$  y  $j$ ,  $y_k^{(i)}$  es la  $k$ -ésima observación en la localización  $i$  y  $f(\cdot, \cdot)$  es la densidad bivalente de un proceso max-estable Fréchet.

Esta estimación por máxima verosimilitud estará especificada parcialmente; es decir, la variable de la que se han extraído las observaciones del modelo,  $\{f(y; \psi), \psi \in \mathbb{R}^d\}$ , tiene una densidad  $g$  desconocida. El modelo quedará especificado si existe  $\psi_* \in \mathbb{R}^d$  tal que  $f(y; \psi_*) = g(y)$  para todo  $y$ .

La estimación del vector de parámetros se distribuirá como

$$\hat{\psi} = N\left(\psi, H(\psi)^{-1} J(\psi) H(\psi)^{-1}\right),$$

donde

$$H(\psi) = n \int \frac{\partial^2 \log f(y; \psi)}{\partial \psi \partial \psi^T} g(y) dy$$
$$J(\psi) = n \int \frac{\partial \log f(y; \psi)}{\partial \psi} \frac{\partial \log f(y; \psi)}{\partial \psi^T}.$$

## 3.6. Selección del modelo

Dados varios modelos a los que se ajustan bien a los datos, será útil disponer de herramientas que nos permitan elegir el más adecuado. Si los modelos a comparar tienen la misma log-verosimilitud, se seleccionará el modelo más simple; sin embargo, si difieren en una mínima cantidad se deberá disponer de algún criterio para discriminar un modelo frente a otro.

Se van a presentar dos criterios de selección de modelos; el Criterio de Información de Takeuchi (Akaike, 1974) y el Estadístico de la Tasa de Verosimilitud (Davison, 2003).

### 3.6.1. Criterio de información de Takeuchi

Este criterio se utiliza para comparar modelos dos a dos.

Se considera una muestra aleatoria  $Y_1, \dots, Y_n$  extraída de una variable con función de densidad  $g$  desconocida.

Se ajustará un modelo estadístico  $f(y; \psi)$  maximizando la log-verosimilitud. La medida de divergencia de Kullback-Leibler (Varin, Vidoni, 2005) determina la discrepancia entre el modelo ajustado  $f$  y el modelo real  $g$ ,

$$D(f_\psi, g) = \int \log \left( \frac{g(y)}{f(y; \psi)} \right) g(y) dy$$

Esta medida siempre será positiva; y se tomará como modelo aquel que minimice  $D(f_\psi, g)$ .

Dado que algunos modelos satisfacen  $D(f_\psi, g) = 0$ , este método no es lo suficientemente discriminante; para solucionar este problema se utiliza el Criterio de Informacion de Takeuchi (TIC),

$$TIC = -2\ell(\hat{\psi}) - 2tr(\hat{J}\hat{H}^{-1}),$$

donde  $\hat{\psi}$  es la estimación de  $\psi$  y  $\hat{H}$  y  $\hat{J}$  son estimaciones consistentes de las matrices  $H(\psi)$  y  $J(\psi)$ , respectivamente.

El mejor modelo se corresponderá con aquel que minimice el TIC.

### 3.6.2. Estadístico de la tasa de verosimilitud

Este criterio es útil para comparar modelos anidados.

Para un modelo estadístico  $\{f(x, \psi)\}$ , con  $\psi^T = (\kappa^T, \phi^T)$ , se quiere comprobar que los datos son consistentes con la hipótesis de que  $\kappa = \kappa_0$ ; para ello, se usará el estadístico de la tasa de verosimilitud  $W(\kappa_0)$ ,

$$W(\kappa_0) = 2 \left\{ \ell(\hat{\kappa}, \hat{\phi}) - \ell(\kappa_0, \hat{\phi}_{\kappa_0}) \right\} \rightarrow \sum_{i=1}^p \lambda_i X_i, \quad n \rightarrow \infty,$$

donde  $\hat{\phi}_{\kappa_0}$  es la estimación por máxima verosimilitud de  $\phi$  bajo la restricción  $\kappa = \kappa_0$ ,  $p$  es la dimensión del parámetro  $\kappa_0$ ,  $X_i$  son variables independientes con distribución  $\chi_1^2$  y  $\lambda_i$  son los autovalores de

$$\{H^{-1}JH^{-1}\}_\kappa [\{H^{-1}\}_\kappa]^{-1}.$$



## Capítulo 4

# Aplicación con R. Extremos espaciales

Se va a trabajar con el paquete de R **SpatialExtremes** y con el conjunto de datos **wind** que recoge los máximos anuales de la velocidad del viento, registrados en 35 estaciones meteorológicas de los Países Bajos en el periodo 1971-2012.



Figura 4.1: Localizaciones de las estaciones meteorológicas

```
maps::map(xlim=c(0,9),ylim=c(50,54),fill=TRUE,col="green3")
points(coord,pch=15)
```

Para el tratamiento de los datos se seguirá el siguiente proceso propuesto por Smith (1990):

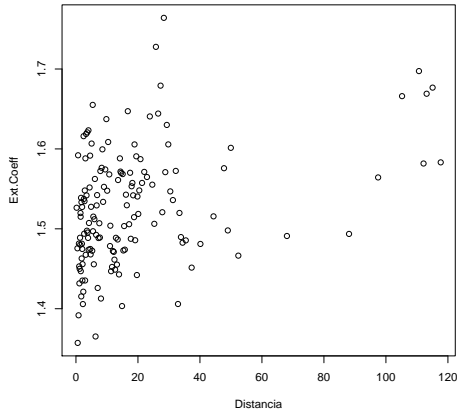
1. Ajuste de la distribución de valores extremos generalizada a los datos.
2. Transformación de los datos a la distribución Fréchet.
3. Ajuste de un proceso max-estable a los datos transformados.
4. Diagnósis y elección del modelo.
5. Predicciones.

#### **4.1. Ajuste de la distribución de valores extremos generalizada a los datos**

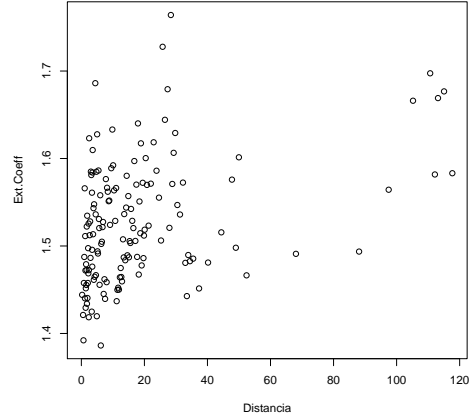
Para el ajuste de la distribución GEV será necesario comprobar si los datos presentan dependencia espacial y localizar tendencias espaciales si las hubiera.

La forma de comprobar la existencia de dependencia espacial es a través de la representación del  $F$ -madograma y el coeficiente extremal.

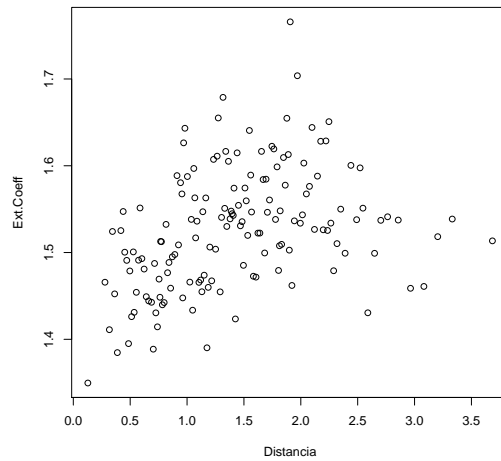




(a) Longitud vs. Altitud



(b) Latitud vs. Altitud



(c) Longitud vs. Latitud

Figura 4.2: Representación del coeficiente extremal frente a la distancia

En estos gráficos no se observa dependencia espacial en función de la altitud, únicamente para las dimensiones longitud y latitud se observa una variación del coeficiente extremal en función de la distancia.

Los siguientes gráficos ayudarán a detectar tendencias espaciales:

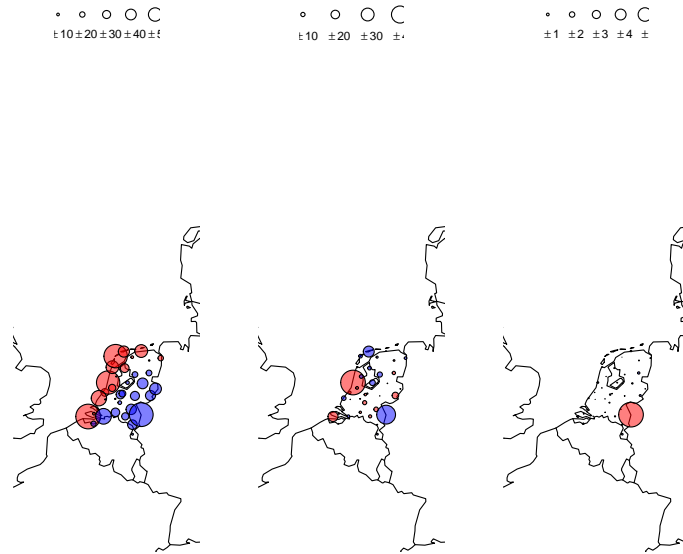


Figura 4.3: Gráficos de tendencias para  $\mu(x)$ ,  $\sigma(x)$  y  $\xi(x)$ , respectivamente.

```

symbolplot(wind, coord, which="mean")
symbolplot(wind, coord)
border<-function(add=FALSE)maps::map(xlim=c(0,9), ylim=c(47.5,57.5), add=add)
symbolplot(wind, coord, plot.border=border, scale=2)

```

En el caso de los parámetros  $\sigma$  y  $\xi$  no se observa ningún patrón. En el gráfico de la izquierda, correspondiente al parámetro de escala  $\mu$ , se observa una separación de los datos, por lo que este parámetro se definirá en función de las coordenadas.

La definición de los parámetros propuesta por Ribatet (2015) es la siguiente:

$$\begin{aligned}\mu(x) &= \beta_{0,\mu} + \beta_{1,\mu}lon(x) + \beta_{2,\mu}lat(x) + \beta_{3,\mu}lon(x)lat(x) \\ \sigma(x) &= \beta_{0,\sigma} \\ \xi(x) &= \beta_{0,\xi}\end{aligned}$$

```
loc.M1<-y lon*lat; scale.M1<-shape.M1<-y 1
```

El ajuste de la distribución GEV a los datos da como resultado el valor de los parámetros estimados junto con sus errores estándar y la matriz de varianzas covarianzas.

```

# Ajuste de la distribución GEV a los datos
M1<-fitspatgev(wind, scale(coord, scale=FALSE), loc.M1, scale.M1, shape.M1);
M1

```

Model: Spatial GEV model  
Deviance: 10735.05  
TIC: 10768.3

Location Parameters:  
locCoeff1 locCoeff2 locCoeff3 locCoeff4  
267.56 -15.96 22.85 2.23

Scale Parameters:  
scaleCoeff1  
33.86

Shape Parameters:  
shapeCoeff1  
-0.09237

Standard Errors  
locCoeff1 locCoeff2 locCoeff3 locCoeff4 scaleCoeff1 shapeCoeff1  
3.12232 1.04371 1.71802 1.58605 1.03977 0.01854

Asymptotic Variance Covariance

	locCoeff1	locCoeff2	locCoeff3	locCoeff4	scaleCoeff1	shapeCoeff1
locCoeff1	9.7488927	-0.5203683	1.5094327	-0.4169546	1.6639010	
locCoeff2	-0.5203683	1.0893215	-0.7382272	0.2126821	-0.0542326	
locCoeff3	1.5094327	-0.7382272	2.9515839	0.9878967	0.3578833	
locCoeff4	-0.4169546	0.2126821	0.9878967	2.5155629	-0.2183085	
scaleCoeff1	1.6639010	-0.0542326	0.3578833	-0.2183085	1.0811170	
shapeCoeff1	-0.0319854	0.0102704	-0.0072285	0.0031509	0.0018621	0.0003439

Optimization Information  
Convergence: successful  
Function Evaluations: 529

El ajuste de este modelo nos permitirá predecir los niveles de retorno para un determinado número de años.

```
x<-seq(min(coord[,1]),max(coord[,1]),length=100)
y<-seq(min(coord[,2]),max(coord[,2]),length=100)
grid<-expand.grid(x,y);
colnames(grid)<-c("lon","lat")
grid[,1]<-grid[,1]-mean(coord[,1]);
grid[,2]<-grid[,2]-mean(coord[,2]);
```

```

ans<-predict(M1,newdata=grid,ret.per=100)$Q100
maps::map(xlim=range(x),ylim=range(y))
image(x,y,matrix(ans,100),add=TRUE,col=cm.colors(64))
contour(x,y,matrix(ans,100),add=TRUE);
maps::map(add=TRUE)

```

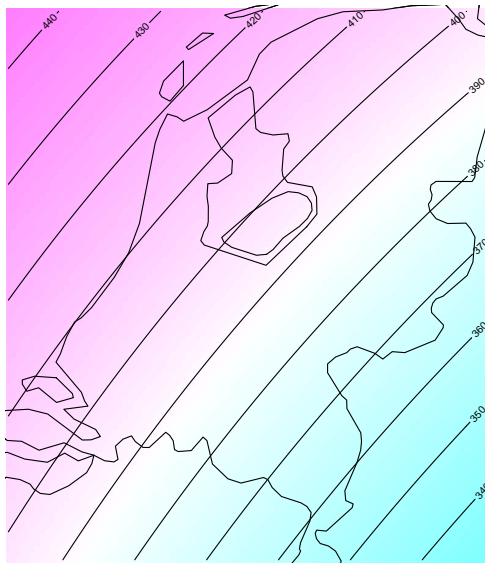


Figura 4.4: Predicción de los niveles de retorno a 100 años

## 4.2. Transformación de los datos a la distribución Fréchet.

En primer lugar se transformarán los datos partiendo de la distribución GEV calculada en la sección anterior; para ello, se tendrán en cuenta los valores de los parámetros ajustados:

```

loc.coeff.M1<-M1$fitted.values[1:4];
sigma.M1<-M1$fitted.values[5];
xi.M1<-M1$fitted.values[6]

mu.M1<-numeric(nrow(coord))
for(i in 1:nrow(coord))

mu.M1[i]<-loc.coeff.M1[1]+loc.coeff.M1[2]*coord[i,1]+loc.coeff.M1[3]*coord[i,2]+
loc.coeff.M1[4]*coord[i,1]*coord[i,2]

```

```

>mu.M1 [1] 1903.066 1928.571 1965.237 1943.246 1992.469 1975.527 2038.791
1976.932 [9] 1987.656 2025.385 2024.461 2074.960 2069.942 2046.741
2127.005 2098.129 [17] 2141.369 2156.563 2123.485 2217.123 2155.867
1798.158 1817.627 1829.274 [25] 1867.950 1872.704 1898.695 1947.514
1934.438 1965.384 1977.807 2014.147 [33] 2003.479 1993.622 2056.935
>sigma.M1
scaleCoeff1
33.85732
>xi.M1
shapeCoeff1
-0.09236842

```

Con el siguiente código y el uso de la función `gev2frec()` se transforma el conjunto de datos original en otro conjunto de datos (`wind.frechet.M1`) del mismo tamaño, en el que cada uno de los vectores tiene marginales Fréchet:

```

wind.frechet.M1<-numeric(length(wind));
wind.frechet.M1<-matrix(wind.frechet.M1,nrow=nrow(wind),ncol=ncol(wind));
for(j in 1:ncol(wind))
for(i in 1:nrow(wind))
wind.frechet.M1[i,j]<-gev2frec(wind[i,j],mu.M1[j],sigma.M1,xi.M1)

```

Además se realizara una segunda transformación en la que los datos serán transformados usando la distribución acumulada empírica, obteniendose el conjunto de datos `wind.frechet.M2`.

```

wind.frechet.M2<-apply(wind,2,gev2frec,emp=TRUE)

```

## 4.3. Ajuste del proceso max-estable a los datos transformados.

### 4.3.1. Modelo 1

Se ajustará sobre el conjunto de datos `wind.frechet.M1`.

De la misma forma que en el ajuste de la distribución GEV, habrá que definir la forma de los parámetros del modelo:

```

symbolplot(log(wind.frechet.M1),coord[,-3],which="mean")
symbolplot(log(wind.frechet.M1),coord[,-3])
border<-function(add=FALSE)maps::map(xlim=c(0,9),ylim=c(47.5,57.5),add=add)
symbolplot(log(wind.frechet.M1),coord[,-3],plot.border=border,scale=10)

```

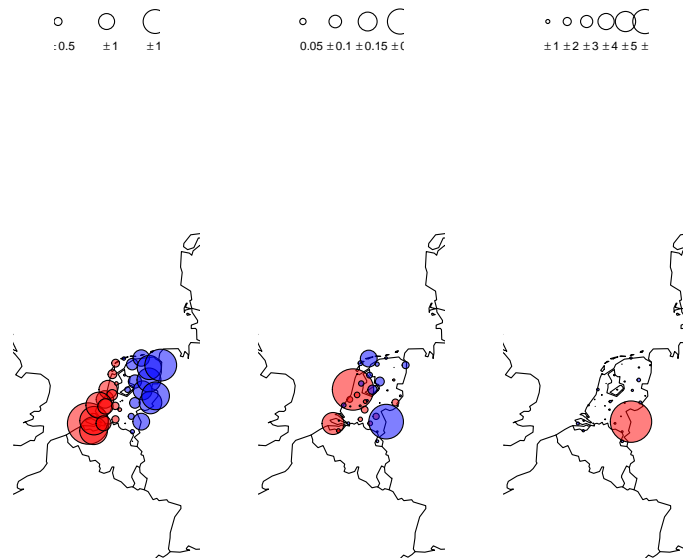


Figura 4.5: Gráficos de tendencias para  $\mu(x)$ ,  $\sigma(x)$  y  $\xi(x)$ , respectivamente.

En este caso, el parámetro de localización va a depender de las coordenadas, mientras que los parámetros de escala y forma serán constantes ya que no se aprecia una tendencia clara en sus respectivos gráficos.

$$\begin{aligned}\mu(x) &= \beta_{0,\mu} + \beta_{1,\mu}lon(x) + \beta_{2,\mu}lat(x) + \beta_{3,\mu}lon(x)lat(x), \\ \sigma(x) &= \beta_{0,\sigma} \\ \xi(x) &= \beta_{0,\xi}\end{aligned}$$

Equivalentemente en R.

```
loc.M1<-y ~ lon*lat; scale.M1<-shape.M1<-y ~ 1;
```

El ajuste del modelo se hará con la función `fitmaxstab()`.

```
M1.maxstab<-fitmaxstab(log(wind.frechet.M1),coord[,-3],"whitmat",nugget=0,
loc.M1,scale.M1,shape.M1)
```

### 4.3.2. Modelo 2

Se repite el mismo proceso para el conjunto de datos `wind.frechet.M2`.

```

symbolplot(wind.frechet.M2,coord[,-3],which="mean")
symbolplot(wind.frechet.M2,coord[,-3])
border<-function(add=FALSE)maps::map(xlim=c(0,9),ylim=c(47.5,57.5),add=add)
symbolplot(wind.frechet.M2,coord[,-3],plot.border=border,scale=3)

```

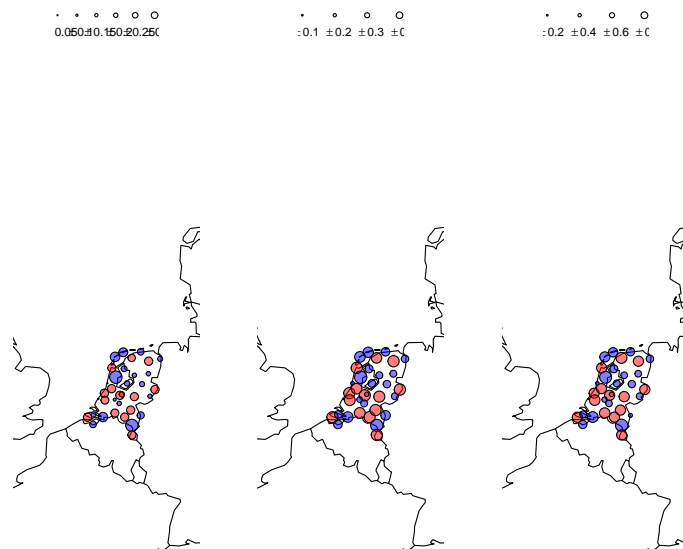


Figura 4.6: Gráficos de tendencias para  $\mu(x)$ ,  $\sigma(x)$  y  $\xi(x)$ , respectivamente.

No se aprecia ningún patrón en la distribución espacial de los puntos, todos los parámetros serán constantes:

```

loc.M2<-scale.M2<-shape.M2<-y 1;
M2.maxstab<-fitmaxstab(wind.frechet.M2,coord[,-3],"whitmat",nugget=0,loc.M2,scale.M2,shape.M2)

```

### 4.3.3. Diagnósis y elección del modelo

Para verificar la validez de cada uno de los modelos se va a analizar que las observaciones estén bien modelizadas en cada una de las localizaciones y que la estructura de dependencia esté bien definida; esto se hará mediante dos funciones: `qqgev()` en el caso del Modelo 1 y `fmadogram()` para el Modelo 2.

#### Modelo 1

```
qqgev(M1.maxstab)
```

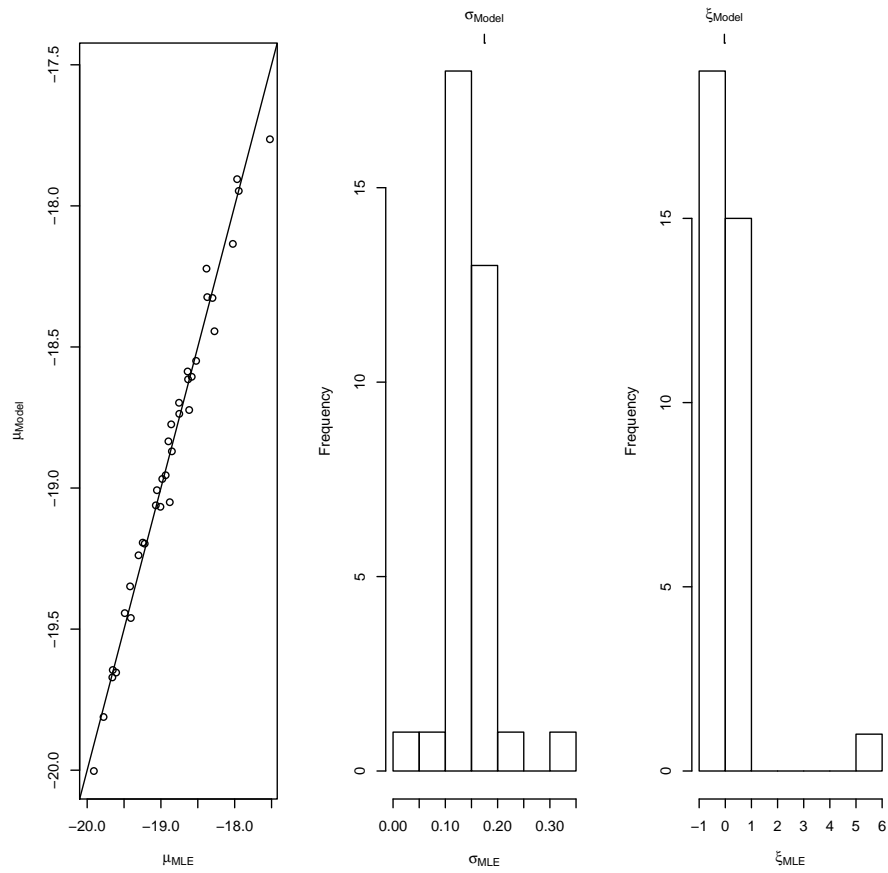


Figura 4.7: Diagnósis de los parámetros del modelo.

Con los gráficos qq se podrá comprobar si los parámetros predichos a través de las tendencias de superficie son relevantes; para ello, se compararán con los estimados por máxima verosimilitud en cada una de las localizaciones:

```
fmadogram(wind.frechet.M1,coord[,-3],M1.maxstab,which=c(".ext"),col=c("grey"))
fmadogram(wind.frechet.M1,coord[,-3],M1.maxstab,n.bins=200,which=c(".ext"),col=c(1,2),m
```



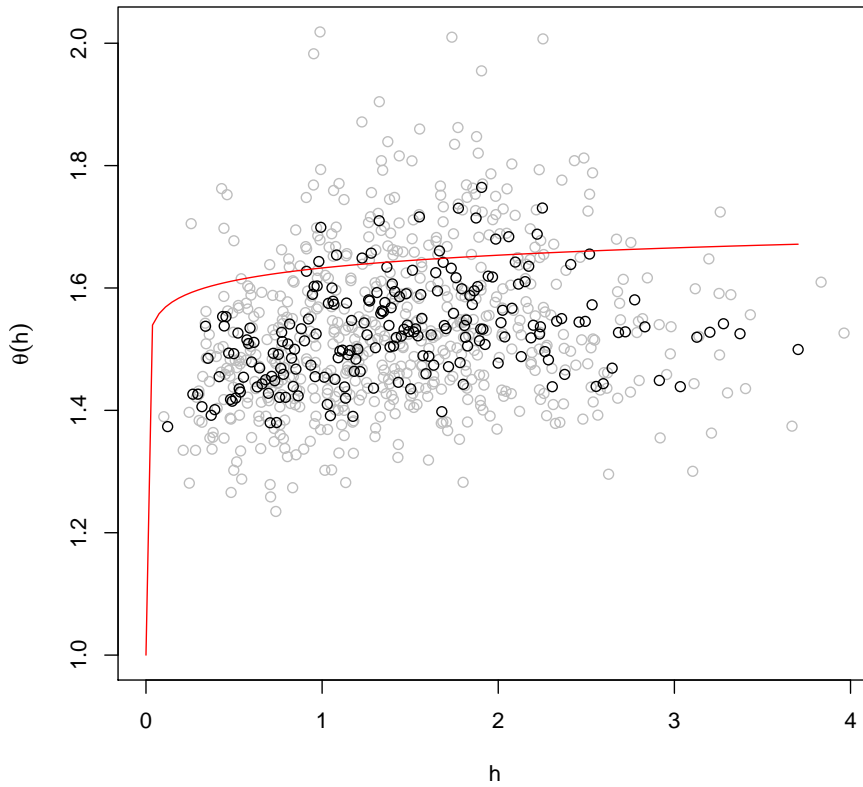


Figura 4.8:  $F$ -madograma Modelo 1.

Mediante el  $F$ -madograma se comprobará si el modelo es capaz de reproducir la estructura de dependencia espacial; con este gráfico se compararán las estimaciones del coeficiente extremo con las predichas por el modelo ajustado.

## Modelo 2

```
fmadogram(wind.frechet.M2, coord[, -3], M2.maxstab, which=c(".ext"), col=c("grey"))
fmadogram(wind.frechet.M2, coord[, -3], M2.maxstab, n.bins=200, which=c(".ext"), col=c(1,2), marge=".emp")
```

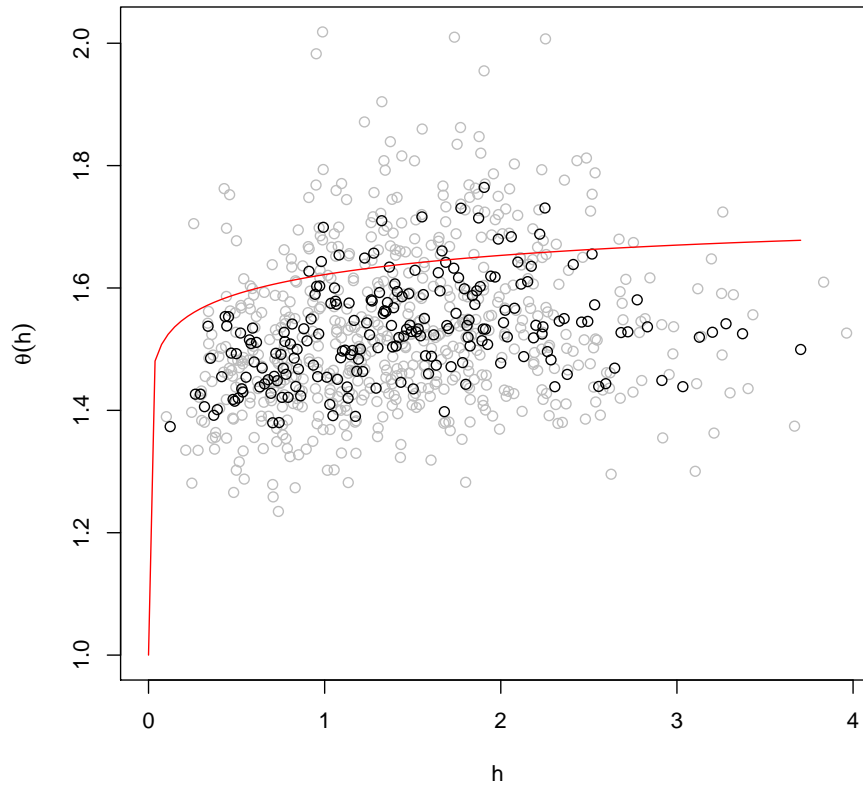


Figura 4.9:  $F$ -madograma.

Para realizar la selección del modelo se podrá usar el estadístico TIC (Criterio de Información de Takeuchi); para los modelos calculados anteriormente se tiene:

```
TIC(M1.maxstab,M2.maxstab)
```

M1.maxstab	M2.maxstab
-14857.13	75049.33

Según este criterio, se elegiría el Modelo 1, ya que es el que minimiza el resultado del TIC; este resultado es previsible, ya que el Modelo 1 es más restrictivo que el Modelo 2 al haber predefinido los parámetros.

Existen otros métodos de selección de modelos que penalizan el número de parámetros, como el Estadístico de la Tasa de Verosimilitud. En R este test se llevará a cabo mediante la función `anova()` usando como método la aproximación de Rotnitzky y Jewell RJ".

```
Eigenvalue(s): 28.69 14.72 9.37
```

```
Analysis of Variance Table
```

	Mdf	Deviance	Df	Chisq	Pr(>sum lambda Chisq)
M2.maxstab	5	74417			
M1.maxstab	8	-15503	3	89920	<2.2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### 4.3.4. Predicciones.

Las predicciones para los modelos espaciales se llevan a cabo mediante la función `predict`, que ya se usó en la sección del ajuste del modelo GEV; sin embargo, habrá que tener en cuenta que estas predicciones se realizan localización a localización, por lo que no se tendrá en cuenta la dependencia espacial.



## Capítulo 5

# Conclusión

El principal objetivo de este trabajo ha sido realizar una revisión de los principales enfoques en el análisis de valores extremos, centrándose principalmente en los procesos max-estables y en la dimensión espacial.

En la última década han surgido muchos avances en materia de extremos, fundamentalmente en el campo de la geoestadística a través del uso de extremos espaciales, siendo los procesos max-estables una solución elegante para modelizar estos valores.

Algunos de los últimos avances en el campo del análisis de los valores extremos son los modelos bayesianos jerárquicos (Ribatet et al., 2012; Reich and Shaby, 2012; Thibaud et al., 2015), que proporcionan algunas mejoras en la definición de las tendencias de superficie y modelos eficientes para las predicciones puntuales; extremos espacio-temporales (Huser, 2013; Turkman and Pereira, 2010) y análisis de valores extremos no estacionarios (Cheng y AghaKuoachak, 2014).

Sin embargo, el desarrollo de esta teoría es relativamente reciente y aún siguen surgiendo nuevos avances teóricos y se esperan progresos en el futuro.



# Bibliografía

- [1] S. A. Padoan A. C. Davison and M. Ribatet. Statistical modeling of spatial extremes. *Institute of Mathematical Statistics, Vol. 27, No. 2, 161-186*, 2012.
- [2] H. Akaike. *Selected Papers of Hirotugu Akaike*, chapter A new look at the statistical model identification, pages 215–222. Springer, 1974.
- [3] J. N. Bacro and C. Gaetan. *Advances and Challenges in Space-time Modelling of Natural.*, chapter 5 - A Review on Spatial Extreme Modelling, pages 103–124. Springer-Verlag, 2012.
- [4] L. Cheng and A. AghaKouchak. Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific Reports, No. 4, 7093*, 2014.
- [5] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.
- [6] S. Coles and J. A. Tawn. Statistics of coastal flood prevention. *Philosophical Transactions of the Royal Society of London, No. 332, 457-476*, 1990.
- [7] S. Coles and J. A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society, No. 53, 377-392*, 1991.
- [8] A. C. Davison. *Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2003.
- [9] L. de Haan. *On Regular Variation and its Applications to the Weak Convergence of Sample Extremes*. PhD thesis, Mathematical Centre Tract, 32, Amsterdam, 1970.
- [10] L. de Haan. A spectral representations for max-stable processes. *Annals of Probability, No. 12, 1194-1204*, 1984.
- [11] L. de Haan and A. Ferreira. *Extreme Value Theory. An Introduction*. Springer, 2006.
- [12] R. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society. No. 24, 180-190*, 1928.
- [13] M. Fréchet. Sur la loi de probabilité de l'écart maximum. *Ann.Soc. Math. Polon. No. 6, 93-116*, 1927.

- [14] R. Huser and A. C. Davison. Space-time modelling of extreme events. *Royal Statistical Society, No. 76, 439-461*, 2013.
- [15] J. Pickand III. Statistical inference using extreme order statistics. *Annals of Statistic, No. 3, 119-131*, 1975.
- [16] M. A. Turkman K. F. Turkman and J. M. Pereira. Asymptotic models and inference for extremes of spatio-temporal data. *Extremes, No. 13, 375*, 2010.
- [17] T. W. Yee L. Thibaut and D. Arnaud. Expectation particle belief propagation. Technical report, University of Oxford. Department of Statistics., 2015.
- [18] A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika, No. 83, 169-187*, 1996.
- [19] B. Lindsay. Composite likelihood methods. *Contemporary Mathematics, No. 80, 220-239*, 1988.
- [20] A. Guillou P. Naveau and D. Cooley. Modelling pairwise dependence of maxima in space. *Biometrika, No. 1, 1-17*, 2009.
- [21] M. D. Penrose. Semi-min-stable processes. *Annals of Probability, No. 3, 1450-1463*, 1992.
- [22] B. J. Reich and B. A. Shaby. A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics, No. 4, 1430-1451*, 2012.
- [23] R. D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhauser, 1997.
- [24] M. Ribatet. Modelling spatial extremes with the spatialextreme package. In *The 9th International Conference on Extreme Value Analysis and Application. University of Montpellier*.
- [25] M. Ribatet. *A User's Guide to the SpatialExtremes Package*. École Polytechnique Fédérale de Lausanne, 2009.
- [26] M. Ribatet. Spatial extremes: Max-stable processes at work. *Journal of Société Française de Statistique, Vol .154, No. 2, 156-177*, 2013.
- [27] M. Ribatet and M. Sedki. Extreme value copulas and max-stable processes. *Journal of Société Française de Statistique, Vol. 152, No. 3, 138-150*, 2012.
- [28] M. Hofmann S. Aulbach, M. Falk and M. Zott. On max-stable processes and the functional d-norm revisited. *Extremes, Vol. 16, No. 3, 255-283*, 2013.
- [29] M. Schlather. Models for stationary max-stable random fields. *Extremes, Vol. 1, No.5, 33-44*, 2002.



- [30] M. Schlather and J. Tawn. A dependence measure for multivariate and spatial extreme values: properties and inference. *Biometrika*, No. 90, 139-156, 2003.
- [31] J. Segers. Max-stable models for multivariate extremes. In *Copula Models and Dependence*. Centre de Recherches Mathématiques, 2011.
- [32] R. L. Smith. Max-stable processes and spatial extremes. Technical report, 1990.
- [33] S. A. Stoev. Max-stable processes: Representations, ergodic properties and statistical applications. Technical report, Department of Statistics, The University of Michigan, 2010.
- [34] C. Varin and P. Vidoni. A note on composite likelihood inference and model selection. *Biometrika*. No. 92, 519-528, 2005.
- [35] R. von Mises. la distribution de la plus grande de n valeurs. *Revue Mathématique de l'Union Interbalcanique*, No. 1, 141-160, 1936.
- [36] Y. Wang and S. A. Stoev. Conditional sampling for spspectral discrete max-stable random fields. *Advances in Applied Probability*, No. 2, 461-483, 2011.
- [37] Z. Zhang. On approximating max-stable processes and constructing extremal copula functions. *Statistical Inference for Stochastic Processes*, No. 12, 89-114, 2009.